RESEARCH CENTER
Sophia Antipolis - Méditerranée

FIELD

Activity Report 2015

Section New Results

Edition: 2016-03-21

ABS Project-Team

6. New Results

6.1. Modeling Interfaces and Contacts

Keywords: docking, scoring, interfaces, protein complexes, Voronoi diagrams, arrangements of balls.

6.1.1. High Resolution Crystal Structures Leverage Protein Binding Affinity Predictions

Participants: Frédéric Cazals, Simon Marillet.

In collaboration with P. Boudinot, Unité de recherche en virologie et immunologie moléculaires, INRA Jouy-en-Josas.

Predicting protein binding affinities from structural data has remained elusive, a difficulty owing to the variety of protein binding modes. Using the structure-affinity-benchmark (SAB, 144 cases with bound/unbound crystal structures and experimental affinity measurements), prediction has been undertaken either by fitting a model using a handfull of pre-defined variables, or by training a complex model from a large pool of parameters (typically hundreds). The former route unnecessarily restricts the model space, while the latter is prone to overfitting.

We design models in a third tier [20], using twelve variables describing enthalpic and entropic variations upon binding, and a model selection procedure identifying the best sparse model built from a subset of these variables. Using these models, we report three main results. First, we present models yielding a marked improvement of affinity predictions. For the whole dataset, we present a model predicting K_d within one and two orders of magnitude for 48% and 79% of cases, respectively. These statistics jump to 62% and 89% respectively, for the subset of the SAB consisting of high resolution structures. Second, we show that these performances owe to a new parameter encoding interface morphology and packing properties of interface atoms. Third, we argue that interface flexibility and prediction hardness do not correlate, and that for flexible cases, a performance matching that of the whole SAB can be achieved. Overall, our work suggests that the affinity prediction problem could be partly solved using databases of high resolution complexes whose affinity is known.

6.1.2. Dissecting Interfaces of Antibody - Antigen Complexes: from Ligand Specific Features to Binding Affinity Predictions

Participants: Frédéric Cazals, Simon Marillet.

In collaboration with: P. Boudinot, Unité de recherche en virologie et immunologie moléculaires, INRA Jouy-en-Josas; M-P. Lefranc, Univ. of Montpellier 2.

B lymphocytes recognize the antigen through their membrane immunoglobulins (IG), that can also be secreted. The diversity of IG-Ag complexes challenges our understanding in terms of binding affinity and interaction specificity.

In this work [21], we dissect the interfaces of IG-Ag complexes from high resolution crystal structures. We show that global interface statistics distinguish ligand types and that interfacial side chains play a key role in the interaction. Our analysis of the relative positions of CDR identifies a remarkably conserved pattern involving seven seams between CDR, with specific variations depending on the ligand type. Finally, we show that structural features of the interface and of the partners yield binding affinity estimates of unprecedented accuracy (median absolute error of 1.02 kcal/mol).

Our findings will be of broad interest, as understanding Ag recognition at the atomic level will help guiding design of better IG targeting Ag for therapeutic or other uses.

6.2. Modeling Macro-molecular Assemblies

Keywords: macro-molecular assembly, reconstruction by data integration, proteomics, modeling with uncertainties, curved Voronoi diagrams, topological persistence.

6.2.1. *Unveiling Contacts within Macro-molecular assemblies by solving Minimum Weight Connectivity Inference Problems*

Participants: Frédéric Cazals, Deepesh Agarwal.

In collaboration with C. Caillouet, and D. Coudert, from the COATI project-team (Inria - I3S (CNRS, University of Nice Sophia Antipolis)).

Consider a set of oligomers listing the subunits involved in sub-complexes of a macro-molecular assembly, obtained e.g. using native mass spectrometry or affinity purification. Given these oligomers, connectivity inference (CI) consists of finding the most plausible contacts between these subunits, and minimum connectivity inference (MCI) is the variant consisting of finding a set of contacts of smallest cardinality. MCI problems avoid speculating on the total number of contacts, but yield a subset of all contacts and do not allow exploiting a priori information on the likelihood of individual contacts. In this context, we present two novel algorithms, `MILP-W` and `MILP-W`$_B$ [14]. The former solves the *minimum weight connectivity inference* (MWCI), an optimization problem whose criterion mixes the number of contacts and their likelihood. The latter uses the former in a bootstrap fashion, to improve the sensitivity and the specificity of solution sets.

Experiments on three systems (yeast exosome, yeast proteasome lid, human eIF3), for which reference contacts are known (crystal structure, cryo electron microscopy, cross-linking), show that our algorithms predict contacts with high specificity and sensitivity, yielding a very significant improvement over previous work, typically a twofold increase in sensitivity.

The software accompanying this paper is made available in the `SBL` , and should prove of ubiquitous interest whenever connectivity inference from oligomers is faced.

6.3. Modeling the Flexibility of Macro-molecules

Keywords: protein, flexibility, collective coordinate, conformational sampling dimensionality reduction.

6.3.1. *Hybridizing Rapidly Growing Random Trees and Basin Hopping Yields an Improved Exploration of Energy Landscapes*

Participants: Frédéric Cazals, Tom Dreyfus, Christine Roth.

In collaboration with C. Robert (IBPC / CNRS, Paris).

The number of local minima of the potential energy landscape (PEL) of molecular systems generally grows exponentially with the number of degrees of freedom, so that a crucial property of PEL exploration algorithms is their ability to identify local minima which are low lying and diverse.

In this work [22], we present a new exploration algorithm, retaining the ability of basin hopping (`BH`) to identify local minima, and that of *transition based rapidly exploring random trees* (`T-RRT`) to foster the exploration of yet unexplored regions. This ability is obtained by interleaving calls to the extension procedures of `BH` and `T-RRT`, and we show tuning the balance between these two types of calls allows the algorithm to focus on low lying regions. Computational efficiency is obtained using state-of-the art data structures, in particular for searching approximate nearest neighbors in metric spaces.

We present results for the BLN69, a protein model whose conformational space has dimension 207 and whose PEL has been studied exhaustively. On this system, we show that the propensity of our algorithm to explore low lying regions of the landscape significantly outperforms those of `BH` and `T-RRT`.

6.4. Algorithmic Foundations

Keywords: computational geometry, Computational topology, Voronoi diagrams, α-shapes, Morse theory, graph algorithm, combinatorial optimization, statistical learning.

6.4.1. Beyond Two-sample-tests: Localizing Data Discrepancies in High-dimensional Spaces

Participants: Frédéric Cazals, Alix Lhéritier.

Comparing two sets of multivariate samples is a central problem in data analysis. From a statistical standpoint, the simplest way to perform such a comparison is to resort to a non-parametric two-sample test (TST), which checks whether the two sets can be seen as i.i.d. samples of an identical unknown distribution (the null hypothesis). If the null is rejected, one wishes to identify regions accounting for this difference. In this paper [17], we presents a two-stage method providing *feedback* on this difference, based upon a combination of statistical learning (regression) and computational topology methods.

Consider two populations, each given as a point cloud in $\mathbb{R}^d$. In the first step, we assign a label to each set and we compute, for each sample point, a discrepancy measure based on comparing an estimate of the conditional probability distribution of the label given a position versus the global unconditional label distribution. In the second step, we study the height function defined at each point by the aforementioned estimated discrepancy. Topological persistence is used to identify persistent local minima of this height function, their *basins* defining regions of points with high discrepancy and in spatial proximity.

Experiments are reported both on synthetic and real data (satellite images and handwritten digit images), ranging in dimension from $d = 2$ to $d = 784$, illustrating the ability of our method to localize discrepancies.

On a general perspective, the ability to provide feedback downstream TST may prove of ubiquitous interest in exploratory statistics and data science.

6.4.2. A Sequential Non-parametric Two-Sample Test

Participants: Frédéric Cazals, Alix Lhéritier.

Given samples from two distributions, a nonparametric two-sample test aims at determining whether the two distributions are equal or not, based on a test statistic. This statistic may be computed on the whole dataset, or may be computed on a subset of the dataset by a function trained on its complement. We propose a third tier [19], consisting of functions exploiting a sequential framework to learn the differences while incrementally processing the data. Sequential processing naturally allows optional stopping, which makes our test the first truly sequential nonparametric two-sample test.

We show that any sequential predictor can be turned into a sequential two-sample test for which a valid p-value can be computed, yielding controlled type I error. We also show that pointwise universal predictors yield consistent tests, which can be built with a nonparametric regressor based on k-nearest neighbors in particular. We also show that mixtures and switch distributions can be used to increase power, while keeping consistency.

ACUMES Team

6. New Results

6.1. Mathematical analysis and control of macroscopic traffic flow models

6.1.1. Vehicular traffic

Participants: Guillaume Costeseque, Paola Goatin, Christophe Chalons [UVST], Simone Göttlich [U Mannheim, Germany], Jerôme Härri [EURECOM], Oliver Kolb [U Mannheim, Germany], Sosina Mengistu-Gashaw [EURECOM], Francesco Rossi [U Aix-Marseille], Stefano Villa [U Milano-Bicocca].

In collaboration with the University of Mannheim and in the framework of the PHC Procope project "Transport Networks Modeling and Analysis", we studied how to manage variable speed limits combined with coordinated ramp metering within the framework of the LWR network model. Following a "first discretize then optimize" approach, we derived the first order optimality system and explained the switch of speeds at certain fixed points in time and the boundary control for the ramp metering as well. Sequential quadratic programming methods are used to solve the control problem numerically. For application purposes, we present experimental setups where variable speed limits are used as a traffic guidance system to avoid traffic jams on highway interchanges and on-ramps, see [35].

The thesis of S. Mengistu-Gashaw, funded by the Labex UCN@Sophia (http://ucnlab.eu/) and co-supervised by P. Goatin and J. Härri, is devoted to understanding and modeling mobility characteristics of scooters and motorcycles for user-centric ITS application. We are currently developing a macroscopic model for heterogeneous traffic including car and motorcycles.

A new traffic flow model has been designed in [44] for taking into account the multiclass and multilane features of real traffic. This model is based on a system of coupled Hamilton-Jacobi PDEs for an appropriate choice of framework that mixes spatial and Lagrangian coordinates. The coupling conditions emerge from the moving bottleneck theory that has been developed in the traffic flow literature several years ago but for which a real mathematical sound basis lacked. Very recently, there were some new results dealing with the existence of a solution under suitable assumptions. However, these results were set for the hyperbolic conservation law in Eulerian coordinates and they are not straightforward to extend to Hamilton-Jacobi equations in different coordinates. Despite that the well-posedness of the problem is still an open problem, a numerical method is developed and it takes advantage of the classical representation formula available for HJ PDEs. This numerical scheme has been proved to provide good qualitative results.

In collaboration with F. Rossi, we proved existence and uniqueness of solutions to a transport equation modelling vehicular traffic in which the velocity field depends non-locally on the downstream traffic density via a discontinuous anisotropic kernel. The result is obtained recasting the problem in the space of probability measures equipped with the ∞-Wasserstein distance. We also show convergence of solutions of a finite dimensional system, which provide a particle method to approximate the solutions to the original problem. See, [45].

Finally, the internship of S. Villa, co-supervised by M. Garavello (U Milano-Bicocca), was devoted to the analytical and numerical study of the Aw-Rascle-Zhang model with moving bottleneck. Two Riemann Solver have been proposed, and two numerical strategies have been developed. A journal article is in preparation in collaboration with C. Chalons.

6.1.2. Crowd motion

Participants: Paola Goatin, Matthias Mimault.

M. Mimault defended hid PhD on December 14th, 2015. The last part of his thesis was devoted to the numerical study of scalar conservation laws with non-local flow in two space dimensions. These equations are meant to model crowd motion, where the movement direction of each pedestrian depends on a weighted mean of the crowd density around him. In particular, he implemented a finite volume numerical scheme which has been used for flow optimization purposes: he applied the adjoint method to compute the gradient for the evacuation time minimization depending on the initial crowd distribution.

6.2. Characterization of model uncertainty for turbulent flows

Participants: Régis Duvigneau, Jérémie Labroquère, Emmanuel Guilmineau [CNRS ECN, Nantes], Marianna Brazza [CNRS IMFT, Toulouse], Mathieu Szubert [CNRS IMFT, Toulouse].

The uncertainty related to turbulence modeling is still a bottleneck in realistic flows simulation. Therefore, some studies have been conducted to quantify this uncertainty for two problems in which turbulence plays a critical role. Firstly, the impact of the model choice has been estimated in the case of a massively detached flow over a 2D backward facing step including an oscillatory active control device, whose parameters are optimized [41], [34]. Secondly, the influence of the transition point location has been investigated, in the case of the 3D flow around a bluff-body, using models ranging from RANS to DES models [40].

6.3. Sensitivity analysis for unsteady flows

Participants: Régis Duvigneau, Dominique Pelletier [Ecole Polytechnique Montreal], Alexander Hay [Ecole Polytechnique Montreal].

Although sensitivity analysis is now commonly used for steady systems, usually on the basis of the adjoint equation method, the application to unsteady problems is still tedious, due to the backward time integration required. Therefore, an alternative approach, namely the sensitivity equation method, has been studied in the framework of the compressible Navier-Stokes equations. A continuous version has preferred to the discrete one for its flexibility and easier implementation. The proposed approach has been verified on several problems of increasing difficulty and the computational efficiency quantified [42].

6.4. Optimization accounting for experimental and numerical uncertainties

Participants: Régis Duvigneau, Olivier Le Maitre [CNRS LIMSI, Orsay], Matthieu Sacher [Ecole Navale, Brest], Alban Leroyer [ECN, Nantes], Patrick Queutey [CNRS ECN, Nantes].

Optimization of real-life applications requires to account for the uncertainties arising during the performance evaluation procedure, that could be either experimental or numerical. A Gaussian-Process based optimization algorithm has been proposed to efficiently determine the global optimum in presence of noise, whose amplitude can be user-defined or inferred from observations. The method has been applied to two very different problems related to performance optimization in sport.

The first case corresponds to the optimization of the shape of a racing kayak. The performance is estimated by coupling Newton's law with Navier-Stokes equations to compute the kayak velocity from the effort of the athlete, considered as input. The proposed method has been used here to filter the noise arising from the numerical simulation.

The second case corresponds to the optimization of a sail trimming, whose performance can be estimated either experimentally in a wind tunnel, or numerically by solving a fluid-structure interaction problem. In the former case, uncertainty has been estimated according to measurements accuracy, while in the latter case the numerical noise has be inferred from a set of observations collected during the optimization.

6.5. High-order numerical schemes for convection-dominated problems

Participants: Régis Duvigneau, Asma Gdhami [ENIT, Tunisia].

The use of high-order numerical schemes is necessary to reduce numerical diffusion in simulations, maintain a reasonable computational time for 3D problems, estimate accurately uncertainties or sensitivities, etc. Consequently, we work to develop high-order numerical schemes for the applications targeted by the team, in particular for convection-dominated problems. More precisely, we intend to include in a unified framework, based on Discontinuous Galerkin approximations, numerical methods accounting for complex geometries (isogeometric methods), uncertainty propagation (high-dimensional cubature) and sensitivity analysis.

6.6. Validation of time dependent diffusion approaches for activated and inhibited cell sheet closure

Participants: Abderrahmane Habbal, Hélène Barelli [Univ. Nice Sophia Antipolis, CNRS, IPMC], Grégoire Malandain [Inria, EPI Morpheme], Boutheina Yahyaoui [PhD, LAMSIN, Univ. Tunis Al Manar], Mekki Ayadi [LAMSIN, Univ. Tunis Al Manar].

We have studied in [21] five MDCK cell monolayer assays in a reference, activated and inhibited migration conditions. Modulo the inherent variability of biological assays, we have shown that in the assay where migration was not exogeneously activated or inhibited, the wound velocity was constant and the Fisher-KPP equation was able to accurately predict, until the final closure of the wound, the evolution of the wound area, the mean velocity of the cell front, and the time at which the closure occurred. When activated or inhibited, the F-KPP equation with constant parameters was unable to reproduce the observed biological cell sheet behavior. We modify the original equation, making the diffusion and proliferation parameters time dependent, following a sigmoid profile. We then set up an optimization loop to identify the sigmoids parameters, by computing a classical error indicator (difference between a computed density and the observed one, obtained through image processing) as done in the cited reference. We then obtain results which convincingly show that our approach is efficient : in both cases, inhibited and activated, the time varying identified parameters allow us to accurately predict until the final closure the evolution of the wound area.

6.7. Game strategies for joint data completion and parameter identification

Participants: Abderrahmane Habbal, Rabeb Chamekh [PhD, LAMSIN, Univ. Tunis Al Manar], Moez Kallel [LAMSIN, Univ. Tunis Al Manar], Nejib Zemzemi [Inria Bordeaux, EPI CARMEN].

We have demonstrated in previous works [22], [23] that Nash game approaches are efficient to tackle ill-posedness for linear second order elliptic Cauchy problems. We next developed a mathematical formulation for the linear elasticity model. The reconstruction is based on data completion and material identification, making it a harsh ill posed inverse problem. Up to now, we have obtained successful results for the Lamé parameter recovery in linear elasticity, using the so-called Kohn and Vogelius functional. Simultaneous data completion and parameter identification is under investigation.

6.8. Revised definition of the Multiple Gradient Descent Algorithm (MGDA)

Participant: Jean-Antoine Désidéri.

The Multiple Gradient Descent Algorithm (MGDA) had been defined originally to identify a descent direction common to a set of gradient vectors. According to a completely general principle, the direction is opposite to the vector of minimum Euclidean norm in the convex hull of the gradients. The Euclidean norm is defined via a general scalar product in $\mathbf{R}^n$. From a theoretical viewpoint, the notion of Pareto-stationarity had been introduced and it was established that if a point is Pareto-optimal and if the objective functions are locally differentiable and convex, then the point is Pareto-stationary. From a computational viewpoint, the descent direction can be determined as the solution of a Quadratic-Programming (QP) formulation. However, when the gradients are linearly independent a direct construction via a Gram-Schmidt orthogonalization process was preferred. We have now generalized the orthogonalization process by the introduction of a hierarchical strategy in the ordering of the subfamily of gradients utilized to construct the orthogonal basis. This strategy aims at making the (multi-dimensional) cone associated with the convex hull of the subfamily as large as possible. As a result, in the case of linearly-dependent gradients, the orthogonalization process not only provides a basis of the spanned subspace, but the subfamily is selected such that its the convex hull is also very representative of a large cone, encompassing in the most favorable cases all the given gradients. By this change in the definition of the algorithm, we were able to reformulate the QP formulation, now stated in a suitable basis, in a way that is well-suited for the treatment of cases where the number of gradients exceeds, possibly vastly, the dimension of the vector space. This revision makes the algorithm much more general and robust [43].

6.9. Multi-point optimization of a time-periodic system of pulsating jets

Participants: Jean-Antoine Désidéri, Régis Duvigneau.

A multi-point optimization exercise governed by the time-dependent compressible Navier-Stokes equations has been solved based on the sensitivity analysis (see above). A system was considered consisting of three pulsating jets acting on a flat-plate boundary layer. As it is well-known, the flow mixing by the jets has the effect of reducing the drag, as this was confirmed by the simulation of the flow in the somewhat arbitrary initial setting of the jets. Then, positions and pulsation frequencies of the jets have been maintained fixed, while their amplitudes and phases, six parameters in total, have been optimized to minimize the drag force. The finite-volume simulation of the time-periodic flow provides the drag force as a function of time over a large number of timesteps (800 for an accurate description of a period). The sensitivity analysis simultaneously provides the derivatives of drag with respect to the six design parameters. These derivatives were averaged over 20 distinct time-intervals, thus yielding 20 averaged gradient vectors of dimension 6. The MGDA was then used to define a descent direction common to the 20 vectors, a descent step was applied to the design parameters, and the process was continued iteratively.

The experiment confirmed the possibility to reduce the drag force at all times of the period, and not only in the average. In contrast, using the average gradient to define the direction of search resulted in a more important reduction of the average drag but at the cost of an increase of drag in a critical portion of the time period. Hence our optimization algorithm is more versatile and powerful than one aiming at minimizing purely statistical functions obtained by time averaging. We also demonstrated the possibility to optimize over a subinterval of the time interval.

6.10. Quasi-Riemannian approach to constrained optimization

Participants: Didier Bailly [Research Engineer, ONERA Department of Applied Aerodynamics, Meudon], Gérald Carrier [Research Engineer, ONERA Department of Applied Aerodynamics, Meudon], Jean-Antoine Désidéri.

In differentiable optimization, the Broyden-Fletcher-Goldfarb-Shanno (BFGS) method is one of the most efficient methods for unconstrained problems. Besides function values, it only requires the specification of the gradient. An approximate Hessian is calculated by successive approximations as part of the iteration, using rank-1 correction matrices. As a result, the iteration has superlinear convergence : when minimizing a quadratic function in n variables, if the one-dimensional minimizations in the calculated directions of search are done exactly, the Hessian matrix approximation is exact after n iterations, and from this, the iteration identifies to Newton's iteration, and produces the exact local optimum in only one additional iteration ($n + 1$ in total).

However the BFGS method does extend to constrained problems very simply. Following Gabay [95] and other authors, Chunhong Qi *et al* [128] have proposed a “Riemannian” variant, RBFGS that indeed incorporates equality constraints in the formulation and actually demonstrates superior convergence rates for problems with a large number of variables. However these Riemannian formulations are non trivial to implement since they require procedures implementing non-trivial differential-geometry operators ('retraction' and 'metric transport') to be developed. In their paper, they assume a formal expression of the constraint to be known. But, in PDE-constrained optimization, many constraints are functional, and it is not clear how can the metric transport operator in particular can be defined.

We are investigating how can a quasi-Riemannian method can be defined based on the sole definition of evaluation procedures for the gradients. By condensing all the equality constraints in one, a purely-explicit approximate retraction operator has been defined that yields a point whose distance to the contraint surface is fourth-order at least. The associated transport operator is currently being examined formally. These techniques will be experimented in the context of constrained optimum-shape design in aerodynamics.

6.11. Unstructured mesh adaptation using an adjoint-based sensor

Participants: Sébastien Bourasseau [Doctoral student, ONERA Department of Computational Fluid Dynamics, Châtillon], Jacques Peter [Research Engineer, ONERA Department of Computational Fluid Dynamics, Châtillon], Jean-Antoine Désidéri.

Mesh adaptation is a powerful tool to obtain accurate aerodynamic simulations at limited cost. When the simulation is aimed at the accurate calculation of aerodynamic outputs (forces, moments) goal-oriented methods based on the adjoint vector of the output of interest are often advocated. The calculation of the total derivative dJ/dX of the aerodynamic function of interest, J, with respect to the volume mesh coordinates, X, has been extended to the case of an unstructured grid. The software developments have been validated for inviscid and laminar viscous flows, and implemented in the ONERA code (elsA). Then a local sensor θ based on dJ/dX was devised to identify areas where the location of the volume mesh nodes has a strong impact on the evaluation of the output J. The sensor has been shown to be adequate in different flow regimes (subsonic, transonic, supersonic), for internal (blade and nozzle) and external (airfoils, wings) aerodynamic configurations. The proposed method has been compared to a well-known goal-oriented method (Darmofal and Venditti, 2001) and to a feature-based method; it yields comparable results at lower cost in simple configurations. A publication is currently subject to minor revisions.

6.12. Multi-fidelity surrogate modeling with application to the optimization of nanophotonic devices

Participants: Cédric Durantin [Doctoral student, CEA LETI Grenoble], Alain Glière [Research Engineer, CEA LETI Grenoble], Jean-Antoine Désidéri.

Multiple models of a physical phenomenon are sometimes available with different levels of approximation, the high fidelity model being more demanding in terms of computational time than the coarse approximation. In this context, including information from the lower fidelity model to build a surrogate model is desirable. A new multi-fidelity metamodeling method, based on Radial Basis Function, the co-RBF, is proposed. The new method is compared with the classical co-kriging on two analytical benchmarks and a realistic validation test, namely the design of a miniaturized photoacoustic gas sensor. The co-RBF method brings better results on high dimensional problem and could be considered as an alternative to co-kriging for multi-fidelity metamodeling.

AOSTE Project-Team

7. New Results

7.1. CCSL as a Logical Clock Calculus Algebra: expressiveness and decidability results

Participants: Robert de Simone, Julien Deantoni, Frédéric Mallet, Qingguo Xu.

CCSL is a language dedicated to the expression of time constraints, based on so-called logical clocks. Its declarative nature is akin to the Lustre or (even closer to) the Signal language, but without values (to clock/event occurrences) and with both synchronous and asynchronous constraints. Solving a set of CCSL constraints amounts to the production of a feasible schedule of the system. While the TimeSquare tool may attempt to generate such a schedule trace by insightful simulation, it is not guaranteed to be complete in its search.So the issue of expressiveness and decidability was left open to this day.

Still, in previous years, we had established the CCSL constraints could be translated into parallel products (extended, transition-labelled) Büchi machines, but some of these machines had to contain integer shift counters, and were thus not fully FSMs. Our (misled) conjecture that CCSL had semilinear, Presburger-arithmetic power was defeated by a new translation expressing (unitary then general) Petri Nets and Vector Addition Systems into CCSL by encoding. The new conjecture that CCSL was then as powerful as Petri Nets was again defeated by a construction interpreting the features of *inhibitor arcs* in CCSL. As such inhibitor arcs extend the expressive power of Petri Nets to become universal (Turing-complete), CCSL enjoys the same universal property (which makes it unfortunately impossible to solve automatically in general).

Despite this negative result we could show that, under natural restrictions such as the assumption that "input" clocks have bounded jitter around a mean rate, and even if those bounds are not exactly known (but may be used as a parameter), then expressiveness remains in the semi-linear, Presburger-arithmetic range.

As a side-effect of this work we provided the translation of CCSL constraints into Büchi components by using a well-defined fragment of the Esterel syntax to express the Buchi automata.

Preliminary results are exposed in a research report. A much more ambitious article is in preparation.

As part of Professor Xu sabbatical in Aoste, we also considered the topic of machine-assisted proof of schedulability using theorem-provers (in our case PVS) [54]).

7.2. Industrial design flow for Embedded System Engineering

Participants: Julien Deantoni, Frédéric Mallet, Marie Agnes Peraldi Frati, Robert de Simone, Ales Mishchenko.

As part of the PIA LEOC Clarity collaborative project we attempt to instill some of our theoretical and methodological ideas into the framework of the (open-source, Polarsys Eclipse) Capella environment. This environment was initially developed inside Thales, under the name ARCADIA/Melody, as a modeling tool flow for System-Level Design in-the-large. As such, several aspects were not fully considered, specially those regarding safe sound simulation semantics at this level, or the role of states and modes in variability regarding both the software applicative and hardware architectural platform models. This research is in part motivated by concrete needs as expressed by end-users such as Airbus, Areva/EDF and Thales.
Results on methodological enhancements are described

7.3. Coordination of heterogeneous Models of Computation as Domain-Specific Languages

Participants: Matias Vara Larsen, Julien Deantoni, Frédéric Mallet.

In the context of the collaborative ANR GEMOC project (9.2.1.2 , we investigated the way the multiview approach generally promoted in Aoste could deal with analysis and simulation of systems specified using multiple heterogeneous languages. Coordinated use of heterogeneous domain specific languages (DSL) led to so-called globalization of modeling language. We wrote a chapter related to these concerns [50], as part of a book dedicated to the challenges of the field, gathering industrial and academic contributors.

This goal was achieved in two steps. First step consisted in specifying a language able to support appropriate information (*i.e.,* the one required for the coordination) in a *Language Behavioral Interface (LBI)*. Second step consisted in using the LBI to define coordination patterns from which the coordination of models can be automatically inferred. Design is supported by an heterogeneous simulation engine that has been developed and integrated in the Gemoc studio environment. Gemoc Studio, enhanced with our new research ideas, won the 9^{th} execution tool contest at ...

We also developed MoCCML (Model of Concurrency and Communication Modeling Language), an imperative extension of the CCSL language in the form of constraint automata [28]. MoCCML defines the concurrent and communication part of the semantics of a language, and is used by the LBI to exhibit internal causalities and synchronizations. Finally, we defined a protocol combining the concurrency aspects and the execution functions (*i.e.,* the rewriting rules) so as to be able to develop, in a modular way, the whole behavioral semantics of a language [30], [31].

Our work this on coordination of heterogeneous languages produced two major results. The first one is the development of BCOoL (Behavioral Coordination Operator Language [33]). BCOoL is a language dedicated to the specification of coordination patterns between heterogeneous languages. It comes with a tool chain allowing the generation of the coordination given a BCOoL operator and specific models. Our second result is the development of an heterogeneous execution engine, integrated to Gemoc studio, to run conjointly different models [44]. Both works were mainly realized by Matias Vara Larsen, as part of his upcoming PhD.

7.4. SoC multiview (meta)modeling for performance, power, and thermal aspects

Participants: Amani Khecharem, Robert de Simone, Emilien Kofman, Julien Deantoni.

In the framework of the ANR HOPE project we progressed the definition of multiview metamodels for the design of Systems-on-Chip) (SoC systems integrating performance, power and thermal aspects. The main concern was to stress regularity and commonality between those views, each developed on "domains" defined as partitions of the original block diagram (clock domains, voltage domains, floorplans,...), and with finite state machine controlers setting the levels of these domains; links between distinct views are originally provided by laws of physics, but then usually identified on discrete allowable values by engineers. The application view, meant to provide typical use-cases to help dimension the SoC platform by abstract simulation, also fits in this framework. This methodological work was presented in the local forum SAME (Sophia-Antipolis MicroElectronices) [53]. It is supposed to work in two ways, both by allowing the appplication of analytic methods to compute an optimized mapping of application tasks onto platform resources, and then to translate these results towards sophisticated simulation environments (such as MCO Platform Architect by Synopsys or ACEplorer by Docea Power/Intel, both partners in the HOPE consortium) which consider non-functional aspects of power and thermal modeling in their simulation environments. The various approaches considered in Aoste to define mapping constraints and solve them algorithmically are presented elsewhere. All this should sonn be rported in Ameni Khacharem PhD document.

7.5. Networks-on-Board: between NoCs and rack connector buses

Participants: Amine Oueslati, Robert de Simone, Albert Savary, Emilien Kofman.

The recent paradigm of Massively Parallel Processor Arrays (MPPA), or more generally manycore Systems-on-Chip, rely on the existence of a high-throughput on-Chip Network (NoC) to interconnect the various cores and processing clusters. Despite its benefits, it requires that all components are put on the same dye, and thus designed monolithically. On the other end, supercomputers are built by assembling racks or blades of processors, connected by fast buses (fast ethernet or infinyband usually), with low predictivity of throughput. A third, intermediate path is explored in the context of the FUI Clistine project, based on a notion of Network-on-Board (or Network-in-Package), aiming at the benefits of NoCs brought to the level of a single PCB board, where the various components can be assembled in a modular fashion. We consider the applicationof our previous expertise on modeling and analysis of NoC-based architecture, with their implications on the optimized mapping of dataflow models of applications onto such interconnects, to adapt them in this new context. The objective is to consider alternative network topologies, and to transate optimal mappings into the concrete network operations on a prototype implementation realized by SynergieCAD, the company heading the project. This topic reflects the PhD thesis of Amine Oueslati, and the engineering work of Albert Savary.

7.6. Solving AAA constraints analytically

Participants: Emilien Kofman, Dumitru Potop Butucaru, Thomas Carle, Raul Gorcitz, Robert de Simone, Mohamed Bergach, Amine Oueslati.

Given two abstract modeling descriptions, one of a dataflow process network for the application, one of a block diagram structure for the computing platform and its interconnects, together with cost functions for the elementary compuations and communications, one is bound to seek optimal mappings pairing the two. Amongst all the possible techniques, an obvious one consists in solving constraint using general solvers (real, integer, or boolean constraint programming, SMT solvers, etc). Given the NP-hard nature of the problem, the issue here is to scale to the dimensions of realistic problems. We conducted extensive experiments on several case studies, with as extra objective the concern of studying how the formulation of constraints, or the exploiation of additional information (in concurency or exclusion of tasks, structural symmetries,...) could impact favorably or negatively the process. Results were compiled in a publication [57].

In the framework of the PhD thesis of Mohamed Bergach, under CIFRE funding with Kontron Toulon, we studied how to adjust a radar application, that typically computes extensively FFT convolutions, on an hybrid CPU/GPU architecture such as IntelCore IvyBridge and Haswell processors. This approach works in two stages: first we considered how to implement a FFT redex as large as possible in exactly one core (either a CPU core or a GPU Execution Unit), so as to make full use of the local register memories and SIMD/vectorial instructions. Not by accident certainly FFT blocks of size exactly 8 and 16 respectively can so be fitted on a GPU (resp. CPU) block. This provides a new "compound" instruction, on which to build modularly and optimization the allocation of larger aplications based on such basic block. This is fully described in Mohamed Bergach PhD document [16].

7.7. Stochastic extension of MARTE/CCSL for CPS modeling

Participant: Frédéric Mallet.

This work was conducted during the sabbatical period of Frédéric Mallet at ECNU Shanghai, in the context of the associated team FM4CPS (9.4.1.1).

As a declarative language, CCSL allows the specification of causal and temporal properties of systems expressed as constraints in a specific syntax. While each constraint reduces the set of possible behaviors, there may still be multiple (schedule) solutions, or none at all. When several solutions remain feasible, our TIMESQUARE tool allows to set up a resolving policy, to choose whether we want to attempt exploring exhaustively all these solutions, or else narrow the solution space according to an auxiliary criterion.

The extension of CCSL with stochastic features and probabilistic information is meant to help provide such an additional criterion, while modeling temporal constraints on the environment which are not necessarily well-known or controllable, specially in the domain of Cyber-Physical Systems. Then, such features should help reducing the set of possible behaviors, narrowing for instance to the most likely ones (in a formal quantitative meaning).

We are currently relying on UPPAAL SMC (Stochastic Model-Checking) toolset as prototype analyzer for the resulting specifications.

7.8. Coupling SystemC and FMI for co-simulation of Cyber-Physical Systems

Participants: Stefano Centomo, Julien Deantoni, Robert de Simone.

In the context of Stefano Centoma master internship, and in collaboration with his global supervisor Professor Davide Quaglia, from the University of Verona, we considered the possibility to build heterogenous, multi-physics co-simulation schemes for hybrid continuous-discrete Cyber-Physical systems. The first step consisted in extracting relevant interface information from IP component described in the SystemC language; it was naturally inspired from some of our former work. But currently IP-XACT is meant to address easy component assembly at the *structural* (static) level, and is not concerned with dynamical aspects of behavior simulation. This extension, and the proper combination with the FMI standard for its purpose, allowing hybrid and multiform co-simulation of SystemC components (and also others describing the continuous physical environment) are the next-step objective being currently tacked.

7.9. Code generation for time-triggered platforms based on Real-Time Scheduling

Participants: Dumitru Potop Butucaru, Raul Gorcitz, Yves Sorel.

We have continued this year the work on real-time scheduling and code generation for time-triggered platforms. Much of this work was carried out as part of a trilateral collaboration with Airbus DS and the CNES, which have funded an (onerous) TTEthernet-based test platform and partly funded the post-doctorate of Raul Gorcitz. The remainder of Raul Gorcitz' post-doc has been funded by the ITEA3 Assume project.

This year, the objective has been to allow code generation on an industry-grade platform comprising ARINC 653-based computers connected through a TTEthernet network. The novelty with respect to previous years comes from the time-triggered TTEthernet network, whose scheduling properties raise new problems. Unlike in classical field buses, resource reservation in a TTEthernet network is done at the level of directed links (physical wires that connect routers and end stations). Each of these links is controlled by an arbiter that determines the scheduling of both time-triggered data transfers and control messages needed to ensure the global time synchronization. This year we have built a model of the TTEthernet network allowing precise real-time scheduling, and worked on code generation aspects. We expect to have a fully running prototype in the next 2 months, and to demonstrate it to our funders. Relevant publications are [18], [38].

For teaching purposes and to achieve a finer understanding of ARINC 653-based operating systems, we have also developed an implementation of the standard on inexpensive RaspberryPi platforms, and published a scientific vulgarization paper [55].

7.10. Real-time systems compilation

Participants: Dumitru Potop Butucaru, Keryan Didier, Mihail Asavoae.

This research line develops over various results of the team over the years, its aim being to develop fully automatic implementation flows going fully automatically from functional and non-functional specification to correct and efficient running implementation. We advocate for a real-time systems compilation approach that combines aspects of both real-time scheduling and compilation of both classical and synchronous languages. Like a classical compiler such as GCC, a real-time systems compiler should use fast and efficient scheduling and code generation heuristics, to ensure scalability. Similarly, it should provide traceability support under the form of informative error messages enabling an incremental trial-and-error design style, much like that of classical application software. This is more difficult than in a classical compiler, given the complexity of the transformation flow (creation of tasks, allocation, scheduling, synthesis of communication and synchronization code, etc.), and requires a full formal integration along the whole flow, including the crucial issue of correct hardware abstraction. A real-time systems compiler should perform precise, conservative timing accounting along the whole scheduling and code generation flow, allowing it to produce safe and tight real-time guarantees. More generally, and unlike in classical compilers, the allocation and scheduling algorithms must take into account a variety of non-functional requirements, such as real-time constraints, criticality, partitioning, preemptability, allocation constraints, etc. As the accent is put on the respect of requirements (as opposed to optimization of a metric, like in classical compilation), resulting scheduling problems are quite different.

We are currently building such a real-time systems compiler, called Lopht. The construction of the Lopht tool, which takes into account complex functional and non-functional specifications is discussed in the corresponding section and in [17].

This year, we have initiated work on two fundamental topics. The first one is sound architecture abstraction – ensuring that the platform models used for real-time scheduling and code generation are conservative abstractions of the real hardware and basic software, allowing the generation of implementations that are functionally and non-functionally correct. This work is performed in the framework of the LEOC Capacites project, which funds the post-doc of Mihail Asavoae. The second line of work aims at formally proving that the output of Lopht is correct with respect to its input models (including functional specification and platform model). This work is performed in the ITEA3 Assume project, which funds the PhD thesis of Keryan Didier. Together with the Parkas team-project we have also considered the implementation of mixed-criticality systems [26].

7.11. Uniprocessor Real-Time Scheduling

Participants: Mamadou Diallo, Yves Sorel, Walid Talaboulma, Robert Davis.

In the context of the master internship of Mamadou Diallo we implemented the offline time trigger scheduler proposed in his PhD thesis by Falou Ndoye on a development board based on an ARM Cortex M4. We used this ARM version since it is better suited to embedded systems, since more predictable, than the ARM 7 we used last year. Especially, it allows to determine more accurately the cost of the scheduler and of the preemptions we use in our offline schedulability analysis. We remind that the schedulability analysis provides a scheduling table which is exploited by the scheduler during the real-time execution of the tasks. This approach allows a low and fixed cost for the scheduler and the preemptions whereas these costs are variable in the case of classical online schedulers. For several task sets we compared the timing diagrams predicted by the schedulability analysis with the real-time timing diagrams measured on the ARM Cortex M4. It turns out that those timings are very close, as expected.

A new direction opened with the arrival of Rob Davis was to consider by studying the impact of the non-preemptivity constraints on the optimality of the schedulers [37], or by considering fixed priorities while scheduling messages in the context of Control Area Networks [36].

7.12. Multiprocessor Real-Time Scheduling

Participants: Aderraouf Benyahia, Laurent George, Salah Eddine Saidi, Yves Sorel, Robert Davis, Liliana Cucu.

In the context of the PhD thesis of Salah Eddine Saidi we considered the co-simulation of several process models specified in continuous time and several controllers models specified in discrete time according to a real-time hardware in the loop approach. These models specified with different tools such as Simulink, AMEsim, Modelica, etc., cooperate according to the FMI standard. They are translated in a dataflow graph that is compliant with the conditioned repetitive dataflow model of our AAA methodology for functional specification. Each model considers the feed-through function as well as the functions which depend of the state, and the state computation itself. In order to meet the real-time constraints of such complex co-simulation we need to execute them on multicore platforms. We studied the limitations of greedy and local search distributed real-time scheduling heuristics we developped in the past for control applications. The first limitation is related to the FMI standard which requires that the functions belonging to a model are allocated to the same core. We first try to introduce additional semaphores in the real-time code generated automatically to avoid these situations. Unfortunately, this solution decreases significantly the acceleration brought by the multicore. Therefore, we started to investigate graph based techniques that add non directed edges to specify the FMI relation and search solutions where some non oriended edges can be oriented to minimize locally the makespan.

In the context of the master internship of Mamadou Diallo we studied the possibilities to extend the offline time trigger scheduler implemented on a uniprocessor to the multiprocessor case. Since the embedded board based on the ARM Cortex M4 we utilize features an ethernet interface, we conducted several experimentations on ethernet switches to measure the end-to-end communication time between several real-time tasks running on such boards with such schedulers.

We completed the work on the gateway with modeling languages for certified code generation carried out in the P FUI project 9.2.2 which ended in June 2015. Mainly, we tested the P modeling language to SynDEx gateway on four industrial use cases provided by AdaCore, Continental and Aboard Engeneering. We specified these applications with the P language and translated them in the SynDEx format. With SynDEx we analysed the schedulability and automatically generated the corresponding code for an Intel 8 cores Xeon ES-1620v2 3.70Ghz. For these applications ranging from 103 to 1403 bloks we obtained an acceleration factor equal to the number of cores.

Thanks in part to the arrival of Rob Davis, our team has participated to the proposition of a new framework in the context of multicore platforms: *Multicore Response Time Analysis framework* [34]. This proposal was made in close collaboration with academic partners such as the University of Luxembourg, Verimag and ISEP Porto. The framework is extensible to different multicore architectures, with various types and arrangements of local memory, and different arbitration policies for the common interconnects. The MRTA framework provides a general approach to timing verification for multicore systems, parametric in the hardware configuration, and so can be used architectural design stage to compare the guaranteed levels of performance that can be obtained with different hardware configurations. The MRTA framework decouples response time analysis from a reliance on context independent WCET values. Instead, the analysis specifies response times directly according to requirements on different hardware resources.

7.13. Probabilistic and statistical temporal analysis

Participants: Liliana Cucu, Robert Davis, Adriana Gogonel, Walid Talaboulma, Dorin Maxim, Cristian Maxim.

Real-time constraint guarantees require worst-case reasoning to provide sound solutions. We have proposed to define and use worst-case reasoning in different contexts: optimal scheduling algorithms, response time analysis, estimation of worst-case execution times. These results have laid the foundations for certifiable probabilistic solutions to real-time systems.

In particular, we have studied the probabilistic response time analysis for systems with multiple probabilistic parameters, either by using bounds based on real-time calculus, extreme value theory, direct calculation or in a context of component-based systems. Generally, probabilistic methods have high complexity cost; using upper-bounds for the input probability distributions we provide conservative(safe) results faster. Worst-case reasoning is also provided for the statical estimation of a task probabilistic worst-case execution time.

Results were published in [22], [24], [58], [56], [42], [46], [23], [42], [43], [40]

7.14. Parametric and Non-Parametric Statistics for Program Performance Analysis and Comparison

Participant: Sid Touati.

This research activity is a continuation of our joint research effort with Julien Worms, Assistant Professor at University of Versailles Saint-Quentin (UVSQ), dealing with statistical program performance analysis and comparison, in presence of performance variability. In the previous study (called Speedup-Test), we gave a rigorous statistical methodology for analysis of program speedups based on mean or median performance metrics: execution time, energy consumption, etc. However mean or median observed performances do not always reflect the user's feeling of performance, especially when the performances are really unstable. In the current study, we propose additional precise performance metrics, based on performance modeling using gaussian mixtures. We explore the difference between parametric and non parametric statistics applied on program performance analysis. Our additional statistical metrics for analysing and comparing program performances give to the user more precise decision tools to select best code versions, not necessarily based on mean or median numbers. Also, we provide a new metric to estimate performance variability based on gaussian mixture model. Our statistical methods are implemented in R, and distributed as open source code. A research report is under completion, before submission as article.

APICS Project-Team

6. New Results

6.1. Inverse problems for Poisson-Laplace equations

Participants: Laurent Baratchart, Sylvain Chevillard, Juliette Leblond, Konstantinos Mavreas, Christos Papageorgakis, Dmitry Ponomarev.

This section is concerned with inverse problems for 3-D Poisson-Laplace equations, among which source recovery issues. Though the geometrical settings differ in Sections 6.1.1 and 6.1.2 , the characterization of silent sources (those giving rise to a vanishing field) is one common problem to both which has been resolved in the magnetization setup [33].

6.1.1. Inverse problems in medical imaging

This work is conducted in collaboration with Jean-Paul Marmorat and Nicolas Schnitzler, together with Maureen Clerc and Théo Papadopoulo from the Athena EPI.

In 3-D, functional or clinical active regions in the cortex are often modeled by pointwise sources that have to be localized from measurements taken by electrodes on the scalp of an electrical potential satisfying a Laplace equation (EEG, electroencephalography). In the works [38][5] on the behavior of poles in best rational approximants of fixed degree to functions with branch points, it was shown how to proceed via best rational approximation on a sequence of 2-D disks cut along the inner sphere, for the case where there are finitely many sources (see Section 4.2).

In this connection, a dedicated software FindSources3D (see Section 3.4.2) is being developed, in collaboration with the team Athena and the CMA. We continued this year algorithmic developments, prompted by a fruitful collaboration with the firm BESA, namely automatic detection of the number of sources (which was left to the user until recently). It appears that, in the rational approximation step, *multiple* poles possess a nice behavior with respect to branched singularities. This is due to the very physical assumptions on the model (for EEG data, one should consider *triple* poles). Though numerically observed in [7], there is no mathematical justification so far why multiple poles generate such strong accumulation of the poles of the approximants. This intriguing property, however, is definitely helping source recovery. It is used in order to automatically estimate the “most plausible” number of sources (numerically: up to 3, at the moment). Further, a modular and ergonomic platform version of the software is under development.

In connection with these and other brain exploration modalities like electrical impedance tomography (EIT), we are now studying conductivity estimation problems. This is the topic of the PhD research work of C. Papageorgakis (co-advised with the Athena project-team and BESA GmbH). In layered models, it concerns the estimation of the conductivity of the skull (intermediate layer). Indeed, the skull was assumed until now to have a given isotropic constant conductivity, whose value can differ from one individual to another. A preliminary issue in this direction is: can we uniquely recover and estimate a single-valued skull conductivity from one EEG recording? This has been established in the spherical setting when the sources are known, see [17]. Situations where sources are only partially known and the geometry is more realistic than a sphere are currently under study. When the sources are unknown, we should look for more data (additional clinical and/or functional EEG, EIT, ...) that could be incorporated in order to recover both the sources locations and the skull conductivity. Furthermore, while the skull essentially consists of hard bone part that may be assumed to have constant electrical conductivity, it also contains spongy bone compartments. These two distinct parts of the skull possess quite different conductivities. The influence of that second value on the overall model is now being studied [19].

6.1.2. Inverse magnetization issues in the thin-plate framework

This work is carried out in the framework of the “équipe associée Inria” IMPINGE, comprising Eduardo Andrade Lima and Benjamin Weiss from the Earth Sciences department at MIT (Boston, USA) and Douglas Hardin, Michael Northington and Edward Saff from the Mathematics department at Vanderbilt University (Nashville, USA).

Localizing magnetic sources from measurements of the magnetic field away from the support of the magnetization has been the fundamental issue under investigation by IMPINGE. The goal was to determine magnetic properties of rock samples (*e.g.* meteorites or stalactites) from fine field measurements close to the sample that can nowadays be obtained using SQUIDs (superconducting quantum interference devices). Currently, rock samples are cut into thin slabs and the magnetization distribution is considered to lie in a plane, which makes for a somewhat less indeterminate framework than EEG as regards inverse problems because “less” magnetizations can produce the same field (for the slab has no inner volume). Note however that EEG data consist of values of the normal current and of the associated potential, while in the present setting only values of the normal magnetic field are measured.

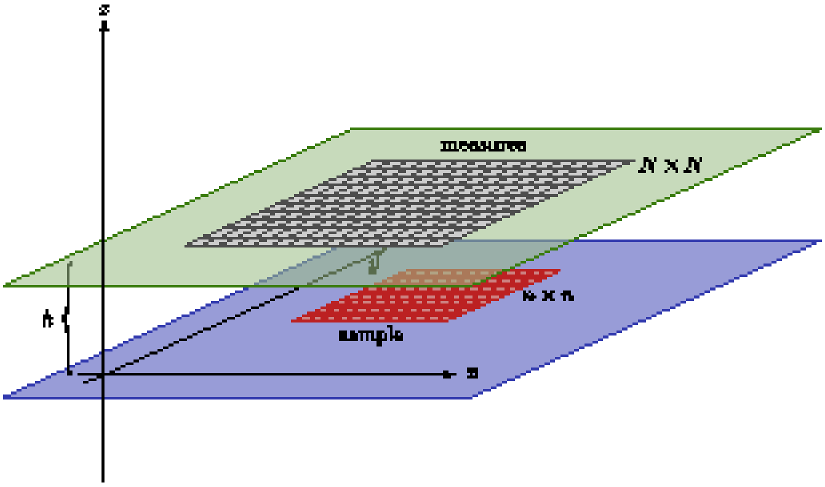


Figure 3. Schematic view of the experimental setup

Figure 3 presents a schematic view of the experimental setup: the sample lie on a horizontal plane at height 0 and its support is included in a rectangle. The vertical component B_3 of the field produced by the sample is measured on points of a horizontal $N \times N$ rectangular grid at height h.

Over the previous years, we mainly focused on developing techniques to recover magnetizations with rather sparse support. To this end, we set up a heuristic procedure to recover sparse magnetizations, based on iterative truncation of the support of the recovered magnetization. In this heuristics, magnetizations were represented by dipoles placed at the points of a regular rectangular $n \times n$, which seemed general enough a model class to correctly approximate the magnetizations commonly encountered in samples.

The procedure turned out to be poor when trying to recover the magnetization itself, due to the severe ill-posedness of the problem and the unexpected existence of magnetizations that produce almost no field at the height where measurements are performed, although the corresponding magnetic distributions strongly differ from truly silent distributions. Nevertheless, whenever the support could be significantly shrunk while keeping the error small (*i.e.*, explaining the data satisfactorily), estimates of the net moment so far, based on the dipolar model obtained by inversion, have been good.

This suggests that recovering the net moment and recovering the magnetization are rather different problems, the first one being less ill-posed than the second. Although the information provided by the net moment of the sample seems to be much weaker than knowing the full magnetic distribution, its importance has been emphasized by the geophysicists at MIT for at least three reasons:

- It yields important geological information on the sample in particular to estimate the magnitude of the ambient magnetic field at the time the rock was formed.
- It can be estimated independently to some extent, using a magnetometer, thereby allowing one to cross-validate the approach.
- From a computation point of view, knowledge of the net moment should lead to numerically stable reconstruction of an equivalent unidirectional magnetization. The support of the latter would provide us with valuable information to test for unidirectionality of the true magnetization, which is an important question to physicists in connection with rocks history and formation.

This year, we addressed the problem of directly recovering the net moment, without recourse to full inversion. Indeed, the latter is rather inefficient as it requires using a cluster and even then, for some samples, days of evaluation in order to obtain only a coarse estimate of the net moment. This research effort led us to investigate three different and complementary approaches.

First, we improved over Fourier based techniques previously designed by reformulating the problem with the help of the Kelvin transform. This gave us an asymptotic expansion of the net moment involving, at the first order, the integrals $\iint B_3(x,y,h)\,\mathrm{d}x\mathrm{d}y$, $\iint x\,B_3(x,y,h)\,\mathrm{d}x\mathrm{d}y$ and $\iint y\,B_3(x,y,h)\,\mathrm{d}x\mathrm{d}y$, computed on a disc with large radius. Although the method is promising, the computations are quite involved and we did not manage yet to obtain higher-order terms. This is a part of D. Ponomarev PhD work.

In parallel, and based on the results obtained with Fourier transform, we investigated a second approach, consisting in directly computing asymptotic expansions of the above integrals, on several domains (namely, the 2-D balls of radius R for the 1, 2 and ∞ norm, that are squares, disks, diamonds). In all cases, we get

$$\iint x\,B_3(x,y,h)\,\mathrm{d}x\mathrm{d}y = \alpha\langle m_1\rangle + \beta(\langle t_1\,m_3\rangle - h\langle m_1\rangle)/R + \mathcal{O}(1/R^3),$$

where $\langle m_1\rangle$ is the moment of the first component m_1 of the magnetization and $\langle t_1\,m_3\rangle$ is the first moment of m_3 with respect to the first variable. The constants α and β depend on the domain where the integral is computed. Therefore, an appropriate linear combination of the integrals computed on the different domains allows us to compute $\langle m_1\rangle$ with an accuracy of $\mathcal{O}(1/R^3)$. Similar results are obtained for $\langle m_2\rangle$ and $\langle m_3\rangle$ with the other integrals. Preliminary numerical experiments confirm the practical usability of these formulas in order to recover the moment of magnetizations. A research report is currently being written to sum up these results.

Finally, a third more ambitious approach has been investigated. As an attempt to generalize the previous expansions, our initial question was: given measurement of B_3, say on a square, find a function $\phi(x,y)$ such that $\iint \phi(x,y)B_3(x,y)\,\mathrm{d}x\mathrm{d}y$ is a the best possible estimate of the net moment components $\langle m_i\rangle$ $(i=1,2,3)$. This problem does not admit a solution because, for any $\epsilon > 0$, there exists a function ϕ_ϵ allowing to estimate the moment with an error bounded by ϵ. However, when ϵ tends to zero, the function ϕ_ϵ is expected to have strong oscillations, which hinders an accurate computation of $\iint \phi(x,y)B_3(x,y)\,\mathrm{d}x\mathrm{d}y$ since B_3 is only known on a discrete grid of points. We therefore expressed the problem as a bounded extremal problem (see Section 3.3.1): to find the best ϕ_ϵ (with the smallest possible error value ϵ) under the constraint that $\|\nabla\phi_\epsilon\|_2 \le M$. Here, M is a user-defined parameter. We proved theoretical results regarding this bounded extremal problem (existence and uniqueness of a solution, characterization of its solution as a solution of integro-differential equation) and we are currently designing a numerical procedure to compute it. An article on this topic is in preparation.

Still in the course of D. Ponomarev's PhD research, the study of a 2D spectral problem for the truncated Poisson operator in planar geometry has been pursued. It is a simplified formulation of the relation between the magnetization and the magnetic potential (of which the magnetic field is the gradient) and is expected to produce an efficient representation basis (the eigenfunctions of the magnetization-to-field operator). This is a long-standing problem. Noteworthy properties of solutions have been obtained through connections with other spectral problems and asymptotic reductions for large and small values of the main parameters (distance h from the measurement plane to the sample support and sample support size), yielding approximate solutions by means simpler integral equations and ODEs.

The year 2015 was the last of our "équipe associée" IMPINGE with the MIT and Vanderbilt University. The final report is available on the web page of the associate team [0]. This collaboration is currently supported in part by a MIT-France seed funding from the US side, and we applied for a three-years extension of the associate team.

6.1.3. Inverse magnetization issues from sparse spherical data

The team APICS is a partner of the ANR project MagLune concerning Lunar magnetism, associated to the Geophysics and Planetology Department of Cerege, CNRS, Aix-en-Provence (see Section 8.2.2). Measurements of the remanent magnetic field of the Moon let geoscientists think that the Moon used to have a magnetic dynamo for some time, but the exact process that triggered and fed this dynamo is not yet understood, much less why it stopped. In particular, the Moon is too small to have a convecting dynamo like the Earth has. The overall goal of the project is to devise models to explain how this dynamo phenomenon was possible on the Moon.

To this end, the geophysicists from Cerege will go to NASA to perform some measurements on samples brought back from the Moon by Apollo missions. The samples are kept inside bags with a protective atmosphere, and geophysicists are not allowed to open the bags, nor to take out the samples from NASA facilities. Therefore, measurements must be performed with some rudimentary instrument and our colleagues from Cerege designed a specific magnetometer. This device allows them to obtain measurements of the components of the magnetic field produced by the sample, at some discrete set of points located on disks belonging to three cylinders (see Figure 4).

This collaboration started this year and some preparatory work was necessary fix conventions used by our colleagues from Cerege in order to handle their measurements. During his Master 2 internship, Konstantinos Mavreas has developed a method based on rational approximation, using the same ideas as those underlying the FindSources3D tool (see Sections 3.4.2 and 6.1.1), for the case where the field produced by the sample can be well explained by a single magnetic dipole, whose position and moment are unknown. See his report [0]. Konstantinos Mavreas is now engaged in a PhD within APICS and will extend these results to the case of several dipoles.

6.2. Matching problems and their applications

Participants: Laurent Baratchart, Martine Olivi, David Martinez Martinez, Fabien Seyfert.

This is collaborative work with Stéphane Bila (Xlim, Limoges, France), Yohann Sence (Xlim, Limoges, France), Thierry Monediere (Xlim, Limoges, France), Francois Torrès (Xlim, Limoges, France).

Filter synthesis is usually performed under the hypothesis that both ports of the filter are loaded on a constant resistive load (usually 50 Ohm). In complex systems, filters are however cascaded with other devices, and end up being loaded, at least at one port, on a non purely resistive frequency varying load. This is for example the case when synthesizing a multiplexer: each filter is here loaded at one of its ports on a common junction. Thus, the load varies with frequency by construction, and is not purely resistive either. Likewise, in an emitter-receiver, the antenna is followed by a filter. Whereas the antenna can usually be regarded as a resistive load at some frequencies, this is far from being true on the whole pass-band. A mismatch between the antenna and the

---

[0] http://www-sop.inria.fr/apics/IMPINGE/

[0] http://www-sop.inria.fr/members/Konstantinos.Mavreas/main.pdf

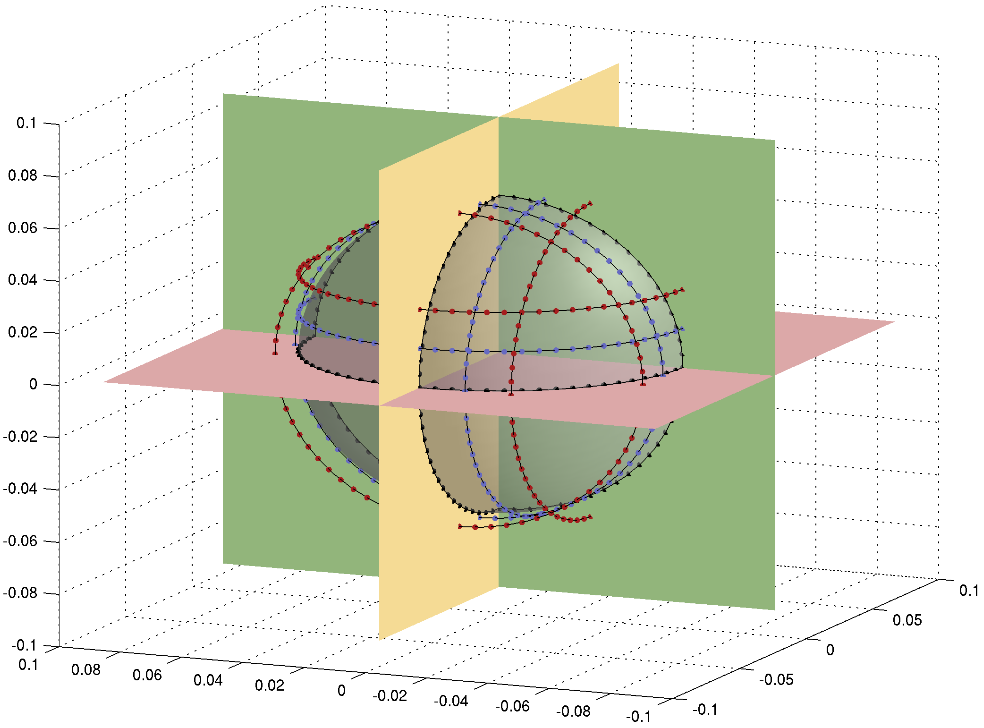


Figure 4. Typical measurements obtained with the instrument of Cerege. Discrete measurements of the field are performed on three cylinders. On each cylinder, the magnetic field $\mathbf{B}$ is expressed as a component B_h co-linear with the axis of the cylinder, and a component $\mathbf{B_s}$ parallel to a section of the cylinder. $\mathbf{B_s}$ is itself decomposed as a tangential component B_τ and a normal component B_n, with respect to the circle given by the intersection of the cylinder with the corresponding section. At black points B_n is measured, at blue points B_h is measured, and at red points B_τ is measured.

filter, however, causes irremediable power losses, both in emission and transmission. Our goal is therefore to develop a method for filter synthesis that allows us to match varying loads on specific frequency bands, while enforcing some rejection properties away from the pass-band.

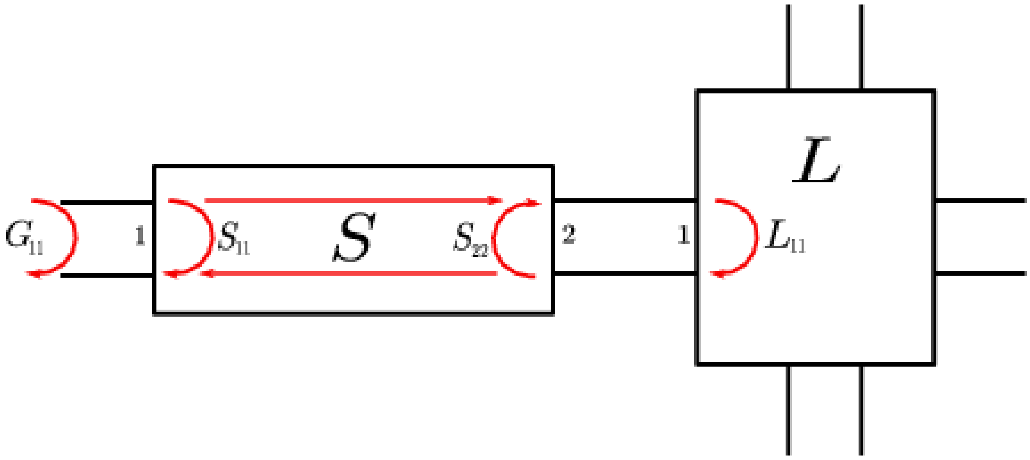


Figure 5. Filter plugged on a system with reflexion coefficient L_{11}

Figure 5 shows a filter with scattering matrix S, plugged at its right port on a frequency varying load with reflexion parameter $L_{1,1}$. If the filter is lossless, simple algebraic manipulations show that on the frequency axis the reflexion parameter satisfies:

$$|G_{1,1}| = \left| \frac{S_{2,2} - \overline{L_{1,1}}}{1 - S_{2,2} L_{1,1}} \right| = \delta(G_{1,1}, S_{2,2}).$$

The matching problem of minimizing $|G_{1,1}|$ amounts therefore to minimize the pseudo-hyperbolic distance δ between the filter's reflexion parameter $S_{2,2}$ and the load's reflexion $L_{1,1}$, on a given frequency band. On the contrary enforcing a rejection level on a stop band, amounts to maintaining the value of $\delta(L_{1,1}, S_{2,2})$ above a certain threshold on this frequency band. For a broad class of filters, namely those that can be modeled by a circuit of n coupled resonators, the scattering matrix S is a rational function of McMillan degree n in the frequency variable. The matching problem thus appears to be a rational approximation problem in the hyperbolic metric.

6.2.1. Approach based on interpolation

When the degree n of the rational function $S_{2,2}$ is fixed, the hyperbolic minimization problem is non-convex and led us to seek methods to derive good initial guesses for classical descent algorithms. To this effect, if $S_{2,2} = p/q$ where p, q are polynomials, we considered the following interpolation problem $\mathcal{P}$: given n frequency points $w_1 \cdots w_n$ and a transmission polynomial r, to find a monic polynomial p of degree n such that:

$$j = 1..n, \qquad \frac{p}{q}(w_j) = \overline{L_{1,1}(w_j)}$$

where q is the unique monic Hurwitz polynomial of degree n satisfying the Feldtkeller equation

$$qq^* = pp^* + rr^*,$$

which accounts for the losslessness of the filter. The frequencies (w_k) are perfect matching points, as $\delta(S_{2,2}(w_k), L_{1,1}(w_k)) = 0$ holds, while the real zeros (x_k) of r are perfect rejection points (i.e. $\delta(S_{2,2}(x_k), L_{1,1}(x_k)) = 1$). The interpolation problem is therefore a point-wise version of our original matching-rejection problem. The monic restriction on p and q ensures the realisability of the filter in terms of coupled resonating circuits. If a perfect phase shifter is added in front of the filter, realized for example with a transmission line on a narrow frequency band, these monic restrictions can be dropped and an interpolation point w_{n+1} added, thereby yielding another interpolation problem $\widehat{\mathcal{P}}$. Our main result, states that $\mathcal{P}$ as well as $\widehat{\mathcal{P}}$ admit a unique solution. Moreover the evaluation map defined by $\psi(p) = (p/q(x_1), \cdots, p/q(x_n))$ is a homeomorphism from monic polynomials of degree n onto $\mathbb{D}^n$ ($\mathbb{D}$ the complex open disk), and ψ^{-1} is a diffeomorphism on an open, connected, dense set of $\mathbb{D}^n$. This last property has shown crucial for the design of an effective computational procedure based on continuation techniques. Current implementation of the latter tackles instances of $\mathcal{P}$ or $\widehat{\mathcal{P}}$ for $n = 10$ in less than $0.1\ sec$, and allows for us a recursive use of this interpolation framework in multiplexer synthesis problems. We presented these techniques at the European Microwave Week 2015 in the workshop dedicated to "Recent Advances in the Synthesis of Microwave Filters and Multiplexers". The detailed mathematical proofs can be found in [21] and will be submitted shortly. On a related topic, namely the de-embedding of filters in multiplexers, our work has been published in [13].

6.2.2. Uniform matching and global optimality considerations

The previous interpolation procedure provides us with a matching/rejecting filtering characteristics at a discrete set of frequencies. This can serve as a starting point for heavier optimization procedures where the matching and rejection specifications are expressed uniformly over the bandwidth. Although the practical results thus obtained have shown to be quite convincing, we have no proof of their global optimality. This led us to seek alternative approaches able to asses, at least in simple cases, global optimality of the derived response. Following the approach of Fano and Youla, we considered the problem of a designing a 2×2 lossless frequency response, under the condition that a specified load can be "unchained" from one of its port. This classically amounts to set interpolation conditions on the response at the transmission zeros of the Darlington extension of the load. When the load admits a rational representation of degree 1, and if the transmission zeros of the overall system are fixed, then we were able to show that the uniform matching problem over an interval reduces to a convex minimization problem with convex constraints over the set of non-negative polynomials of given degree. In this case, which is already of some practical interest for antenna matching (antenna usually exhibit a single resonance in their matching band which is reasonably approximated at order 1), it is therefore possible to perform filter synthesis with a guarantee on the global optimality of the obtained characteristics. Procedures to derive the solution are currently being investigated, and lie at the heart of our contribution to the ANR-project Cocoram.

6.3. Unambiguous de-embedding of filters

Participants: Matthias Caenepeel, Martine Olivi, Fabien Seyfert.

This work was conducted in collaboration with Yves Rolain (VUB, Brussels, Belgium)

Coupling topologies that admit multiple realizations may lead to ambiguous de-embedding tuning procedures where distinct coupled resonator circuits are identified from the same measurements. This is for example the case of the well-known coupling topologies in triplets, quadruplets and extended boxed. If no additional measurements are performed on the DUT (device under tuning), the different solutions to the coupling matrix synthesis problem are undistinguishable, as they yield similar scattering responses. We therefore studied specific tuning strategies to discriminate among them. The later uses a sequence of measurements of the DUT, obtained after varying some discriminating tuning parameters of the filter and testing for coherence of the extracted circuits. This work was presented by Matthias Caenepeel at IMS 2015 in Phoenix [15] and at the

EuMC 2015 in Paris [16]. In a similar vein Matthias is currently developing techniques taking advantage of the differential information provided by EM solvers in order to compute the Jacobian matrix of the identified coupling matrix(ces) with respect to the geometrical parameters of the filter.

6.4. Orthogonal Polynomials

Participant: Laurent Baratchart.

We studied this year the asymptotic behavior of the orthonormal polynomials P_n with respect to a non-negative weight w on a simply connected planar domain Ω:

$$\int_\Omega P_n \overline{P}_k w \, dm = \delta_{n,k},$$

with $\delta_{n,k}$ the Kronecker symbol. We proved that if Ω has boundary $\partial\Omega$ of class $C^{1,\alpha}$, $\alpha > 0$, and if w converges in some appropriate sense to a boundary function $w_1 \in L^p(\partial\Omega)$ while not vanishing "too much" at the boundary, then

$$P_n(z) = \left(\frac{n+1}{\pi}\right)^{1/2} z^n S^-_{w_1}(\Phi(z))\Phi^n(z)\Phi'(z)\{1+o(1)\}$$

outside the convex hull of Ω, with Φ the conformal map from the complement of Ω onto the complement of the unit disk normalized so that $\Phi'(\infty) = \infty$, and $S^-_{w_1}$ the so-called exterior Szegő function of w_1.

This generalizes considerably known asymptotics on analytic domains with Hölder smooth non vanishing weights [10]. The proof rests on some Hardy space theory, conformal mapping and $\overline{\partial}$ techniques. An exposition of the result was given at the conference *Orthogonal and Multiple Orthogonal Polynomials*, August 9-14 2015, Oaxaca (Mexico). An article is being written to report on this result.

6.5. Asymptotics of Rational Approximants

Participant: Laurent Baratchart.

This is joint work with M. Yattselev (IUPUI).

We studied best rational approximants in the *sup* norm to an analytic function f on compact set K of the analyticity domain Ω with connected complement. We showed that if the function can be continued analytically except over a set of logarithmic capacity zero comprising at most finitely many branchpoints, then the n-th root of the approximation error converges as n goes large to $e^{-2/C}$, with C the minimal Green capacity in $\mathbb{C} \setminus K$ of a compact set E outside of which f is single valued. Moreover, if $C > 0$, the normalized counting measure of the poles converges to the Green equilibrium distribution on E. We are currently considering the case of infinitely many branchpoints so as to get a somewhat final result on weak asymptotics in rational approximation to functions with polar singular set.

The proof rests on a blend of AAK-theory and potential theory.

ASCLEPIOS Project-Team

6. New Results

6.1. Medical Image Analysis

6.1.1. Longitudinal Analysis and Modeling of Brain Development

Participants: Mehdi Hadj-Hamou [Correspondent], Xavier Pennec, Nicholas Ayache, Hervé Lemaitre [Inserm U1000], Jean-Luc Martinot [Inserm U1000].

This work is partly funded through the ERC Advanced Grant MedYMA 2011-291080 (on Biophysical Modeling and Analysis of Dynamic Medical Images).

Processing pipeline - brain development - adolescence - longitudinal analysis - non-rigid registration algorithm - extrapolation

1. We proposed and detailed a deformation-based morphometry computational framework, called Longitudinal Log-Demons Framework (LLDF), which estimates the longitudinal brain deformations from image data series, transports them in a common space and performs statistical group-wise analyses (see Fig. 1). This processing pipeline is based on freely available softwares and relies on the LCC log-Demons non-linear diffeomorphic registration algorithm with an additional modulation of the similarity term using a confidence mask to increase robustness with respect to brain boundary intensity artifacts.
2. The LLDF framework is applied to the study of longitudinal trajectories during adolescence, for which little is known. The aim of this project is to provide models of brain development during adolescence based on diffeomorphic registration parametrised by SVFs. Our study focused particularly on the link between sexual dimorphism and the longitudinal evolution of the brain. This work was done in collaboration with J.L. Martinot et H. Lemaître (Inserm U1000).

6.1.2. Inter-Operative Relocalization in Flexible Endoscopy

Participants: Anant Suraj Vemuri [Correspondent], Stéphane Nicolau, Luc Soler, Nicholas Ayache.

This work has been performed in collaboration with IHU Strasbourg and IRCAD, France.

Computer Assisted Intervention, Barrett's Esophagus, Biopsy Relocalization, Electromagnetic tracking

Oesophageal adenocarcinoma arises from Barrett's oesophagus, which is the most serious complication of gastro-oesophageal reflux disease. Strategies for screening involve periodic surveillance and tissue biopsies. A major challenge in such regular examinations is to record and track the disease evolution and relocalization of biopsied sites to provide targeted treatments.

In an earlier paper, we introduced the first approach to inter-operative relocalization using electromagnetic tracking system. In [21], we propose three incremental experiments to our approach. First, we analyse the error bounds of our system on synthetic data with a realistic noise model. Second, we provide a pseudo ground-truth on *in-vivo* pig data using an optical tracking system. Accuracy results obtained were consistent with the synthetic experiments despite uncertainty introduced due to breathing motion, and remain inside acceptable error margins according to medical experts. Finally, a third experiment was designed using data from pigs to simulate a real task of biopsy site relocalization, and evaluated by ten experts. It clearly demonstrated the benefit of our system towards assisted guidance by improving the biopsy site retrieval rate from 47.5% to 94%.

This inter-operative relocalization framework was then further extended in [53] to provide a constrained image based search as shown in Fig. 2 to obtain the best view point match to the live view. Within this context, we investigate the effect of (a) the choice of feature descriptors and colour-space, (b) filtering of uninformative frames and (c) endoscopic modality, for view point localization. Our experiments indicate an improvement in the best view-point retrieval rate to [92%, 87%] from [73%, 76%] (in our previous approach) for Narrow band imaging and white-light endoscopic image modalities.

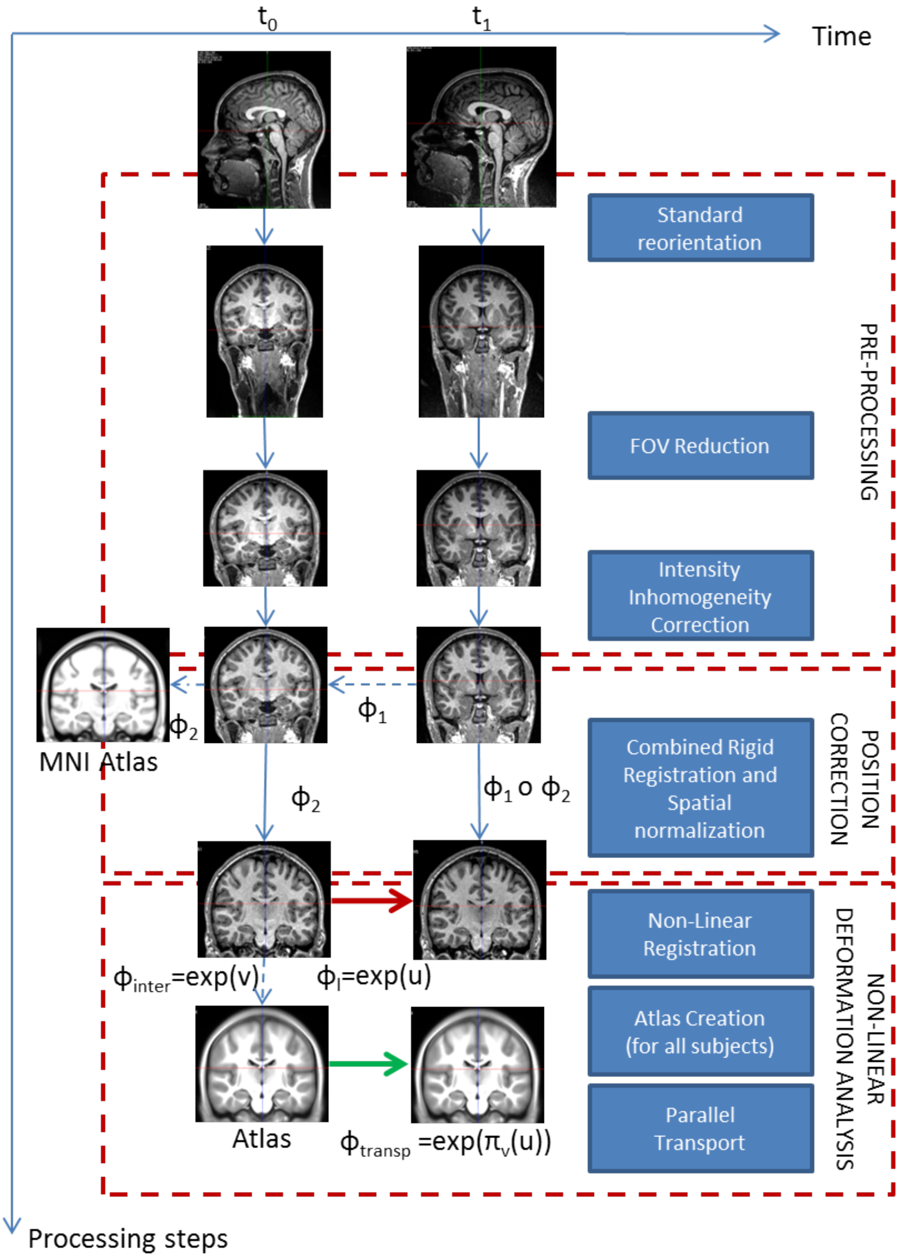


Figure 1. Proposed processing pipeline for longitudinal analysis: the pipeline is composed of three major steps. Starting with raw images, we first pre-process them, then correct the spatial position differences to end up with the longitudinal deformations for each subject in the atlas space.

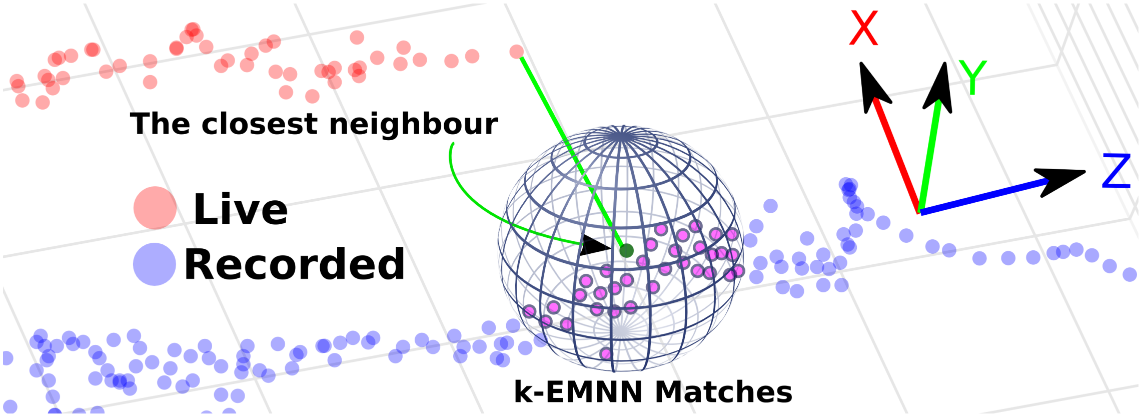


Figure 2. ElectroMagnetic tracker Nearest Neighbour (EMNN) matches. Firstly, the EMNN is obtained using the 3D position match. Then, in a radius r around this match, all the points on the trajectory are considered as the k-EMNN matches. For the images in the k-EMNN matches, scene matching is performed.

6.1.3. Segmentation and anatomic variability of the cochlea and other temporal bone structures from medical images

Participants: Thomas Demarcy [Correspondent], Hervé Delingette, Clair Vandersteen [IUFC, Nice], Dan Gnansia [Oticon Medical], Nicholas Ayache.

This work is supported by the National Association for Research in Technology (ANRT) through the CIFRE Grant 2013-1165 and Oticon Medical (Vallauris). Part of this work is also funded by the European Research Council through the ERC Advanced Grant MedYMA 2011-291080 (on Biophysical Modeling and Analysis of Dynamic Medical Images). This work is done in collaboration with the Department of Ear Nose Throat Surgery (IUFC, Nice) and the Nice University Hospital (CHU).

image segmentation ; surgery planning ; shape modelling ; anatomic variability ; cochlear implant ; temporal bone

- We designed a parametric shape model of the intracochlear anatomy with anatomical prior learned from temporal bones high-resolution images, see Fig. 3 .
- We evaluated the cochleostomy location regarding two surgical approaches (endaural compared to conventional posterior tympanotomy) [20].

6.1.4. Structured sparse Bayesian modelling for non-rigid registration and cardiac motion tracking

Participants: Loic Le Folgoc [Correspondent], Hervé Delingette, Antonio Criminisi, Nicholas Ayache.

This work has been partly supported by the Inria – Microsoft Research Joint Center and by the European Research Council through the ERC Advanced Grant MedYMA 2011-291080 (on Biophysical Modeling and Analysis of Dynamic Medical Images).

Non-rigid Registration - Structured Sparse Bayesian Learning - Automatic Relevance Determination - Reversible-jump Markov Chain Monte Carlo - Cardiac Motion Tracking - Uncertainty Quantification

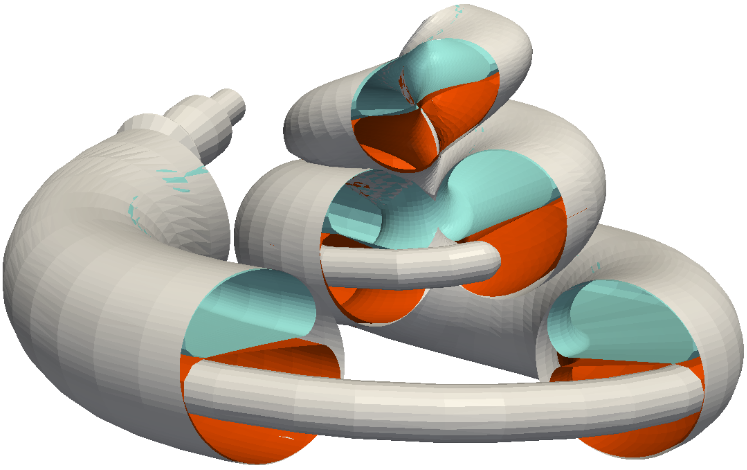


Figure 3. Cochlear implant electrode-array with respect to scala tympani (red) and scala vestibuli (blue).

We developed a generic structured sparse Bayesian model of image registration with three main contributions: an extended image similarity term, the automated tuning of registration parameters and uncertainty quantification. We proposed an approximate inference scheme that is tractable on 4D clinical data. We demonstrated the performance of our approach on cine MR, tagged MR and 3D Ultra Sound cardiac images, and showed state-of-the-art results on benchmark datasets evaluating accuracy of motion and strain.

Moreover, we evaluated the quality of uncertainty estimates returned by the approximate inference scheme. We compare the predictions of the approximate scheme with those of an inference scheme developed on the grounds of reversible jump Markov Chain Monte-Carlo [94](see Fig. 4). We provided more insight into the theoretical properties of the sparse structured Bayesian model and into the empirical behaviour of both inference schemes.

This work is described in the PhD manuscript of Loïc Le Folgoc, defended at Université Nice Sophia Antipolis, 2015 [6].

6.1.5. Image Segmentation and Synthesis of brain tumor MR images

Participants: Nicolas Cordier [correspondent], Hervé Delingette, Nicholas Ayache.

Part of this work was funded by the European Research Council through the ERC Advanced Grant MedYMA (on Biophysical Modeling and Analysis of Dynamic Medical Images).

Brain, MRI, Glioma, Patch-based Segmentation, Image synthesis

The segmentation of glioblastoma, the most severe case of brain tumors, is a crucial step for diagnostic assessment and therapy planning. In order to perform the manual delineation of the tumor compartments, the clinicians have to concurrently screen multi-channel 3D MRI, which makes the process both time-consuming and subject to inter-expert delineation variability.

We have developed 2 contributions for the analysis of MR brain tumor images:

- A patch-based multi-atlas automatic glioma segmentation algorithm[13]. Unlike prior work on patch-based multi-atlas segmentation, our approach does not assume any prior knowledge about the location of pathological structures (no local search window).

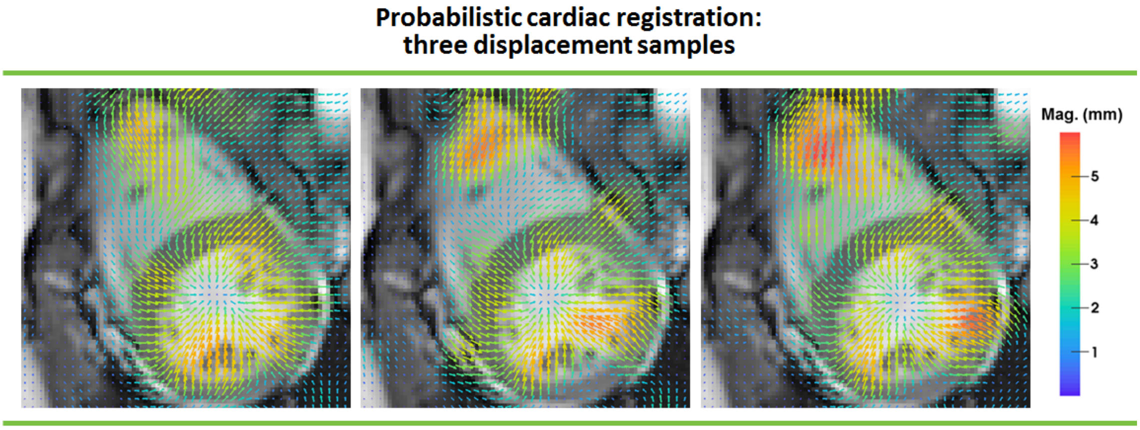


Figure 4. Three displacement fields sampled from the same posterior distribution of coefficients associated with the registration of two cardiac images.

- A patch-based image synthesis algorithm (see Fig.5) [4], which generates multi-sequence MR images of the brain with glioma from a single label image. The synthesis of images may be useful to benchmark segmentation algorithms or to increase the size of annotated medical image databases.

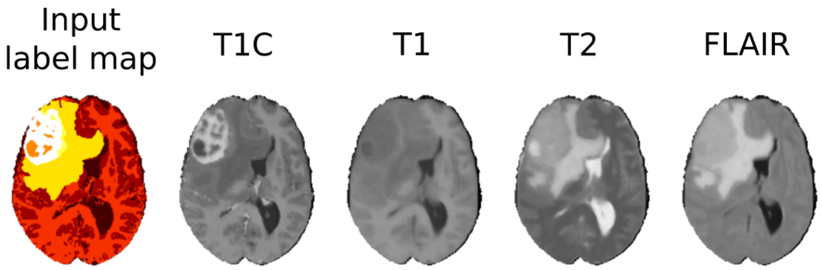


Figure 5. Synthesis of high-grade glioma MR image from a single label map.

6.1.6. Infarct localization from myocardial deformation

Participants: Nicolas Duchateau [Correspondent], Maxime Sermesant.

This work received the partial support from the European Union 7th Framework Programme (VP2HF FP7-2013-611823) and the European Research Council (MedYMA ERC-AdG-2011-291080).

Myocardial infarct, Computer-aided diagnosis, Dimensionality reduction, Biomechanical modeling

- We investigate new methods for predicting the location of myocardial infarcts from local wall deformation [31], which is useful for risk stratification from routine examinations such as 3D echocardiography. In a broader perspective, this project also aims at determining relevant biomarkers to study cardiac function [54], and eventually at combining several of those markers in an efficient manner [59].

- Non-linear dimensionality reduction aims at estimating the Euclidean space of coordinates encoding deformation patterns, and is combined with multi-scale kernel regressions to infer the low-dimensional coordinates and the infarct location of new cases.
- These concepts were tested on 500 synthetic cases with infarcts of random extent, shape, and location, generated from a realistic electromechanical model. Our prediction goes beyond the current diagnosis of infarct either achieved at the global or segmental level, and significantly outperforms the clinically-used thresholding of the deformation patterns.

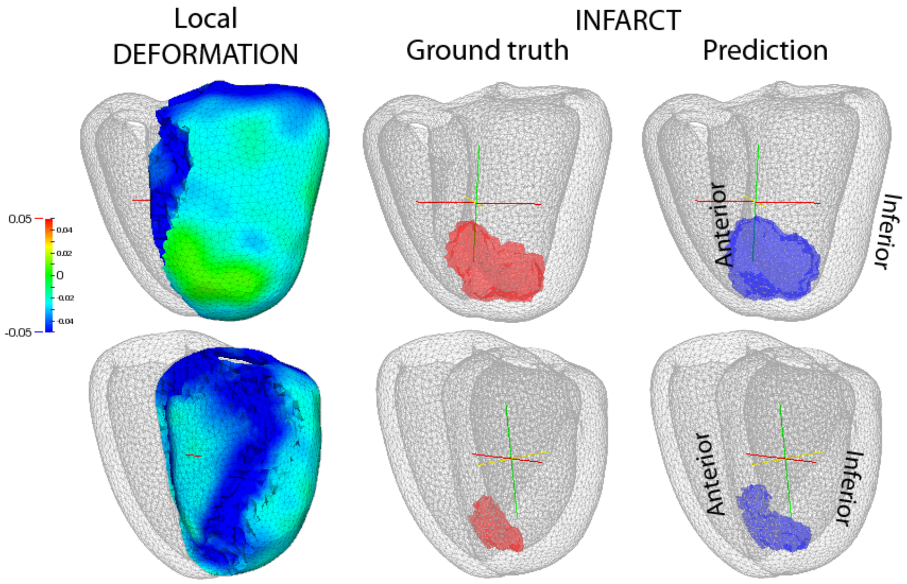


Figure 6. Examples of myocardial deformation patterns, ground truth infarct location, and estimated infarct location.

6.2. Computational Anatomy

6.2.1. Geometric generative model of organ shapes: statistical properties of template shape estimation

Participants: Nina Miolane [Correspondent, Inria - Stanford], Xavier Pennec [Inria], Susan Holmes [Stanford].

This work is conducted jointly with the Department of Statistics of Stanford, in the context of the associated team GeomStats and the FSCIS (France-Stanford Center for Interdisciplinary Studies) fellowship of Nina Miolane.

template, atlas, consistency, estimation theory, Expectation-Maximization algorithm, shapes, quotient space, lie group, sub-Riemannian, in-painting, neuro-geometry, visual cortex, diffusion

This work focuses on the interaction between statistics and geometry, for applications in Medical Imaging. The first part deals with a generative model of (organ) shapes and, more precisely, on the estimation of the mean shape or template. The second part of this work surveys and unveils the mathematical framework needed to extend Neurogeometry, used in 2D Computer Vision, to applications in 3D imaging.

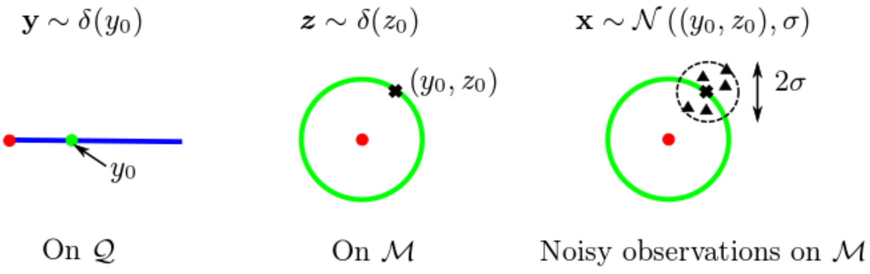


Figure 7. A schematic representation of the generative model of organ shapes.

- In the first part, we define a geometric statistical framework of an organ shapes generative model (see Figure 7). This is done through the differential geometry of quotient spaces.
- Then, we interpret the computation of the mean organ shape (or template) through the max-max algorithm, as an approximate maximum likelihood estimation in this framework.
- Finally, we study the statistical properties of the template computed with this procedure. More precisely, we show that the estimation is inconsistent and that the inconsistency cannot be neglected when the real template is close to the singularity of the quotient space at the scale of the ambient noise on the images [44].
- In the second part, the particularities of a 3D neurogeometry are highlighted with respect to the 2D case. They rely on the fact that 2D neurogeometry is inspired by the primary visual cortex, which codes for our 2D visual field (our retina is 2D). Imagining a 3D visual field or a 3D retina would give rise to a 3D neurogeometry.
- The conceptual framework of a 3D neurogeometry is more subtle, and a new level of mathematical structures arises (see Figure 8). Thus, inpainting (sub-Riemannian diffusion) have to be generalized.
- Applications for in-painting or super-resolution in 3D medical images are described [43].

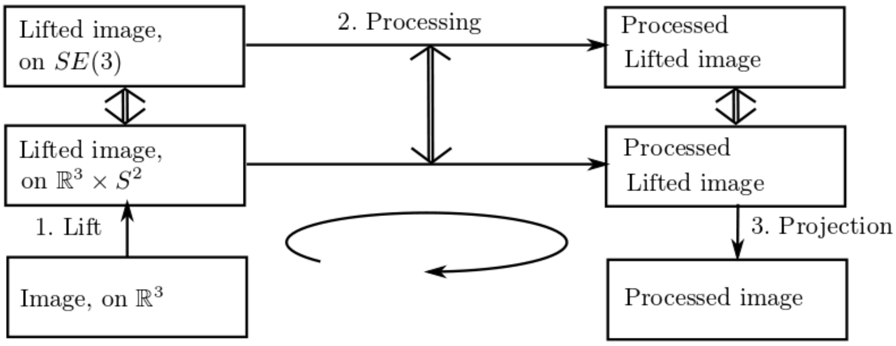


Figure 8. The 3 steps of an image processing pipeline in 3D neurogeometry.

6.2.2. Compact representation of longitudinal deformations

Participants: Raphaël Sivera [Correspondent], Hervé Delingette, Nicholas Ayache.

This work is supported by a PhD fellowship from the University Nice Sophia Antipolis and by the European Research Council through the ERC Advanced Grant MedYMA 2011-291080 (on Biophysical Modeling and Analysis of Dynamic Medical Images).

Longitudinal modeling, Learning in manifolds, Structured sparsity.

The analysis of dynamic or longitudinal series of medical images is important to better understand the observed evolutions of the organs but also to provide robust computer aided diagnosis tools. This analysis can be performed through a reduced representation of geometric transformations capturing the deformation between 2 time points.

In the context of cardiac motion analysis, we proposed a framework to represent arbitrary diffeomorphisms described as Stationary Velocity Fields (SVF) in a low dimensional linear space (see fig. 9).

To this end, we first improved the Inverse Scaling and Squaring (ISS) algorithm from [83] to transform displacement fields into SVFs. Second, through a structured sparse decomposition of these deformations over the cardiac cycle, we provided a preliminary approach for comparing trajectories of cine-MR images between two patients.

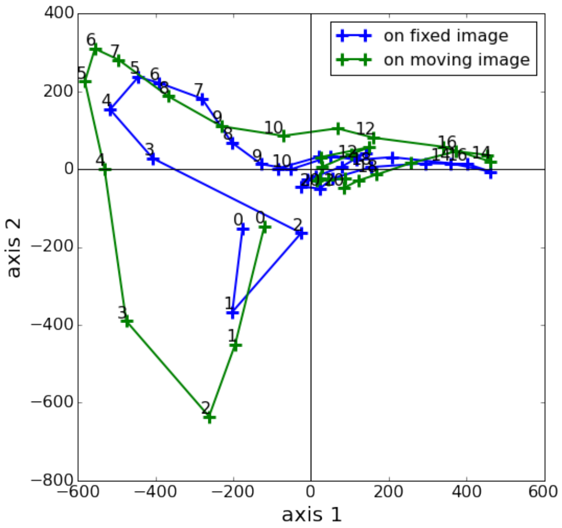


Figure 9. Trajectories of two registered cardiac cycles projected on a 2D space using dimensionality reduction tools.

6.2.3. Statistical analysis of heart motion

Participants: Marc-Michel Rohé [Correspondent], Nicolas Duchateau, Maxime Sermesant, Xavier Pennec.

This work is partly supported by the FP7 European project MD-Paedigree and by the ERC Advanced Grant MedYMA 2011-291080 (on Biophysical Modeling and Analysis of Dynamic Medical Images).

Statistical analysis, Registration, Reduced order models, Machine learning

This work aims at developing statistical tools to analyze cardiac motion. In particular, we are interested in approximating complex motion models with few parameters or modes that are clinically relevant (reduced models). To this end, we have introduced a polyaffine cardiac motion model that reduces the deformation parameters to a few interpretable parameters, and the most important modes to represent the variability seen in a population are automatically selected. We then performed a group-wise statistical analysis, which relates the model parameters to clinical indices specific to a given pathology. This method was used to classify a population of healthy/infarcted hearts [48] (see Fig. 10), as well as to study cardiac motion of adolescents with cardiomyopathies within the European project "MD-Paedigree".

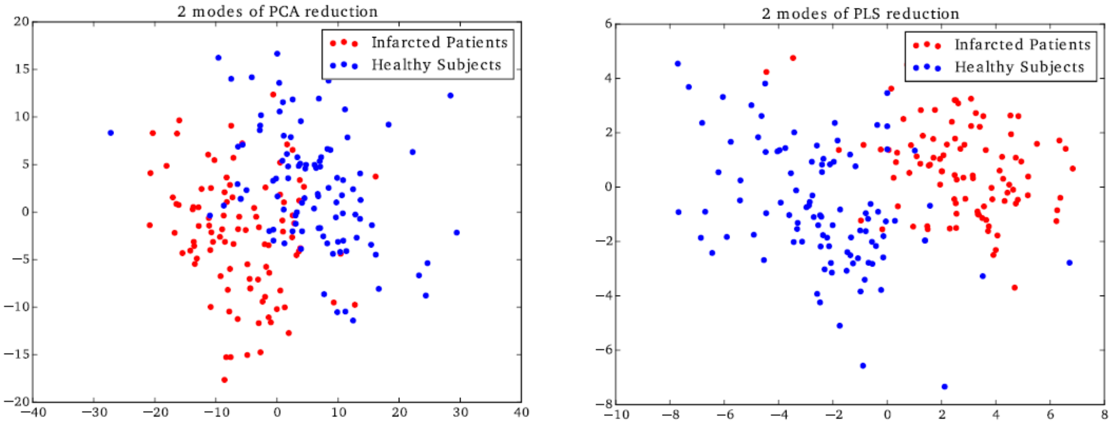


Figure 10. Projection of healthy/infarcted patients cardiac motions on two modes extracted from PCA (left) and PLS (right) methods.

6.2.4. Statistical Learning via Synthesis of Medical Images

Participants: Hervé Lombaert [Correspondent], Héloise Bleton, Hervé Delingette, Nicholas Ayache, Antonio Criminisi.

This work is partly supported by a grant from Microsoft Research-Inria Joint Centre and by the ERC Advanced Grant MedYMA 2011-291080 (on Biophysical Modeling and Analysis of Dynamic Medical Images).

statistical learning, synthesis

Machine learning approaches typically require large training datasets in order to capture as much variability as possible. Application of conventional learning methods on medical images is difficult due to the large variability that exists among patients, pathologies and image acquisitions. The project aims at exploring how realistic image synthesis could be used to improve existing machine learning methods.

We tackled the problem of better exploiting existing training sets, via a smart modeling of the image space, and applying conventional random forests using guided bagging [99]. Synthesis of complex data, such as cardiac diffusion images (DTI), was also done, with a refined version of [98].

Then, we tackled the problem of exploiting *Geometry in Data*, via intrinsic representations of shapes and data [27]. Spectral decomposition (Fig. 11) of shapes provides a new intrinsic framework for synthesizing complex shapes such as cerebral surfaces [35], and describing functions efficiently on these complex surfaces. This framework establishes the basics for machine learning of surface data [36]. An early application was conducted on retinotopy [57] (the study of functions in the visual cortex).

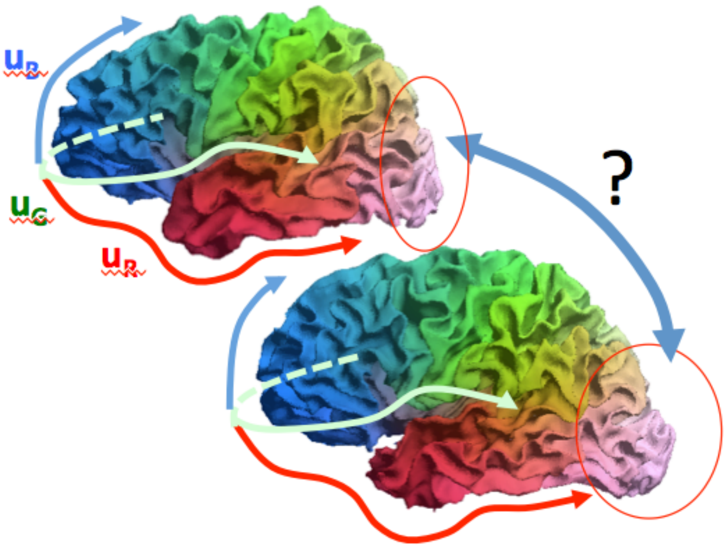


Figure 11. Spectral Representation – How to find an intrinsic representation of shapes for finding fast point correspondences and for learning surface data.

6.2.5. Consistency of the estimation of the template in quotient spaces

Participants: Loïc Devilliers [Correspondent], Stéphanie Allassonnière [Ecole Polytechnique], Xavier Pennec.

Template estimation, Fréchet mean, quotient spaces

In [24], we studied the estimation of the template (the mean shape of our data) when the data is transformed by unknown group elements. In the case of a finite group acting isometrically on a linear space, we proved that the estimation of the template using the Fréchet mean in the quotient space is not always consistent.

6.3. Computational Physiology

6.3.1. Computational modeling of radiofrequency ablation for the planning and guidance of abdominal tumor treatment

Participants: Chloé Audigier [Correspondent], Hervé Delingette, Tommaso Mansi [Siemens], Nicholas Ayache.

This PhD work was carried out between the Asclepios research group, Inria Sophia Antipolis, France and Medical Imaging Technologies, Healthcare Technology Center, Siemens Medical Solutions USA, Princeton, NJ.

Radiofrequency Ablation Modeling, Patient-Specific Simulation, Lattice Boltzmann Method, Computer Model, Computational Fluid Dynamics, Heat Transfer, Cellular Necrosis, Parameter Estimation, Therapy Planning, Liver, Pre-clinical Study, Medical Imaging

Radio Frequency Ablation (RFA) is a minimally invasive therapy suited for liver tumor ablation. However, a patient-specific predictive tool is needed to plan and guide the treatment.

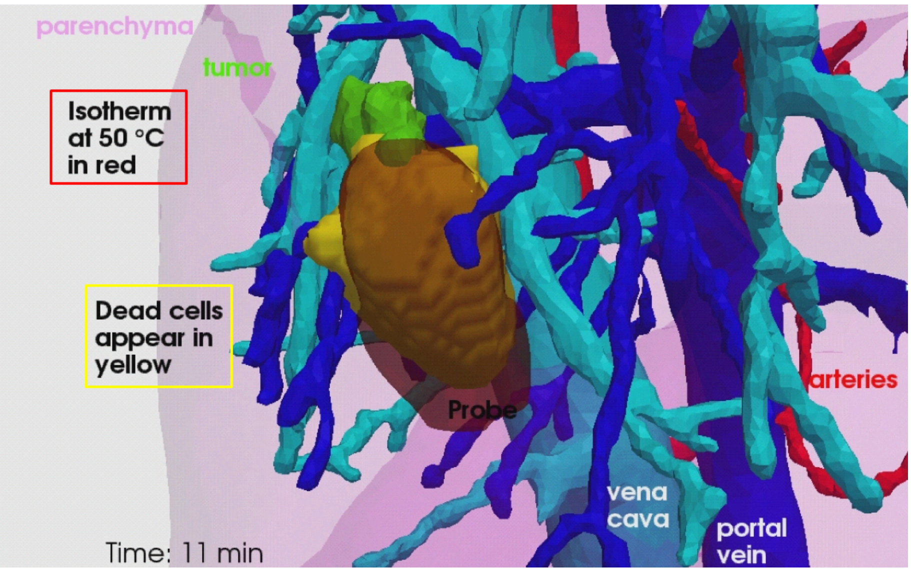


Figure 12. Computed isotherm at 50°C and computed necrosis in a subject-specific geometry.

We developed a computational framework for patient-specific planning of RFA, which includes the following contributions:

- A detailed computational model of the biophysical mechanisms (heat transfer, cellular necrosis, hepatic blood flow) involved in RFA of abdominal tumors based on patient images.
- A new implementation of the bio-heat equations coupled with a cellular necrosis model using the Lattice Boltzmann Method (LBM) on Graphics Processing Units (GPU), which allows near real-time computation.
- A Computational Fluid Dynamics (CFD) and porous media solver using LBM algorithm to compute the patient-specific blood flow in the hepatic circulatory system and the blood flow distribution inside the parenchyma.
- A complete patient-specific geometry including hepatic venous and arterial circulation system.
- The automatic estimation of the main parameters of the model. Two personalization strategies tested and evaluated on clinical and pre-clinical data.
- The evaluation of the proposed model on a clinical dataset of ten patients (see Fig. 12).
- The evaluation on a preclinical dataset of five swines from a comprehensive experimental set-up specially designed for RFA model validation.

The proposed RFA model and its evaluation on clinical data are presented in [10], and the evaluation of the RFA model on pre-clinical data is presented in [25]. The proposed model, its personalisation and its evalutation against clinical and preclinical data are presented in Chloé Audigier's PhD thesis [1].

6.3.2. Learning Cardiac Ablation Targets from Image Data and Simulation

Participants: Rocio Cabrera Lozoya [Correspondent], Maxime Sermesant, Nicholas Ayache.

This work was supported by the ERC Advanced Grant MedYMA 2011-291080 (on Biophysical Modeling and Analysis of Dynamic Medical Images).

Cardiac electrophysiology modeling, Intracardiac electrogram modeling, Machine learning, Radiofrequency ablation planning, electroanatomical mapping, local abnormal ventricular activities (LAVA)

Ventricular radiofrequency ablation can have a critical impact on preventing sudden cardiac arrest but it is challenging due to a highly complex arrhythmogenic substrate. We used advanced delayed enhanced-MR image characteristics in a machine learning framework to predict the presence of local abnormal ventricular activities (LAVA). Furthermore, we enriched these predictions through MR image-based patient-specific electrophysiology simulations and the modeling of normal and LAVA-like intracardiac electrograms using the dipole approach and their incorporation in the learning framework (see Fig. 13). Confidence maps can then be generated and analyzed prior to RFA to guide the intervention.

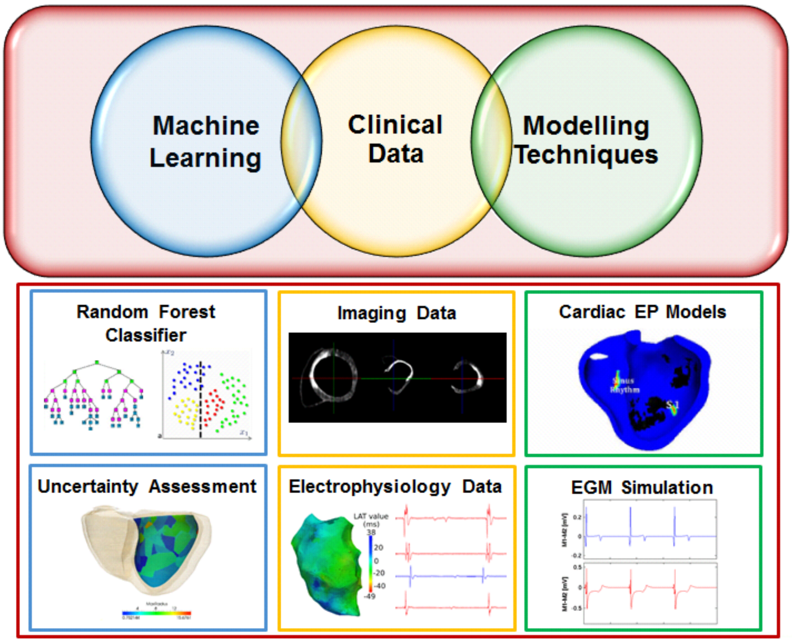


Figure 13. Coupled learning and simulation framework for LAVA identification.

6.3.3. Biophysical Modeling and Simulation of Longitudinal Brain MRIs with Atrophy in Alzheimer's Disease

Participants: Bishesh Khanal [Correspondent], Nicholas Ayache, Xavier Pennec.

This work has been partly supported by the European Research Council through the ERC Advanced Grant MedYMA (on Biophysical Modeling and Analysis of Dynamic Medical Images).

Alzheimer's Disease (AD), modeling brain deformation, biophysical model, simulation

- We developed a framework to generate patient specific multiple time-point images based on our biophysical model of brain deformation due to atrophy in Alzheimer's Disease (AD)[34]. From two time-point brain MRIs of a patient, we used the framework to simulate a new time-point brain MRI with the personalized atrophy for the patient (see Fig. 14).
- The framework can be used to evaluate methods that study the temporal relationships, ordering and co-evolution of atrophy in different structures of the brain.

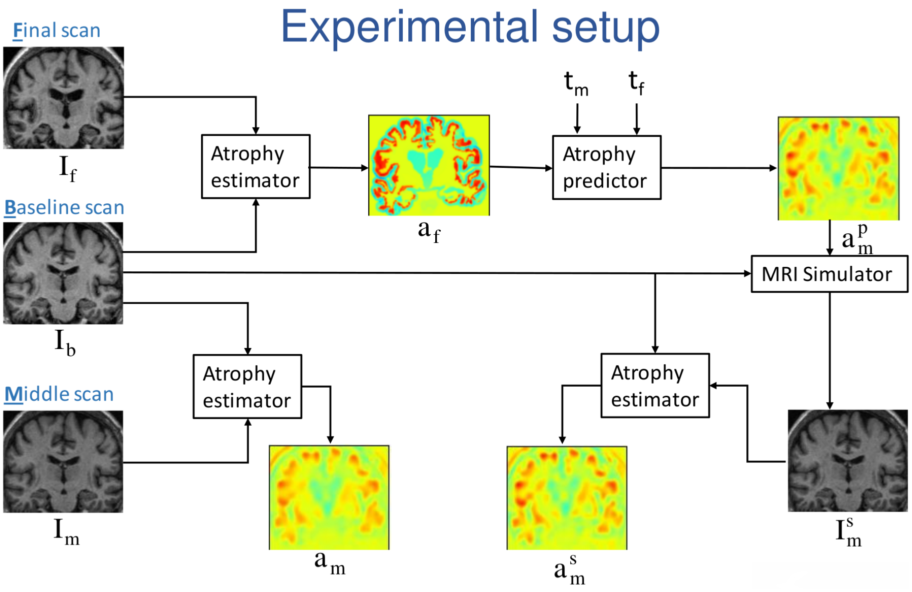


Figure 14. From a given baseline scan and a final scan at time t_f, a middle scan at time t_m is predicted and simulated. An experimenatal setup that allows comparing the atrophy (a^s_m) estimated from a simulated middle-scan against the atrophy (a_m) estimated from the real middle scan is shown.

6.3.4. Brain Tumor Growth Personalization and Segmentation Uncertainty

Participants: Matthieu Lê [Correspondent], Hervé Delingette, Jan Unkelbach, Nicholas Ayache.

This work is carried out between Asclepios research group, Inria Sophia Antipolis, France and the Department of Radiation Oncology of the Massachusetts General Hospital, Boston, USA. It is supported by the ERC Advanced Grant MedYMA 2011-291080 (on Biophysical Modeling and Analysis of Dynamic Medical Images).

Tumor growth, radiotherapy, modeling, personalization, segmentation, uncertainty, Bayesian

- We developed a method for the Bayesian personalization of a brain tumor growth model based on clinical MRIs [37] (see Fig. 15 Left).
- We proposed an algorithm for the sampling of several plausible segmentations, based on a single clinical segmentation (see Fig. 15 Right). This allows the uncertainty quantification of the radiotherapy plan based on several sample clinical target volumes [38]. This paper received the Young Scientist Award at the 2015 MICCAI conference in Munich, Germany.

6.3.5. Uncertainty quantification in personalised Cardiac models. Application to myocardial fiber uncertainty.

Participants: Roch-Philippe Molléro [Correspondent], Dominik Neumann [Siemens], Marc-Michel Rohé, Hervé Delingette, Maxime Sermesant, Xavier Pennec, Nicholas Ayache, Tommaso Mansi [Siemens].

This work was partly supported by the FP7 European project MD-Paedigree and was done in collaboration with Siemens Corporate Technology, Erlangen, Germany and Siemens Corporate Research, Princeton, New Jersey.

Heart Modeling - Myocardial Fibers - Biophysical Simulation - Uncertainty Quantification

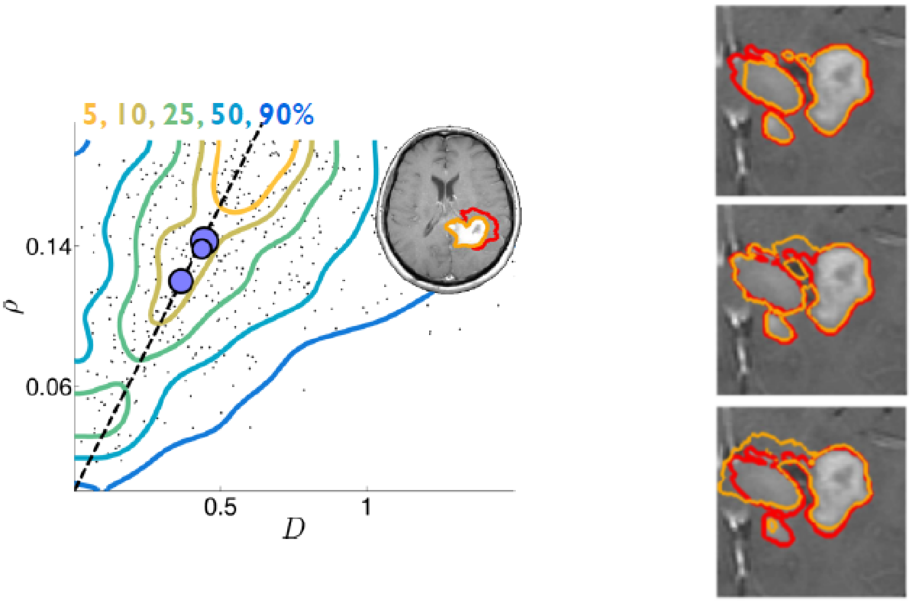


Figure 15. (Left) Bayesian personalization of a glioblastoma patient. Isocontours of the posterior probability of the diffusion parameter D and the proliferation parameter ρ; (Right) Different sampled plausible segmentations in orange based on the clinical segmentation in red.

Computational models of the heart are of increasing interest for clinical applications due to their discriminative and predictive power. However, the personalisation step to go from a generic model to a patient-specific one is still challenging. In particular, it is still difficult to quantify the uncertainty on the estimated parameters and predicted values.

We developped a pipeline (see Fig. 16) to evaluate the impact of myocardial fibre uncertainty on the personalisation of an electromechanical model of the heart from ECG and medical images:

- We studied how to estimate the variability of the fibre architecture among a given population (from a myocardial fiber atlas).
- Then, we showed the variability of the personalised simulations, in electrophysiology (EP) and in biomechanics, with respect to the principal variations of the fibres.
- Finally, we discussed how the variations in this population of fibres impact the parameters of the personalised simulations.

This work led to a paper at FIMH 2015 conference in Maastricht, The Netherlands [45].

6.3.6. Non-invasive personalisation of the electrical heart model

Participants: Sophie Giffard-Roisin [Correspondent], Maxime Sermesant, Nicholas Ayache, Hervé Delingette.

This work has been supported by the European Project FP7 under grant agreement VP2HF (no 611823) and the ERC Advanced Grant MedYMA (on Biophysical Modeling and Analysis of Dynamic Medical Images).

Cardiac Modelling, Personalised Simulation, Electrical Simulation

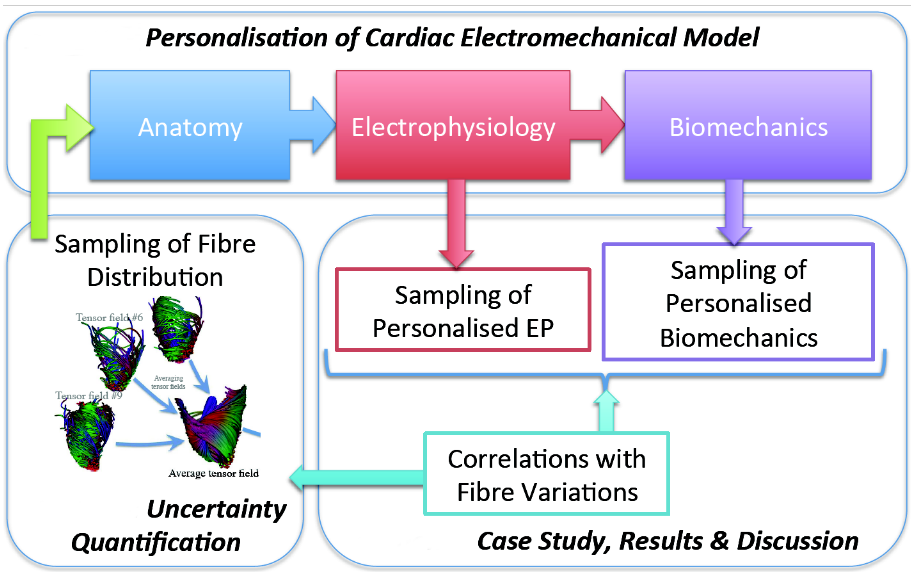


Figure 16. Global scheme of fibre variability propagation along the personalisation pipeline.

Non-invasive cardiac electrical data has been acquired at St Thomas' Hospital, London. It consists in Body Surface Potential Mapping (BSPM), which are recordings of the electrical potential on several locations on the surface of the torso (see Fig. 17). From BSPMs and MRI data of the heart, we aim at personalizing the electrical propagation model of the heart previously developed within the Asclepios team.

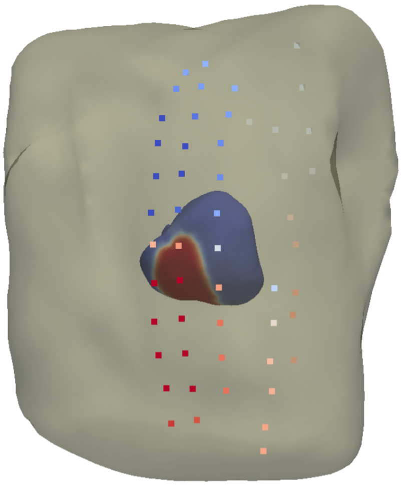


Figure 17. Torso representation used for personalizing cardiac electrical parameters from non-invasive observations.

ATHENA Project-Team

6. New Results

6.1. Modeling in Diffusion MRI

6.1.1. Improving fiber alignment in HARDI by combining contextual PDE flow with constrained spherical deconvolution

Participants: Jorg Portgegies [Department of Mathematics and Computer Science, Eindhoven University of Technology], Rutger Fick, Gonzalo Sanguinetti [Department of Mathematics and Computer Science, Eindhoven University of Technology], Shephan Meesters [Department of Mathematics and Computer Science, Eindhoven University of Technology], Gabriel Girard [Athena, Inria Sophia-A-M & SCIL Lab., Sherbrooke University], Remco Duits [Department of Mathematics and Computer Science, Eindhoven University of Technology].

We propose two strategies to improve the quality of tractography results computed from diffusion weighted magnetic resonance imaging (DW-MRI) data. Both methods are based on the same PDE framework, defined in the coupled space of positions and orientations, associated with a stochastic process describing the enhancement of elongated structures while preserving crossing structures. In the first method we use the enhancement PDE for contextual regularization of a fiber orientation distribution (FOD) that is obtained on individual voxels from high angular resolution diffusion imaging (HARDI) data via constrained spherical deconvolution (CSD). Thereby we improve the FOD as input for subsequent tractography. Secondly, we introduce the fiber to bundle coherence (FBC), a measure for quantification of fiber alignment. The FBC is computed from a tractography result using the same PDE framework and provides a criterion for removing the spurious fibers. We validate the proposed combination of CSD and enhancement on phantom data and on human data, acquired with different scanning protocols. On the phantom data we find that PDE enhancements improve both local metrics and global metrics of tractography results, compared to CSD without enhancements. On the human data we show that the enhancements allow for a better reconstruction of crossing fiber bundles and they reduce the variability of the tractography output with respect to the acquisition parameters. Finally, we show that both the enhancement of the FODs and the use of the FBC measure on the tractography improve the stability with respect to different stochastic realizations of probabilistic tractography. This is shown in a clinical application: the reconstruction of the optic radiation for epilepsy surgery planning.

This work has been published in [19]

6.1.2. Sparse reconstruction challenge for diffusion MRI: Validation on a physical phantom to determine which acquisition scheme and analysis method to use?

Participants: Lipeng Ning [Brigham and Women's Hospital, Harvard Medical School, Boston], Frederik Laun [German Cancer Research Institute], Yogesh Rathi [Brigham and Women's Hospital, Harvard Medical School, Boston], Thinhinane Megherbi [ParIMed Team, LRPE, USTHB, Algiers], Mario Zuccheli [Dpt of Computer Science, University of Verona], Gloria Menegaz [Dpt of Computer Science, University of Verona], Maxime Descoteaux [SCIL Lab., Sherbrooke University], Aurobrata Ghosh, Rutger Fick, Rachid Deriche.

Diffusion magnetic resonance imaging (dMRI) is the modality of choice for investigating in-vivo white matter connectivity and neural tissue architecture of the brain. The diffusion-weighted signal in dMRI reflects the diffusivity of water molecules in brain tissue and can be utilized to produce image-based biomarkers for clinical research. Due to the constraints on scanning time, a limited number of measurements can be acquired within a clinically feasible scan time. In order to reconstruct the dMRI signal from a discrete set of measurements, a large number of algorithms have been proposed in recent years in conjunction with varying sampling schemes, i.e., with varying b-values and gradient directions. Thus, it is imperative to compare the performance of these reconstruction methods on a single data set to provide appropriate guidelines to neuroscientists on making an informed decision while designing their acquisition protocols. For this purpose, the SPArse Reconstruction Challenge (SPARC) was held along with the workshop on Computational Diffusion MRI (at MICCAI 2014) to validate the performance of multiple reconstruction methods using data acquired from a physical phantom. A total of 16 reconstruction algorithms (9 teams) participated in this community challenge. The goal was to reconstruct single b-value and/or multiple b-value data from a sparse set of measurements. In particular, the aim was to determine an appropriate acquisition protocol (in terms of the number of measurements, b-values) and the analysis method to use for a neuroimaging study. The challenge did not delve on the accuracy of these methods in estimating model specific measures such as fractional anisotropy (FA) or mean diffusivity, but on the accuracy of these methods to fit the data. This work presents several quantitative results pertaining to each reconstruction algorithm. The conclusions in this work provide a valuable guideline for choosing a suitable algorithm and the corresponding data-sampling scheme for clinical neuroscience applications.

This work has been published in [18].

6.1.3. A Unifying framework for spatial and temporal diffusion in dMRI

Participants: Rutger Fick, Demian Wassermann, Marco Pizzolato, Rachid Deriche.

We propose a novel framework to simultaneously represent the diffusion-weighted MRI (dMRI) signal over diffusion times, gradient strengths and gradient directions. Current frameworks such as the 3D Simple Harmonic Oscillator Reconstruction and Estimation basis (3D-SHORE) only represent the signal over the spatial domain, leaving the temporal dependency as a fixed parameter. However, microstructure- focused techniques such as Axcaliber and ActiveAx provide evidence of the importance of sampling the dMRI space over diffusion time. Up to now there exists no generalized framework that simultaneously models the dependence of the dMRI signal in space and time. We use a functional basis to fit the 3D+t spatio-temporal dMRI signal, similarly to the 3D-SHORE basis in three dimensional 'q-space'. The lowest order term in this expansion contains an isotropic diffusion tensor that characterizes the Gaussian displacement distribution, multiplied by a negative exponential. We regularize the signal fitting by minimizing the norm of the analytic Laplacian of the basis. The continuous 3D+t signal representation can provide new insights on the anomalous nature of the dMRI signal in human tissues, i.e., when mean-squared molecular displacements varies slower than linearly with the diffusion time. From the fitting one can also estimate the axon radius distribution parameters along any direction using approaches similar to AxCaliber. We validate our technique on synthetic data generated using the theoretical model proposed by Callaghan et al. We show that our method is robust to noise and can accurately describe the restricted spatio-temporal signal decay originating from tissue models such as cylindrical pores. Moreover, we apply our method on real data from an ActiveAx acquisition. Overall our approach allows to represent the complete 3D+t dMRI signal which should prove helpful in understanding normal and pathologic nervous tissue.

This work has been published in [26]

6.1.4. Exploiting the phase in dMRI for microstructure recovery: Towards axonal tortuosity via asymmetric diffusion processes

Participants: Marco Pizzolato, Demian Wassermann, Timothé Boutelier [Olea Medical, La Ciotat], Rachid Deriche.

Microstructure recovery procedures via Diffusion-Weighted Magnetic Resonance Imaging (DW-MRI) usually discard the signal's phase, assuming symmetry in the underlying diffusion process. In this work, we propose to recover the Ensemble Average Propagator (EAP) directly from the complex DW signal in order to describe also eventual diffusional asymmetry, thus obtaining an asymmetric EAP. The asymmetry of the EAP is then related to tortuosity of undulated white matter axons, which are found in pathological scenarios associated with axonal elongation or compression. We derive a model of the EAP for this geometry and quantify its asymmetry. Results show that the EAP obtained when accounting for the DW signal's phase provides useful microstructural information in such pathological scenarios. Furthermore, we validate these results in-silico through 3D Monte-Carlo simulations of white matter tissue that has experienced different degrees of elongation/compression.

This work has been published in [35]

6.1.5. *A temperature phantom to probe the Ensemble Average Propagator asymmetry: an in-silico study*

Participants: Marco Pizzolato, Demian Wassermann, Tanguy Duval [Institute of Biomedical Engineering, Polytechnique Montréal, Montréal], Jennifer Campbell [Montreal Neurological Institute, MCGill University], Timothé Boutelier [Olea Medical, La Ciotat], Julien Cohen-Adad [Institute of Biomedical Engineering, Polytechnique Montréal, Montréal], Rachid Deriche.

The detection and quantification of asymmetry in the Ensemble Average Propagator (EAP) obtained from the Diffusion-Weighted (DW) signal has been shown only for theoretical models. EAP asymmetry appears for instance when diffusion occurs within fibers with particular geometries. However the quantification of EAP asymmetry corresponding to such geometries in controlled experimental conditions is limited by the difficulty of designing fiber geometries on a micrometer scale. To overcome this limitation we propose to adopt an alternative paradigm to induce asymmetry in the EAP. We apply a temperature gradient to a spinal cord tract to induce a corresponding diffusivity profile that alters locally the diffusion process to be asymmetric. We simulate the EAP and the corresponding complex DW signal in such a scenario. We quantify EAP asymmetry and investigate its relationship with the applied experimental conditions and with the acquisition parameters of a Pulsed Gradient Spin-Echo sequence. Results show that EAP asymmetry is sensible to the applied temperature-induced diffusivity gradient and that its quantification is influenced by the selected acquisition parameters.

This work has been published in [36]

6.1.6. *How to get more out of a clinically feasable 64 gradient dMRI acquisition: multi-shell versus single-shell*

Participants: Rutger Fick, Mario Zuccheli [Dpt of Computer Science, University of Verona], Gabriel Girard [SCIL Lab., Sherbrooke University], Maxime Descoteaux [SCIL Lab., Sherbrooke University], Gloria Menegaz [Dpt of Computer Science, University of Verona], Rachid Deriche.

For clinical applications the number of diffusion MRI (dMRI) samples that can be obtained is often limited by scanner time and patient comfort. For this reason one often uses short scanning protocols that acquire just 32 or 64 gradient directions using a single b-value to obtain diffusion measures such as the fractional anisotropy from Diffusion Tensor Imaging (DTI) or to estimate the white matter orientation using Constrained Spherical Deconvolution (CSD). Using 3D-SHORE and MAP-MRI, we show that by spreading the same number of dMRI samples over different b-shells (sampling angularly and radially) we can estimate not only the directionality of the white matter using the ODF, but also the radially dependent higher order diffusion measures that SHORE and MAP-MRI provide. This approach lends itself well for situations where acquisition time is limited, and is therefore particularly well suited for clinical applications.

This work has been published in [29].

6.2. Tissue Microstructures features recovery & applications

6.2.1. *Laplacian-regularized MAP-MRI : Improving axonal caliber estimation*

Participants: Rutger Fick, Demian Wassermann, Gonzalo Sanguinetti, Rachid Deriche.

In diffusion MRI, the accurate description of the entire diffusion signal from sparse measurements is essential to enable the recovery of microstructural information of the white matter. The recent Mean Apparent Propagator (MAP)-MRI basis is especially well suited for this task, but the basis fitting becomes unreliable in the presence of noise. As a solution we propose a fast and robust analytic Laplacian regularization for MAP-MRI. Using both synthetic diffusion data and human data from the Human Connectome Project we show that (1) MAP-MRI has more accurate microstructure recovery com- pared to classical techniques, (2) regularized MAP-MRI has lower signal fitting errors compared to the unregularized approach and a positivity constraint on the EAP and (3) that our regularization improves axon radius recovery on human data.

This work has been published in [27]

6.2.2. Using 3D-SHORE and MAP-MRI to obtain both tractography and microstructural contrasts from a clinical dMRI acquisition

Participants: Rutger Fick, Mario Zuccheli [Dpt of Computer Science, University of Verona], Gabriel Girard [Athena, Inria Sophia-A-M & SCIL Lab., Sherbrooke University], Maxime Descoteaux [SCIL Lab., Sherbrooke University], Gloria Menegaz [Dpt of Computer Science, University of Verona], Rachid Deriche.

Diffusion MRI (dMRI) is used to characterize the directionality and microstructural properties of brain white matter (WM) by measuring the diffusivity of water molecules. In clinical practice the number of dMRI samples that can be obtained is limited, and one often uses short scanning protocols that acquire just 32 to 64 different gradient directions using a single gradient strength (b-value). Such 'single shell' scanning protocols restrict one to use methods that have assumptions on the radial decay of the dMRI signal over different b-values, which introduces estimation biases. In this work, we show, that by simply spreading the same number of samples over multiple b-values (i.e. multi-shell) we can accurately estimate both the WM directionality using 3D-SHORE and characterize the radially dependent diffusion microstructure measures using MAP-MRI. We validate our approach by undersampling both noisy synthetic and human brain data of the Human Connectome Project, proving this approach is well-suited for clinical applications.

This work has been published in [28]

6.2.3. A sensitivity analysis of Q-space Indices with respect to changes in axonal diameter, dispersion and tissue composition

Participants: Rutger Fick, Marco Pizzolato, Demian Wassermann, Mario Zuccheli [Dpt of Computer Science, University of Verona], Gloria Menegaz [Dpt of Computer Science, University of Verona], Rachid Deriche.

In Diffusion MRI, q-space indices are scalar quantities that describe properties of the ensemble average propagator (EAP). Their values are often linked to the axonal diameter – assuming that the diffusion signal originates from inside an ensemble of parallel cylinders. However, histological studies show that these assumptions are incorrect, and axonal tissue is often dispersed with various tissue compositions. Direct interpretation of these q-space indices in terms of tissue change is therefore impossible, and we must treat them as as scalars that only give non-specific contrast – just as DTI indices. In this work, we analyze the sensitivity of q-space indices to tissue structure changes by simulating axonal tissue with changing axonal diameter, dispersion and tissue compositions. Using human connectome project data we then predict which indices are most sensitive to tissue changes in the brain. We show that, in both multi-shell and single-shell (DTI) data, q-space indices have higher sensitivity to tissue changes than DTI indices in large parts of the brain. Based on these results, it may be interesting to revisit older DTI studies using q-space indices as a marker for pathology.

This work has been accepted at the conference ISBI 2016.

6.2.4. MAPL: Tissue microstructure estimation using Laplacian-regularized MAP-MRI and its application to HCP data

Participants: Rutger Fick, Demian Wassermann, Emanuel Caruyer, Rachid Deriche.

The recovery of microstructure-related features of the brain's white matter is a current challenge in diffusion MRI. To robustly estimate these important features from diffusion MRI data, we propose to analytically regularize MAP-MRI's coefficient estimation using the norm of the Laplacian of the reconstructed signal. We first compare our approach, which we call MAPL, with competing state-of-the-art functional basis approaches. We show that it outperforms the original MAP-MRI implementation and the recently proposed modified Spherical Polar Fourier (mSPF) basis with respect to signal fitting, EAP and ODF reconstruction in noisy, sparsely sampled data of a physical phantom with reference gold standard data. Then, to reduce the variance of parameter estimation using multi-compartment tissue models, we propose to use MAPL's signal fitting and extrapolation as a preprocessing step. We study the effect of MAPL on the estimation of axon diameter using a simplified Axcaliber model and axonal dispersion using the Neurite Orientation Dispersion and Density Imaging (NODDI) model. We show the positive effect of using it as a preprocessing step in estimating and reducing the variances of these parameters in the Corpus Callosum of six different subjects of the MGH Human Connectome Project. Finally we correlate the estimated axon diameter, dispersion and restricted volume fractions with Fractional Anisotropy (FA) and clearly show that changes in FA significantly correlate with changes with all estimated parameters. Overall, we illustrate the potential of using a well-regularized functional basis together with multi-compartment approaches to recover important microstructure tissue parameters with much less variability, thus contributing to the challenge of better understanding microstructure-related features of the brain's white matter.

This work has been submitted to the journal NeuroImage.

6.3. Towards microstructural based tractography

6.3.1. AxTract: Microstructure-driven tractography based on the Ensemble Average Propagator

Participants: Gabriel Girard [Athena, Inria Sophia-A-M & SCIL Lab., Sherbrooke University], Rutger Fick, Maxime Descoteaux [SCIL Lab., Sherbrooke University], Demian Wassermann, Rachid Deriche.

In this work, we propose a novel method to simultaneously trace brain white matter (WM) fascicles and estimate WM microstructure characteristics. Recent advancements in diffusion-weighted imaging (DWI) allow multi-shell acquisitions with b-values of up to 10,000 s/mm2 in human subjects, enabling the measurement of the ensemble average propagator (EAP) at distances as short as 10 micro-meters. Coupled with continuous models of the full 3D DWI signal and the EAP such as Mean Apparent Propagator (MAP) MRI, these acquisition schemes provide unparalleled means to probe the WM tissue in vivo. Presently, there are two complementary limitations in tractography and microstructure measurement techniques. Tractography techniques are based on models of the DWI signal geometry without taking specific hypotheses of the WM structure. This hinders the tracing of fascicles through certain WM areas with complex organization such as branching, crossing, merging, and bottlenecks that are indistinguishable using the orientation-only part of the DWI signal. Microstructure measuring techniques, such as AxCaliber, require the direction of the axons within the probed tissue before the acquisition as well as the tissue to be highly organized. Our contributions are twofold. First, we extend the theoretical DWI models proposed by Callaghan et al. to characterize the distribution of axonal calibers within the probed tissue taking advantage of the MAP-MRI model. Second, we develop a simultaneous tractography and axonal caliber distribution algorithm based on the hypothesis that axonal caliber distribution varies smoothly along a WM fascicle. To validate our model we test it on in-silico phantoms and on the HCP dataset

This work has been published in [23]

6.3.2. Studying white matter tractography reproducibility through connectivity matrices

Participants: Gabriel Girard [Athena, Inria Sophia-A-M & SCIL Lab., Sherbrooke University], Kevin Whittingstall [SCIL Lab., Sherbrooke University], Maxime Descoteaux [SCIL Lab., Sherbrooke University], Rachid Deriche.

Diffusion-weighted imaging is often used as a starting point for in vivo white matter (WM) connectivity to reconstruct potential WM pathways between brain areas. In this study, we investigate the reproducibility of the connectivity matrix, resulting from different tractography parameters. We vary the number of streamlines used to construct the matrix in cortical to cortical connectivity and analyze its effects. We also compare the effect of probabilistic and deterministic local streamline tractography algorithms, seeding both from the WM and from WM-grey matter interface.

This work has been published in [31]

6.3.3. *Structural connectivity reproducibility through multiple acquisitions*

Participants: Gabriel Girard [Athena, Inria Sophia-A-M & SCIL Lab., Sherbrooke University], Kevin Whittingstall [SCIL Lab., Sherbrooke University], Maxime Descoteaux [SCIL Lab., Sherbrooke University], Rachid Deriche.

dMRI is often used to reconstruct white matter pathways between brain areas for in vivo brain connectivity. In this study, we investigate the reproducibility and the specificity of connectivity matrices in cortico-cortical connectivity using probabilistic and deterministic streamline tractography, seeding from both the white matter and the white matter-grey matter interface.

This work has been published in [30]

6.4. Computational Diffusion MRI

6.4.1. *Robust and efficient linear registration of white-matter fascicles in the space of streamlines*

Participants: Eleftherios Garyfallidis [SCIL Lab., Sherbrooke University], Omar Cepeda [SCIL Lab., Sherbrooke University], Demian Wassermann, Maxime Descoteaux [SCIL Lab., Sherbrooke University].

The neuroscientific community is very much interested in analyzing specific white matter bundles like the arcuate fasciculus, the corticospinal tract, or the recently discovered Aslant tract to study sex differences, lateralization and many other connectivity applications. For this reason, experts spend time manually segmenting these fascicles and bundles using streamlines obtained from diffusion MRI tractography. However, to date, there are very few computational tools available to register these fascicles directly so that they can be analyzed and their differences quantified across populations. In this work, we introduce a novel, robust and efficient framework to align bundles of streamlines directly in the space of streamlines. We call this framework Streamline-based Linear Registration. We first show that this method can be used successfully to align individual bundles as well as whole brain streamlines. Additionally, if used as a piecewise linear registration across many bundles, we show that our novel method systematically provides higher overlap (Jaccard indices) than state-of-the-art nonlinear image-based registration in the white matter. We also show how our novel method can be used to create bundle-specific atlases in a straightforward manner and we give an example of a probabilistic atlas construction of the optic radiation. In summary, Streamline-based Linear Registration provides a solid registration framework for creating new methods to study the white matter and perform group-level tractometry analysis.

This work has been published in [14]

6.4.2. *Cortical surface parcellation via dMRI using mutual nearest neighbor condition*

Participants: Brahim Belaoucha, Maureen Clerc, Théodore Papadopoulo.

In this work, we present a method that aims at parcellating the cortical surface from individual anatomy. The parcellation is obtained using the mutual nearest neighbor criteria to obtain regions that have similar fiber distribution. The later is obtained by applying a probabilistic tractography on the diffusion MRI (dMRI), a non-invasive modality allowing the access to the structural information of the cortical surface. The proposed algorithm is compared to some of the atlases that can be found in the literature. We show that these atlases have lower similarity of fibers distributions than the proposed algorithm.

This work has been accepted at the conference ISBI 2016.

6.5. Clinical and Neurocognitive Applications of Diffusion MRI

6.5.1. *Plasticity of left perisylvian white-matter tracts is associated with individual differences in math learning brain structure and function*

Participants: Dietsje Jolles [Stanford University & Leiden University], Demian Wassermann, Ritika Chokhani [Stanford University], Jennifer Richardson [Stanford University], Caitlin Tenison [Stanford University], Roland Bammer [Stanford University], Lynn Fuchs [Vanderbilt University], Kaustubh Supekar [Stanford University], Vinod Menon [Stanford University].

Plasticity of white matter tracts is thought to be essential for cognitive development and academic skill acquisition in children. However, a dearth of high-quality diffusion tensor imaging (DTI) data measuring longitudinal changes with learning, as well as methodological difficulties in multi-time point tract identification have limited our ability to investigate plasticity of specific white matter tracts. Here, we examine learning- related changes of white matter tracts innervating inferior parietal, prefrontal and temporal regions following an intense two-month math tutoring program. DTI data were acquired from 18 third grade children, both before and after tutoring. A novel fiber tracking algorithm based on a White Matter Query Language (WMQL) was used to identify three sections of the superior longitudinal fasciculus (SLF) linking frontal and parietal (SLF-FP), parietal and temporal (SLF-PT) and frontal and temporal (SLF-FT) cortices, from which we created child-specific probabilistic maps. The SLF-FP, SLF-FT, and SLF-PT tracts identified with the WMQL method were highly reliable across the two time points and showed close correspondence to tracts previously described in adults. Notably, individual differences in behavioral gains after two months of tutoring were specifically correlated with plasticity in the left SLF-FT tract. Our results extend previous findings of individual differences in white matter integrity, and provide important new insights into white matter plasticity related to math learning in childhood. More generally, our quantitative approach will be useful for future studies examining longitudinal changes in white matter integrity associated with cognitive skill development.

This work has been published in [16]

6.5.2. *Prefrontal cortex white matter tracts in prodromal Huntington disease*

Participants: Joy T. Matsui [Iowa University], Jatin G. Vaidya [Iowa University], Demian Wassermann [Iowa University], Regina Eunyoung Kim [Iowa University], Vincent A. Magnotta [Iowa University], Hans J. Johnson [Iowa University], Jane S. Paulsen [Iowa University], Predict-Hd Investigators And Coordinators Of The Huntington Study Group [NIH].

Huntington disease (HD) is most widely known for its selective degeneration of striatal neurons but there is also growing evidence for white matter (WM) deterioration. The primary objective of this research was to conduct a large-scale analysis using multi-site diffusion-weighted imaging (DWI) tractography data to quantify diffusivity properties along major prefrontal cortex WM tracts in prodromal HD. Fifteen international sites participating in the PREDICT-HD study collected imaging and neuropsychological data on gene-positive HD subjects without a clinical diagnosis (i.e. prodromal) and gene-negative control subjects. The anatomical prefrontal WM tracts of the corpus callosum (PFCC), anterior thalamic radiations (ATR), inferior fronto-occipital fasciculi (IFO), and uncinate fasciculi (UNC) were identified using streamline tractography of DWI. Within each of these tracts, tensor scalars for fractional anisotropy, mean diffusivity, radial diffusivity, and axial diffusivity coefficients were calculated. We divided prodromal HD subjects into three CAG-age product (CAP) groups having Low, Medium, or High probabilities of onset indexed by genetic exposure. We observed significant differences in WM properties for each of the four anatomical tracts for the High CAP group in comparison to controls. Additionally, the Medium CAP group presented differences in the ATR and IFO in comparison to controls. Furthermore, WM alterations in the PFCC, ATR, and IFO showed robust associations with neuropsychological measures of executive functioning. These results suggest that long-range tracts essential for cross-region information transfer show early vulnerability in HD and may explain cognitive problems often present in the prodromal stage.

This work has been published in [17]

6.6. Perfusion MRI

6.6.1. *Perfusion MRI deconvolution with delay estimation and non-negativity constraints*

Participants: Marco Pizzolato, Auro Ghosh, Timothé Boutelier [Olea Medical, La Ciotat], Rachid Deriche.

Perfusion MRI deconvolution aims to recover the time-dependent residual amount of indicator (residue function) from the measured arterial and tissue concentration time-curves. The deconvolution is complicated by the presence of a time lag between the measured concentrations. Moreover the residue function must be non-negative and its shape may become non-monotonic due to dispersion phenomena. We introduce Modified Exponential Bases (MEB) to perform deconvolution. The MEB generalizes the previously proposed exponential approximation (EA) by taking into account the time lag and introducing non-negativity constraints for the recovered residue function also in the case of non-monotonic dispersed shapes, thus overcoming the limitation due to the non-increasing assumtion of the EA. The deconvolution problem is solved linearly. Quantitative comparisons with the widespread block-circulant Singular Value Decomposition show favorable results in recovering the residue function.

This work has been published in [34]

6.6.2. Elucidating dispersion effects in perfusion MRI by means of dispersion-compliant bases

Participants: Marco Pizzolato, Rutger Fick, Timothé Boutelier [Olea Medical, La Ciotat], Rachid Deriche.

Dispersion effects in perfusion MRI data have a relevant influence on the residue function computed from deconvolution of the measured arterial and tissular concentration time-curves. Their characterization allows reliable estimation of hemodynamic parameters and can reveal pathological tissue conditions. However the time-delay between the measured concentration time-curves is a confounding factor. We perform deconvolution by means of dispersion-compliant bases, separating dispersion from delay effects. In order to characterize dispersion we introduce shape parameters, such as the dispersion time and index. We propose a new formulation for the dispersed residue function and perform in-silico experiments that validate the reliability of our approach against the block-circulant Singular Value Decomposition. We successfully apply the approach to stroke MRI data and show that the calculated parameters are coherent with physiological considerations, highlighting the importance of dispersion as an effect to be measured rather than discarded.

This work has been accepted at the conference ISBI 2016.

6.6.3. Unveiling the dispersion kernel in DSC-MRI by means of dispersion-compliant bases and control point interpolation techniques

Participants: Marco Pizzolato, Rutger Fick, Timothé Boutelier [Olea Medical, La Ciotat], Rachid Deriche.

In DSC-MRI the presence of dispersion affects the estimation, via deconvolution, of the residue function that characterizes the perfusion in each voxel. Dispersion is descibed by a Vascular Transport Function (VTF) which knolewdge is essential to recover a dispersion-free residue function. State-of-the-art techniques aim at characterizing the VTF but assume a specific shape for it, which in reality is unknown. We propose to estimate the residue function without assumptions by means of Dispersion-Compliant Bases (DCB). We use these results to find which VTF model better describes the in-vivo data for each tissue type by means of control point interpolation approaches.

This work has been submitted to the conference ISMRM 2016.

6.6.4. Improved vascular transport function characterization in DSC-MRI via deconvolution with dispersion-compliant bases

Participants: Marco Pizzolato, Rutger Fick, Timothé Boutelier [Olea Medical, La Ciotat], Rachid Deriche.

Bolus dispersion phenomena affect the residue function computed via deconvolution of DSC-MRI data. Indeed the obtained effective residue function can be expressed as the convolution of the true one with a Vascular Transport Function (VTF) that characterizes the dispersion. The state-of-the-art technique CPI+VTF allows to estimate the actual residue function by assuming a model for the VTF. We propose to perform deconvolution representing the effective residue function with Dispersion-Compliant Bases (DCB) without assumptions on the VTF, and then apply the CPI+VTF on DCB results. We show that DCB improve robustness to noise and allow to better characterize the VTF.

This work has been submitted to the conference ISMRM 2016.

6.7. MEG, EEG and cochlear modeling

6.7.1. MEM-diffusion MRI framework to solve MEEG inverse problem

Participants: Brahim Belaoucha, Jean-Marc Lina, Maureen Clerc, Théodore Papadopoulo.

In this work, we present a framework to fuse information coming from diffusion magnetic resonance imaging (dMRI) with Magnetoencephalography (MEG)/ Electroencephalography (EEG) measurements to reconstruct the activation on the cortical surface. The MEG/EEG inverse-problem is solved by the Maximum Entropy on the Mean (MEM) principle and by assuming that the sources inside each cortical region follow Normal distribution. These regions are obtained using dMRI and assumed to be functionally independent. The source reconstruction framework presented in this work is tested using synthetic and real data. The activated regions for the real data is consistent with the literature about the face recognition and processing network.

This work was published in the proceedings of the conference EUSIPCO 2015 [22].

6.7.2. MEG/EEG reconstruction in the reduced source space

Participants: Brahim Belaoucha, Théodore Papadopoulo.

Obtaining the brain activity with the distributed source model from MEG or EEG measurements is ill-posed problem due to the high number of unknowns compared to the number of measurements. The idea of ths work is to reduce the solution space size from the number of sources to a smaller space. Assuming that sources inside each functional region have equal activation allows us to reduce the number of columns in the leadfield matrix from the number of nodes S required to model the cortex to a number of regions K, which is much smaller. These regions are obtained from a dMRI parcellation-based region growing algorithm. A region is assumed to contain sources that have similar fibers distribution. To obtain a sparse solution, we assume that only a few regions are active simultaneously. BIC1 is used to obtain the optimal number of regions (Kp) that explains the MEG/EEG data.

We compared the results of the proposed method to the ones from Minimum Norm Estimate (MNE) and LASSO. The first gives a smooth solution and the second gives a sparse solution. To test the accuracy of the reconstruction, we activated simultaneously from two to five regions in both hemispheres with synthetic low SNR signals (10 dB). Our approach could detect the right number of activated regions and provided more accurate reconstructions compared to MNE and LASSO.

Our approach assumes that few regions are active simultaneously which allows us to reduce the space to a few unkowns. It can be seen as an approximation to the l0 norm. Even though assuming a constant activation in each functional region is a hard constraint, it allows us to reduce the space size from S to K. The obtained solution can be used to detect extended sources (e.g epileptic activity) or as an initialization step to other approaches to obtain more detailed solutions in the active regions.

This work was presented at the conference BaCI 2015 [24].

6.7.3. Realistic simulation of electric potential distributions of different stimulation modes in an implanted cochlea

Participants: Kai Dang, Maureen Clerc, Clair Vandersteen [Institut Universitaire de la Face et du Cou, Nice], Nicolas Guevara [Institut Universitaire de la Face et du Cou, Nice], Dan Gnansia [Oticon Medical/Neurelec].

Simulation of the intracochlear potentials is an important approach to study the activation of auditory nerve fibers under electrical stimulations. However, it is still unclear to which extent the simulation results are affected by precision in reproducing the exact cochlear geometry. In this study, we address to this question by comparing the actual electric potential measured from implanted human specimen with the simulation outputs from two different parametric 3D cochlear models. One of the models is created from the default values while the other is adapted to the micro-CT scan data of the implanted cochlea.

This work was presented at the Association of Research in Otolaryngology 38th MidWinter Meeting, Feb 2015, Baltimore, United States [38].

We also made an in situ validation of electrical models: Cochlear implants have been proved to be an effective treatment for patients with sensorineural hearing loss. Among all the approaches that have been developed to design better cochlear implants, 3D model-based simulation stands out due to its detailed description of the electric field which helps reveal the electrophysiological phenomena inside the cochlea. With the advances in the cochlear implant manufacturing technology, the requirement on simulation accuracy increases. Improving the simulation accuracy relies on two aspects: 1) a better geometrical description of the cochlea that is able to distinguish the subtle differences across patients; 2) a comprehensive and reliable validation of the created 3D model. In this paper, targeting at high precision simulation, we propose a parametric cochlea model which uses micro-CT images to adapt to different cochlea geometries, then demonstrate its validation process with multi-channel stimulation data measured from a implanted cochlea. Comparisons between the simulation and validation data show a good match under a variety of stimulation configurations. The results suggest that the electric field distribution is affected by the geometric characteristics of each individual cochlea. These differences can be correctly reflected by simulations based on a 3D model tuned with personalized data.

This work was presented at the 7th International IEEE EMBS Conference on Neural Engineering, Apr 2015, Montpellier, France [25].

6.7.4. Influence of skull modelling on conductivity estimation for EEG source analysis

Participants: Christos Papageorgakis, Maureen Clerc, Benjamin Lanfer [BESA GmbH].

The skull conductivity strongly influences the accuracy of EEG source localization methods. As the conductivity of the skull has strong inter-individual variability, conductivity estimation techniques are required. Typically, conductivity estimation is performed on data from a single event-related stimulation paradigm, which can be explained by one dipole source. A conductivity value for the skull can be estimated as the value for which the single dipole source provides the best goodness of fit to the data. This conductivity value is then used to analyse the actual data of interest. It is known that the optimal local skull conductivity when modelling the skull as one compartment depends on the amount of spongiosa present locally. The research question arising is: Is conductivity estimation based on data from a single paradigm meaningful without accounting for the internal skull structure ?

This work was presented at the conference BaCI 2015 [33], and is submitted for journal publication.

6.7.5. Dictionary learning for M/EEG multidimensional data

Participants: Christos Papageorgakis, Sebastian Hitziger, Théodore Papadopoulo.

Signals obtained from magneto- or electroencephalography (M/EEG) are very noisy and inherently multidimensional, i.e. provide a vector of measurements at each single time instant. To cope with noise, researchers traditionally acquire measurements over multiple repetitions (trials) and average them to classify various patterns of activity. This is not optimal because of trial-to-trial variability (waveform variation, jitters). The jitter-adaptive dictionary learning method (JADL) has been developed to better handle for this variability (with a particular emphasis on jitters). JADL is a data-driven method that learns a dictionary (prototype pieces) from a set of signals, but is currently limited to a single channel, which restricts its capacity to work with very noisy data such as M/EEG. We propose an extension to the jitter-adaptive dictionary learning method, that is able to handle multidimensional measurements such as M/EEG.

This work was presented at the conference BaCI 2015 [32].

6.8. Brain Computer Interfaces

6.8.1. Decoding covert shifts of attention induced by ambiguous visuospatial cues

Participants: Romain Trachel, Maureen Clerc, Thomas Brochier [Institut de Neurosciences de la Timone, Marseille].

Simple and unambiguous visual cues (e.g., an arrow) can be used to trigger covert shifts of visual attention away from the center of gaze. The processing of visual stimuli is enhanced at the attended location. Covert shifts of attention modulate the power of cerebral oscillations in the alpha band over parietal and occipital regions. These modulations are sufficiently robust to be decoded on a single trial basis from electroencephalography (EEG) signals. It is often assumed that covert attention shifts are under voluntary control, and that they also occur in more natural and complex environments, but there is no direct evidence to support this assumption. We address this important issue by using random-dot stimuli to cue one of two opposite locations, where a visual target is presented. We contrast two conditions, one in which the random-dot motion is predictive of the target location, and the other, in which it provides ambiguous information. Behavioral results show attention shifts in anticipation of the visual target, in both conditions. In addition, using the common spatial patterns (CSPs) algorithm, we extract EEG power features in the alpha-band (around 10 Hz) that best discriminate the attended location in single trials. We obtain a significant decoding accuracy in 7/10 subjects using a cross-validation procedure applied in the predictive condition. Interestingly, similar accuracy (significant in 5/10 subjects) is obtained when the CSPs trained in the predictive condition are tested in the ambiguous condition. In agreement with this result, we find that the CSPs show very similar topographies in both conditions. These results shed a new light on the behavioral and EEG correlates of visuospatial attention in complex visual environments. This study demonstrates that alpha-power features could be used in brain computer interfaces to decode covert attention shifts in an environment containing ambiguous spatial information.

This work was published in Frontiers in Human Neurosciences [20].

6.8.2. Online extraction and single trial analysis of regions contributing to erroneous feedback detection

Participants: Eoin Thomas, Matthew Dyson, Laurence Casini, Boris Burle.

Understanding how the brain processes errors is an essential and active field of neuroscience. Real time extraction and analysis of error signals provide an innovative method of assessing how individuals perceive ongoing interactions without recourse to overt behaviour. This area of research is critical in modern Brain–Computer Interface (BCI) design, but may also open fruitful perspectives in cognitive neuroscience research. In this context, we sought to determine whether we can extract discriminatory error-related activity in the source space, online, and on a trial by trial basis from electroencephalography data recorded during motor imagery. Using a data driven approach, based on interpretable inverse solution algorithms, we assessed the extent to which automatically extracted error-related activity was physiologically and functionally interpretable according to performance monitoring literature. The applicability of inverse solution based methods for automatically extracting error signals, in the presence of noise generated by motor imagery, was validated by simulation. Representative regions of interest, outlining the primary generators contributing to classification, were found to correspond closely to networks involved in error detection and performance monitoring. We observed discriminative activity in non-frontal areas, demonstrating that areas outside of the medial frontal cortex can contribute to the classification of error feedback activity.

This work was published in NeuroImage [13].

AYIN Team

7. New Results

7.1. Markov Random Fields

7.1.1. New hierarchical joint classification method of SAR and optical multiresolution remote sensing datas

Participants: Ihsen Hedhli, Josiane Zerubia [contact].

This work was carried out in collaboration with Prof. Gabriele Moser and Prof. Sebastiano Serpico from DITEN departement [www.diten.unige.it/], University of Genoa, Italy.

Nowadays, a wide variety of remote sensing images is available. Therefore, it becomes more and more important to be able to analyze compound data sets consisting of different types of images acquired by different sensors, as they allow a spatially distributed and temporally repetitive view of the monitored area at the desired spatial scales. In particular, the opportunity of joint availability of synthetic aperture radar (SAR) and optical images offers high resolution (HR), all-weather, day/night, short revisit time data, as well as polarimetric and multifrequency acquisition capabilities. Similarly, the strong differences in terms of wavelength range (microwave vs. visible and near infrared), sensitivity to cloud cover and sun illumination (strong for optical imagery vs. almost negligible for SAR), and noise-like properties (speckle in SAR vs. generally low noise variance in current HR optical sensors) make the joint use of HR optical and SAR imagery especially interesting for many applications to environmental monitoring and natural risk management. Within this framework, there is a definite need for classification methods that automatically correlate different sets of images taken on the same area from different sensors and at different resolutions. This year we developed a novel classification approach for multiresolution, multisensor (optical and synthetic aperture radar), and/or multiband images. Accurate and time-efficient classification methods are particularly important tools to support rapid and reliable assessment of the ground changes. Given the huge amount and variety of data available currently from last-generation satellite missions, the main difficulty is to develop a classifier that can benefit from multiband, multiresolution, and multisensor input imagery. As shown in Figure 1 , the proposed method addresses the problem of multisensor fusion of SAR with optical data for classification purposes, and allows input data collected at multiple resolutions and additional multiscale features derived through wavelets to be fused. The proposed approach formalizes a supervised Bayesian classifier within a multiple quadtree topology that combines a class-conditional statistical model for pixel-wise information and a hierarchical Markov random field (MRF) for multisensor and multiresolution contextual information.

7.2. Marked point processes

7.2.1. Integrating RJMCMC and Kalman filters for multiple object tracking

Participants: Paula Craciun, Josiane Zerubia [contact].

This work has been done in collaboration with Dr. Mathias Ortner from Airbus D&S [http://www.space-airbusds.com/fr/]

Recently, we have proposed a new spatio-temporal marked point process model for tracking small, rigid objects in high resolution images. We have shown very good detection and tracking results for synthetic biological data as well as remotely sensed sequences. The model is based on defining a dedicated energy function that is highly non-convex. The solution is found by minimizing this energy function using a suitable batch-optimization scheme based on Reversible Jump Markov Chain Monte Carlo (RJMCMC) sampler. This approach is motivated by the low temporal frequency of the sequences (1Hz).

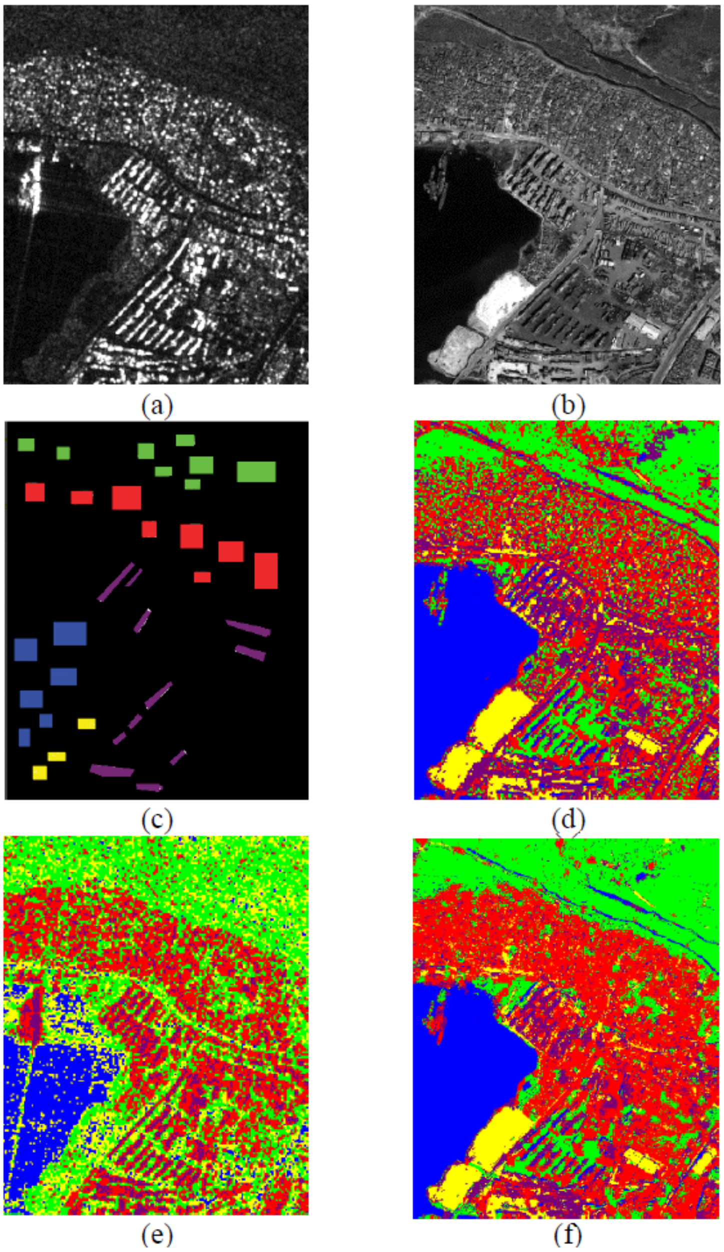


Figure 1. (a) SAR image (© ASI), (b) one channel from the optical image (© GeoEye), (c) the available ground truth, (d) hierarchical MRF-based classification obtained from the optical image, using Laferté method, (e) hierarchical MRF-based classification obtained for the SAR image, using Laferté method, (f) hierarchical MRF-based classification obtained by the proposed method.

Sequential filters have proven to provide relatively fast and reliable tracking performances in particular for single target tracking. We have efficiently exploited the properties of sequential filters within the RJMCMC sampling scheme. The filter is used to generate more meaningful perturbation proposals which are then evaluated using an appropriate Green acceptance ratio. Better perturbation proposals increase the acceptance probability of the overall RJMCMC sampling scheme which in turn leads to a faster convergence.
Figure 2 shows the detection and tracking results on two synthetic biological sequences as well as on two sequences of simulated satellite images of Toulon by courtesy of Airbus Defence & Space, France. The evolution of the energy with the number of iterations for the standard RJMCMC sampler and the proposed sampler is also shown. The proposed sampler is depicted blue.

7.2.2. Initialization and estimation of parameters for marked point processes applied to automatic object detection on satellite images

Participants: Aurélie Boisbunon, Josiane Zerubia [contact].

This work has been done in collaboration with Dr. Rémi Flamary (Université de Nice Sophia Antipolis), Prof. Alain Rakotomamonjy (Université de Rouen) et Alain Giros (CNES). It was partially funded by the French Spatial Agency CNES [http://www.cnes.fr].

Sparse representations, large scale, stochastic algorithms, machine learning, image processing Marked point processes (MPP) strongly rely on parameters, whose estimation affects both computation time and performances. In this work, we proposed two approaches: the first one consists in initializing MPPs with a first coarse solution obtained very quickly from sparse regularization methods, while the second one estimates the parameters by the Stochastic Approximation Expectation-Maximization (SAEM) algorithm. We give details on both approaches below. The first coarse solution is obtained from a deterministic sparse regularization method. This method is based on the representation of an image with objects as a sum of convolutions between atoms of a dictionary and matrices of positions of the objects in the image. The atoms of the dictionary are fixed in advance and correspond to different instances of the objects (scales, angles, shapes, etc). This way, we transform the problem of object detection into the problem of estimating extremely sparse matrices. The algorithm we derived for solving the associated optimization problem is both parallelized and very efficient. This work started last year, and continued this year by conducting more tests.

7.2.3. Generic curvilinear structure modeling via marked point process theory

Participants: Seong-Gyun Jeong, Yuliya Tarabalka, Josiane Zerubia [contact].

This work has been done in collaboration with Dr. Nicolas Nisse (COATI team [https://team.inria.fr/coati/], Inria-SAM) and Dr. Yuliya Tarabalka (Titane team [https://team.inria.fr/titane/team/], Inria-SAM)

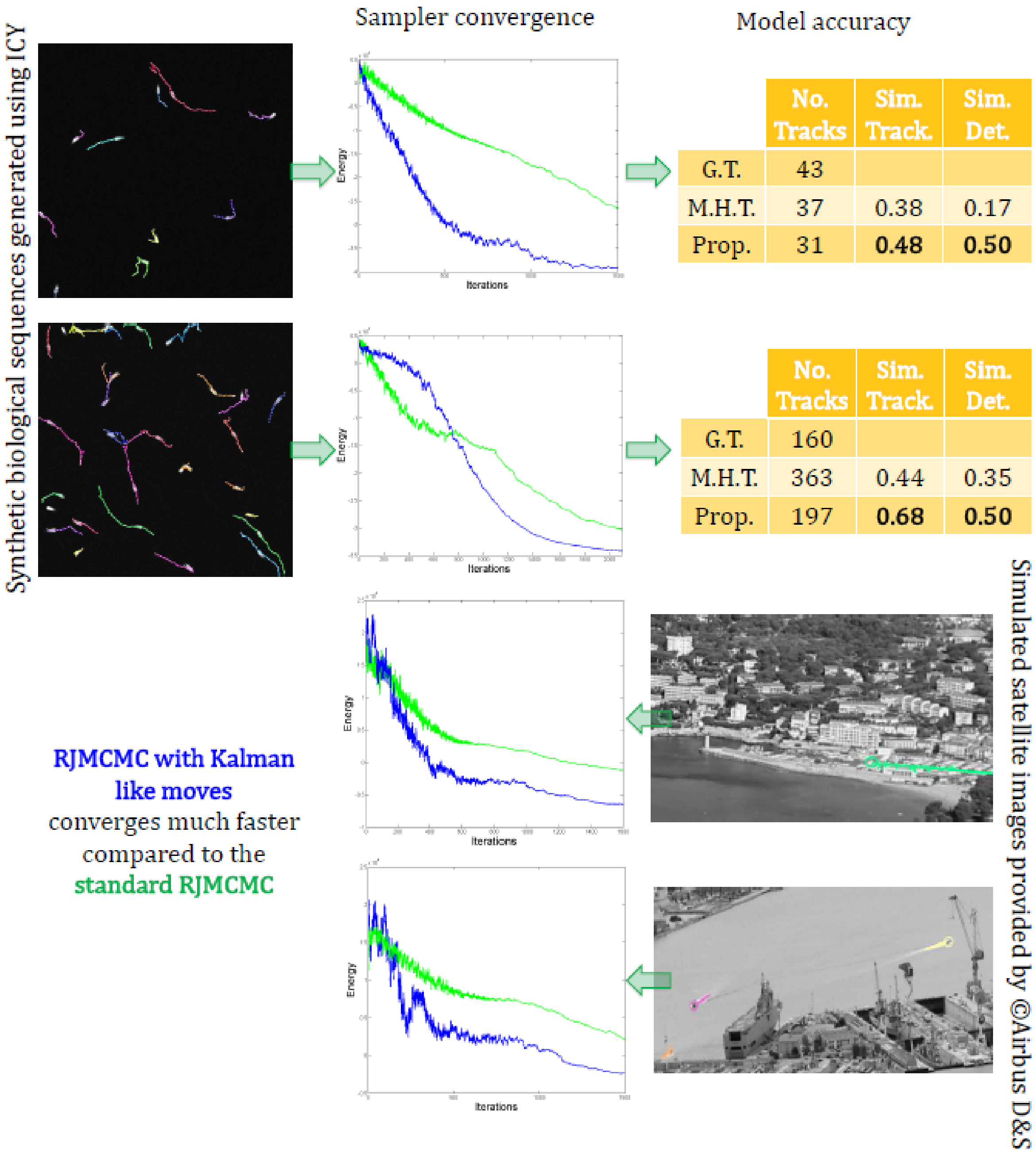
We propose a novel curvilinear structure reconstruction algorithm based on ranking learning and graph theory. In this work we reconstruct the curvilinear structure as a set of small line segments (via MPP). Specifically, we infer the structured output ranking of the line segments via Structured Support Vector Machine(SSVM). To predict the existence of the curvilinear structure, we measure oriented image gradient maps and morphological profiles. We propose an orientation-aware curvilinear feature descriptor and a feature grouping operator to improve the structural consistency for learning system. In order to provide topological information, we develop a graph-based curvilinear structure reconstruction algorithm. The proposed algorithm builds a graph based on the output ranking scores and searches the longest geodesic paths which are associated with the latent curvilinear structure. Experimental results (see Figure 3) show that the proposed algorithm faithfully detects the curvilinear structures and preserves topological information compared with the competing algorithms.

7.3. Other approaches

7.3.1. Acne detection on polarized or non-polarized images

Participants: Zhao Liu, Josiane Zerubia [contact].

This work is in collaboration with Dr. Queille-Roussel and Prof. Bahadoran in CHU Nice, France. Now Dr. Zhao Liu is a post-doc at Manchester University [www.manchester.ac.uk/], Manchester, UK.



	No. Tracks	Sim. Track.	Sim. Det.
G.T.	43		
M.H.T.	37	0.38	0.17
Prop.	31	**0.48**	**0.50**

	No. Tracks	Sim. Track.	Sim. Det.
G.T.	160		
M.H.T.	363	0.44	0.35
Prop.	197	**0.68**	**0.50**

Figure 2. Tracking results and sampler convergence on two synthetic biological sequences (generated using ICY [http://icy.bioimageanalysis.org/], a free software offered by the Quantitative Analysis Unit from the Pasteur Institute, France) as well as two sequences of simulated satellite images of Toulon (by courtesy of Airbus Defence & Space, France). The RJMCMC sampler with Kalman like moves (shown in blue) requires a significantly lower number of iterations until convergence as compared to the standard RJMCMC.

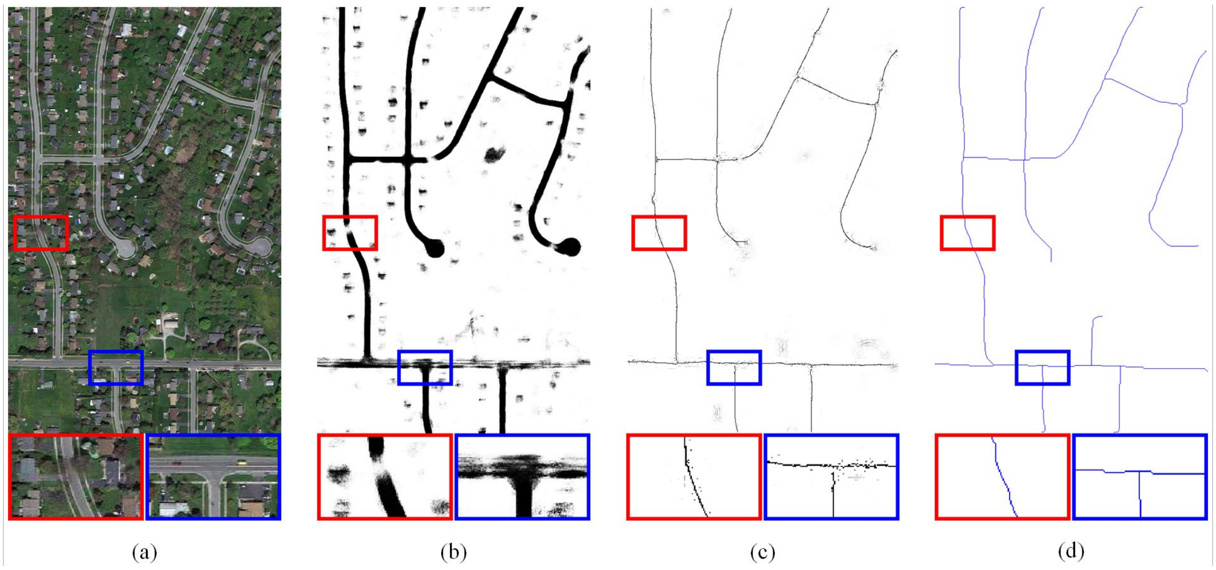


Figure 3. Compared with (b) the segmentation and (c) the centerline detection methods, (d) the proposed algorithm well represents topological features of the curvilinear structure. Setting a threshold value yields to lose correlated information of the pixels on the reconstructed curvilinear structure. In this example, road network is partially occluded by trees or cars, so that the local measure often fails to detect the underlying curvilinear structures. Although the centerline is able to quantify scale (width) of curvilinear structure, it is inaccurate to classify pixels around junctions. In this work we learn spatial patterns of the curvilinear structures with structured output ranking scores. We also propose a graph-based representation algorithm to obtain the topological information.

This work is in collaboration with Dr. Queille-Roussel and Prof. Bahadoran in CHU Nice, France.

Acne vulgaris, a highly prevalent skin disease, has a significant life quality impact on patients. It is generally believed that this type of skin disorder results from proliferation of propionibacterium acnes in pilosebaceous units, which can lead to inflammatory lesions due to increase of oxyhemoglobin level. So far there is no golden standard for acne diagnosis in clinics. It entirely depends on dermatologists' experience for acne assessment. But significant variability among individual diagnosis may lead to less trustworthy results, and less reproducibility of human evaluation makes the comparison of acne follow-up difficult. This work, incorporating the knowledge of optical characteristics of human skin, identifies cutaneous chromophore distribution using bilateral decomposition. Then the inflammatory acne lesions are detected by a Markov random field (MRF) model associating the chromophore descriptors. Experimental results (see Figure 4) show that the proposed method is robust to large dynamic range intensity, and the derived automatic segmentation of inflammatory acne appears to be highly consistent to human visual assessment. This research work was started in 2013. This year, more tests have been conducted on polarized and non-polarized images.

Figure 4. Acne detection using proposed method. (Left) Original image provided by CHU Nice. (Right) Acne detection result.

7.3.2. Finer registration of facial wrinkles in time series images

Participants: Nazre Batool, Josiane Zerubia [contact].

Dr. Batool was funded by the Inria-DPEI fellowship for the period Feb. 2014 – May 2015. Currently she is a postdoc researcher at CMIV, Linköping University [www.liu.se/cmiv], Linköping, Sweden.

The goal of this work is to evaluate quantitatively the subtle variations in facial wrinkles for the same subject in response to treatment using image-based analysis. Any image analysis technique for the analysis of such subtle image variations would require high accuracy and precision for good performance. As in other imaging problems geared towards detection of temporal changes, accurate registration of key image features (wrinkles) is mandatory as a first step. We propose to compare image features in key wrinkle sites only while excluding the noise introduced by changes in surrounding skin texture. Therefore, previously we proposed a 2-step registration algorithm where the initial registration was based on the alignment of facial landmarks such as corners of eyes, nose, and mouth. Then a method based on Large Deformation Diffeomorphic Metric Mapping (LDDMM) was used to achieve finer local registration for wrinkles. However, the LDDMM algorithm had the shortcoming of the unavailability of time invariant finer facial landmarks and that the deformations were guided by image intensities which were varying among images as well due to subtle changes in skin texture. The deformation of skin due to underlying movement can be categorized loosely as locally rigid because the local skin texture remains constant but globally non-rigid because of the movement of skin areas due to slight expression and misalignment. Due to this dual nature of deformation, registration schemes such as thin plate spline or affine transformations are not applicable. Our improved approach is to guide the LDDMM registration on skin features with higher intensity gradients only (such as due to moles, wrinkles, rough surface) which have the higher probability of being constant and detected across temporal changes. First we detect key landmarks and landmark correspondences using the Gabor feature images where the phase correlation is used to find estimates of landmark correspondences. The phase correlation is based on the well-known Fourier shift property i.e. a shift in the spatial domain of two images results in a linear phase difference in the frequency domain of their respective Fourier Transforms. Figure 5 shows Gabor features of two images captured 4 weeks apart in (a) and (b). Figure 5 (c) shows key landmarks placed at high Gabor amplitude sites and (d) shows their corresponding landmarks detected using Fourier phase correlation.

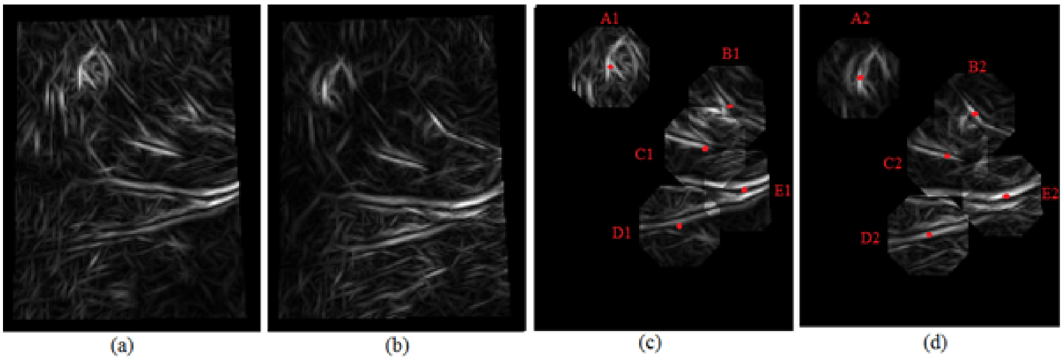


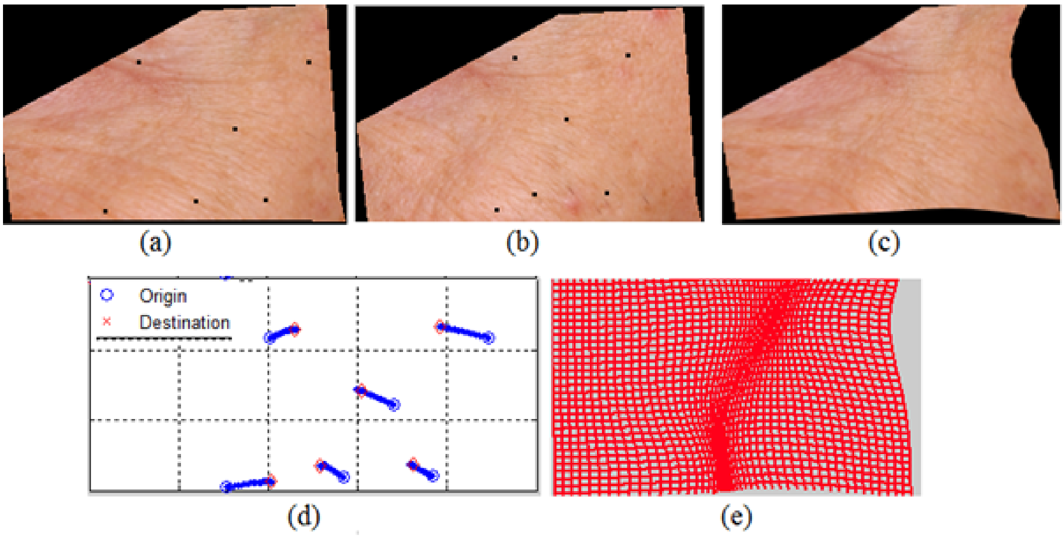
Figure 5. Detection of landmarks and correspondences. (a) Gabor response for source image. (b) Gabor response for target image. (c) 5 landmarks with circular templates in source image. (d) Corresponding detected landmarks with circular templates in target image.

Then, as a next step, the detected key landmarks and their corresponding positions are used in the landmark based LDDMM algorithm to find locally non-rigid deformations between two images. Figure 6 shows an example where the corresponding landmarks are shown as black dots in (a) and (b). Fig. 5 (c) shows the image in (a) wrapped to (b) using LDDMM based on landmark correspondences. In (d) the drifts of landmarks are shown during the LDDMM algorithm and (e) shows the non-rigid deformation of underlying image grid. In the future, the proposed wrinkle registration algorithm will be used to compare wrinkle intensities in time series of images to quantify very minute changes in wrinkles in response to dermatological treatments.

7.3.3. Hyperspectral Image Processing for Detection and Grading of Skin Erythema

Participants: Ali Madooei, Josiane Zerubia [contact].

Ali Madooei worked at Inria Sophia Antipolis on an internship funded by the Canadian Mitacs Globalink Research Award & Inria. He is currently in his last year of PhD at Simon Fraser [www.sfu.ca/] University, Canada. This work has been conducted in collaboration with Ramy M. Abdlaty, Lilian Doerwald-Munoz, Dr. Joseph Hayward and Prof. Qiyin Fang from Mc Master university [http://future.mcmaster.ca/]/Juravinsky cancer center [www.jcc.hhsc.ca/], Canada, and Prof. Joseph Hayward from Simon Fraser University, Canada.



(a) (b) (c)

(d) (e)

Figure 6. Finer registration of micro features in skin images. (a) Source image with landmarks shown as black dots. (b) Target image with landmarks. (c) Source image wrapped to target image. (d) Path of landmarks during LDDMM registration. (e) Deformation of underlying grid during LDDMM registration.

Acute skin erythema is a common side effect in patients undergoing radiotherapy treatment. It displays itself as an increase in skin redness and irritation. Erythema has been reported to correlate to individual patient response to radiation and therefore may be useful to guide and modify courses of treatment in a timely manner. Currently, upon visual examination, a qualitative score can be assigned to characterize the severity of erythema, which then may be used for assessing radiation response. Due to the subjective nature of this method, additional non-invasive techniques are needed for more accurate evaluation. Previous studies have mainly focused on tissue reflectance spectroscopy or imaging photography. The former retrieves spectral information from point measurements while the latter is obtained with conventional Red, Green, Blue(RGB) colour cameras. Photography has the advantage of offering spatial information but this comes at the cost of losing much of spectral information. We use hyperspectral imaging (HSI) which provides both spatial and spectral representation of the affected area. A hyperspectral camera effectively divides the spectrum into very many thin image slices (the actual number depending on the camera and application see Fig.7). This fine-grained slicing reveals spectral structure that may not be evident to the eye or to an RGB camera but can provide a rich set of information for image processing. As an emerging imaging modality for medical applications, the combination of HSI devices with adequate image processing techniques offers the perfect landscape for developing new methods for noninvasive disease monitoring and diagnosis.

The purpose of our study was to investigate the possibility of monitoring the degree of erythema using HSI data. To this aim, we proposed an image processing pipeline and conducted controlled experiments to demonstrate the efficacy of the proposed approach for (1) reproducing clinical assessments, and (2) outperforming RGB imaging data. We combined the problem of erythema detection and grading into a multi-class classification problem where each pixel is classified as one of the four erythema classes or a non-erythema class. We used a weighted LDA (linear discriminant analysis) classifier to deal with noisy labels. Moreover, we devised pre-processing steps to deal with noisy measurements. We evaluated the system against the clinical assessment of an experienced clinician. We also compared the performance to that of using digital photography (instead of HSI). The results from this preliminary study are encouraging and indicate that hyperspectral image data do contain relevant information, and indeed outperform imaging photography. In the future, we want to

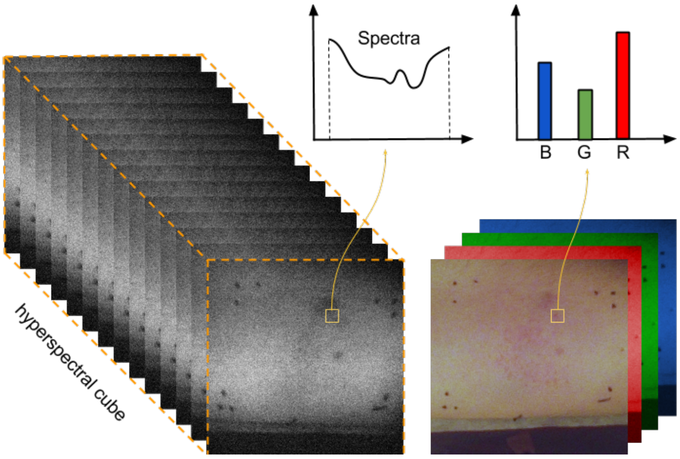


Figure 7. A schematic representation of hyperspectral vs. RGB image data.

extend the technique to further detect other skin responses to radiation (such as dry/moist desquamation, skin necrosis, etc.) and also to experiment with real patients undergoing radiotherapy. Our ultimate objective is to build a system for monitoring radiation response in individuals using HSI technology and image processing.

7.3.4. SAR data classification using generalized Gamma mixture model

Participants: Vladimir Krylov, Josiane Zerubia [contact].

Vladimir Krylov is a former AYIN post-doc, now post-doc at DITEN department, University of Genoa [www.diten.unige.it/], Italy. This work has been performed in collaboration with Prof. Heng-Chao Li, Prof. Ping-Zhi Fan (Southwest Jiaotong University, Chengdu [english.swjtu.edu.cn/], China) and Prof. William Emery (University of Colorado [www.colorado.edu/], Boulder, USA).

The accurate statistical modeling of synthetic aperture radar (SAR) images is a crucial problem in the context of effective SAR image processing, interpretation and application. In this work a semi-parametric approach is designed within the framework of finite mixture models based on the generalized Gamma distribution (GΓD) in view of its flexibility and compact analytical form. Specifically, we have developed a generalized Gamma mixture model (GΓMM) to implement an effective statistical analysis of high-resolution SAR images and proved the identifiability of such mixtures. A low-complexity unsupervised estimation method has been derived by combining the proposed histogram-based expectation-conditional maximization algorithm and the Figueiredo-Jain mixture estimation algorithm. This resulted in a numerical maximum likelihood (ML) estimator that can simultaneously determine the ML estimates of component parameters and the optimal number of mixture components. The state-of-the-art performance of the proposed method has been validated experimentally on a wide range of high-resolution SAR amplitude and intensity images.

In Fig. 8 we demonstrate a typical result of the developed statistical modeling technique on a portion of a 2 meter resolution L-band image acquired by an airborne EMISAR system. The unsupervised GΓMM estimate contains five components and reports a very accurate result that outperforms the considered benchmark statistical modeling methods. In order to visualize the estimated five statistical components we also report a maximum likelihood classification map.

7.3.5. Multitemporal image change detection with a False Discovery Rate approach

Participants: Vladimir Krylov, Josiane Zerubia [contact].

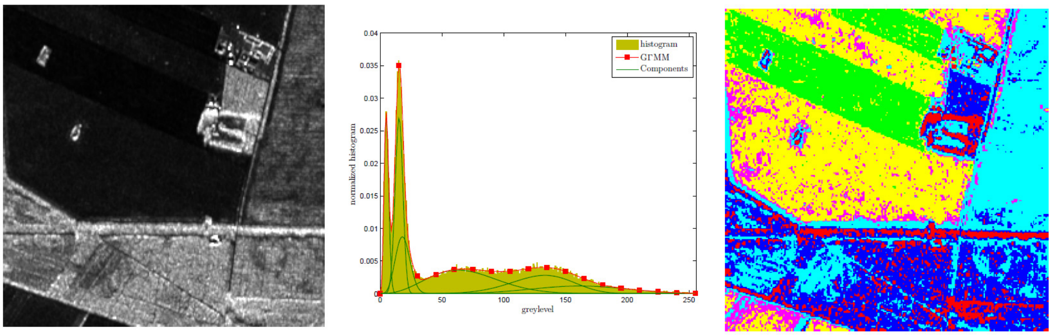


Figure 8. Statistical modeling of a EMISAR (©ESA) image (left) by generalized Gamma mixture model (middle) and its visualization by maximum likelihood classification (right).

This work has been performed in collaboration with Prof. Sebastiano Serpico and Prof. Gabriele Moser, DITEN department, University of Genoa [www.diten.unige.it/], Italy.

Multitemporal change detection is one of the fundamental image processing problems and multiple detection, monitoring and tracking applications rely on its accurate and timely performance. In this work we address the problem of unsupervised change detection on two or more coregistered images of the same object or scene at several time instants. The designed method is appropriate for short image sequences with a relatively small amount of changes. Such analysis is instrumental in various applications where acquisitions are relatively sparse and report limited meaningful changes, in particular, in remote sensing and medical image processing. We develop a novel patch-based hypothesis testing approach which is based on a false discovery rate formulation for statistical significance testing. This alternative error metric allows to adjust the family-wise error rate by imposing control over the proportion of the false positives in the detection. The designed change detector allows the use of various statistical features. The appropriate choice of the latter enables the detector to address application-specific detection problems with a particular set of disturbance factors, like noise, illumination variation, etc. In particular, we demonstrate the use of two rank-based statistics for change detection on image pairs and one multisample statistic for the analysis of image sequences. The experiments with remotely sensed radar, dermatological, and still camera surveillance imagery demonstrate competitive performance and flexibility of the proposed method.

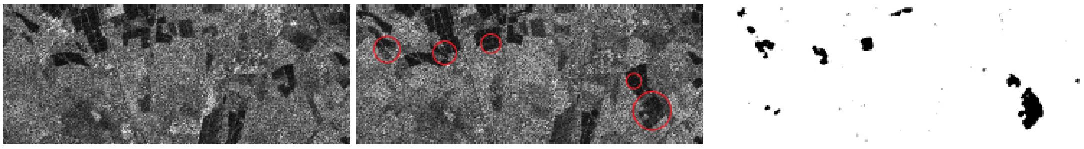


Figure 9. Change detection on a pair of 15 meter resolution XSAR images (first and second) obtained with a false discovery rate error metric based on the Cramer-von Mises statistic. The changes are highlighted with red circles (second), and the unsupervised detection result is reported in black (third).

A typical result obtained with the proposed change detection technique is reported in Fig. 9 . The proposed approach gives a unified statistical thresholding procedure to perform change detection based on statistical features that have a known distribution under the no-change hypothesis. This approach is essentially non-parametric and is highly parallelizable.

BIOCORE Project-Team

7. New Results

7.1. Mathematical methods and methodological approach to biology

7.1.1. Mathematical analysis of biological models

7.1.1.1. Mathematical study of semi-discrete models

Participants: Jean-Luc Gouzé, Frédéric Grognard, Ludovic Mailleret, Pierre Bernhard, Elsa Rousseau, Nicolas Bajeux.

Semi-discrete models have shown their relevance in the modeling of biological phenomena whose nature presents abrupt changes over the course of their evolution [96]. We used such models and analyzed their properties in several practical situations that are developed in Section 7.2.3 , some of them requiring such a modeling to describe external perturbations of natural systems, and others to take seasonality into account. External perturbations of interacting populations occur when some individuals are introduced or removed from a natural system, which occurs frequently in pest control applications, either through the direct removal of pests, or through the introduction of biological control agents [71],[27]. Seasonality is an important property of most agricultural systems in temperate environments since the year is divided into a cropping season and a 'winter' season, where the crop is absent, as in our analysis of the sustainable management of crop resistance to pathogens [25] or in the dynamics of plant pathogens [50].

7.1.1.2. Model reduction and sensitivity analysis

Participants: Suzanne Touzeau, Jean-Luc Gouzé, Stefano Casagranda, Victor Bernal Arzola.

Analysis and reduction of biochemical models. Dynamic models representing complex biological systems with numerous interactions can reach high dimensions and include complex nonlinearities. A model reduction method based on process weighing and pruning was developed and implemented on various models (ERK signaling pathway, circadian rythms in Drosophila) [41]. A global sensitivity analysis was performed to check the method robustness against parameter uncertainty and variability. This work is part of Stefano Casagranda's ongoing PhD thesis and is also a collaboration with Bayer (Sophia-Antipolis).

Parameter identification in compartmental systems. In collaboration with F. Dayan (R&D Manager, Dassault Systèmes), we worked on practical problems of identifiability of parameters in linear pharmacokinetic models. This was the subject of the internship of V. A. Bernal [58].

7.1.2. Metabolic and genomic models

Participants: Jean-Luc Gouzé, Madalena Chaves, Ismail Belgacem, Olivier Bernard, Stefano Casagranda, Francis Mairet, Sofia Almeida.

7.1.2.1. Continuous models analysis

Piecewise quadratic systems for studying growth rate in bacteria. These new systems (first introduced in [82]) use an expression for growth rate that may depend on any number of variables and have several quadratic modes. Relative to the "classical" piecewise affine systems, this new formulation allows the existence of sliding motion as well as oscillatory behaviour for solutions at the thresholds where the vector fields are opposing [21].

Transcription and translation models in bacteria. We study detailed models of transcription and translation for genes in a bacterium, in particular the model of gene expression of RNA polymerase. We also study other models of the global cellular machinery. This is part of the PhD theses of Ismael Belgacem [11] and Stefano Casagranda, and done in collaboration with Inria IBIS project-team, in particular with D. Ropers.

Design of a bistable switch to control cellular uptake. In a joint work with Diego Oyarzún (Imperial College), we analyse the construction of a synthetic bistable system using an unbranched metabolic chain with a global enzyme regulator, as an application of [109]. Bistability can be achieved by choosing an appropriate pattern of regulation. Robustness conditions are given in terms of the promoter dynamic ranges to guarantee a bistable uptake flux [35].

A reduced model for the mammalian cell cycle. We focused on identifying and analyzing the main mechanisms behind the cell cycle and proposed a mathematical model to describe them. This reduced model successfully reproduces oscillatory behaviors including the progress towards a mitosis phase, and then mitosis itself, characterized by an increase in cyclin B. The model was the topic of a poster at the Signalife Workshop [68]. This is a collaboration with F. Delaunay (Ibv Nice) in the framework of Labex Signalife.

7.1.2.2. Hybrid models analysis

Attractor computation using interconnected Boolean networks. During the visit of Daniel Figueiredo, we have worked on an extension of the method proposed in [83]. The idea is to not only use the attractors but also an appropriate set of strongly connected components in the computation of the asymptotic graph [115]. Numerical simulations show a great improvement in the problem of discarding spurious attractors.

Periodic orbits in non monotonic negative feedback circuits. We study the occurrence of periodic solutions in an n-dimensional class of negative feedback systems defined by smooth vector fields with a window of not necessarily monotonic activity. By circumscribing the smooth system by two piecewise linear ones, we show there exists an invariant toroidal region which contains a periodic orbit of the original smooth system [37].

7.1.2.3. Estimation and control

Optimal allocation of resources in a bacterium. We study by techniques of optimal control the optimal allocation between metabolism and gene expression during growth of bacteria [85], in collaboration with Inria IBIS project-team.

Control of a model of synthesis of a virulence factor. In collaboration with J.-A. Sepulchre (INLN Nice), we model the production of a virulence factor by a bacterium in a continuous stirred tank reactor. The production of this enzyme is genetically regulated, and degrades a polymeric external substrate into monomers. A nonlinear control is built [48].

7.2. Fields of applications

7.2.1. Bioenergy

7.2.1.1. Modelling microalgae production

Participants: Olivier Bernard, Antoine Sciandra, Frédéric Grognard, Ghjuvan Grimaud, Quentin Béchet, David Demory, Hubert Bonnefond, Jean-Philippe Steyer, Francis Mairet.

Experimental developments

Experiments have been carried out to study the effects of nitrogen limitation on the lipid production in microalgae [23] and support model development. These experiments have been carried out in the Lagrangian simulator, under constant or periodic light and temperature, varying the total amount of light dose in the day. The response in terms of storage carbon (triglycerides and carbohydrates) has been measured and correlated to the environment fluctuations.

Other experiments were carried out to reproduce the light signal percept by a cell in a raceway pond [84], derived from hydrodynamical studies [92]. An electronic platform was developed to reproduce this high frequency light signal. The experiments show that the microalgae adapt their pigments to the average light that they have received [23]. Experiments with coloured light demonstrated that the growth rate results from the absorbed light, whatever its wavelength.

A new methodology to measure cell viability has been set up. This approach is very promising to distinguish between net and gross growth rate [20]. It was used in the models to assess the impact of temperature on growth and mortality. The mortality turns out to increase exponentially with temperature. The effect of a short term temperature stress was also tested to understand the consequences of a temperature peak in a cultivation system. Finally, it was shown that microalgae can bear with temperature peaks above $T_{\max}$ if they do not last too long [57].

On top of this, we set up a new experimental platform to carry out pilot experiments with solar light. The platform includes four raceways and the equipment to inoculate and harvest the microalgae [60]. We tested the impact of coloured film mimicking possible photovoltaic material. The collected data were used to calibrate models integrating the light spectrum [64].

These works have been carried out in collaboration with A. Talec, S. Rabouille, and E. Pruvost (CNRS/UPMC -Oceanographic Laboratory of Villefranche-sur-Mer LOV).

In collaboration with the IFREMER-PBA team (Nantes) we contributed to a study on the efficiency of dyes (BODIPY and Nile red) to quantify lipid content in microalgae [38].

Metabolism of carbon storage and lipid production

A macroscopic model for lipid production by oleaginous microalgae [7] has been previously proposed. This model describes the accumulation of neutral lipids (which can be turned into biofuel), carbohydrates and structural carbon. A review of the microalgal metabolism reconstruction [15] together with the associated metabolic models has been carried out [14]. A metabolic model has been set up and validated for the microalgae *Isochrysis luthea*. It predicts carbohydrate and lipid accumulation, under conditions of light/dark cycles and/or nitrogen deprivation [72], [1]. A model was developed to represent heterotrophic growth on a mixture of acetate and butyrate [39]. A metabolic model was set up, on the basis of the DRUM framework [1], in order to simulate autotophic, heterotropic and mixotrophic growth, and to determine how to reduce substrate inhibition.

Modelling the coupling between hydrodynamics and biology

In collaboration with the Inria ANGE team, a model coupling the hydrodynamics of the raceway (based on multilayer Saint Venant system) with microalgae growth was developed [79]. This model is supported by the work of ANGE aiming at improving the multi-layer Saint-Venant approach to more finely represent the hydrodynamics of the raceway [54].

Modelling the photosynthesis response to fast fluctuating light

The impact of hydrodynamics on the light perceived by a single cell was studied thanks to fluid dynamics simulations of a raceway pond [90]. The light signals that a cell experiences at the Lagrangian scale, depending on the fluid velocity, were then estimated. A Droop-Han model was used to assess the impact of light fluctuation on photosynthesis. A new model accounting for photoacclimation was also proposed [34]. Single cell trajectories were simulated, and the effect on photosynthesis efficiency was assessed using models of photosynthesis [91]. These results were compared to experimental measurements where the high frequency light was reproduced [84].

Modeling microalgae production processes

The integration of different models developed within BIOCORE [54], [19], [7] was performed to represent the dynamics of microalgae growth and lipid production in raceway systems, on the basis of the dynamical model developed to describe microalgal growth under light and nitrogen limitations. The strength of this model is that it takes into account the strong interactions between the biological phenomena (effects of light and nitrogen on growth, photoacclimation ...), temperature effect [78], [111] and the radiative transfer in the culture (light attenuation due to the microalgae).

Using these approaches, we have developed a model which predicts lipid production in raceway systems under varying light, nutrients and temperature [107]. This model is used to predict lipid production in the perspective of large scale biofuel production [54]. It was also used to assess the microalgal production potential in France,

when taking into account the actual meteorology on a 2.5 degree grid, for 2012, the use of lands, slope, proximity of nutrients and CO_2 [93].
In the framework of the ANR project Purple Sun, we developed a thermic model of a raceway pond within a greenhouse in order to estimate the culture temperature. We also included in the microalgae model the effect of light wavelength. This model has been calibrated on experimental data from LOV and has been used to support lighting strategy in order to optimize microalgal productivity (a patent on this process has been submitted).

Nitrogen fixation by diazotrophs

The fixation of nitrogen by *Croccosphera watsonii* was represented with a macro metabolic model [87] quantifying te main fluxes of carbon and nitrogen in the cell. The model was calibrated and validated with the data of three experiments carried out with different duration of the light period and daily dose. Extension of the model were studied to include the effect of temperature [61].

This work is done in collaboration with Sophie Rabouille (CNRS-Oceanographic Laboratory of Villefranche-sur-Mer LOV).

Modelling thermal adaptation in microalgae

An extended statistical analysis was carried out on a database representing the temperature response of more than 200 microalgal species. First the model proposed by [78] turned out to properly reproduce the temperature response. A model was then extracted to predict the observed link between the cardinal temperatures. This lead to the reduction of the parameter number down to 2, with still a good prediction capability.

We have used Adaptive Dynamics theory to understand how temperature drives evolution in microalgae. For a constant temperature, we have shown that the optimal temperature trait tends to equal the environment temperature. We then studied the case where temperature is periodically fluctuating [88]. We now use this method at the scale of the global ocean, validating our approach with experimental data sets from 194 species [42], [49].

7.2.1.2. Control and Optimization of microalgae production

On-line monitoring

Interval observers give an interval estimation of the state variables, provided that intervals for the unknown quantities (initial conditions, parameters, inputs) are known [86]. Several developments were carried out in this direction to improve the design and performances of interval observers, and accounting for a specific structure (*i.e.* triangular) or property (*i.e.* Input to State Stable), [104]. Interval observers were designed for the estimation of the microalgae growth and lipid production within a production process [101][54] and validated experimentally [100][29].

Optimization of the bioenergy production systems

Based on simple microalgae models, analytical optimization strategies were proposed. We assessed strategies for optimal operation in continuous mode using the detailed model for raceways [106], [107]. We first solved numerically an optimal control problem in which the input flow rate of the raceway is calculated such that the productivity in microalgae biomass is maximized on a finite time horizon. Then, we re-analysed the optimization problem and derived a simplified strategy to reach biomass productivities very near to the maximal productivities obtained with the optimal control. These approaches were extended to outdoor cultivation, considering a possible variable culture depth. The optimal strategy for both depth and dilution rate was proposed in order to better manage the process inertia and finally avoid over warming periods. This work was done during the doctoral stay of Riccardo de Luca (Univ. Padova).

We also propose a nonlinear adaptive controller for light-limited microalgae culture, which regulates the light absorption factor (defined by the ratio between the incident light and the light at the bottom of the reactor). We show by numerical simulation that this adaptive controller can be used to obtain near optimal productivity under day-night cycles [31].

Interactions between species

Large scale culture of microalgae for bioenergy involves a large biodiversity. Control of such systems requires to consider the interactions between the different species. Such systems involve bacteria and microalgae, and the competition between these organisms can have several equilibrium points, which can be studied with Monod, Contois and Droop type models [28].

In the framework of the ANR Facteur 4 project, we propose to drive this competition exploring different strategies in order to select species of interest.

We had formerly proposed an adaptive controller which regulates the light at the bottom of the reactor [102]. When applied for a culture with n species, the control law allows the selection of the strain with the maximum growth rate for a given range of light intensity. This is of particular interest for optimizing biomass production as species adapted to high light levels (with low photoinhibition) can be selected. We have also proposed a strategy based on light stresses in order to penalize the strains with a high pigment content and finally select microalgae with a low Chlorophyll content [12]. This characteristic is of particular interest for maximizing biomass production in dense culture. The strategy has been carried out at the LOV and eventually the productivity of *Tisochrysis lutea* was improved by 75% [62]. A patent on this strategy is under submission. Strategies to improve the temperature response have also been proposed. First we modelled the adaptive dynamics for a population submitted to a variable temperature [88]. This was used at the LOV to design experiments with periodic temperature stresses during 200 days aiming at enlarging the thermal niche of *Tisochrysis lutea*. It resulted in an increase by 2 degrees of the thermal niche [12].
Finally, in a more theoretical framework, we studied how to select as fast as possible a given species in a chemostat with two species at the initial instant. Using the Pontryagin maximum principle, we have shown that the optimal strategy is to maintain the substrate concentration to the value maximizing the difference between the growth rates of two species [73]. We now try to extend this result for n species with mutations.

7.2.2. Biological depollution

7.2.2.1. Control and optimization of bioprocesses for depollution

Participants: Olivier Bernard, Francis Mairet, Jean-Luc Gouzé.

We have considered the problem of global stabilization of an unstable bioreactor model (e.g. for anaerobic digestion), when the measurements are discrete and in finite number ("quantized"). These measurements define regions in the state space, wherein a constant dilution rate is applied We show that this quantized control may lead to global stabilization: trajectories have to follow some transitions between the regions, until the final region where they converge toward the reference equilibrium [30].

Although bioprocesses involve an important biodiversity, the design of bioprocess control laws are generally based on single-species models. In [98], we have proposed to define and study the multispecies robustness of bioprocess control laws: given a control law designed for one species, what happens when two or more species are present? We have illustrated our approach with a control law which regulates substrate concentration using measurement of growth activity. Depending on the properties of the additional species, the control law can lead to the correct objective, but also to an undesired monospecies equilibrium point, coexistence, or even a failure point. Finally, we have shown that, for this case, the robustness can be improved by a saturation of the control.

Moreno [105] have proposed an optimal strategy for fed-batch bioreactor with substrate inhibition. Thanks to the Pontryagin maximum principle and the Hamilton-Jacobi equation, we have shown that the same strategy is still optimal when mortality is included in the model [75]. We have also studied the problem when the singular arc is non-necessary admissible everywhere (i.e. the singular control can take values outside the admissible control set). We have pointed out the existence of a frame point on the singular arc above which any singular trajectory is not globally optimal. Moreover, we have provided an explicit way for computing numerically the switching curves and the frame point [17].

7.2.2.2. Coupling microalgae to anaerobic digestion

Participants: Olivier Bernard, Antoine Sciandra, Jean-Philippe Steyer, Frédéric Grognard, Francis Mairet.

The coupling between a microalgal pond and an anaerobic digester is a promising alternative for sustainable energy production and wastewater treatment by transforming carbon dioxide into methane using light energy. The ANR Phycover project is aiming at evaluating the potential of this process [113], [112].

In a first stage, we developed models for anaerobic digestion of microalgae. Two approaches were used: first, a dynamic model has been developed trying to keep a low level of complexity so that it can be mathematically tractable for optimisation [97] . On the other hand, we have tested the ability of ADM1 [114] (a reference model which considers 19 biochemical reactions) to represent the same dataset. This model, after modification of the hydrolysis step [99] has then been used to evaluate process performances (methane yield, productivity...) and stability though numerical simulations.

We have proposed and analysed a three dimensional model which represent the coupling of a culture of microalgae limited by light and an anaerobic digester. We first prove the existence and attraction of periodic solutions. Applying Pontryagin's Maximum Principle, we have characterized optimal controls, including the computation of singular controls, in order to maximize methane production. Finally, we have determined numerically optimal trajectories by direct and indirect methods [74].

Finally, we have studied the coupling between three ecosystems: an anaerobic digester, a wastewater treatment pond (with microalgae and nitrifiers) and a microalgal pond. Different possible coupling configurations were tested in simulation. A numerical optimization was carried out to identify, depending on the choice of the objective function (energy production, pollution removal) the optimal arrangement between the three processes. The optimal volume for each process was then determined. This work has been carried out in the framework of the Phycover ANR project and was the subject of the internship of Ignacio Lopez (Universidad de Chile).

7.2.2.3. Life Cycle Assessment

Participants: Olivier Bernard, Jean-Philippe Steyer.

This work is the result of a collaboration with Arnaud Helias of INRA-LBE (Laboratory of Environmental Biotechnology, Narbonne) and Pierre Collet (IFPEN).

In the sequel of the pioneering life cycle assessment (LCA) work of [94], we continued to identify the obstacles and limitations which should receive specific research efforts to make microalgae production environmentally sustainable.
The improvements due to technological breakthrough (leading to higher productivities) have been compared to the source of electricity. It turns out that the overall environmental balance can much more easily be improved when renewable electricity is produced on the plant [36]. As a consequence, a new paradigm to transform solar energy (in the large) into transportation biofuel is proposed, including a simultaneous energy production stage. This motivated the design of the purple sun ANR-project where electricity is produced by semi transparent photovoltaic panels [77] under which microalgae are growing.

Finally, some work are aiming at normalising LCA for microalgae and proposing guidelines to make the LCA more easily comparable [22].

These works have been carried out in collaboration with E. Latrille and B. Sialve (INRA-LBE).

7.2.3. Design of ecologically friendly plant production systems

7.2.3.1. Controlling plant pests

Participants: Frédéric Grognard, Ludovic Mailleret, Suzanne Touzeau, Nicolas Bajeux.

Optimization of biological control agent introductions

The question of how many and how frequently natural enemies should be introduced into crops to most efficiently fight a pest species is an important issue of integrated pest management. The topic of optimization of natural enemies introductions has been investigated for several years [6] [108], unveiling the crucial influence of within-predator density dependent processes. Since parasitoids may be more prone to exhibit positive density dependent dynamics rather than negative ones, which are prevalent among predatory biocontrol agents, the current modeling effort consists in studying the impact of positive predator-predator interactions on the optimal introduction strategies (PhD of Nicolas Bajeux, [70], [71]).

The influence of the spatial structure of the environment on biological control efficacy has also been investigated; first results indicate that spatial structure tends to influence it in quite a same way as intra-specific competition does [27].

Connected research on the influence of space on the establishment of biological control agents is also being pursued both through computer simulations and laboratory experiments on parasitoids of the genus *Trichogramma*. This is the topic of the PhD thesis of Thibaut Morel Journel (UMR ISA) [13]; in particular, we showed how landscape connectivity or spatial heterogeneity shape establishment dynamics in spatially structured environments [33], [51], [40]. This research linked to invasion biology also led some of us to contribute with opinion or review contributions to a special issue on biological invasions, in connexion with the GdR Invabio [26], [32].

7.2.3.2. *Controlling plant pathogens*

Participants: Frédéric Grognard, Ludovic Mailleret, Suzanne Touzeau, Elsa Rousseau, Mélanie Bonneault.

Sustainable management of plant resistance

Because plants can get sick, we studied other plant protection methods dedicated to fight plant pathogens. One such method is the introduction of plant strains that are resistant to one pathogen. This often leads to the appearance of virulent pathogenic strains that are capable of infecting the resistant plants. It is therefore necessary to find ways to protect the durability of such resistances, which are a natural exhaustible resource. We looked at landscape scale spatial deployment strategies of resistant crops able to maximize crop yield [25], allowing for the modification of the spatial arrangement of resistant crops over cropping seasons, showing dramatic increases in crop yield in particular epidemic situations [25].

Experiments were also conducted in INRA Avignon, followed by high-throughput sequencing (HTS) to identify the dynamics of virus strains competing within host plants. Different plant genotypes were chosen for their contrasted effects on genetic drift and selection they induce on virus populations. Those two evolutionary forces can play a substantial role on the durability of plant resistance. Therefore we fitted a mechanistic-statistical model to these HTS data in order to disentangle the relative role of genetic drift and selection during within-host virus evolution [53], [69], [43], [44]. A stochastic model was also produced to simulate the effect of drift on the virus epidemiological dynamics and on the durability of qualitative resistances [59]. This is the topic of Elsa Rousseau's PhD thesis, and is done in collaboration with Frédéric Fabre (INRA Bordeaux) and Benoît Moury (INRA Avignon).

We also developed an epidemiological model describing the dynamics of root-knot nematodes in a protected vegetable cropping system, to design optimal management strategies of crop resistance [110]. The model was fitted to experimental and field data. Preliminary results show that alternating susceptible and resistant crops not only increased the resistance durability, but reduced the disease intensity over time [63].

Finally we developed an epidemiological model including non-conventional gene-for-gene interactions in crops, based on the phoma stem canker of oilseed rape, to assess the durability of crop resistance in the field and design efficient deployment strategies [65]. This ongoing work is part of the K-Masstec project, which also incorporates experimental and field studies in collaboration with BIOGER (INRA Grignon).

Eco-evolutionary dynamics of plant pathogens in seasonal environments

Understanding better pathogen evolution also requires to understand how closely related plant parasites may coexist. Such coexistence is widespread and is hardly explained through resource specialization. We showed that, in agricultural systems in temperate environments, the seasonal character of agrosystems is an important force promoting evolutionary diversification of plant pathogens [89]. The plant parasites reproduction mode may also strongly interact with seasonality. In this context, we investigated the special case of oak powdery mildew, an oak disease which is actually caused by a complex of two different species, combining original plant epidemic data with the semi-discrete seasonal plant epidemic model we introduced a few years ago [50] [95]

This work has been done in collaboration with Frédéric Hamelin (Agrocampus Ouest) during Anne Bisson's internship, Marie Laure Desprez Loustau and Frederic Fabre (INRA Bordeaux).

7.2.3.2.1. Optimality/games in population dynamics

Participants: Frédéric Grognard, Ludovic Mailleret, Pierre Bernhard.

Optimal foraging and residence times variations

A continued collaboration with Vincent Calcagno (UMR ISA) has yielded paper where we reanalyzed the so-called Marginal Value Theorem (MVT), first published in 1976 [80], [81]. Ongoing work aims at pointing how this latter theorem has been misused in some biological literature.

We also investigated the problem in foraging theory of evaluating the expected harvest of an animal when conspecifics may arrive on the same patch of resource in a stochastic fashion, specificaly according to a Poisson process or a Bernoulli process. A joint article with Frédéric Hamelin (Agrocampus Ouest) has been submitted for publication.

With Marc Deschamps, similar questions were studied in theoretical economy in the context of a Cournot competition on a single market. Again, an article has been submitted for publication.

The handicap paradox

We have investigated the question of "how could evolution have reached a state characterized by the handicap paradox?" with the tools of adaptive dynamics. We have reached the conclusion that, if one accepts adaptive dynamics as a model of evolution, and our model of sexual selection, the handicap paradox equilibrium is indeed the limit state of evolution [18].

This work was conducted with Frédéric Hamelin (Agrocampus Ouest).

CASTOR Project-Team

5. New Results

5.1. Plasma boundary reconstruction

Participants: Jacques Blum, Cédric Boulbe, Blaise Faugeras.

A new fast and stable algorithm has been developped for the reconstruction of the plasma boundary from discrete magnetic measurements taken at several locations surrounding the vacuum vessel. The resolution of this inverse problem takes two steps. In the first one we transform the set of measurements into Cauchy conditions on a fixed contour Γ_O close to the measurement points. This is done by least square fitting a truncated series of toroidal harmonic functions to the measurements. The second step consists in solving a Cauchy problem for the elliptic equation satisfied by the flux in the vacuum and for the overdetermined boundary conditions on Γ_O previously obtained with the help of toroidal harmonics. It is reformulated as an optimal control problem on a fixed annular domain of external boundary Γ_O and fictitious inner boundary Γ_I. A regularized Kohn-Vogelius cost function depending on the value of the flux on Γ_I and measuring the discrepancy between the solution to the equation satisfied by the flux obtained using Dirichlet conditions on Γ_O and the one obtained using Neumann conditions is minimized. The method presented here has led to the development of a software, called VacTH-KV, which enables plasma boundary reconstruction in any Tokamak (see [14]).

5.2. Free boundary - Transport Solver - Controller coupling

Participants: Cédric Boulbe, Blaise Faugeras, Jean François Artaud [IRFM CEA Cadarache], Vincent Basiuk [IRFM CEA Cadarache], Emiliano Fable [Max-Planck-Institut für Plasmaphysik, Garching], Philippe Huyn [IRFM CEA Cadarache], Eric Nardon [IRFM CEA Cadarache], Jakub Urban [IPP, Academy of Sciences of the Czech Republic, Prague].

Last year, a first version of the workflow coupling a free boundary equilibrium code, the European transport solver ETS and a plasma shape and position controller had been developped. In 2015, this new tool has been tested and improved. An experiment realized on the Tokamak TCV and called "yoyo" shot has been successfully simulated. This work has been realised in the framework of the Eurofusion Work Package: Code Development for integrated modelling project.

5.3. A finite element method with overlapping meshes for free-boundary toroidal plasma equilibria in realistic geometry

Participants: Holger Heumann, Francesca Rappetti.

Existing finite element implementations for the computation of free-boundary toroidal plasma equilibria approximate the flux function by piecewise polynomial, globally continuous functions. Recent numerical results for the self-consistent coupling of equilibrium and resistive diffusion in the spirit of Grad-Hogan suggest the necessity of higher regularity. Enforcing continuity of the gradient in finite elements methods on triangular meshes, leads to a drastic increase in the number of unknowns, since the degree of the polynomial approximation needs to be increased beyond four. Therefore existing implementations for the fixed boundary problem resort to (curvilinear) quadrilateral meshes and approximation spaces based on cubic Hermite splines. Fine substructures in the realistic geometry of a tokamak, such as air-gaps, passive structures and the vacuum vessel prevent the use of quadrilateral meshes for the whole computational domain, as it would be necessary for the free-boundary problem.

In this work we propose a finite element method that employs two meshes, one of quadrilaterals in the vacuum domain and one of triangles outside, which *overlap* in a narrow region around the vacuum domain. This approach gives the flexibility to achieve easily and at low cost higher order regularity for the approximation of the flux function in the domain covered by the plasma, while preserving accurate meshing of the geometric details exterior to the vacuum. The continuity of the numerical solution in the region of overlap is weakly enforced by relying on the mortar projection. A publication is in preparation.

5.4. Inverse transient plasma equilibrium problem

Participants: Holger Heumann, Jacques Blum.

The inverse transient plasma equilibrium problem aims at precomputing the trajectories of externally applied voltages in the poloidal field coils of a tokamak. A basic implementation of this problem in 2011/2012 in CEDRES++ during Holger Heumann's PostDoc at Inria, provided first insight into the capabilities and also difficulties of this approach. Application engineers are highly interested in this application, but realistic cases will require more evolved numerical methods to reduce the computational time and memory requirements. In 2014 we implemented the inverse transient plasma equilibrium problem in FEEQS.M to facilitate our search for better algorithms. In 2015 we started working on realistic test cases for the upcoming WEST tokamak. In order to make such problems accessible by the current version of our code, we had to split the time interval of interest into 5 subintervals, on which we solve 5 inverse problems. Only by the initial condition the problem on a subinterval is connected to its predecessor. Next we faced some serious convergence problem of the optimisation algorithms for some of these problems. These led us to do extensive benchmark runs with different optimisation algorithms and implementation, including both Gradient and SQP-type methods, either with handcoded or MATLAB-native implementations. As a result we envisage for 2016 the incorporation of the SQP implementation of Jean-Charles Gilbert, which seems to be perfectly adapted to optimal control problems such as ours. Another improvement was achieved in reducing the actual number of free control parameters and to replace piecewise linear control trajectories with high order polynomials.

5.5. High order for the axisymmetric magnetohydrodynamic equilibrium problem

Participants: Holger Heumann, Lukas Drescher [TU Berlin], Kersten Schmidt [TU Berlin].

We implemented a higher order finite element method (FEM) for solving numerically axisymmetric magnetohydrodynamic (MHD) equilibrium problems. The focus is on high accuracy and the capabilities of high-order FEM implementations for faster calculations. High order FEM for elliptic problems, such as the considered MHD equilibrium problem, is well established and understood. This work uses the hp-FEM software CONCEPTS developed at ETH Zürich/TU Berlin. Further, we developed a novel method for computing accurately the so-called flux surface averages, that are important in transient MHD calculations. This new method circumvents the expensive and very technical computation of line-integrals and fits seamlessly into the high order finite element method. A publication is in preparation.

5.6. Towards automated magnetic divertor design for optimal heat exhaust

Participants: Holger Heumann, Maarten Blommaert [FZ, Jülich (Germany)], Martine Baelmans [KU Leuven (Belgium)], Nicolas R. Gauger [TU, Kaiserslautern (Germany)], Detlev Reiter [FZ, Jülich (Germany)].

Avoiding excessive structure heat loads in future fusion tokamaks is regarded as one of the greatest design challenges. In this joint effort, we aim at developing a tool to study how the severe divertor heat loads can be mitigated by reconfiguring the magnetic confinement. For this purpose, the free boundary equilibrium code FEEQS.M was integrated with a plasma edge transport code to work in an automated fashion. A practical and efficient adjoint based sensitivity calculation was proposed to evaluate the sensitivities of the integrated code. The sensitivity calculation was applied to a realistic test case and compared with finite difference sensitivity calculations.

The integration of the free boundary equilibrium solver FEEQS.M allowed to assess the validity of a previous simplified model introduced by M. Bloomaert. It was found that the absence of plasma response currents significantly limits the accuracy of this simplified model.

The novel procedure was applied to obtain first results for the new WEST (Tungsten Environment in Steady-state Tokamak) divertor currently under construction in the Tore Supra tokamak at CEA. The sensitivities and the related divertor optimization paths are strongly affected by the extension of the magnetic model (see [24]).

5.7. Bohm boundary conditions

Participants: Richard Pasqueti, Sebastian Minjeaud.

Focusing on a minimal model proposed in the late 2000's by the IRFM (Cadarache), an algorithm has been proposed to enforce at the plates the inequality $M \geq 1$, where M is the parallel Mach number. The algorithm is implemented in the FBGKI code, but still requires improvements to enhance the robustness of the numerical method (see [18]).

5.8. High order approximation of dispersive equations and conservation of invariants

Participants: Richard Pasqueti, Sebastian Minjeaud.

Focusing on the Korteweg-de Vries (KdV) equation, algorithms have been proposed to handle high order derivative terms (third order for KdV) with C^0 elements and to preserve invariants (mass and momentum for KdV) through the time-scheme (see [33]).

5.9. Taylor-Galerkin stabilized Finite Element

Participants: José Costa, Boniface Nkonga.

The theoretical part of Taylor-Galerkin/Variational multi-scales (TG/VMS) strategy applied to MHD and reduced MHD modeling has been achieved last year. The final method amounts to adding in the finite element formulation, a self-adjoint operator associated to the most critical hyperbolic component of the system to be solved. The design of the critical contours and the identification of associated waves to be stabilized is problem dependent and related to the Jacobian matrix. This year we have continued the investigations for the design and improvement of the stabilization started in 2015. For application to plasma configurations with X-point, we have designed a numerical strategy that preserved the initial equilibrium without flows. The Bohm boundary condition on open flux walls has been formulated and is now under validation.

5.10. Toward full MHD numerical modeling with $C1$ finite element.

Participants: José Costa, Giorgio Giorgiani, Hervé Guillard, Boniface Nkonga.

In this context the single fluid full MHD model is considered and the divergence free constraint on the magnetic field is achieved by introduction of a vector potential . The use of the vector potential has the additional advantage that the toroidal component is the magnetic flux of the Grad-Shafranov equilibrium. However, using the vector potential as variable introduces higher order derivatives in the system and classical C0 finite elements cannot be directly applied. This is why our finite element strategies uses shape/test functions whose derivatives have global continuity in space (smooth finite elements). The global approach uses cross product shape/test functions between poloidal(2D) and the Toroidal(1D). In the 2D poloidal plane, discretization uses either quadrangular or triangular elements. In order to derive efficient strategies for the full MHD in the vector potential formulation, the Gauge condition on the vector potential and the boundary conditions have been enforced by penalization. For the Gauge condition it gives rise to element contributions but also boundary integrals that should be computed on curved surfaces that sometime fitted the magnetic surfaces. Equations are formulated in semi-conservative form such as to apply integration by parts. Therefore, boundary conditions can be viewed as evolution of fluxes or variables. Integral formulation on the boundary is very useful for higher order finite elements and also easier for the treatment of corners. Indeed in this context the boundary conditions are edge/surfaces oriented and boundary corners are driven by the neighborhood edge penalization. This strategy is the one that will be used for future developments.

2D Quadrangular Cubic Bezier Finite Elements:
This finite element has been used for a while for reduced MHD models in the software Jorek. Reduced MHD uses the projection of the momentum equation in a space orthogonal to the equilibrium. When full MHD models are used, the momentum equation needs to be projected in the equilibrium space and this projection should be consistent with the Grad-Shafranov equilibrium that is used to compute the initial state. This has been achieved by a proper computation of the JxB contribution in the momentum equation, taking into account the poloidal variation of the toroidal component of the magnetic field. After a detailed analysis, we have performed this year some implementations and numerical validations. An Inria report is under preparation.

2D Triangular Powell-Sabin Finite Elements:
In order to avoid some mesh singularities when using quadrangular meshes for complex geometries and flux surfaces shapes, triangular elements are a possible option. It is not so easy to derive smooth finite elements on triangles with reduced number of degrees of freedom (ddl). The Bell reduced-quintic finite elements we have considered in the previous years have too many unknowns (6 per vertex). Powell-Sabin splines are piecewise quadratic polynomials with a global C1-continuity and 3 unknowns per vertex, they have a local support, they form a convex partition of unity, they are stable, and they have a geometrically intuitive interpretation involving control triangles. Construction of the Powell-Sabin splines needs some geometrical tools that have been developed: Minimum area enclosing triangle of a set of control points (sequential and parallel). This construction is applied to each vertex of the triangular mesh and used to derive the local shape/test functions. These Powell-Sabin splines have been used successfully in the area of computer aided geometric design for the modeling and fitting of surfaces. We have used the Powell-Sabin (PS) splines for the approximation of elliptic partial differential equations (including Grad-Shafranov) in a rectangular domain. In this context, the optimal rate of convergence (order 3) has been recovered. This year, validations have been performed for hyperbolic 2D Euler equations with VMS stabilization. The context of the 3D toroidal geometries has been considered and implemented. Preliminary validations are satisfactory. An Inria report is also under preparation.

5.11. Genuinely multidimensional Riemann Solver

Participants: Jeaniffer Vides, Boniface Nkonga.

Multidimensional Riemann solvers were pioneered by Abgrall. Abgrall, Maire, Nkonga, Després and Loubere have extensively developed them especially as node-solvers for Lagrangian hydrodynamics. Another strain of work comes from explorations by Wendroff and Balsara who took a space-time approach. In this work, the resolved state is obtained via space-time integration over a wave model, just as was done by Wendroff and Balsara. However, an algebraic approach is used for the development of the fluxes. It is, therefore, shown that the multidimensional fluxes can be obtained by application of jump conditions at the boundaries of the wave model. The problem is of course over determined with the result that the shock jump conditions are only satisfied approximately in a least squares sense. Even so, this work gives us new perspective on multidimensional Riemann solvers. The literal satisfaction of the shock jump conditions (up to least squares approximation) makes it easier to understand multidimensional Riemann solvers as a natural extension of the one-dimensional Riemann solvers. Contributions have also been made on the development of a minimalist wave model, which might help in reducing dissipation. Further innovations are reported on the assembling of fluxes based on the structure of the wave model, and those innovations are potentially useful. For MHD the CT approach consists of constraining the transport of magnetic field so that the divergence is always kept zero. The method relies on exploiting the dualism between the flux components and the electric field. Since the electric field is needed at the edges of the mesh, the multidimensional Riemann solver can also provide the electric field. By running an extensive set of simulations, it is shown that the multidimensional Riemann solver is robust and can be used to obtain divergence-free formulations for MHD that perform well on several stringent calculations. The work performed this year was to improve this strategy by enriching of sub-structures the description of the strongly interaction of waves. These improvements were done in collaboration with the invited professor D. Balsara. This work has resulted in an article to be published in the Journal of Computational Physics in 2016.

5.12. Multi scales approximations of "Shallow water" flows.

Participants: Jeaniffer Vides, Boniface Nkonga, Sergey Gavrilyuk, Kseniya Ivanova.

The terminology "Shallow water" is used to characterize thin flows on curved surfaces. It is customary for this type of flows; to use the incompressible Navier-Stokes equations to asymptotically derive reduced models for the evolution of the depth integrated speed and the thickness of the flow. Reduced models are mainly hyperbolic and finite volume method are often used for their numerical approximation. Approximation strategies are generally structured as follow:

- Construction of a global coordinate system associated with an assumed analytical surface.
- Reduction of the model relatively to the global coordinate system
- Approximation of the surface by a finite number of elements.
- Approximation of the reduced model using the discrete surface.

In the context of real applications, it is presumptuous to expect an analytical formulation of the surface. From the data provided by observation satellites, we can usually extract a discrete description of the surfaces that drives thin flow. Therefore, it is more practical to use the discrete description as the starting point of the resolution strategy. This is the angle of approach that we have considered. We locally define two mesh scales: the element scale and the cell scale. The discrete mapping and the reduced model are defined at the element scale and the average values that evolve in time are defined at the cell scale. First applications have been successfully performed. Our efforts have been extended to include relevant physics at each scale, including sheared flows. We have used a multi-dimensional formulation. An Inria report is under preparation.

5.13. Asymptotic theory of reduced MHD models

Participant: Hervé Guillard.

In the study of fusion plasma, one of the fundamental model used for stability studies is the magnetohydrodynamic (MHD) model. Many theoretical and numerical works in this field use specific approximations of this model known as *reduced* MHD models. The derivation of these reduced MHD models has been formulated as a special instance of the theory of singular limit of hyperbolic system of partial differential equations with a large operator. This formulation allows to use the general results of this theory and to prove rigorously that reduced MHD models are valid approximations of the full MHD equations [29]. In particular, it is proven that the solutions of the full MHD system converge to the solutions of an appropriate reduced model. These results substantiate the intuitive physical idea that in the presence of a strong magnetic field, motion in the plane perpendicular to the plasma is nearly incompressible.

5.14. Finite volume approximations for fusion plasma

Participants: Hervé Guillard, Afeintou Sangam, Elise Estibals.

The MHD model used for plasma studies in tokamak is very often based on the magnetic vector potential form of the equations where the vector potential satisfies $\nabla \times \mathbf{A} = \mathbf{B}$ with $\mathbf{B}$ the magnetic field and only a small number of numerical models use the conservative formulation based on $\mathbf{B}$. One of the shortcomings of this latter formulation is the necessity to enforce numerically the divergence free constraint on the magnetic field that can be difficult to achieve and/or computationally costly. Another difficulty is that the equilibrium solution of the MHD equation given by the Grad-Shafranov equation is not an exact solution of the discrete equation.

We have begun to investigate the use of the $\mathbf{B}$ formulation for tokamak studies. The divergence free constraint is taken into account by a projection at each time step on a rotated gradient field. This step ensures a strict respect of the divergence free constraint while being extremely cheap since the scalar field is simply advected by the flow. Preliminary numerical experiments show that this approach can have some interest. The design of a well-balanced solver will be the next step of these studies.

COATI Project-Team

7. New Results

7.1. Network Design and Management

Participants: Jean-Claude Bermond, Christelle Caillouet, David Coudert, Frédéric Giroire, Frédéric Havet, Nicolas Huin, Alvinice Kodjo, Fatima Zahra Moataz, Joanna Moulierac, Nicolas Nisse, Stéphane Pérennes.

7.1.1. Wireless Networks

7.1.1.1. Dimensioning Microwave Wireless Networks

In [47], we aim at dimensioning fixed broadband microwave wireless networks under unreliable channel conditions. As the transport capacity of microwave links is prone to variations due to, e.g., weather conditions, such a dimensioning requires special attention. It can be formulated as the determination of the minimum cost bandwidth assignment of the links in the network for which traffic requirements can be met with high probability, while taking into account that transport link capacities vary depending on channel conditions. The proposed optimization model represents a major step forward since we consider dynamic routing. Experimental results show that the resulting solutions can save up to 45% of the bandwidth cost compared to the case where a bandwidth over-provisioning policy is uniformly applied to all links in the network planning. Comparisons with previous work also show that we can solve much larger instances in significantly shorter computing times, with a comparable level of reliability.

7.1.1.2. Data Gathering and Personalized Broadcasting in Radio Grids with Interference

In the gathering problem, a particular node in a graph, the base station, aims at receiving messages from some nodes in the graph. At each step, a node can send one message to one of its neighbors (such an action is called a call). However, a node cannot send and receive a message during the same step. Moreover, the communication is subject to interference constraints, more precisely, two calls interfere in a step, if one sender is at distance at most dI from the other receiver. Given a graph with a base station and a set of nodes having some messages, the goal of the gathering problem is to compute a schedule of calls for the base station to receive all messages as fast as possible, i.e., minimizing the number of steps (called makespan). The gathering problem is equivalent to the personalized broadcasting problem where the base station has to send messages to some nodes in the graph, with same transmission constraints. In [24], we focus on the gathering and personalized broadcasting problem in grids. Moreover, we consider the non-buffering model: when a node receives a message at some step, it must transmit it during the next step. In this setting, though the problem of determining the complexity of computing the optimal makespan in a grid is still open, we present linear (in the number of messages) algorithms that compute schedules for gathering with $dI \in \{0, 1, 2\}$. In particular, we present an algorithm that achieves the optimal makespan up to an additive constant 2 when $dI = 0$. If no messages are “close” to the axes (the base station being the origin), our algorithms achieve the optimal makespan up to an additive constant 1 when $dI = 0$, 4 when $dI = 2$, and 3 when both $dI = 1$ and the base station is in a corner. Note that, the approximation algorithms that we present also provide approximation up to a ratio 2 for the gathering with buffering. All our results are proved in terms of personalized broadcasting.

7.1.2. Elastic Optical Networks

7.1.2.1. On Spectrum Assignment in Elastic Optical Tree-Networks

To face the explosion of the Internet traffic, a new generation of optical networks is being developed; the Elastic optical Networks (EONs). The aim with EONs is to use the optical spectrum efficiently and flexibly. The benefit of the flexibility is, however, accompanied by more difficulty in the resource allocation problems. In [54], [51], [14], we study the problem of Spectrum Allocation in Elastic Optical Tree-Networks. In trees, even though the routing is fixed, the spectrum allocation is NP-hard. We survey the complexity and approximability results that have been established for the SA in trees and prove new results for stars and binary trees.

7.1.3. Fault Tolerance

7.1.3.1. Shared Risk Link Group

The notion of Shared Risk Link Groups (SRLG) captures survivability issues when a set of links of a network may fail simultaneously. The theory of survivable network design relies on basic combinatorial objects that are rather easy to compute in the classical graph models: shortest paths, minimum cuts, or pairs of disjoint paths. In the SRLG context, the optimization criterion for these objects is no longer the number of edges they use, but the number of SRLGs involved. Unfortunately, computing these combinatorial objects is NP-hard and hard to approximate with this objective in general. Nevertheless some objects can be computed in polynomial time when the SRLGs satisfy certain structural properties of locality which correspond to practical ones, namely the star property (all links affected by a given SRLG are incident to a unique node) and the span 1 property (the links affected by a given SRLG form a connected component of the network). The star property is defined in a multi-colored model where a link can be affected by several SRLGs while the span property is defined only in a mono-colored model where a link can be affected by at most one SRLG. In [59], we extend these notions to characterize new cases in which these optimization problems can be solved in polynomial time. We also investigate the computational impact of the transformation from the multi-colored model to the mono-colored one. Experimental results are presented to validate the proposed algorithms and principles.

In [22], we investigate the k-diverse routing problem which is to find a set of k pairwise SRLG-disjoint paths between a given pair of end nodes of the network. This problem has been proven NP-complete in general and some polynomial instances have been characterized. We consider more specifically the case where the SRLGs are localized and satisfy the star property. We first provide counterexamples to the polynomial time algorithm proposed by X. Luo and B. Wang (DRCN'05) for computing a pair of SRLG-disjoint paths in networks with SRLGs satisfying the star property, and then prove that this problem is in fact NP-complete. We then characterize instances that can be solved in polynomial time or are fixed parameter tractable, in particular when the number of SRLGs is constant, the maximum degree of the vertices is at most 4, and when the network is a directed acyclic graph. Finally we consider the problem of finding the maximum number of SRLG-disjoint paths in networks with SRLGs satisfying the star property. We prove that this problem is NP-hard to approximate within $O(|V|^{1-\varepsilon})$ for any $0<\varepsilon<1$, where V is the set of nodes in the network. Then, we provide exact and approximation algorithms for relevant subcases.

7.1.3.2. Design of Fault-tolerant On-board Networks with Variable Switch Sizes

In [29], we focus on designing networks that are capable, in the presence of faulty output ports, of rerouting input signals to operational output ports. Since the components of a satellite cannot be repaired, redundant amplifiers are added, and the interconnection network satisfies the following fault tolerance property: the network connects the set of input ports with the set of output ports, and for any set of at most k output port failures, there exists a set of edge-disjoint paths connecting the input ports to the operational output ports. Since each switching device is expensive, these interconnection networks are constructed using the fewest possible switches, or at least a number of switches close to the minimum value. The networks are controlled centrally from Earth. Each time an amplifier in use develops a fault, the controller sends messages to the switches to change their settings, so as to ensure that the inputs remain connected to functioning amplifiers.

Current switches have four ports. Obviously, the larger the number of ports, the more expensive will be the switches, but then fewer will be required. So the cost of such a network involves a trade-off between the total number of switches and their unit cost. In order to determine the minimum-cost network, we give some bounds on the minimum number $\mathcal{N}(n,k,r)$ of $2r$-port switches in interconnection networks with n inputs and $n+k$ outputs.

We first show $\mathcal{N}(n,k,r) \le \left\lceil \frac{k+2}{2r-2} \right\rceil \left\lceil \frac{n}{2} \right\rceil$. When $r \ge k/2$, we prove a better upper bound: $\mathcal{N}(n,k,r) \le \frac{r-2+k/2}{r^2-2r+k/2} n + O(1)$. Next, we establish some lower bounds. We show that if $k \ge r$, then $\mathcal{N}(n,k,r) \ge \frac{3n+k}{2r}$. We improve this bound when $k \ge 2r$: $\mathcal{N}(n,k,r) \ge \frac{3n+2k/3-r/2}{2r-2+\frac{3r}{\lfloor \frac{k}{r} \rfloor}}$. Finally, we determine $\mathcal{N}(n,k,r)$ up to additive constants for $k \le 6$.

7.1.4. Reducing Networks' Energy Consumption

Due to the increasing impact of ICT (Information and Communication Technology) on power consumption and worldwide gas emissions, energy efficient ways to design and operate backbone networks are becoming a new concern for network operators. Recently, energy-aware routing (EAR) has gained an increasing popularity in the networking research community. The idea is that traffic demands are redirected over a subset of the network links, allowing other links to sleep to save energy. We studied variant of this problems.

7.1.4.1. Robust Energy-aware Routing with Redundancy Elimination

In [31], we propose GreenRE – a new EAR model with the support of data redundancy elimination (RE). This technique, enabled within routers, can virtually increase the capacity of network links. Based on real experiments on a Orange Labs platform, we show that performing RE increases the energy consumption for routers. Therefore, it is important to determine which routers should enable RE and which links to put into sleep mode so that the power consumption of the network is minimized. We model the problem as a Mixed Integer Linear Program and propose greedy heuristic algorithms for large networks. Simulations on several network topologies show that the GreenRE model can gain further 37% of energy savings compared to the classical EAR model. In [27], we introduce an extended model of the classical multi-commodity flow problem with compressible flows which is also robust with fluctuation of traffic demand and compression rate. An heuristic built on this model allows for 16-28% extra energy saving.

7.1.4.2. Optimizing IGP Link Weights for Energy-efficiency in Multi-period Traffic Matrices

To guarantee QoS while implementing EAR, all traffic demands should be routed without violating capacity constraints and the network should keep its connectivity. From the perspective of traffic engineering, we argue that stability in routing configuration also plays an important role in QoS. In details, frequent changes in network configuration (link weights, slept and activated links) to adapt with traffic fluctuation in daily time cause network oscillations. In [35], we propose a novel optimization method to adjust the link weights of Open Shortest Path First (OSPF) protocol while limiting the changes in network configurations when multi-period traffic matrices are considered.

7.1.4.3. Energy Efficient Content Distribution

Recently, there is a trend to introduce content caches as an inherent capacity of network equipment, with the objective of improving the efficiency of content distribution and reducing the network congestion. In [18], we study the impact of using in-network caches and content delivery network (CDN) cooperation on an EAR. We formulate this problem as Energy Efficient Content Distribution, we propose an integer linear program (ILP) and a heuristic algorithm to solve it. The objective of this problem is to find a feasible routing, so that the total energy consumption of the network is minimized while the constraints given by the demands and the link capacity are satisfied. We exhibit for which the range of parameters (size of caches, popularity of content, demand intensity, etc.) it is useful to use caches. Experimental results show that by placing a cache on each backbone router to store the most popular content, along with well choosing the best content provider server for each demand to a CDN, we can save about 20% of power in average of all the backbone networks considered.

7.1.5. Routing Theory and Forwarding Index

Motivated by finding the best set of links that should be on for energy efficiency, we study the problem of determining the minimum forwarding index of a graph. The (edge) forwarding index of a graph is the minimum, over all possible routings of all the demands, of the maximum load of an edge. This metric is of a great interest since it captures the notion of global congestion in a precise way: the lesser the forwarding-index, the lesser the congestion. This parameter has been studied for different graph classes in the literature. In [42], we determine, for different numbers of edges, the best spanning graphs of a square grid, namely those with a low forwarding index. In [61], [43], we study the following design question: Given a number e of edges and a number n of vertices, what is the least congested graph that we can construct? and what forwarding-index can we achieve? We answer here these questions for different families of graphs: general graphs, graphs with bounded degree, sparse graphs with a small number of edges by providing constructions, most of them asymptotically optimal. Doing so, we partially answer the practical problem that initially motivated our work: If an operator wants to power only e links of its network, in order to reduce the energy consumption (or wiring cost) of its networks, what should be those links and what performance can be expected?

On the complexity of equal shortest path routing.

Additionally, we studied the complexity of configuring the OSPF-ECMP (for Open Shortest Path First-Equal Cost Multiple Path) protocol. In [32], we show that the problem of maximizing even a single commodity flow for the OSPF-ECMP protocol cannot be approximated within any constant factor ratio. Besides this main theorem, we derive some positive results which include polynomial-time approximations and an exponential-time exact algorithm.

7.1.6. Routing in Software Defined Networks (SDN)

Software Defined Networking (SDN) is gaining momentum with the support of major manufacturers. While it brings flexibility in the management of flows within the data center fabric, this flexibility comes at the cost of smaller routing table capacities. In [50], we investigate compression techniques to reduce the forwarding information base (FIB) of SDN switches. We validate our algorithm, called MINNIE, on a real testbed able to emulate a 20 switches fat tree architecture. We demonstrate that even with a small number of clients, the limit in terms of number of rules is reached if no compression is performed, increasing the delay of all new incoming flows. MINNIE, on the other hand, reduces drastically the number of rules that need to be stored with a limited impact on the packet loss rate. We also evaluate the actual switching and reconfiguration times and the delay introduced by the communications with the controller. In parallel, we considered the algorithmic problem of compressing bidimensional routings table with priorities on the rules. We carry out in [40] a study of the problem complexity, providing results of NP-completeness, of Fixed-Parameter Tractability and approximation algorithms. In [44], we then propose green routing schemes performing simultaneously the selection of the routes, the compression of the routing tables, and decide to put in sleep mode unused links. These algorithms are tested on networks from the SNDLib library.

7.1.7. Video Streaming

7.1.7.1. Study of Repair Protocols for Live Video Streaming Distributed Systems

In [41], we study distributed systems for live video streaming. These systems can be of two types: structured and un-structured. In an unstructured system, the diffusion is done opportunistically. The advantage is that it handles churn, that is the arrival and departure of users, which is very high in live streaming systems, in a smooth way. On the opposite, in a structured system, the diffusion of the video is done using explicit diffusion trees. The advantage is that the diffusion is very efficient, but the structure is broken by the churn. In this paper, we propose simple distributed repair protocols to maintain, under churn, the diffusion tree of a structured streaming system. We study these protocols using formal analysis and simulation. In particular, we provide an estimation of the system metrics, bandwidth usage, delay, or number of interruptions of the streaming. Our work shows that structured streaming systems can be efficient and resistant to churn.

7.2. Graph Algorithms

Participants: Nathann Cohen, David Coudert, Frédéric Giroire, Fatima Zahra Moataz, Benjamin Momège, Nicolas Nisse, Stéphane Pérennes.

COATI is also interested in the algorithmic aspects of Graph Theory. In general we try to find the most efficient algorithms to solve various problems of Graph Theory and telecommunication networks. We use graph theory to model various network problems. We study their complexity and then we investigate the structural properties of graphs that make these problems hard or easy. In particular, we try to find the most efficient algorithms to solve the problems, sometimes focusing on specific graph classes from which the problems are polynomial-time solvable. Many results introduced here are presented in detail in the PhD thesis of F. Z. Moataz [14].

7.2.1. Graph Hyperbolicity

The Gromov hyperbolicity is an important parameter for analyzing complex networks which expresses how the metric structure of a network looks like a tree (the smaller gap the better). It has recently been used to provide bounds on the expected stretch of greedy-routing algorithms in Internet-like graphs, and for various applications in network security, computational biology, the analysis of graph algorithms, and the classification of complex networks.

7.2.1.1. Exact Algorithms for Computing the Gromov Hyperbolicity

The best known theoretical algorithm computing this parameter runs in $O(n^{3.69})$ time, which is prohibitive for large-scale graphs. In [26], we propose an algorithm for determining the hyperbolicity of graphs with tens of thousands of nodes. Its running time depends on the distribution of distances and on the actual value of the hyperbolicity. Although its worst case runtime is $O(n^4)$, it is in practice much faster than previous proposals as observed in our experimentations on benchmark instances. We also propose a heuristic algorithm that can be used on graphs with millions of nodes.

In [37], we provide a more efficient algorithm: although its worst-case complexity remains in $O(n^4)$, in practice it is much faster, allowing, for the first time, the computation of the hyperbolicity of graphs with up to $200,000$ nodes. We experimentally show that our new algorithm drastically outperforms the best previously available algorithms, by analyzing a big dataset of real-world networks. We have also used the new algorithm to compute the hyperbolicity of random graphs generated with the Erdös-Renyi model, the Chung-Lu model, and the Configuration Model.

7.2.1.2. Hyperbolicity of Particular Graph Classes

Topologies for data center networks have been proposed in the literature through various graph classes and operations. A common trait to most existing designs is that they enhance the symmetric properties of the underlying graphs. Indeed, symmetry is a desirable property for interconnection networks because it minimizes congestion problems and it allows each entity to run the same routing protocol. However, despite sharing similarities these topologies all come with their own routing protocol. Recently, generic routing schemes have been introduced which can be implemented for any interconnection networks. The performances of such universal routing schemes are intimately related to the hyperbolicity of the topology. Motivated by the good performances in practice of these new routing schemes, we propose in [56] the first general study of the hyperbolicity of data center interconnection networks. Our findings are disappointingly negative: we prove that the hyperbolicity of most data center topologies scales linearly with their diameter, that it the worst-case possible for hyperbolicity. To obtain these results, we introduce original connection between hyperbolicity and the properties of the endomorphism monoid of a graph. In particular, our results extend to all vertex and edge-transitive graphs. Additional results are obtained for de Bruijn and Kautz graphs, grid-like graphs and networks from the so-called Cayley model.

In [57], we investigate more specifically on the hyperbolicity of bipartite graphs. More precisely, given a bipartite graph $B = (V_0 \cup V_1, E)$ we prove it is enough to consider any one side V_i of the bipartition of B to obtain a close approximate of its hyperbolicity $\delta(B)$ — up to an additive constant 2. We obtain from this result the sharp bounds $\delta(G) - 1 \leq \delta(L(G)) \leq \delta(G) + 1$ and $\delta(G) - 1 \leq \delta(K(G)) \leq \delta(G) + 1$ for every graph G, with $L(G)$ and $K(G)$ being respectively the line graph and the clique graph of G. Finally, promising extensions of our techniques to a broader class of intersection graphs are discussed and illustrated with the case of the biclique graph $BK(G)$, for which we prove $(\delta(G) - 3)/2 \leq \delta(BK(G)) \leq (\delta(G) + 3)/2$.

7.2.2. Tree-decompositions

We study the computational complexity of different variants of tree-decompositions. We also study their relationship with various pursuit-evasion games.

7.2.2.1. Diameter of Minimal Separators in Graphs (structure vs metric in graphs)

In [39], we establish general relationships between the topological properties of graphs and their metric properties. For this purpose, we upper-bound the diameter of the *minimal separators* in any graph by a function of their sizes. More precisely, we prove that, in any graph G, the diameter of any minimal separator S in G is at most $\lfloor \frac{\ell(G)}{2} \rfloor \cdot (|S|-1)$ where $\ell(G)$ is the maximum length of an isometric cycle in G. We refine this bound in the case of graphs admitting a *distance preserving ordering* for which we prove that any minimal separator S has diameter at most $2(|S|-1)$. Our proofs are mainly based on the property that the minimal separators in a graph G are connected in some power of G. Our result easily implies that the *treelengtht*$l(G)$ of any graph G is at most $\lfloor \frac{\ell(G)}{2} \rfloor$ times its *treewidtht*$w(G)$. In addition, we prove that, for any graph G that excludes an *apex graph*H as a minor, $tw(G) \leq c_H \cdot tl(G)$ for some constant c_H only depending on H. We refine this constant when G has bounded genus. As a consequence, we obtain a very simple $O(\ell(G))$-approximation algorithm for computing the treewidth of n-node m-edge graphs that exclude an apex graph as a minor in $O(nm)$-time.

7.2.2.2. *Minimum Size Tree-decompositions*

Tree-decompositions are the cornerstone of many dynamic programming algorithms for solving graph problems. Since the complexity of such algorithms generally depends exponentially on the width (size of the bags) of the decomposition, much work has been devoted to compute tree-decompositions with small width. However, practical algorithms computing tree-decompositions only exist for graphs with treewidth less than 4. In such graphs, the time-complexity of dynamic programming algorithms is dominated by the size (number of bags) of the tree-decompositions. It is then interesting to minimize the size of the tree-decompositions. In [48], [14], we consider the problem of computing a tree-decomposition of a graph with width at most k and minimum size. We prove that the problem is NP-complete for any fixed $k \geq 4$ and polynomial for $k \leq 2$; for $k = 3$, we show that it is polynomial in the class of trees and 2-connected outerplanar graphs.

7.2.2.3. *Non-deterministic Graph Searching in Trees*

Non-deterministic graph searching was introduced by Fomin et al. to provide a unified approach for pathwidth, treewidth, and their interpretations in terms of graph searching games. Given $q \geq 0$, the q-limited search number, $s_q(G)$, of a graph G is the smallest number of searchers required to capture an invisible fugitive in G, when the searchers are allowed to know the position of the fugitive at most q times. The search parameter $s_0(G)$ corresponds to the pathwidth of a graph G, and $s_\infty(G)$ to its treewidth. Determining $s_q(G)$ is NP-complete for any fixed $q \geq 0$ in general graphs and $s_0(T)$ can be computed in linear time in trees, however the complexity of the problem on trees has been unknown for any $q > 0$. We introduce in [16] a new variant of graph searching called restricted non-deterministic. The corresponding parameter is denoted by rs_q and is shown to be equal to the non-deterministic graph searching parameter s_q for $q = 0, 1$, and at most twice s_q for any $q \geq 2$ (for any graph G). Our main result is a polynomial time algorithm that computes $rs_q(T)$ for any tree T and any $q \geq 0$. This provides a 2-approximation of $s_q(T)$ for any tree T , and shows that the decision problem associated to s_1 is polynomial in the class of trees. Our proofs are based on a new decomposition technique for trees which might be of independent interest.

7.2.2.4. *k-Chordal Graphs: from Cops and Robber to Compact Routing via Treewidth*

Cops and robber games, introduced by Winkler and Nowakowski and independently defined by Quilliot, concern a team of cops that must capture a robber moving in a graph. We consider in [34] the class of k-chordal graphs, i.e., graphs with no induced (chordless) cycle of length greater than k, $k \geq 3$. We prove that $k-1$ cops are always sufficient to capture a robber in k-chordal graphs. This leads us to our main result, a new structural decomposition for a graph class including k-chordal graphs. We present a polynomial-time algorithm that, given a graph G and $k \geq 3$, either returns an induced cycle larger than k in G, or computes a tree-decomposition of G, each bag of which contains a dominating path with at most $k-1$ vertices. This allows us to prove that any k-chordal graph with maximum degree Δ has treewidth at most $(k-1)(\Delta-1)+2$, improving the $O(\Delta(\Delta-1)k-3)$ bound of Bodlaender and Thilikos (1997). Moreover, any graph admitting such a tree-decomposition has small hyperbolicity. As an application, for any n-vertex graph admitting such a tree-decomposition, we propose a compact routing scheme using routing tables, addresses and headers of size $O(k \log \Delta + \log n)$ bits and achieving an additive stretch of $O(k \log \Delta)$. As far as we know, this is the first routing scheme with $O(k \log \Delta + \log n)$-routing tables and small additive stretch for k-chordal graphs.

7.2.2.5. *Connected Surveillance Game*

The surveillance game [68] models the problem of web-page prefetching as a pursuit evasion game played on a graph. This two-player game is played turn-by-turn. The first player, called the observer, can mark a fixed amount of vertices at each turn. The second one controls a surfer that stands at vertices of the graph and can slide along edges. The surfer starts at some initially marked vertex of the graph, its objective is to reach an unmarked node before all nodes of the graph are marked. The surveillance number $sn(G)$ of a graph G is the minimum amount of nodes that the observer has to mark at each turn ensuring it wins against any surfer in G. Fomin et al. also defined the connected surveillance game where the observer must ensure that marked nodes always induce a connected subgraph. They ask what is the cost of connectivity, i.e., is there a constant $c > 0$ such that the ratio between the connected surveillance number $csn(G)$ and $sn(G)$ is at most c for any graph G. It is straightforward to show that $csn(G) \leq \Delta sn(G)$ for any graph G with maximum degree Δ. Moreover, it has been shown that there are graphs G for which $csn(G) = sn(G) + 1$. In [30], we investigate the question of the cost of the connectivity. We first provide new non-trivial upper and lower bounds for the cost of connectivity in the surveillance game. More precisely, we present a family of graphs G such that $csn(G) > sn(G) + 1$. Moreover, we prove that $csn(G) \leq \sqrt{sn(G)n}$ for any n-node graph G. While the gap between these bounds remains huge, it seems difficult to reduce it. We then define the online surveillance game where the observer has no a priori knowledge of the graph topology and discovers it little-by-little. This variant, which fits better the prefetching motivation, is a restriction of the connected variant. Unfortunately, we show that no algorithm for solving the online surveillance game has competitive ratio better than $\Omega(\Delta)$. That is, while interesting, this variant does not help to obtain better upper bounds for the connected variant. We finally answer an open question [68] by proving that deciding if the surveillance number of a digraph with maximum degree 6 is at most 2 is NP-hard.

7.2.3. Distributed Algorithms

7.2.3.1. Allowing each Node to Communicate only once in a Distributed System: Shared Whiteboard Models

In [21] we study distributed algorithms on massive graphs where links represent a particular relationship between nodes (for instance, nodes may represent phone numbers and links may indicate telephone calls). Since such graphs are massive they need to be processed in a distributed way. When computing graph-theoretic properties, nodes become natural units for distributed computation. Links do not necessarily represent communication channels between the computing units and therefore do not restrict the communication flow. Our goal is to model and analyze the computational power of such distributed systems where one computing unit is assigned to each node. Communication takes place on a whiteboard where each node is allowed to write at most one message. Every node can read the contents of the whiteboard and, when activated, can write one small message based on its local knowledge. When the protocol terminates its output is computed from the final contents of the whiteboard. We describe four synchronization models for accessing the whiteboard. We show that message size and synchronization power constitute two orthogonal hierarchies for these systems.We exhibit problems that separate these models, i.e., that can be solved in one model but not in a weaker one, even with increased message size. These problems are related to maximal independent set and connectivity. We also exhibit problems that require a given message size independently of the synchronization model.

7.2.3.2. Computing on Rings by Oblivious Robots: a Unified Approach for Different Tasks

A set of autonomous robots have to collaborate in order to accomplish a common task in a ring-topology where neither nodes nor edges are labeled (that is, the ring is anonymous). In [36], we present a unified approach to solve three important problems: the exclusive perpetual exploration, the exclusive perpetual clearing, and the gathering problems. In the first problem, each robot aims at visiting each node infinitely often while avoiding that two robots occupy a same node (exclusivity property); in exclusive perpetual clearing (also known as searching), the team of robots aims at clearing the whole ring infinitely often (an edge is cleared if it is traversed by a robot or if both its endpoints are occupied); and in the gathering problem, all robots must eventually occupy the same node. We investigate these tasks in the Look-Compute-Move model where the robots cannot communicate but can perceive the positions of other robots. Each robot is equipped with visibility sensors and motion actuators, and it operates in asynchronous cycles. In each cycle, a robot takes a snapshot of the current global configuration (Look), then, based on the perceived configuration, takes a decision to stay idle or to move to one of its adjacent nodes (Compute), and in the latter case it eventually moves to this neighbor (Move). Moreover, robots are endowed with very weak capabilities. Namely, they are anonymous, asynchronous, oblivious, uniform (execute the same algorithm) and have no common sense of orientation. In this setting, we devise algorithms that, starting from an exclusive and rigid (i.e. aperiodic and asymmetric) configuration, solve the three above problems in anonymous ring-topologies.

7.2.4. Miscellaneous

7.2.4.1. Finding Paths in Grids with Forbidden Transitions

A transition in a graph is a pair of adjacent edges. Given a graph $G = (V, E)$, a set of forbidden transitions $F \subseteq E \times E$ and two vertices $s, t \in V$, we study in [64], [45], [46], [14] the problem of finding a path from s to t which uses none of the forbidden transitions of F. This means that it is forbidden for the path to consecutively use two edges forming a pair in F. The study of this problem is motivated by routing in road networks in which forbidden transitions are associated to prohibited turns as well as routing in optical networks with asymmetric nodes, which are nodes where a signal on an ingress port can only reach a subset of egress ports. If the path is not required to be elementary, the problem can be solved in polynomial time. On the other side, if the path has to be elementary, the problem is known to be NP-complete in general graphs [69]. In [45], we study the problem of finding an elementary path avoiding forbidden transitions in planar graphs. We prove that the problem is NP-complete in planar graphs and particularly in grids. In addition, we show that the problem can be solved in polynomial time in graphs with bounded treewidth. More precisely, we show that there is an algorithm which solves the problem in time $O((3\Delta(k+1))2k+4n))$ in n-node graphs with treewidth at most k and maximum degree Δ.

7.3. Graph theory

Participants: Nathann Cohen, Frédéric Havet.

7.3.1. Graph Colouring

7.3.1.1. Steinberg-like Theorems for Backbone Colouring

Motivated by some channel assignment problem, we study the following variation of graph colouring problem. A function $f : V(G) \to \{1, \cdots, k\}$ is a (proper) k-colouring of G if $|f(u) - f(v)| \geq 1$, for every edge $uv \in E(G)$. The chromatic number $\chi(G)$ is the smallest integer k for which there exists a proper k-colouring of G. Given a graph G and a subgraph H of G, a circular q-backbone k-colouring c of (G, H) is a k-colouring of G such that $q \leq |c(u) - c(v)| \leq k - q$, for each edge $uv \in E(H)$. The circular q-backbone chromatic number of a graph pair (G, H), denoted $CBC_q(G, H)$, is the minimum k such that (G, H) admits a circular q-backbone k-colouring. In [19], we first show that if G is a planar graph containing no cycle on 4 or 5 vertices and $H \subseteq G$ is a forest, then $CBC_2(G, H) \leq 7$. Then, we prove that if $H \subseteq G$ is a forest whose connected components are paths, then $CBC_2(G, H) \leq 6$.

7.3.1.2. Complexity of Greedy Edge-colouring

The Grundy index of a graph $G = (V, E)$ is the greatest number of colours that the greedy edge-colouring algorithm can use on G. In [33], we prove that the problem of determining the Grundy index of a graph $G = (V, E)$ is NP-hard for general graphs. We also show that this problem is polynomial-time solvable for caterpillars. More specifically, we prove that the Grundy index of a caterpillar is $\Delta(G)$ or $\Delta(G) + 1$ and present a polynomial-time algorithm to determine it exactly.

7.3.1.3. Proper Orientation Number

An *orientation* of a graph G is a digraph D obtained from G by replacing each edge by exactly one of the two possible arcs with the same endvertices. For each $v \in V(G)$, the *indegree* of v in D, denoted by $d^-_D(v)$, is the number of arcs with head v in D. An orientation D of G is *proper* if $d^-_D(u) \neq d^-_D(v)$, for all $uv \in E(G)$. The *proper orientation number* of a graph G, denoted by $\overrightarrow{\chi}(G)$, is the minimum of the maximum indegree over all its proper orientations. It is well-known that $\chi(G) \leq \overrightarrow{\chi}(G) + 1 \leq \Delta(G) + 1$, for every graph G, where $\chi(G)$ and $\Delta(G)$ denotes the chromatic number and the maximum degree of G. In other words, the proper orientation number (plus one) is an upper bound on the chromatic number which is tighter than the maximum degree.

In [17], we ask whether the proper orientation number is really a more accurate bound than the maximum degree in the following sense : does there exists a positive ϵ and such that $\vec{\chi}(G) \leq \epsilon \cdot \chi(G) + (1-\epsilon)\Delta(G)$. As an evidence to this, we prove that if G is bipartite (i.e. $\chi(G) \leq 2$) then $\vec{\chi}(G) \leq \left(\Delta(G) + \sqrt{\Delta(G)}\right)/2 + 1$.

However, the proper orientation number has the drawback to be difficult to compute. We prove in [17] that deciding whether $\vec{\chi}(G) \leq \Delta(G) - 1$ is already an NP-complete problem on graphs with $\Delta(G) = k$, for every $k \geq 3$. We also show that it is NP-complete to decide whether $\vec{\chi}(G) \leq 2$, for planar *subcubic* graphs G. Moreover, we prove that it is NP-complete to decide whether $\vec{\chi}(G) \leq 3$, for planar bipartite graphs G with maximum degree 5.

Nevertheless, it might be interesting to bound the proper orientation number on some graph families. In particular, if we prove that for a graph with treewidth at most t, the proper orientation number is bounded by a function of t, this would imply that finding the proper orientation number of a graph with bounded treewidth is polynomial-time solvable. In [17] we prove $\vec{\chi}(G) \leq 4$ if G is a tree (or equivalently a graph with treewidth at most 1). In [53], we study the cacti which is a special class of graphs with treewidth at most 2. We prove that $\vec{\chi}(G) \leq 7$ for every cactus. We also prove that the bound 7 is tight by showing a cactus having no proper orientation with maximum indegree less than 7. We also prove that any planar claw-free graph has a proper orientation with maximum indegree at most 6 and that this bound can also be attained.

7.3.2. Subdivisions of Digraphs

An important result in the Roberston and Seymour minor theory is the polynomial-time algorithm to solve the so-called Linkage Problem. This implies in particular, that for any fixed graph H, deciding whether a graph G contains a subdivision of H as a subgraph can be solved in polynomial time.

We consider the directed analogue F-subdivision problem, which is an analogue for directed graphs (i.e. digraphs). Given a directed graph D, does it contain a subdivision of a prescribed digraph F? In [20], we give a number of examples of polynomial instances, several NP-completeness proofs as well as a number of conjectures and open problems. In [62], we give further support to several open conjectures and speculations about algorithmic complexity of finding F-subdivisions. In particular, up to 5 exceptions, we completely classify for which 4-vertex digraphs F, the F-subdivision problem is polynomial-time solvable and for which it is NP-complete. While all NP-hardness proofs are made by reduction from some version of the 2-linkage problem in digraphs, some of the polynomial-time solvable cases involve relatively complicated algorithms.

7.4. Applications to Other Domains

Participants: Christelle Caillouet, David Coudert, Nicolas Nisse.

7.4.1. Unveiling Contacts within Macro-molecular assemblies by solving Minimum Weight Connectivity Inference Problems

Consider a set of oligomers listing the subunits involved in sub-complexes of a macro-molecular assembly, obtained e.g. using native mass spectrometry or affinity purification. Given these oligomers, connectivity inference (CI) consists in finding the most plausible contacts between these subunits, and minimum connectivity inference (MCI) is the variant consisting in finding a set of contacts of smallest cardinality. MCI problems avoid speculating on the total number of contacts, but yield a subset of all contacts and do not allow exploiting a priori information on the likelihood of individual contacts. In this context, we present in [15] two novel algorithms, MILP-W and MILP-WB. The former solves the minimum weight connectivity inference (MWCI), an optimization problem whose criterion mixes the number of contacts and their likelihood. The latter uses the former in a bootstrap fashion, to improve the sensitivity and the specificity of solution sets. Experiments on three systems (yeast exosome, yeast proteasome lid, human eiF3), for which reference contacts are known (crystal structure, cryo electron microscopy, cross-linking), show that our algorithms predict contacts with high specificity and sensitivity, yielding a very significant improvement over previous work, typically a twofold increase in sensitivity. The software accompanying this paper is made available, and should prove of ubiquitous interest whenever connectivity inference from oligomers is faced.

7.4.2. Recovery of Disrupted Airline Operations using k-Maximum Matching in Graphs

When an aircraft is approaching an airport, it gets a short time interval (called *slot*) that it can use to land. If the landing of the aircraft is delayed (because of bad weather, or if it arrives late, or if other aircrafts have to land first), it loses its slot and Air traffic controllers have to assign it a new slot. However, slots for landing are a scarce resource of the airports and, to avoid that an aircraft waits too much time, Air traffic controllers have to regularly modify the assignment of the slots of the aircrafts. Unfortunately, for legal and economical reasons, Air traffic controllers can modify the slot-assignment only using two kind of operations: either assign to aircraft A a slot that was free, or give to A the slot of another aircraft B and assign to B a free slot. The problem is then the following. Let $k \geq 1$ be an odd integer and let G be a graph and M be a matching (set of pairwise disjoint edges) of G. What is the maximum size of a matching that can be obtained from M by using only augmenting paths of length at most k? Moreover, how to compute such a maximum matching? This problem has already been studied in the context of wireless networks, mainly because it provides a simple approximation for the classical matching problem. We prove in [65], [49] that this problem can be solved in polynomial-time when $k \leq 3$. Then, we show that, for any odd integer $k \geq 5$, the problem is NP-complete in planar bipartite graphs with maximum degree at most 3.

7.4.3. Inference of Curvilinear Structure based on Learning a Ranking Function and Graph Theory

To detect curvilinear structures in natural images, we propose in [63] a novel ranking learning system and an abstract curvilinear shape inference algorithm based on graph theory. We analyze the curvilinear structures as a set of small line segments. In this work, the rankings of the line segments are exploited to systematize the topological feature of the curvilinear structures. Structured Support Vector Machine is employed to learn the ranking function that predicts the correspondence of the given line segments and the latent curvilinear structures. We first extract curvilinear features using morphological profiles and steerable filtering responses. Also, we propose an orientation-aware feature descriptor and a feature grouping operator to improve the structural integrity during the learning process. To infer the curvilinear structure, we build a graph based on the output rankings of the line segments. We progressively reconstruct the curvilinear structure by looking for paths between remote vertices in the graph. Experimental results show that the proposed algorithm faithfully detects the curvilinear structures within various datasets.

7.4.4. Web Transparency for Complex Targeting: Algorithms, Limits, and Tradeoffs

Big Data promises important societal progress but exacerbates the need for due process and accountability. Companies and institutions can now discriminate between users at an individual level using collected data or past behavior. Worse, today they can do so in near perfect opacity. The nascent field of web transparency aims to develop the tools and methods necessary to reveal how information is used, however today it lacks robust tools that let users and investigators identify targeting using multiple inputs. In [67], [38], we formalize for the first time the problem of detecting and identifying targeting on combinations of inputs and provide the first algorithm that is asymptotically exact. This algorithm is designed to serve as a theoretical foundational block to build future scalable and robust web transparency tools. It offers three key properties. First, our algorithm is service agnostic and applies to a variety of settings under a broad set of assumptions. Second, our algorithm's analysis delineates a theoretical detection limit that characterizes which forms of targeting can be distinguished from noise and which cannot. Third, our algorithm establishes fundamental tradeoffs that lead the way to new metrics for the science of web transparency. Understanding the tradeoff between effective targeting and targeting concealment lets us determine under which conditions predatory targeting can be made unprofitable by transparency tools.

COFFEE Project-Team (section vide)

DEMAR Project-Team

5. New Results

5.1. Modelling and identification of the sensory-motor system

5.1.1. Implementation and Validation of a Stride Length Estimation Algorithm, Using a Single Basic Inertial Sensor on Healthy Subjects and Patients Suffering from Parkinson's Disease

Participants: Christine Azevedo Coste, Benoît Sijobert, Mourad Benoussaad [ENIT, Tarbes, France], Christian Geny [CHU Montpellier, Neurology, France], Jennifer Denys [stagiaire M2 STIC SANTE - DEMAR].

Providing a clinical oriented solution, our study presented a gyrometer and accelerometer based algorithm for stride length estimation. Compared to most of the numerous existing works where only an averaged stride length is computed from several IMU, or where the use of the magnetometer is incompatible with everyday use, our challenge here has been to extract each individual stride length in an easy-to-use algorithm requiring only one inertial sensor attached to the subject shank. Our results were validated on healthy subjects and patients suffering from Parkinson's disease (PD). Estimated stride lengths were compared to GAITRite walkway system data: the mean error over all the strides was less than 6 percents for healthy group and 10.3 percents for PD group. This method provides a reliable portable solution for monitoring the instantaneous stride length and opens the way to promising applications ([27]).

5.1.2. Dynamic mapping of upper limb tremor by muscle ultrasonography

Participants: Olivier Tassaert [stagiaire M1 - DEMAR / ICAR], Benjamin Gilles, Olivier Strauss [LIRMM], Christian Geny [CHU Montpellier, Neurology, France], Christine Azevedo Coste.

Focal treatment of action tremor by botulinum toxin injections has been inadequately investigated and at best provides modest relief with significant muscle weakness. Complexity of multi-joint tremulous movements results in non-individualized dosing regimens. Tremor is complex, especially in the upper extremity, and its manifestation can change depending on posture, task, and bodypart. Proper characterization of the tremor based on visual inspection alone is a daunting task for the clinician. Identification of the main trembling muscles task disturbing is challenging because many upper limb muscles are bi-functional. The performance of electromyographic(EMG) pattern-recognition based method in classifying movements strongly depends on arm positions and needs multiple measurements. High density-surface EMG (HD-sEMG) is a non-invasive promising technique to measure electrical muscle activity but has not been used for tremor research because deep muscles could not be investigated Quantification of tremor dynamics by kinematics may be a feasible assessment and guidance tool which can be used to optimize injection conditions for focal tremor therapy. This approach is limited by the redundancy of the upper limb muscle organization Contribution of synergistic muscles toward specific movements over multi joint systems may change with varying position of distal or proximal joints. The choice of injected muscles remains highly subjective and variable. In the study of Rahami, ten different arm or forearm muscles have been injected and improvement was mild and delayed and associated with muscle weakness. In recent years, muscle ultrasonography has become a promising tool for diagnosing neuromuscular disorders . This technique is a non-invasive, low-cost, imaging modality that may be used to characterize normal and pathological muscle tissue but also subtle muscular activity (fasciculatios) in amyotrophic lateral sclerosis. The frequency of tremor remains stable during movement (3 to 8 Hz). We have initiated the investigation of the use of standard ultrasound as a technique to identify muscle groups responsible of upper limb tremor in patient with essential tremor or Parkinson's disease. The feasibility of the overall procedure has been validated: the acquisition procedure on patients, the possibility to track and segment the apparent motion in images using optical flow, and the ability to segment muscle groups by registering a 3D anatomical template.

5.1.3. Understanding electrophysiological effects of direct electrical stimulation of the brain during wide awake surgery

Participants: Marion Vincent, Olivier Rossel, Mitsuhiro Hayashibe, Hugues Duffau [CHU Montpellier], David Guiraud, François Bonnetblanc.

Direct electrical stimulation (DES) have been recently introduced in the neurosurgery of slow-growing and infiltrative brain tumors to guide the resection. By generating transient perturbations, this method allows the real-time identification of both cortical areas and subcortical networks that are essential for the function. Thus, as much as possible, non-functional tissue can be removed while minimizing the sequelae. However, there is much controversy as to whether the use of DES during wideawake surgery is the gold standard for studying the brain function. It is sometimes wrongly assumed that electrical microstimulation (EMS) and DES induce similar effects in the nervous tissues and have comparable behavioural consequences. Both of them are used to perform functional brain mapping: EMS for animal fundamental neuroscience experiments, and DES for neurosurgery patients. We tried to shed new light on electrical stimulation (ES) techniques in brain mapping by comparing EMS and DES [1]. In fact, their effects cannot be directly compared - especially in the electrophysiological domain. There is a gap between theory and practice for ES of the brain. We do not know exactly how ES and especially DES influence the electrophysiological state of networks in the brain; a strong biophysical rationale is lacking. In contrast, the gap between EMS and DES highlights the potential for new experimental paradigms in electrical stimulation for functional brain mapping. In view of this gap and recent technical developments in stimulator design, it may now be time to move towards alternative, innovative protocols. Moreover, the understanding of the electrophysiological effects of DES remains an open and key question. Intra-operative EEG (iEEG) recordings were studied to analyze if and how stimulation currents spread at distant sites. Data were collected during an awake brain surgery for one patient. We observed significant changes in the frequency content at different iEEG sites during DES [2]. Subcortical DES led to neuromodulation at more sites than cortical DES (Figure 4). This may be due to (i) a better conduction and propagation following the direct stimulation of large, myelinated axons and (ii) the greater current intensity in subcortical DES. Further research will have to characterize these aspects more carefully and apply cortical and subcortical DES with identical current intensities [30], [31].

5.1.4. Functional Connectivity Analysis of Motor Imagery EEG signal for Brain-computer Interfacing Application: A preliminary study

Participants: Saugat Bhattacharyya, Poulami Ghosh [Jadavpur University, India], Ankita Mazumder [Jadavpur University, India], D.n. Tibarewala [Jadavpur University, India], Mitsuhiro Hayashibe.

The human brain can be considered as a graphical network having different regions with specific functionality and it can be said that a virtual functional connectivity are present between these regions. These regions are regarded as nodes and the functional links are regarded as the edges between them. The intensity of these functional links depend on the activation of the lobes while performing a specific task(e.g. motor imagery tasks, cognitive tasks and likewise). The analysis of these networks are performed by using a very useful mathematical tool called graph theory. Graph theory basically represents the entire functional network with a number of nodes and edges between them and the amount of connectivity existing between two nodes is depicted by assigning weights to the edges between them. In this study we have tried to utilize functional connectivity between different parts of the human brain for classifying a motor imagery task.

Brain connectivity patterns can be determined by using two types of measures, namely, Bivariate and Multivariate. Here we have considered a multivariate measure known as multivariate autoregressive (MVAR) model. One of the most widely investigated connectivity measure is the Directed Transfer Function (DTF). This function basically computes the directional influences between any two given nodes. There are a number of theoretical indices for defining a graph. In this preliminary work, two indices, namely node strength and network density are measured from the DTF values. In the current study, the BCI competition Dataset III is used for computing different multivariate measures.

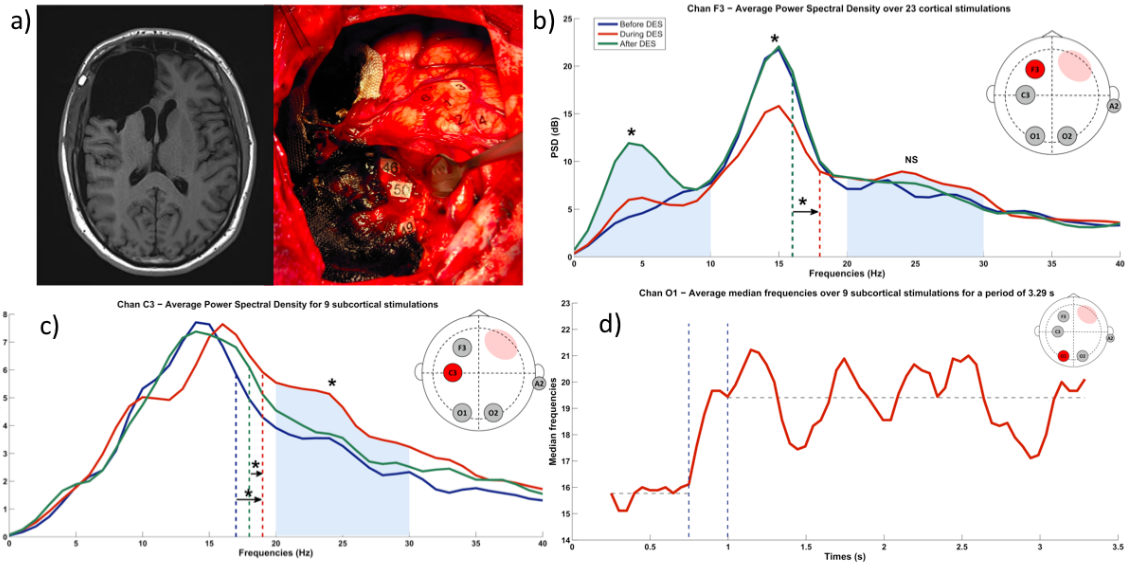
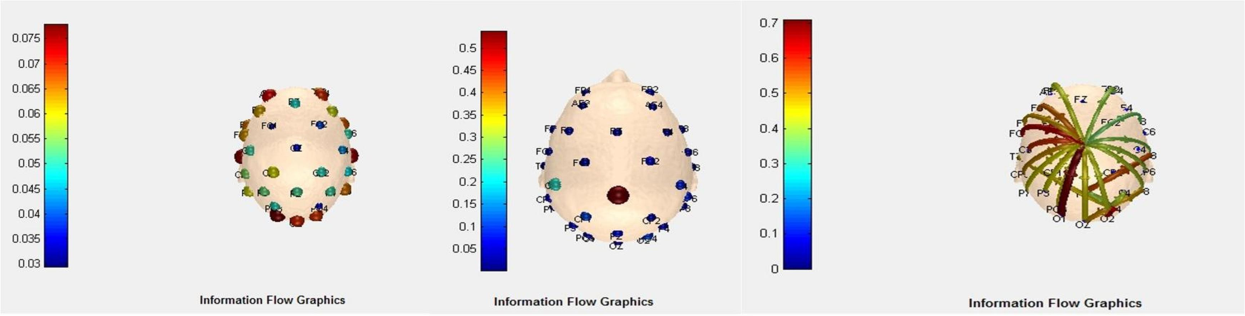


Figure 4. (a) Post-operative MRI of the patientâs brain, showing the right frontal cavity and an intraoperative view of the brain with the main anatomical landmarks. (b) The mean PSD of the iEEG signal, on F3, before, during and after each period of cortical DES. (c) The mean PSD of the iEEG signal, on C3, before during and after each period of subcortical DES. (d) The moving window median frequency averaged over nine subcortical DESs periods for PSD measured at O1.

The inflow-outflow graph of subject 1 while imagining right hand movement in the first training set are given in Fig.5 . Fig. 5 (a) describes the amount of inflow of functional connectivity going out of all the 32 electrodes and these are color coded to indicate the intensity of these inflows. From Fig. 5 (a) it is quite evident that the inflows are maximum in the frontal, temporal and occipital lobes. Figure 5 (b) depicts the functional outflow from the nodes and in contrast to Fig.5 (a) it shows that the outflows are maximum from the Central lobe(Cz). In Fig 5 (c), the direction of the flow between different nodes are shown and it can be seen clearly that majority of the paths are going from Cz to different nodes of the frontal, parietal and temporal lobes.



(a) (b) (c)

Figure 5. (a) Inflow graph, (b) Outflow graph and (c) Out to inflow graph of the functional connectivity network of the brain while imagining right hand movement.

5.1.5. A Generic Transferable EEG Decoder for Online Detection of Error Potential in Target Selection

Participants: Saugat Bhattacharyya, Amit Konar [Jadavpur University, India], D.n. Tibarewala [Jadavpur University, India], Mitsuhiro Hayashibe.

Detection of error from electroencephalography (EEG) signals as feedback while performing a discrete target selection task is beneficial for general Brain-computer Interfacing (BCI) systems including rehabilitative application. Error Related Potentials (ErrP) are EEG signals which occur when the participant observes an erroneous feedback from the system.

In this study, we have designed a novel scheme for detection of error feedback directly from the EEG signal. For this purpose, we have used a P300-speller dataset from the ‘BCI Challenge @ NER 2015’ competition hosted at Kaggle. The task involves the subject to select a letter of a word which is followed by a feedback period. The feedback period displays the letter selected and if the selection is wrong, the subject perceives it by the generation of ErrP signal. Our proposed system is designed to detect whether the feedback is erroneous or not. The decoder designed for this task is an ensemble of linear discriminant analysis, quadratic discriminant analysis and logistic regression classifier. The decoder is also transferable in nature as it is should work with single-trial on new subject without any prior subject-specific training.

The block diagram of the BCI system adopted for online ErrP detection from input EEG signals is shown in Fig.6 . The system implements three main processes: i) Pre-processing of the signal, i.e., temporal filtering in the bandwidth [0.1, 10]Hz, ii) Extraction of relevant features corresponding to the mental state from the signal using savitzsky-golay filter and meta-data of the features, and iii) Classification of the features, using our proposed decoder, to detect the intention of the participant from two given states: *Error* and *No-Error*. A switch is incorporated in the design to detect the beginning of feedback period in the trials, which is marked in the datasets. We have tested the online functionality of the BCI system on the test dataset provided in the website. To simulate a real-time condition, the EEG is continuously streamed until an onset of the feedback period is detected. On detection of the feedback period, the system extracts a pre-defined length of signal for further processing and the rest are rejected.

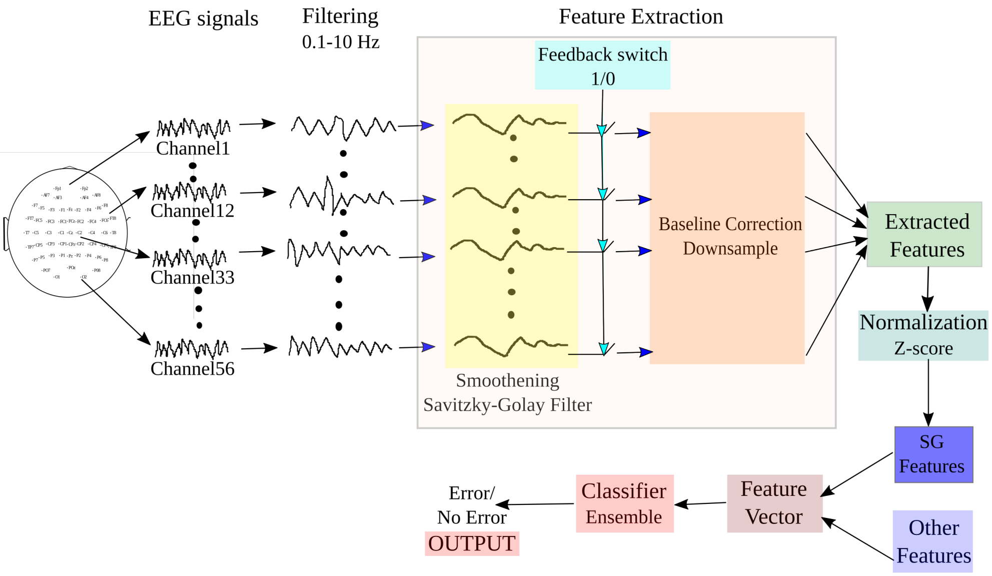


Figure 6. Block diagram of the BCI system adopted for online detection of Error Related Potentials from the input EEG

5.2. Synthesis and Control of Human Functions

5.2.1. FES-cycling and participation to Cybathlon competition

Participants: Christine Azevedo Coste, Benoît Sijobert, Charles Fattal [CRF DIVIO, Dijon, France], Antonio Padilha [UNB, Brasilia, Brazil], Emerson Fachin Martins [UNB, Brasilia, Brazil], David Andreu.

DEMAR and University of Brasilia will jointly participate with two SCI pilots to Cybathlon - FES-Bike competition. Cybathlon intends to promote assistive technologies during a competition. Two trikes will be adapted, an original control strategy will be proposed and two paraplegic individuals (one from Brazil and one from France) will be trained during the upcoming year. The protocol will be submitted to CPP ethical committee for agreement in the coming weeks (http://freewheels.inria.fr/).

5.2.2. PersoStim: A Personalized Closed-loop FES Control of Muscle Activation with Evoked EMG Feedback

Participants: Mitsuhiro Hayashibe, Zhan Li [University of Electronic Science and Technology of China], David Andreu, David Guiraud.

Functional electrical stimulation (FES) is a useful technique for restoring motor functions for spinal cord injured (SCI) patients. Muscle contractions can be artificially driven through delivery of electrical pulses to impaired muscles, and the electrical activity of contracted muscles under stimulus recorded by electromyography (EMG) is called M-wave. The FES-induced muscle activation which is represented by evoked EMG recordings can indicate the muscle state. Accurate control of muscle activation level by FES is the first step toward achieving more complicated FES control tasks.

A new FES closed-loop control strategy, EMG-feedback predictive control (EFPC), was developped to adaptively control stimulation pattern compensating to time-varying muscle state changes such as muscle fatigue and stimulation electrode detachment, along with the consideration of the personalized muscle responses to the electrical stimulation. This software manages a real-time FES system for control of muscle activation by online modulating pulse width of stimulus. The excitation muscle dynamics is modelled by Hammerstain system with stimulus pulse width and eEMG as input and output respectively. The model predictive control strategy is adopted to systematically produce the pulse width command of the stimulator. It is implemented together with Vivaltis portable stimulator. Four reference muscle activation patterns are provided to test and validate the real-time closed-loop FES control system. Real-time control results show promising control performances.

Recently, this software was demonstrated at the event of Rencontre Inria-Industrie 13/10/2015 at Bordeaux. https://www.inria.fr/centre/bordeaux/innovation/rii-sante/demonstrations2

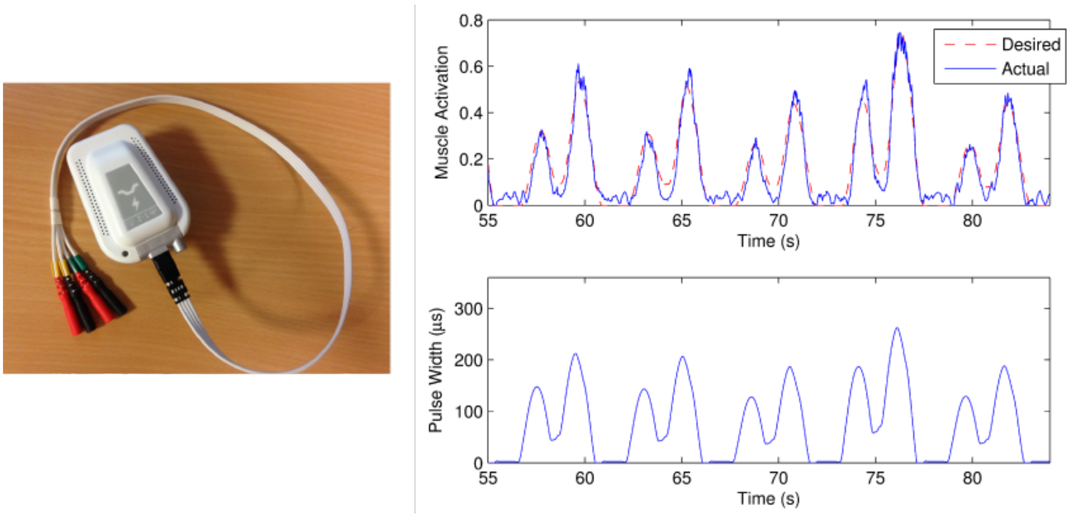


Figure 7. Left: Vivaltis portable stimulator, Right: Real-time control performance of muscle activation with desired dual sinusoidal shaped muscle activation pattern (red dash line is desired muscle activation trajectory and blue solid line is the measured muscle activation under the muscle activation control by FES). The lower plot is the corresponding computed stimulation pulse width.

5.2.3. Direct spinal stimulation for rehabilitation of bladder, bowel and sexual functions in spinal cord injury

Participants: Christine Azevedo Coste, Luc Bauchet [CHU Montpellier], Claire Delleci [CHU Bordeaux], Charles Fattal [CRF DIVIO, Dijon, France], Thomas Guiho, David Guiraud, Jean-Rodolphe Vignes [CHU Bordeaux].

Complete spinal cord injury results in loss of movement and sensory sensations but also in function of organs. For example, nearly all spinal cord injured subjects lose their bladder control and are prone to kidney failure if they do not apply intermittent (self-) catheterization. Electrical stimulation of the sacral spinal roots with an implantable neuroprosthesis is one option besides self-catheterization to become continent and control micturition. However, many persons do not ask for this neuroprosthesis since deafferentation and loss of sensory functions and reflexes are serious side effects. Spinal cord stimulation (SCS) is a general term which includes both epidural and intradural stimulation. Originally associated with the treatment of chronic neurogical pain (in the 1970ies), SCS led also to immediate and profound improvements of sensory and motor functions in recent studies both on SCI patients (only on very few case studies) and rodents. Despite these promising results some limitations have still to be overcome. Among them, the use of small animal models, the empirical aspect of the stimulation procedure and the impact of these protocols on intestinal and urinary functions are critical. To counteract these limits, we want to explore intradural and epidural stimulations in an intermediate model- the house pig- and assess their impact on bladder, guts and genitals. In order to evaluate our approach, we will record EMG signals of lower limbs and sphincters (both urethral and anal), and simultaneously, we will monitor bladder and rectal pressure.

Already preliminary experimental explorations were performed with direct spinal cord stimulation in June (on 2 animals). Experiments were conducted under neurosurgeons involved in the project and urodynamics was recorded together with rectum pressure and sphincters EMG during each stimulation session.

5.3. Neuroprostheses and technology

5.3.1. Selectivity of nerve stimulation using a 12 pole multipolar cuff

Participants: Wafa Tigra, Olivier Rossel, Thomas Guiho, David Guiraud, Christine Azevedo Coste, Hubert Taillades [UM].

Experimentations were performed on 5 rabbits (New Zealand white). A multipolar cuff electrode (12 poles, diameter 3 mm, length 20 mm, 12 oblong contacts of 5mm length) was placed around the sciatic nerve of the rabbit 3 cm above the tibiofibular bifurcation. The nerve was stimulated with increasing intensity. The protocol consisted of the activation of one or more channels of the electrode, the input is a biphasic asymmetric stimulation and the pulse width is modulated in intensity (up to 2.4 mA) and fixed in length (250 μsec, 100 μsec interstim). A stimulus (4 Hz) is used for 2 seconds. 48 configurations of stimulation were tested. Needle electrodes were inserted on the lateral and medial gastrocnemius, soleus, tibialis and extensor digitorum muscles to record EMG signals and were used to evaluate the selectivity capacities of given cuff electrode configuration. The rabbit foot was also attached to a force platform. Inter-fascicular selectivity was observed for the 5 animals. Intra fascicular selectivity was also observed in 3 animals. Placed at a single location, our cuff electrode is capable to activate, selectively, some muscles. Experiments were performed under ethical comittee agreement at the "Plateau Technique Chirurgie Exprimentale" (Montpellier).

5.3.2. A novel EMG interface for individuals with quadriplegia to pilot robot hand grasping

Participants: Wafa Tigra, Benjamin Navarro [LIRMM], Andrea Cherubini [LIRMM], Xavier Gorron [???], Anthony Gelis [PROPARA], Charles Fattal [CRF DIVIO, Dijon, France], David Guiraud, Christine Azevedo Coste.

We have developed and validated a new human-machine interface dedicated to individuals with quadriplegia. We investigated the feasibility of online processing sus-lesional muscle responses, to pilot an assistive device. The ability to voluntary contract a set of selected muscles was assessed in five spinal cord injured subjects through electromyography analysis. Two subjects have also been asked to use the EMG interface to control palmar and lateral grasping of a robot hand (fig.8). These preliminary results sound very promising and open the way to new interface solutions for high level spinal cord injured patients(fig.8).

5.3.3. Wearable 56-pole stimulator

Participants: Arthur Hiairrassary, David Andreu, David Guiraud, Olivier Rossel, Thomas Guiho.

In the context of the EPIONE European project, we have designed and developed, with Axonic, a wearable multichannel stimulator (fig.9), to face phantom limb pain (PLP). This 56-pole neural stimulator is based on four 16-pole stimulation units (each one being connected to a 16-pole TLIFE intra-fascicular electrode) connected to a real-time controller by means of an embedded deterministic network. This controller, in charge of executing FES functions (threshold determination, sensation characterization, etc.), pilots the stimulation units and allows for real-time modulation of the multisite stimulation. The controller can be remotely configured and exploited by the practitioner, by means of a dedicated software (Synergy Neuromodulation Software). But it has been also connected to the controller of the EPFL's artificial hand, in order to link hand touch sensors with neural stimulation to induce natural, meaningful sensations to the amputee.

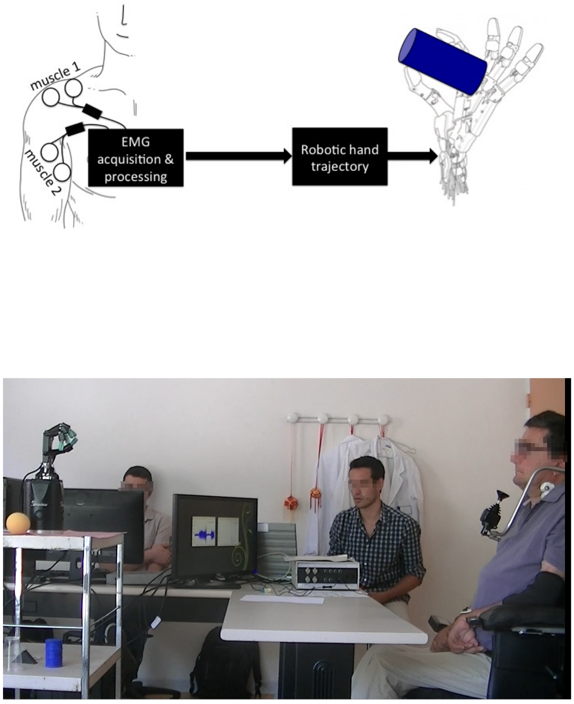


Figure 8. Principle of robot hand control through EMG signals; Setup description and upper arm positioning during EMG recordings.

This stimulator has been deeply validated through animal experiments (rats and pigs, respectively with UAB Barcelone and SMI Alboorg) and is currently used on human at UCBM Rome (http://project-epione.eu/).

Figure 9. 56-pole neural stimulator

5.3.4. CORAIL: Neural Stimulation Integrated Circuit

Participants: Jérémie Salles, Guy Cathébras, Milan Demarcq, David Guiraud, Guillaume Souquet, David Andreu.

DEMAR is currently finishing the development of CORAIL (Current Output Reconfigurable Asic Interface Low power), a new ASIC dedicated to electric neural stimulation.

Its main analog characteristics are:

- 12 independent current output channels;
- a full-scale current of 5 mA with a quantum of 1.3 μA;
- a symmetrical power supply ($\pm$ VHT with ground middlepoint).

This front-end integrated circuit is designed to perform multipolar electrical stimulation of the nerve with highly configurable waveforms in order to achieve selective activation of organs or muscles. In comparison with previously developed current output ASICs, the CORAIL IC embeds new features such as the storage of multiple electrode configurations or the possibility to internally combine poles. These specific aspects of CORAIL and the fact that its elaboration benefited from clinical experience in the team will allow enhanced integration within the whole electrical stimulation environment.

The resulting stimulation ASIC aims to be part of an implanted distributed stimulation system, composed of multiple stimulation units spread across the body, in which CORAIL will be the front-end entity in charge of delivering the current to the electrode. Thus, special care has been paid to its integration in such a network with an emphasis on low power consumption for which different mechanisms have been implemented.

The ASIC is currently undergoing the last phases of its development and a first version is due for fabrication in February 2016.

5.3.5. *Tele-Rehabilitation Platform for Gait Training*

Participants: Mitsuhiro Hayashibe, Antonio P.l. Bo [Universidade de Brasilia, Brasil], Leslie Casas [Pontificia Universidad Catolica del Peru, Peru], Gonzalo Cucho [Pontificia Universidad Catolica del Peru, Peru], Dante Elias [Pontificia Universidad Catolica del Peru, Peru].

Throughout the world there is an increasing need for better technologies for rehabilitation and assistance. These new solutions must present improved performance in terms of therapy effectiveness, while at the same time minimizing the corresponding costs. In this scenario, computer-aided methods represent a promising alternative for the challenges currently faced by the rehabilitation domain. A tele-rehabilitation platform for gait training in intercontinental circumstances is developed under STIC-AmSud program. This project was joint program 2012-2013 among Inria France, UnB (University of Brasilia) and PCUP (Pontifical Catholic University of Peru) for tele-rehabilitation framework. This system has two mode: Self-modulation control in which the subject can control the speed of the motion therapy with his comfortable training speed and Guidance control mode in which the motion transfer is performed from one therapist to one patient. Guidance control can be performed both with local data transmission and intercontinental data transmission. Fig. 10 shows the case where the motion transfer regarding foot placement was performed with local data transmission. The test with intercontinental data transmission was also realized between France and Peru.

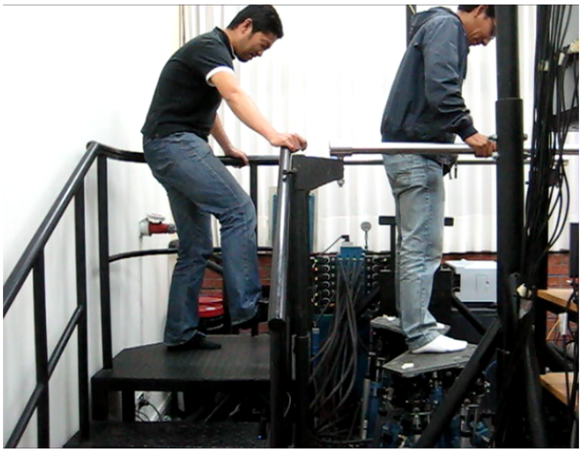


Figure 10. Tele-rehabilitation platform for gait training: Guidance control mode.

5.3.6. *Control and scheduling co-design for stimulation systems*

Participants: Daniel Simon, David Andreu.

Functional Electrical Stimulation (FES) is used in therapy for rehabilitation or substitution for disabled people. They are control systems using electrodes to interface a digital control system with livings. Hence the whole system gathers continuous-time (muscles and nerves) and discrete-time (controllers and communication links) components. During the design process, realistic simulation remains a precious tool ahead of real experiments to check without danger that the implementation matches the functional and safety requirements. To this aim we are developing a real-time open software simulation system, dedicated to the analysis of FES systems deployed over distributed execution resources and wireless links. The simulation tool is especially devoted to the joint design and analysis of control loops and real-time features.

Realistic simulations are effective tools to design and tune complex systems whose analysis cannot be provided only by theory. Several simulation steps can be explored, from simple functional analysis to HIL, to design, test, tune and validate both the single components of the system and their interactions in a distributed architecture. Simulations are precious, as they allow for non-destructive trials, which must be considered in any domain but this is of particular interest for bio-engineering [42].

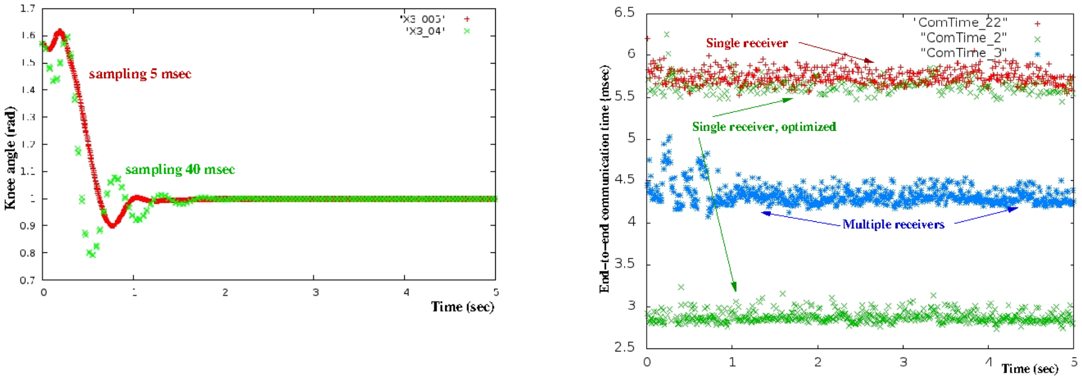


Figure 11. Simulation traces : knee position w.r.t. sampling and communication delays

It is expected that this particular simulator may provide inputs in two main directions. Firstly it allows for preliminary testing and tuning new FES protocols without needing for real experiments with patients, and may help for writing the ethical protocols needed for any experiments involving livings. Secondly it can be used to preliminary evaluation of new technologies or implementations, without costly reworking of existing electronic chips or certified components.

The simulation software is open, so that enhancements w.r.t. to the original release can be added upon request of various designers and to fulfill various objectives.

5.3.7. Control loops design principles for autonomic computing

Participants: Daniel Simon, Eric Rutten [Inria Grenoble Rhône-Alpes], Nicolas Marchand [GIPSA-lab].

Computing systems are becoming more and more dynamically reconfigurable or adaptive, to be flexible w.r.t. their environment and to automate their administration. Autonomic computing proposes a general structure of feedback loop to take this into account. We are particularly interested in approaches where this feedback loop is considered as a case of control loop where techniques stemming from Control Theory can be used to design efficient safe, and predictable controllers. This approach is emerging, with separate and dispersed effort, in different areas of the field of reconfigurable or adaptive computing, at software or architecture level.

We aim at conveying to Computer Scientists the interest and advantages of adopting a Control Theory perspective for the efficient and predictable design of autonomic systems. Compared with open-loop, closed-loop control provides adaptability and robustness, allowing for the design of fault-tolerant systems against varying and uncertain operating conditions. However, there still is a deep need for research in the problems of mapping from high-level objectives in terms of Quality of Service (QoS) or Service Level Objectives (SLO) and abstract models towards lower-levels effective actions on the managed systems [46].

5.4. Others

5.4.1. Do doors opening affect the air contamination in clean surgery? A Prospective, Cross-sectional Study (the ARIBO Project)

Participants: Gabriel Birgand [APHP], Christine Azevedo Coste, Stephane Rukly [INSERM], Roger Pissard-Gibollet [Inria Grenoble Rhône-Alpes], Jean-Christophe Lucet [APHP].

Inappropriate staff behaviours can lead to environmental contamination in the operating room (OR) and subsequent surgical site infection (SSI). This study focused on the continued assessment of OR staff behaviours using doors sensors and their impact on the SSI risk during surgical procedures. This multicentre observational study included 13 ORs in 10 hospitals, 5 University hospitals and 5 private hospitals. Two specialties of clean surgery with cutaneous approach were included: cardiac surgery with procedures requiring a full median sternotomy (CABG or valve replacement surgery); and planned orthopaedic surgery for total hip (THR) or knee replacement (TKR). For each surgical specialty involved, the observed ORs were randomly selected. Doors opening were observed by means of wireless inertial sensors fixed on the doors. For each surgical procedure, 3 microbiological air counts, continuous particles counts of 0.3, 0.5 and 5μm particles, and one bacteriological sample of the wound before skin closure were performed. We collected informations on the OR staff, surgical procedures and surgical environment characteristics. Statistics were performed using univariate and multivariate analysis to adjust on aerolic and architectural characteristics of the OR. We included 34 orthopaedic and 26 cardiac procedures. The mean duration of intervention, from patient entry to exit in the OR, was 5.3 (SD 1.1) h. in cardiac and 2.6 (0.7) h. in orthopaedic surgery. The median number of doors opening was 146 (IQR: 121-183; Min-Max: 86-319) per intervention and 29 (IQR: 23-36; Min-Max: 17, 54) per h. in cardiac surgery and 71.5 (IQR: 58-92; Min-Max: 54-136) per intervention and 29 (IQR: 25-34; Min-Max: 16-65) per h. in orthopaedic procedures. Doors stayed open in average 43 minutes (Min-Max: 19-115) in cardiac and 36 (8-199) in orthopaedic, representing respectively 13.5 percents and 23 percents of the duration of intervention. The highest frequency of doors opening was observed between wound closure and patient exit, median 20.1 openings/h (12.5-32.3) and from patient entry to the incision 13.2 openings/h (8-19). The number and duration of doors opening was significantly different between centres (higher in university hospital, p<0.01)). High frequency of openings was observed for doors that should stay closed during procedures (materials store, decontamination room). The number of doors opening from skin incision to wound closure affected significantly the 0.5 and 5μm particles count (p<0.01 and 0.02 respectively). This study based on automatic observation suggests a large heterogeneity of doors openings between types of interventions, ORs and hospitals. Data give a standard of doors opening for CABG, THR and TKR. Door openings affected air contamination, potentially jeopardizing operating room sterility. The causes and influences of behaviours in the OR must be evaluated to identify ways to reduce the associated risks.

DIANA Project-Team

6. New Results

6.1. Service Transparency

6.1.1. From Network-level Measurements to Expected QoE: the Skype Use Case

Participants: Thierry Spetebroot, Nicolas Aguilera, Damien Saucez and Chadi Barakat.

Modern Internet applications rely on rich multimedia contents making the quality of experience (QoE) of end users sensitive to network conditions. Several models were developed in the literature to express QoE as a function of measurements carried out on the traffic of the applications themselves. In this contribution, we propose a new methodology based on machine learning able to link expected QoE to network and device level measurements outside the applications' traffic. This direct linking to network and device level measurements is important for the prediction of QoE. We prove the feasibility of the approach in the context of Skype. In particular, we derive and validate a model to predict the Skype QoE as a function of easily measurable network performance metrics. One can see our methodology as a new way of performing measurements in the Internet, where instead of expressing the expected performance in terms of network and device level measurements that only specialists can understand, we express performance in clear terms related to expected quality of experience for different applications. More details on this approach and on our application ACQUA can be found in section 5.1 , in the paper summarizing the results [16] and on the application web page http://team.inria.fr/diana/acqua/.

6.1.2. Towards a General Solution for Detecting Traffic Differentiation at the Internet Access

Participants: Ricardo Ravaioli and Chadi Barakat.

In recent years network neutrality has been widely debated from both technical and economic points of view. Various cases of traffic differentiation at the Internet access have been reported throughout the last decade, in particular aimed at bandwidth consuming traffic flows. In this contribution we present a novel application-agnostic method for the detection of traffic differentiation, through which we are able to correctly identify where a shaper is located with respect to the user and evaluate whether it affected delays, packet losses or both. The tool we propose, ChkDiff, replays the user's own traffic in order to target routers at the first few hops from the user. By comparing the resulting flow delays and losses to the same router against one other, and analyzing the behaviour on the immediate router topology spawning from the user end-point, ChkDiff manages to detect instances of traffic shaping. This contribution is published in [15] where we provide a detailed description of the design of the tool for the case of upstream traffic, the technical issues it overcomes and a validation in controlled scenarios. It is the result of collaboration with the SIGNET group at I3S in the context of a PhD thesis funded by the UCN@SOPHIA Labex.

6.1.3. A Diagnostic Tool for Content-Centric Networks

Participant: Thierry Turletti

In collaboration with our colleagues at NICT, Japan, we have proposed the Contrace tool for Measuring and Tracing Content-Centric Networks (CCNs). CCNs are fundamental evolutionary technologies that promise to form the cornerstone of the future Internet. The information flow in these networks is based on named data requesting, in-network caching, and forwarding – which are unique and can be independent of IP routing. As a result, common IP-based network tools such as ping and traceroute can neither trace a forwarding path in CCNs nor feasibly evaluate CCN performance. We designed "contrace," a network tool for CCNs (particularly, CCNx implementation running on top of IP) that can be used to investigate 1) the Round-Trip Time (RTT) between content forwarder and consumer, 2) the states of in-network cache per name prefix, and 3) the forwarding path information per name prefix. We report a series of experiments conducted using contrace on a CCN topology created on a local testbed and the GEANT network topology emulated by the Mini-CCNx emulator.

The results confirm that contrace is not only a useful tool for monitoring and operating a network, but also a helpful analysis tool for enhancing the design of CCNs. Further, contrace can report the number of received interests per cache or per chunk on the forwarding routers. This enables us to estimate the content popularity and design more effective cache control mechanisms in experimental networks (see our publication in the IEEE Communication Magazine [9]).

6.1.4. An efficient packet extraction tool for large experimentation traces

Participants: Thierry Turletti and Walid Dabbous

Network packet tracing has been used for many different purposes during the last few decades, such as network software debugging, networking performance analysis, forensic investigation, and so on. Meanwhile, the size of packet traces becomes larger, as the speed of network rapidly increases. Thus, to handle huge amounts of traces, we need not only more hardware resources, but also efficient software tools. However, traditional tools are inefficient at dealing with such big packet traces. In this work, we propose pcapWT, an efficient packet extraction tool for large traces. PcapWT provides fast packet lookup by indexing an original trace using a Wavelet Tree structure. In addition, it supports multi-threading for avoiding synchronous I/O and blocking system calls used for file processing, and it is particularly efficient on machines with SSD disks. PcapWT shows remarkable performance enhancements in comparison with traditional tools such as tcpdump and most recent tools such as pcapIndex in terms of index data size and packet extraction time. Our benchmark using large and complex traces shows that pcapWT reduces the index data size down below 1% of the volume of the original traces. Moreover, packet extraction performance is 20% better than with pcapIndex. Furthermore, when a small amount of packets are retrieved, pcapWT is hundreds of times faster than tcpdump. This work has been done in collaboration with our colleagues at Universidad Diego Portales (UDP) and Universidad de Chile and has been published in the Computer Networks journal [10].

6.1.5. Social Clicks: What and Who Gets Read on Twitter?

Participants: Maksym Gabielkov and Arnaud Legout

Online news domains increasingly rely on social media to drive traffic to their website. Yet we know surprisingly little about how social media conversation mentioning an online article actually generates a click to it. Posting behaviors, in contrast, have been fully or partially available and scrutinized over the years. While this has led to multiple assumptions on the diffusion of information, each were designed or validated while ignoring this important step. We made a large scale, validated and reproducible study of social clicks, that is also the first data of its kind, gathering a month of web visits to online resources that are located in 5 leading news domains and that are mentioned in the third largest social media by web referral (Twitter). Our dataset amounts to 2.8 million posts, together responsible for 75 billion potential views on this social media, and 9.6 million actual clicks to 59,088 unique resources. We design a reproducible methodology, carefully corrected its biases, enabling data sharing, future collection and validation. As we prove, properties of clicks and social media Click-Through-Rates (CTR) impact multiple aspects of information diffusion, all previously unknown. Secondary resources, that are not promoted through headlines and are responsible for the long tail of content popularity, generate more clicks both in absolute and relative terms. Social media attention is actually long-lived, in contrast with temporal evolution estimated from posts or impressions. The actual influence of an intermediary or a resource is poorly predicted by their posting behavior, but we show how that prediction can be made more precise. The results are reported in an article under submission, no report available yet.

6.1.6. ReCon: Revealing and Controlling PII Leaks in Mobile Network Traffic

Participant: Arnaud Legout

It is well known that apps running on mobile devices extensively track and leak users' personally identifiable information (PII); however, these users have little visibility into PII leaked through the network traffic generated by their devices, and have poor control over how, when and where that traffic is sent and handled by third parties. In this work, we present the design, implementation, and evaluation of ReCon: a cross-platform system that reveals PII leaks and gives users control over them without requiring any special privileges or custom OSes. ReCon leverages machine learning to reveal potential PII leaks by inspecting network traffic, and provides a visualization tool to empower users with the ability to control these leaks via blocking or substitution of PII. We evaluate ReCon's effectiveness with measurements from controlled experiments using leaks from the 100 most popular iOS, Android, and Windows Phone apps, and via an user study with 92 participants. In this study, that was approved by the Inria Ethical Board (COERELE), we show that ReCon is accurate, efficient, and identifies a wider range of PII than previous approaches. The results are reported in an article under submission, no report available yet.

6.2. Open Network Architecture

6.2.1. Storage on Wheels: Offloading Popular Contents Through a Vehicular Cloud

Participants: Luigi Vigneri and Chadi Barakat.

The increasing demand for mobile data is overloading the cellular infrastructure. Small cells and edge caching is being explored as an alternative, but installation and maintenance costs for sufficient coverage are significant. In this work, we perform a preliminary study of an alternative architecture based on two main ideas: (i) using vehicles as mobile caches that can be accessed by user devices; compared to small cells, vehicles are more widespread and require lower costs; (ii) combining the mobility of vehicles with delayed content access to increase the number of cache hits (and reduce the load on the infrastructure). Contrary to standard DTN-type approaches, in our system max delays are guaranteed to be kept to a few minutes (beyond this deadline, the content is fetched from the infrastructure). We first propose an analytical framework to compute the optimal number of content replicas that one should cache, in order to minimize the infrastructure load. We then investigate how to optimally refresh these caches to introduce new contents, as well as to react to the temporal variability in content popularity. Simulations suggest that our vehicular cloud considerably reduces the infrastructure load in urban settings, assuming modest penetration rates and tolerable content access delays. This work is currently under submission; it is the result of collaboration with the Mobile Communications Department at Eurecom in the context of a PhD thesis funded by the UCN@SOPHIA Labex.

6.2.2. Geographically Fair In-Network Caching for Mobile Data Offloading

Participant: Chadi Barakat

Data offloading from the cellular network to low-cost WiFi has been the subject of several research works in the last years. In-network caching has also been studied as an efficient means to further reduce cellular network traffic. In this contribution, done jointly with the Maestro project-team, we consider a scenario where mobile users can download popular contents (e.g., maps of a city, shopping information, social media, etc.) from WiFi-enabled caches deployed in an urban area. We study the optimal distribution of contents among the caches (i.e., what contents to put in each cache) to minimize users' access cost in the whole network. We argue that this optimal distribution does not necessarily provide geographic fairness, i.e., users at different locations can experience highly variable performance. In order to mitigate this problem, we propose two different cache coordination algorithms based on gossiping. These algorithms achieve geographic fairness while preserving the minimum access cost for end users. More details on this contribution can be found in [12].

6.2.3. Virtual Service Providers (vSP)

Participant: Damien Saucez

The ability of SOHO networks to connect to the Internet through several Internet service providers, gives high potential to enable rich cloud-based network services for enterprises. Nevertheless, it remains a huge challenge for SOHOs to leverage such multi-homing and cloud networking capabilities. For such a reason, we introduced the vSP concept (virtual Service Provider). The idea of vSP is to hide the technical complexity inherent to multi-homing and allow SOHOs to seamlessly use their cloud resources. The role of the vSP is to orchestrate traffic between the different Internet Services Providers (ISPs) in order to maximize the cloud service performance without requiring any intervention of the SOHO network administrator. This ongoing work is done in collaboration with Telecom ParisTech, Ericsson, LISPERS.net, and Cisco Systems and is presented in two papers [19], [20] and detailed in one IETF Internet-draft [19].

6.2.4. Rules Placement Problem in OpenFlow Networks

Participants: Xuan Nam Nguyen, Damien Saucez, Chadi Barakat and Thierry Turletti

Software-Defined Networking (SDN) abstracts low-level network functionalities to simplify network management and reduce costs. The OpenFlow protocol implements the SDN concept by abstracting network communications as flows to be processed by network elements. In OpenFlow, the high-level policies are translated into network primitives called rules that are distributed over the network. While the abstraction offered by OpenFlow allows to potentially implement any policy, it raises the new question of how to define the rules and where to place them in the network while respecting all technical and administrative requirements. We proposed a comprehensive study of the so-called OpenFlow rules placement problem with a survey of the various proposals intending to solve it [11] and developed an offline optimization framework for this problem with a polynomial time approximation in [13].

6.3. Experimental Evaluation

6.3.1. Automating ns-3 Experimentation in Multi-Host Scenarios

Participants: Alina Ludmila Quereilhac, Damien Saucez, Thierry Turletti and Walid Dabbous

ns-3 is a flexible simulator whose capabilities go beyond running purely synthetic simulations in a local desktop. Due to its ability to run unmodified Linux applications, to execute in real time mode, and to exchange traffic with live networks, ns-3 can be combined with live hosts to run distributed simulations or to transparently integrate live and simulated networks. Nevertheless, setting up ns-3 multi-host experiment scenarios might require considerable manual work and advanced system administration skills. The NEPI experiment management framework is capable of automating deployment, execution, and result collection of experiment scenarios that combine ns-3 with multiple hosts in various ways, reducing the burden of manual scenario set up. We proved that this approach can be used to seamlessly running parallel simulations on a cluster of hosts, running distributed simulation spanning multiple hosts, and integrating live and simulated networks. This work has been published in [18] and has been awarded as the best paper of the workshop.

6.3.2. DiG: Emulating Data Centers and Cloud Architectures in a Grid Network

Participants: Hardik Soni, Thierry Turletti, Damien Saucez

We are witnessing a considerable amount of research work related to data-center and cloud infrastructures but evaluations are often limited to small scale scenarios as very few researchers have access to a real infrastructure to confront their ideas to reality. We have designed an experiment automation tool, called DiG (Data-centers in the Grid), which explicitly allocates physical resources in grids to emulate data-center and cloud networks. DiG allows one to utilize grid infrastructures to evaluate research ideas pertaining to data-centers and cloud environments at massive scale and with real traffic workload. We have automated the procedure of building target network topologies while respecting effective performance capacity of available physical resources in the grid against the demand of links and hosts in the experiment. We demonstrate a showcase where DiG automatically builds a large data-center topology composed of hundreds of servers executing various Hadoop intensive workloads (see our demo abstract at IEEE NFV/SDN 2015 in [24]).

ECUADOR Project-Team

6. New Results

6.1. AD-adjoints and C dynamic memory management

Participants: Laurent Hascoët, Raphaël Couronné, Sri Hari Krishna Narayanan [Argonne National Lab. (Illinois, USA)], Mathieu Morlighem [University of California at Irvine (USA)].

One of the current frontiers of AD research is the definition of an adjoint AD model that can cope with dynamic memory management. This research is central in our ongoing effort towards adjoint AD of C, and more remotely towards AD of C++. This research is conducted in collaboration with the MCS department of Argonne National Lab. Our partnership is formalized by joint participation in the Inria joint lab JLESC, and partly funded by the Partner University Fund (PUF) of the French embassy in the USA.

Adjoint AD must reproduce in reversed order the control decisions of the original code. In languages such as C, allocation of dynamic memory and pointer management form a significant part of these control decisions. Reproducing memory (de)allocation in reverse means reallocating memory, possibly receiving a different memory chunk. Reproducing pointer addresses in reverse thus require to convert addresses in the former memory chunks into equivalent addresses in the new reallocated chunks. Together with Krishna Narayanan from Argonne, we experiment on real applications to find the most efficient solution to this address conversion problem. We jointly develop a library (called ADMM, ADjoint Memory Management) whose primitives can be used in AD adjoint code to handle this address conversion. Using this library together with Tapenade, we could obtain a correct AD adjoint code of a medium-size industrial code (“Multibody", structural mechanics) that exhibits a typical usage of C pointer arithmetic. This year, the same effort was conducted with the OpenAD AD tool, leading us to an ADMM library less dependent on the particular target AD tool. A joint publication with our colleagues from Argonne is in preparation.

In parallel, we investigate alternative implementation strategies for ADMM, one of which could be to build our own memory (de)allocation mechanism, This should ultimately rely on the standard C library. As a result, management of adjoint memory addresses could be done deeper in the system and therefore with a smaller overhead, at the cost of some additional portability issues.

We pursue our objective of improving reliability of the AD adjoint model for C codes to a similar level as achieved for Fortran. To this end we apply Tapenade to increasingly larger and complex C codes. In addition to the already mentioned “Multibody" application, we initiated differentiation of two new complex applications:

- “BLN" is a code developed by the Inria team ABS, that computes the potential energy of possible conformations of a macromolecule. Its gradient is used to explore the local minima in the energy landscape of these conformations. The AD adjoint of a Fortran implementation of BLN has been built by Tapenade and successfully validated. The adjoint of the C implementation is a challenge that helps us clarify the adjoint AD model that we use in Tapenade. The C version of BLN that we are considering is actually a (partly mechanical) translation of the actual C++ source. This makes this code an even more appealing and challenging test case. This work was mostly conducted by Raphaël Couronné as a part of his summer internship with us.
- “SEISM" is a code developed by Mathieu Morlighem from UC Irvine, jointly with Eric Larour from JPL. This is a glaciology code closely related to the larger “ISSM" code, in C++. One objective, addressed mostly by Mathieu Morlighem, is to clarify recommendations on the C programming style (again very much inspired here from the C++ style) that allows AD to perform better. The other objective, adressed mostly by our team, it to experiment with quite intricate data structures, where Tapenade’s static pointer destination analysis is used intensively.

6.2. AD-adjoints of MPI-parallel codes

Participants: Laurent Hascoët, Ala Taftaf, Sri Hari Krishna Narayanan [Argonne National Lab. (Illinois, USA)].

We have a long-standing collaboration with Argonne National Lab on the question of adjoint AD of message-passing parallel codes. We continued joint development of the Adjoinable-MPI library (AMPI) that provides efficient tangent and adjoint AD for MPI-parallel codes, independently of the AD tool used (now AdolC, dco, OpenAD, Tapenade).

During her PhD work, Ala Taftaf is considering the question of checkpointing applied to the AD-adjoint of an MPI-parallel code. Checkpointing is a memory/runtime tradeoff which is essential for adjoint AD of large codes, in particular parallel codes. However, for MPI codes this question has always been addressed by ad-hoc hand manipulations of the differentiated code, and with no formal assurance of correctness. Ala Taftaf is investigating the assumptions implicitly made during past experiments, to clarify and generalize them. On one hand we propose an extension of the adjoint of MPI point-to-point communication primitives, so that the semantics of an adjoint program is preserved for any placement of checkpoints. On the other hand, we propose an alternative extension of these adjoint communications, more efficient but that requires a number of restrictions on the placement of checkpoints. We shall try to provide proof of correctness of these strategies, and in particular demonstrate that they cannot introduce deadlocks. Tradeoffs between the two extensions should be investigated. Ala Taftaf presented her research on "'Adjoint-Checkpointing on MPI-parallel codes" at the EuroAD workshop in Paderborn, Germany, december 1-2. A conference article has been submitted to Eccomas 2016 in Crete.

6.3. AD-adjoints of Iterative Processes

Participants: Laurent Hascoët, Ala Taftaf, Sri Hari Krishna Narayanan [Argonne National Lab. (Illinois, USA)], Daniel Goldberg [University of Edinburgh, UK].

Adjoint codes naturally propagate partial gradients backwards from the result of the simulation. However, this uses the data flow of the simulation in reverse order, at a cost that increases with the length of the simulation. In the special case of iterative Fixed-Point loops, only the final converged result should be used: the "initial guess" and the intermediate non-converged states should not be considered by the adjoint calculation, and this remark brings enormous gain in memory use. We selected the strategy proposed by Bruce Christianson [22] and this year we continued its application to medium-size testcases provided by Queen Mary University for the AboutFlow project. We also simplified the user interface provided to trigger this special strategy extension in Tapenade. Ala Taftaf presented her results at the ECCOMAS Eurogen conference in Glasgow [15], september 14-16.

In parallel, we collaborated with Krishna Narayanan from ANL and Dan Goldberg from University of Edinburgh (UK) to implement the same strategy into the OpenAD tool, in view of applying it to a glaciology configuration of the MIT GCM code. A joint article describing the results has been submitted for publication.

6.4. AD-adjoints of large real codes

Participants: Laurent Hascoët, Valérie Pascual, Raphaël Couronné, Fabrice Zaoui [EDF R&D, LNHE].

In collaboration with EDF, Valérie Pascual is applying Tapenade to the hydrographic code "Mascaret". Both tangent and adjoint diferentiated codes have been built and validated. Application of the tangent differentiated Mascaret for data assimilation on two real cases is described in a joint publication [14].

During his summer internship, Raphaël Couronné has applied Tapenade to the MIT "GCM", a reference code in the Earth Sciences community. We have obtained a valid adjoint for a recommended configuration of this very large Fortran code. This test showed some maturity of the Tapenade tool for Fortran, as it turned out that no modification nor debug of the tool was needed. We are now discussing with the MIT team to schedule further collaboration.

In cooperation with the partners of the FP7 project UMRIDA, the team has assisted Alenia-Aermacchi (Filomena Cariglino and Nicola Ceresola) in the efficient differentiation of their Euler/Navier Stokes code “UNS3D" in tangent mode, dealing in particular with its use of MPI.

The team has assisted Marcin Wyrozebski from Warsaw University of Technology, to apply Tapenade to a CFD software from WUT.

6.5. Resolution of linearised systems and efficiency

Participants: Olivier Allain [Lemma], Gautier Brèthes, Alain Dervieux, Bruno Koobus [Université Montpellier 2], Emmanuelle Itam [Université Montpellier 2], Vincent Levasseur [Lemma], Stephen Wornom [Lemma].

For Fluid Mechanics as well as for Structural Mechanics, an implicit time-advancing is mandatory. It can be applied efficiently if the large systems involved are solved with a good parallel algorithm. In the 90’s, a generation of solution algorithms was devised on the basis of Domain Decomposition Methods (DDM). For complex models (compressible flows...), Schwarz DDM were combined with quasi-Newton algorithms such as GMRES. These are for example Restrictive Additive Schwarz (RAS), which is used in our platform AIRONUM. RAS was developed by Cai, Farhat and others. RAS is an ancestor of the widely used class of Newton-Krylov-Schwarz (NKS) algorithms. For hundreds of processors many versions of NKS, and in particular RAS, are almost scalable (convergence rate independant of the number of processors). But scalability vanishes for a medium-large number of processors (thousands). In the ANR ECINADS, coordinated by Ecuador, a Coarse-Grid Deflated RAS was developped: iteration-wise scalability holds for all parts, except for the coarse grid direct solver, which concerns a much smaller problem. Effective Convergence Scalability (ECS) was confirmed up to 2048 processors. Beyond this level the asymptotic complexity of the coarse-grid direct solver becomes predominant and ECS is lost. In other words, with a Coarse-Grid Deflated RAS, the size of the coarse grid problem which is solved by a direct algebraic solver must be limited in order to enjoy ECS. For finer meshes, the coarse system cannot be finer, and the efficiency is lower. It is then natural to consider intermediate meshes on which iterative solvers will be applied. In the ANR MAIDESC, Gautier Brèthes has defined a multi-mesh Full MultiGrid (FMG) algorithm adapted to anisotropic mesh adaptation. In 2015, the method has been extended to MPI-based massive parallelism, in cooperation with the Lemma team for the computation of incompressible flows. As a perspective, our parallel MG can be complemented with the previous version of the solver (deflated RAS) for a higher degree of scalability.

A second issue which we addressed is the use of explicit time advancing. Many unsteady flows have to be computed with explicit time advancing. A single explicit time step is of a low cost and can be highly accurate. Explicit time advancing is mandatory for wave propagation: blast shocks of vortices in wakes. However the meshes used may involve small regions in which the explicit time step should be very small and large regions in which such a small time step is a waste. The family of time-advancing methods in which unsteady phenomena are computed using different time steps in different regions is called the multirate methods. In our cooperation with University of Montpellier, a novel multirate method using cell-agglomeration has been designed and developed in our AIRONUM platform. An article is in preparation. This work takes place in the ANR MAIDESC programme.

6.6. Control of approximation errors

Participants: Gautier Brèthes, Eléonore Gauci, Alain Dervieux, Adrien Loseille [GAMMA team, Inria-Rocquencourt], Frédéric Alauzet [GAMMA team, Inria-Rocquencourt], Stephen Wornom [Lemma], Anca Belme [university of Paris 6].

The study of combination of full multigrid (FMG) with anisotropic mesh adaption (AA), started with the thesis of Gautier Brèthes, has been published [13].

Further studies of mesh adaptation for viscous flows are currently performed and a journal paper, joint with Inria team Gamma3 and University of Paris 6 (Anca Belme) is in preparation.

An important novelty in mesh adaption is the norm-oriented AA method. The method relies on the definition of ad hoc correctors. It has been developed in the academic platform “FMG" for elliptic problems. Gautier Brèthes gave several presentations in conferences, a journal article has been submitted. The introduction of the norm-oriented idea considerably amplifies the impact of adjoint-based AA. The applied mathematician and the engineer now have methods when faced to mesh adaptation for the simulation of a complex PDE system, since they can specify which error norm level they wish, and for which norm [12], [16]. Another version is developed jointly with Inria team Gamma3 for the compressible Euler model [19].

A cooperation has started between Gautier Brèthes et Thierry Coupez (Ecole Centrale de Nantes) concerning discrete metrics. This takes place in the ANR MAIDESC program. An article is in preparation.

Éléonore Gauci started last year a thesis (co-advised by Frédéric Alauzet) on the study of norm-oriented criteria for CFD and coupled CSM-CFD systems. She gave a presentation at the “Coupled Problems" symposium.

Post-doc Guilherme Cunha did a study (in cooperation with Lemma) on the combination of mesh adaptation and coefficient identification for unsteady phenomena.

The theoretical studies are supported by an ANR project MAIDESC coordinated by ECUADOR and Gamma3, which deals with meshes for interfaces, third-order accuracy, meshes for boundary layers, and curved meshes.

CFD application are supported by the European FP7 project UMRIDA which deals with the application of AA to approximation error modelling and control.

6.7. Turbulence models

Participants: Alain Dervieux, Bruno Koobus [University of Montpellier 2], Emmanuelle Itam [University of Montpellier 2], Marianna Braza [CNRS-IMFT at Toulouse], Stephen Wornom [Lemma], Bruno Sainte-Rose [Lemma].

The purpose of our work in hybrid RANS/LES is to develop new approaches for industrial applications of LES-based analyses. In the applications targetted (aeronautics, hydraulics), the Reynolds number can be as high as several tenth millions, far too high for pure LES models. However, certain regions in the flow can be better predicted with LES than with usual statistical RANS (Reynolds averaged Navier-Stokes) models. These are mainly vortical separated regions as assumed in one of the most popular hybrid model, the hybrid Detached Eddy Simulation model. Here, “hybrid” means that a blending is applied between LES and RANS. An important difference between a real life flow and a wind tunnel or basin is that the turbulence of the flow upstream of each body is not well known.

This year, we have continued the evaluation of a dynamic formulation of Piomelli-Germano type for the Variational-multiscale model. We have also modified the integration of the boundary layer by adding the so-called Menter correction imposing the Bradshaw law. We have studied these improvements on multiple-body flows. An emblematic case is the interaction between two parallel cylinders, one being in the wake of the other. A flow around a space probe at high Reynolds number is also studied [18], [17].

FOCUS Project-Team

7. New Results

7.1. Service-oriented computing

Participants: Maurizio Gabbrielli, Elena Giachino, Saverio Giallorenzo, Claudio Guidi, Mario Bravetti, Cosimo Laneve, Ivan Lanese, Michael Lienhardt, Jacopo Mauro, Fabrizio Montesi, Gianluigi Zavattaro.

7.1.1. Orchestrations

Orchestration models and languages in the context of Service-Oriented Architectures (SOA) are used to describe the composition of services focusing on their interactions. Concrete web services are connected to abstract service definitions for the aim of service discovery. In [16] we study a natural notion of compliance between clients and services in terms of their bpel (abstract) descriptions. The induced preorder shows interesting connections with the must preorder and has normal form representatives that are parallel-free finite-state activities, called contracts. Moreover, in [22] we focus on advancements of the orchestration language Jolie aiming at the development of dynamically adaptable orchestrated systems.

7.1.2. Choreographies

Choreographies are high-level descriptions of distributed interacting systems featuring as basic unit a communication between two participants. A main feature of choreographies is that they ensure deadlock-freedom by construction. From a choreography one can automatically derive a description of the behaviour of each participant using a notion of projection. Choreographies can be used both at the level of types (multiparty session types) or as a programming language. In [19] we surveyed our results about verification of adaptable processes, focusing in particular on distributed adaptability, where a process can update part of a protocol (specified by a choreography) by performing dynamic distributed updates over a set of protocol participants. In [14] we illustrate our approach to develop and verify distributed, adaptive software systems. The cornerstone of our framework is the use of choreography languages, which allow us to obtain correctness by construction. Moreover, in [36] we present DIOC, a language for programming distributed applications that are free from deadlocks and races by construction. A DIOC program describes a whole distributed application as a unique entity (choreography). DIOC allows the programmer to specify which parts of the application can be updated. At runtime, these parts may be replaced by new DIOC fragments from outside the application. DIOC programs are compiled, generating code for each site, in a lower-level language called DPOC. As a consequence DPOC applications are free from communication deadlocks and races, even in presence of runtime updates.

7.2. Models for reliability

Participants: Elena Giachino, Ivan Lanese.

7.2.1. Reversibility

We have continued the study of causal-consistent reversibility started in the past years. In [42] we defined the causal-consistent reversible semantics (both controlled and uncontrolled) of muKlaim, a formal coordination language based on distributed tuple spaces, by adapting the approach developed for message passing calculi in the past years. A major novelty is that the muKlaim read primitive allows two processes to access a shared resource independently, giving rise to a causality structure which is not found in message passing calculi.

In [31] we studied the issue of compliance of a client w.r.t. a server in a reversible setting using behavioural contracts. In particular, when an agreement cannot be reached, the client and the server can synchronously rollback to the last point of choice, looking for alternatives. As a main result, we showed that compliance is decidable even for recursive contracts.

7.3. Cloud Computing and Deployment

Participants: Elena Giachino, Saverio Giallorenzo, Claudio Guidi, Cosimo Laneve, Michael Lienhardt, Jacopo Mauro, Gianluigi Zavattaro.

7.3.1. Cloud application deployment

Configuration and management of applications in the cloud is a complex task that requires advanced methodologies and tools. A foundational study of the problem has been carried out in [21] where we have identified the critical tasks to be solved, quantified their computational complexity, and proposed simplifications to the problem with the idea of limiting the computational complexity at the cost of having approximated (but acceptable, in most cases) solutions. Our attention has been dedicated to the implementation of a tool for the efficient solution of one of these tasks, namely, the automatic planning of the management actions needed to properly configure a cloud application [17]. This tool, called Metis, has been already exploited in the realization of an integrated platform for the automatic deployment of the cloud application called Blender [39] as well as in the context of the ABS modeling language [37] in order to be able to support the automatic reasoning about deployment costs already during the early phases of application design and development. We have also performed a foundational study of the problem of reconfiguring an application instead of deploying it from scratch. Our foundational study allowed us, on the one hand, to quantify the computational complexity of the problem (PSpace-Complete) and, on the other hand, to precisely identify the source of such complexity (the presence of legacy components that cannot be re-deployed from scratch).

7.3.2. Cloud resource management

One of the main challenges in the management of cloud applications is the quantification of the computing resources needed by the applications to be deployed. More precisely, it is important to quantify upper bounds to the number of needed computing resources in order to either previously acquire them or have a precise quantification of the costs for executing an application. In [40] a static analysis technique is proposed that computes upper bounds of virtual machine usages in a concurrent language with explicit acquire and release operations of virtual machines. See the section on deadlock analysis for more details.

7.4. Resource Control and Probabilities

Participants: Michele Alberti, Martin Avanzini, Flavien Breuvart, Alberto Cappai, Ugo Dal Lago, Simone Martini, Giulio Pellitta, Alessandro Rioli, Davide Sangiorgi, Marco Solieri, Valeria Vignudelli.

7.4.1. Resource Control

7.4.1.1. Time Complexity Analysis of Concurrent and Higher-Order Functional Programs

We have extensively studied the problem of automatically analysing the complexity of programs. We first of all studied the problem for concurrent object-oriented programs [41]. To determine this complexity we have used intermediate abstract descriptions that record relevant information for the time analysis, called behavioural types. Behavioural types are then translated into so-called cost equations, making parallelism explicit. Cost equations are finally fed into an automatic off-the-shelf solver for obtaining the actual time complexity. The same problem has been also analysed when the underlying program is functional [29]. We showed how the complexity of higher-order functional programs can be analysed automatically by applying program transformations to a defunctionalized version of them, and feeding the result to existing tools for the complexity analysis of first-order term rewrite systems. This is done while carefully analysing complexity preservation and reflection of the employed transformations such that the complexity of the obtained term rewrite system reflects on the complexity of the initial program. This approach turns out to work well in practice, in particular since off-the-shelf complexity tool for first-order rewrite systems matured to a state where they are both fast and powerful. However, the implementation of such tools is quite sophisticated. To ensure correctness of the obtained complexity bounds, we extended CeTA, a certified proof checker for rewrite tools, with the formalisation of various complexity techniques underlying state-of-the-art complexity tools [30]. This way, we detected conflicts in theoretical results as well as bugs in existing complexity provers.

7.4.1.2. Function Algebras and Implicit Complexity

A fundamental result about ramified recurrence, one of the earliest systems in implicit complexity, has been proved [28]. This has been obtained through a careful analysis on how the adoption of an evaluation mechanism with sharing and memoization impacts the class of functions which can be computed in polynomial time. We have first shown how a natural cost model in which lookup for an already computed result has no cost is indeed invariant. As a corollary, we have then proved that the most general notion of ramified recurrence is sound for polynomial time.

7.4.1.3. Geometry of Interaction

We see the the geometry of interaction as a foundational framework in which the efficiency of higher-order computation can be analyzed. This has produced some very interesting results, also stimulated by the bilateral Inria project CRECOGI, which started this year. We have first of all studied the geometry of interaction of the resource lambda-calculus, a model of linear and nondeterministic functional languages. In a strictly typed restriction of the resource lambda-calculus, we have studied the notion of path persistence, and defined a geometry of interaction that characterises it [18]. Furthermore, we have carried out our work on multitoken machines, started in 2014. More specifically, we have studied multitoken interaction machines in the context of a very expressive linear logical system with exponentials, fixpoints and synchronization [34]. On the one hand, we have proved that interaction is guaranteed to be deadlock-free. On the other hand, the resulting logical system has been proved to be powerful enough to embed PCF and to adequately model its behaviour, both when call-by-name and when call-by-value evaluation are considered.

7.4.2. Probabilistic Models

7.4.2.1. Applicative Bisimilarity

Notions of equivalences for probabilistic programming languages have been studied and analysed, together with their relationships with context equivalence. More specifically, we have studied how applicative bisimilarity behaves when instantiated on a call-by-value probabilistic lambda-calculus, endowed with Plotkin's parallel disjunction operator [20]. We have proved that congruence and coincidence with the corresponding context relation hold for both bisimilarity and similarity, the latter known to be impossible in sequential languages. We have also shown that applicative bisimilarity works well when the underlying language of programs takes the form of a linear lambda-calculus extended with quantum data [35]. The main results are proofs of soundness for the obtained notion of equivalence.

7.4.2.2. From Equivalences to Metrics

The presence of probabilistic (thus quantitative) notions of observation makes equivalence relations too coarse-grained as ways to compare programs. This opens the way to metrics in which, indeed, not all non-equivalent programs are at the same distance. We have studied the problem of evaluating the distance between affine lambda-terms [33]. A natural generalisation of context equivalence has been shown to be characterised by a notion of trace distance, and to be bounded from above by a coinductively defined distance based on the Kantorovich metric on distributions. A different, again fully-abstract, tuple-based notion of trace distance has been shown to be able to handle nontrivial examples. A similar thing has been done in a calculus for probabilistic polynomial time computation [32], thus paving the way towards getting effective proof methodologies for computational indistinguishability, a key notion in modern cryptography.

7.5. Verification techniques for extensional properties

Participants: Daniel Hirschkoff, Elena Giachino, Michael Lienhardt, Cosimo Laneve, Jean-Marie Madiot, Davide Sangiorgi.

Extensional properties are those properties that constrain the behavioural descriptions of a system (i.e., how a system looks like from the outside). Examples of such properties include classical functional correctness, deadlock freedom and resource usage. Related to techniques for extensional properties are the issues of decidability (the problem of establishing whether certain properties are computationally feasible).

7.5.1. Static analysis of deadlock freedom and resource usage

Deadlock detection in concurrent programs that create networks with an arbitrary number of nodes is extremely complex and solutions either give imprecise answers or do not scale. To enable the analysis of such programs, we have studied an algorithm for detecting deadlocks in a basic concurrent object-oriented language. The algorithm (i) associates behavioural types, called *lam*, to programs by means of a type inference system and (ii) uses an ad-hoc verification technique highlighting circular dependencies in lam [15]. The algorithm has been prototyped and has been extended to a full-fledged programming language, called `ABS`.

A technique similar to [15] has been used for computing upper bounds of resource usages in [40]. In particular, the metaphor in this paper has been virtual machines usage in a concurrent language with explicit acquire and release operations. The problematic issue in such languages is when the release is delegated to other (ad-hoc or third party) concurrent codes (by passing them as arguments of invocations) – a feature that is currently used in `Amazon Elastic Cloud Computing` or in the `Docker FiWare`. As for deadlock analysis, the technique is modular and consists of (i) a type system associating programs with behavioural types that records relevant information for resource usage (creations, releases, and concurrent operations), (ii) a translation function that takes behavioural types and returns cost equations, and (iii) an automatic off-the-shelf solver for the cost equations. A soundness proof of the type system establishes the correctness of the above technique with respect to the cost equations. The technique has also been experimentally evaluated and the experiments show that it allows us to derive bounds for programs that are better than other techniques, such as those based on amortized analysis.

Another technique for enforcing program correctness is the one used in [36], [14], where the programming of distributed applications is guaranteed to be free from communication deadlocks and races by means of *choreographies*. Choreographies are behavioural types which allow one to obtain correctness by construction (more details on this paper in Section 7.1).

7.5.2. Name mobility

The article [44] studies the behavioural theory of $\pi\mathtt{P}$, a π-calculus featuring restriction as the only binder. In contrast with calculi such as Fusions and Chi, reduction in $\pi\mathtt{P}$ generates a preorder on names rather than an equivalence relation. Two characterisations of barbed congruence in $\pi\mathtt{P}$ are analyzed: the first is based on a compositional LTS, and the second is an axiomatisation. The results in this paper bring out basic properties of $\pi\mathtt{P}$, mostly related to the interplay between the restriction operator and the preorder on names.

7.5.3. Coinductive techniques

Coinductive techniques, notably those based on bisimulation, are widely used in concurrency theory to reason about systems of processes. The bisimulation proof method can be enhanced by employing "bisimulations up-to" techniques. A comprehensive theory of such enhancements has been developed for first-order (i.e., CCS-like) LTSs and bisimilarity, based on the notion of compatible function for fixed-point theory.

A proof method different from bisimulation is investigated in [46], [23]. This method is based on unique solution of special forms of inequations called contractions, and inspired by Milner's theorem on unique solution of equations. The method is as powerful as the bisimulation proof method and its "up-to context" enhancements. The definition of contraction can be transferred onto other behavioural equivalences, possibly contextual and non-coinductive. This enables a coinductive reasoning style on such equivalences, either by applying the method based on unique solution of contractions, or by injecting appropriate contraction preorders into the bisimulation game. The technique can be applied both to first-order languages and to higher-order languages.

7.5.4. Expressiveness and decidability in actor-like systems

In [48], the limit of classical Petri nets is studied by discussing when it is necessary to move to the so-called Transfer nets, in which transitions can also move to a target place all the tokens currently present in a source place. More precisely, we consider a simple calculus of processes that interact by generating/consuming messages into/from a shared repository. For this calculus classical Petri nets can faithfully model the process behavior. Then we present a simple extension with a primitive allowing processes to atomically rename all the data of a given kind. We show that with the addition of such primitive it is necessary to move to Transfer nets to obtain a faithful modeling.

7.6. Constraint Programming

Participants: Roberto Amadini, Maurizio Gabbrielli, Jacopo Mauro.

The Constraint Programming (CP) paradigm is a general and powerful framework that enables to express relations between different entities in form of constraints that must be satisfied. The concept of constraint is ubiquitous and not confined to the sciences: constraints appear in every aspect of daily life in the form of requirements, obligations, or prohibitions. Historically, the FOCUS group has always had an interest in CP, see e.g., [53], [54]. The possible applications of CP are in fact numerous and disparate. As an example, CP can be used for the deployment of services in the cloud [21], [39].

CP essentially consists of two layers: (i) a modeling level, in which a real-life problem is identified, examined, and formalized into a mathematical model by human experts; (ii) a solving level, aimed at resolving as efficiently and comprehensively as possible the model defined in (i) by means of software agents called constraint solvers. Over the last years we dealt with a particular aspect of CP, that is, the so called portfolio approaches [12], [27], [10]. In a nutshell, a portfolio approach in CP can be seen as the problem of predicting which is (are) the best constraint solver(s) —among a portfolio of available solvers— for solving a given CP problem. A constraint solver that relies on a portfolio of underlying, individual solvers is also dubbed a portfolio solver.

Our studies on portfolio approaches lead to development of the SUNNY-CP portfolio solver [26], [25]. SUNNY-CP relies on underlying state-of-the-art constraint solvers for solving a given CP problem encoded in the MiniZinc language, nowadays a de-facto standard for modeling CP problems. Initially developed as a sequential solver [26], SUNNY-CP has been later on enhanced by enabling the simultaneous execution of its solvers on different cores [25]. This extension allowed SUNNY-CP to win the gold medal in the open track of 2015 MiniZinc Challenge [cite], the annual competition for CP solvers.

However, we did not restrict the work on portfolio approaches to the CP field only. Indeed, we also performed some preliminary studies for evaluating SUNNY (i.e., the algorithm on which SUNNY-CP relies) in other application domains like, e.g., Boolean satisfiability (SAT), Quantified Boolean Formula (QBF), and Answer-Set Programming (ASP) [47], [24].

GALAAD2 Team

6. New Results

6.1. Certifying isolated singular points and their multiplicity structure

Participant: Bernard Mourrain.

The paper [4] presents two new constructions related to singular solutions of polynomial systems. The first is a new deflation method for an isolated singular root. This construction uses a single linear differential form defined from the Jacobian matrix of the input, and defines the deflated system by applying this differential form to the original system. The advantages of this new deflation is that it does not introduce new variables and the increase in the number of equations is linear instead of the quadratic increase of previous methods. The second construction gives the coefficients of the so-called inverse system or dual basis, which defines the multiplicity structure at the singular root. We present a system of equations in the original variables plus a relatively small number of new variables. We show that the roots of this new system include the original singular root but now with multiplicity one, and the new variables uniquely determine the multiplicity structure. Both constructions are "exact", meaning that they permit one to treat all conjugate roots simultaneously and can be used in certification procedures for singular roots and their multiplicity structure with respect to an exact rational polynomial system.

Joint work with Agnes Szanto, Department of Mathematics, North Carolina State University, Raleigh, USA; Jonathan D. Hauenstein, Department of Applied and Computational Mathematics and Statistics, University of Notre Dame, USA.

6.2. On the construction of general cubature formula by flat extensions

Participants: Marta Abril-Bucero, Bernard Mourrain.

We describe a new method to compute general cubature formulae [5]. The problem is initially transformed into the computation of truncated Hankel operators with flat extensions. We then analyse the algebraic properties associated to flat extensions and show how to recover the cubature points and weights from the truncated Hankel operator. We next present an algorithm to test the flat extension property and to additionally compute the decomposition. To generate cubature formulae with a minimal number of points, we propose a new relaxation hierarchy of convex optimization problems minimizing the nuclear norm of the Hankel operators. For a suitably high order of convex relaxation, the minimizer of the optimization problem corresponds to a cubature formula. Furthermore cubature formulae with a minimal number of points are associated to faces of the convex sets. We illustrate our method on some examples, and for each we obtain a new minimal cubature formula.

This is a joint work with C. Bajaj (Univ. of Austin, Texas, USA).

6.3. A moment matrix approach to computing symmetric cubatures

Participants: Mathieu Collowald, Evelyne Hubert.

A quadrature is an approximation of the definite integral of a function by a weighted sum of function values at specified points, or nodes, within the domain of integration. Gaussian quadratures are constructed to yield exact results for any polynomial of degree $2r-1$ or less by a suitable choice of r nodes and weights. Cubature is a generalization of quadrature in higher dimension. Constructing a cubature amounts to find a linear form

$$\Lambda : \mathbb{R}[x] \to \mathbb{R}, p \mapsto \sum_{j=1}^{r} a_j \, p(\xi_j)$$

from the knowledge of its restriction to $\mathbb{R}[x]_{\leq d}$. The unknowns are the number of nodes r, the weights a_j and the nodes ξ_j.

In [7] we use a basis-free version of an approach to cubatures based on moment matrices in terms of the Hankel operator $\mathcal{H}$ associated to Λ. The existence of a cubature of degree d with r nodes boils down to conditions of ranks and positive semidefiniteness on $\mathcal{H}$. We then recognize the nodes as the solutions of a generalized eigenvalue problem.

Standard domains of integration are symmetric under the action of a finite group. It is natural to look for cubatures that respect this symmetry. Introducing adapted bases obtained from representation theory, the symmetry constraint allows to block diagonalize the Hankel operator $\mathcal{H}$. We then deal with smaller-sized matrices both for securing the existence of the cubature and computing the nodes. The sizes of the blocks are furthermore explicitly related to the orbit types of the nodes with the new concept of the matrix of multiplicities of a finite group. It provides preliminary criteria of existence of a cubature with a given organisation of the nodes in orbit types.

The Maple implementation of the presented algorithms allows to determine, with moderate computational efforts, all the symmetric cubatures of a given degree. We present new relevant cubatures.

6.4. Invariantization of symmetric polynomial systems

Participants: Mathieu Collowald, Evelyne Hubert.

Assuming the variety of a set of polynomials is invariant under a group action, we provide a set of invariants that define the same variety. The contribution is about infinite algebraic groups, the case of finite group being previously known. We introduce for those a new concept of algebraic invariantization. It is based on the construction of rational invariants by Hubert and Kogan [14], a construction for which we provide here new simplified proofs.

6.5. Effective criterions for bigraded birational maps

Participant: Laurent Busé.

A rational map $\mathcal{F}: \mathbb{P}^m \dashrightarrow \mathbb{P}^n$ between projective spaces is defined by a collection of homogeneous polynomials $\mathbf{f} := (f_0, ..., f_n)$ in $m+1$ variables of the same degree. The problem of deciding or providing sufficient conditions for such a map $\mathcal{F}$ to be birational have attracted a lot of interest in the past and it is still an active area of research. Methods that are based of some properties of the syzygies of $\mathbf{f}$ are definitely the more adapted for computational purposes in the sense that they make the problem of birationality effectively computable in the usual implementation of the Gröbner basis algorithm. The goal of this work is to extend these syzygies-based methods and techniques to the context of rational maps whose source is a product of two projective spaces $\mathbb{P}^r \times \mathbb{P}^s$ instead.

An important motivation for considering bi-graded rational maps comes from the field of geometric modeling. Indeed, the geometric modeling community uses almost exclusively bi-graded rational maps for parameterizing curves, surfaces or volumes under the name of rational tensor-product Bézier parameterizations. It turns out that an important property is to guarantee the birationality of these parameterizations onto their images. An even more important property is to preserve this birationality property during a design process, that is to say when the coefficients of the defining polynomials are continuously modified. As a first attempt to tackle these difficult problems, we analyze in detail birational maps from $\mathbb{P}^1 \times \mathbb{P}^1$ to $\mathbb{P}^2$ in low bi-degree by means of syzygies.

This work is done in the context of the SYRAM project which is funded by the MathAmSud programme. It is a collaboration with N. Botbol (University of Buenos Aires), M. Chardin (University of Paris 6), H. Hassanzadeh (University of Rio de Janeiro), A. Simis (University de Pernambuco) and Q. H. Tran (University of Paris 6). A paper is in preparation.

6.6. Orthogonal projection of points on Bézier curves and surfaces

Participant: Laurent Busé.

In this work, we introduce a new method for computing the orthogonal projections of a point onto a Bézier curve or surface. It is based on the concept of matrix representation we have introduced and developed in some previous works, which is here applied to the parameterizations of the normal planes or lines of a curve or surface, respectively. It consists in the computation of a matrix depending of the ambient space variables, which is done in a pre-processing step, and then the use of tools from numerical linear algebra for a fast and accurate solving of each instance of the problem.

This is an on going work done in the context of the SYRAM project which is funded by the MathAmSud programme. It is a collaboration with N. Botbol (University of Buenos Aires) and M. Chardin (University of Paris 6).

6.7. Extraction of cylinders and cones from minimal point sets

Participants: Laurent Busé, André Galligo, Jiajun Zhang.

The extraction of geometric primitives from 3D point clouds is an important problem in reverse engineering. These 3D point clouds are typically obtained by means of accurate 3D scanners and there exists several methods for performing the 3D geometric primitives extraction. An important category among these method are based on a RANSAC method. For such methods, the primitives are directly extracted for the input point cloud. The basic idea is to extract a particular elementary type of shape, such as planes, spheres, cylinders, cones or tori, from the smallest possible set of points and then to judge if this extracted primitive is relevant to the full point cloud. Therefore, for this category of methods it is very important to compute a particular type of shape through the smallest possible number of points, including normals or not. If the extraction of planes and spheres is easy to treat, the cases of cylinders, cones and tori are more involved. In this work, we aim at developing methods for extracting these geometric primitives from the smaller possible number of points (counting multiplicities if normals are taken into account). Another objective is also to provide methods for extraction without using estimated normals in order to improve the accuracy of the extracted geometric primitive, or to use mixed data depending of the applied context (some points with normals and some other points without normals). A paper is in preparation.

6.8. Discriminant of a complete intersection space curve

Participant: Laurent Busé.

In this work, we develop the formalism of the discriminant of a complete intersection curve in the three dimensional projective space, that is to say a curve which is represented as the zero locus of two homogeneous polynomials in four variables. Our main objective is to provide a new computational approach to this object without relying on the so-called "Cayley trick" for which it is necessary to introduce new variables. We also aim at getting a universal definition of this discriminant over the integers so that it holds under any specialization of the coefficients to an arbitrary commutative ring. Another aspect of this work is to explore properties of this discriminant, typically invariance, covariance and change of basis properties.

This is an on going work which is done in collaboration with Ibrahim Nonkane (University of Ouagadougou, Burkina Faso).

6.9. Resultants, flexes, and the generalization of Salmon's formula

Participant: Laurent Busé.

Given an algebraic variety $S \subset \mathbb{P}^n$ and a point $p \in S$, the osculation order of the point p is the maximum of the multiplicity of intersection at p of S with any line through p. We denote it by μ_p and define $Flex(S) = \{p \in \mathbb{P}^n | \mu_p > n\}$.

If $n = 2$, it is known that if C is a plane algebraic curve of degree d then $Flex(C)$ is the intersection of C with its Hessian, this latter being of degree $3d - 6$. A famous generalization of this result to the case $n = 3$ has been obtained by Salmon in 1860: for a general variety S, $Flex(S)$ is the intersection of S with another hypersurface of degree $11d - 24$. In this work, we are studying the generalization of this formula to arbitrary dimension n. We proved that given $S \subset \mathbb{P}^n$ of degree d, $Flex(S)$ is obtained by intersecting S with another hypersurface of degree

$$d\left(\sum_{k=1}^{n} \frac{n!}{k}\right) - n!$$

We are also looking for an explicit expression of an equation of this latter hypersurface.

This is a work in progress which is done in the context of a PICS collaboration funded by CNRS. It is a joint work with M. Chardin (University Paris 6), C. D'Andrea (University of Barcelona), M. Sombra (University of Barcelona) and M. Weiman (University of Caen).

6.10. Computer Algebra Applied to a Solitary Waves Study

Participant: André Galligo.

In [3], we apply Computer algebra techniques, such as algebraic computations of resultants and discriminants, certified drawing (with a guaranteed topology) of plane curves, to a problem in Fluid dynamics: We investigate "capillary-gravity" solitary waves in shallow water, relying on the framework of the Serre-Green-Naghdi equations. So, we deal with 2 dimensional surface waves, propagating in a shallow water of constant depth. By a differential elimination process, the study reduces to describing the solutions of an ordinary non linear first order differential equation, depending on two parameters. The paper is illustrated with examples and pictures computed with the computer algebra system Maple.

Joint work with Didier Clamond (University of Nice, France) and Denys Dutykh (University of Le Bourget, France).

6.11. H1-parameterizations of plane physical domains with complex topology in Isogeometric analysis

Participants: André Galligo, Bernard Mourrain, Meng Wu.

Isogeometric analysis (IGA) is a method for solving geometric partial differential equations(PDEs). Generating parameterizations of a PDE's physical domain is a basic and important issue within IGA framework. In [13] , we present a global H1-parameterization method for a planar physical domain with complex topology.

Joint work with B. NKonga, Univeristy of Nice - Sophia Antipolis and EPI CASTOR, Inria.

GEOMETRICA Project-Team

7. New Results

7.1. Mesh Generation and Geometry processing

7.1.1. Discrete Derivatives of Vector Fields on Surfaces An Operator Approach

Participants: Frédéric Chazal, Maksim Ovsjanikov.

In collaboration with O. Azencot, M. Ben Chen (Technion, Israel Institute of Technology).

Vector fields on surfaces are fundamental in various applications in computer graphics and geometry processing. In many cases, in addition to representing vector fields, the need arises to compute their derivatives, for example, for solving partial differential equations on surfaces or for designing vector fields with prescribed smoothness properties. In this work, we consider the problem of computing the Levi-Civita covariant derivative, that is, the tangential component of the standard directional derivative, on triangle meshes. This problem is challenging since, formally, tangent vector fields on polygonal meshes are often viewed as being discontinuous, hence it is not obvious what a good derivative formulation would be. We leverage the relationship between the Levi-Civita covariant derivative of a vector field and the directional derivative of its component functions to provide a simple, easy-to-implement discretization for which we demonstrate experimental convergence. In addition, we introduce two linear operators which provide access to additional constructs in Riemannian geometry that are not easy to discretize otherwise, including the parallel transport operator which can be seen simply as a certain matrix exponential. Finally, we show the applicability of our operator to various tasks, such as fluid simulation on curved surfaces and vector field design, by posing algebraic constraints on the covariant derivative operator.

7.1.2. Isotopic Meshing within a Tolerance Volume

Participant: David Cohen-Steiner.

In collaboration with M. Mandad, P. Alliez (Titane Project-team).

We give an algorithm [22] that generates from an input tolerance volume a surface triangle mesh guaranteed to be within the tolerance, intersection free and topologically correct. A pliant meshing algorithm is used to capture the topology and discover the anisotropy in the input tolerance volume in order to generate a concise output. We first refine a 3D Delaunay triangulation over the tolerance volume while maintaining a piecewise-linear function on this triangulation, until an isosurface of this function matches the topology sought after. We then embed the isosurface into the 3D triangulation via mutual tessellation, and simplify it while preserving the topology. Our approach extends to surfaces with boundaries and to non-manifold surfaces. We demonstrate the versatility and efficacy of our approach on a variety of data sets and tolerance volumes.

7.1.3. CGALmesh: A Generic Framework for Delaunay Mesh Generation

Participants: Jean-Daniel Boissonnat, Clément Jamin, Mariette Yvinec.

In collaboration with P. Alliez (Titane Project-team).

CGALmesh [21] is the mesh generation software package of the Computational Geometry Algorithm Library (CGAL). It generates isotropic simplicial meshes—surface triangular meshes or volume tetrahedral meshes—from input surfaces, 3D domains, and 3D multidomains, with or without sharp features. The underlying meshing algorithm relies on restricted Delaunay triangulations to approximate domains and surfaces and on Delaunay refinement to ensure both approximation accuracy and mesh quality. CGALmesh provides guarantees on approximation quality and on the size and shape of the mesh elements. It provides four optional mesh optimization algorithms to further improve the mesh quality. A distinctive property of CGALmesh is its high flexibility with respect to the input domain representation. Such a flexibility is achieved through a careful software design, gathering into a single abstract concept, denoted by the oracle, all required interface features between the meshing engine and the input domain. We already provide oracles for domains defined by polyhedral and implicit surfaces.

7.2. Topological and Geometric Inference

7.2.1. Subsampling Methods for Persistent Homology

Participants: Frédéric Chazal, Bertrand Michel.

In collaboration with B.T. Fasy, F. Lecci, A. Rinaldo and L. Wasserman (Carnegie Mellon University).

Persistent homology is a multiscale method for analyzing the shape of sets and functions from point cloud data arising from an unknown distribution supported on those sets. When the size of the sample is large, direct computation of the persistent homology is prohibitive due to the combinatorial nature of the existing algorithms. We propose to compute the persistent homology of several subsamples of the data and then combine the resulting estimates. We study the risk of two estimators and we prove that the subsampling approach carries stable topological information while achieving a great reduction in computational complexity.

7.2.2. Efficient and Robust Persistent Homology for Measures

Participants: Frédéric Chazal, Steve Oudot.

In collaboration with M. Buchet (Ohio State University) and Donald Sheehy (University of Connecticut).

A new paradigm for point cloud data analysis has emerged recently, where point clouds are no longer treated as mere compact sets but rather as empirical measures. A notion of distance to such measures has been defined and shown to be stable with respect to perturbations of the measure. This distance can eas- ily be computed pointwise in the case of a point cloud, but its sublevel-sets, which carry the geometric infor- mation about the measure, remain hard to compute or approximate. This makes it challenging to adapt many powerful techniques based on the Euclidean distance to a point cloud to the more general setting of the distance to a measure on a metric space. We propose [28] an efficient and reliable scheme to approximate the topological structure of the family of sublevel-sets of the distance to a measure. We obtain an algorithm for approximating the persistent homology of the distance to an empirical measure that works in arbitrary metric spaces. Precise quality and complexity guarantees are given with a discussion on the behavior of our approach in practice.

7.2.3. Topological analysis of scalar fields with outliers

Participants: Frédéric Chazal, Steve Oudot.

In collaboration with M. Buchet, T.K. Dey, F. Fan, Y. Wang (Ohio State University).

Given a real-valued function f defined over a manifold M embedded in Euclidean space, we are interested in recovering structural information about f from the sole information of its values on a finite sample P [27]. Existing methods provide approximation to the persistence diagram of f when the noise is bounded in both the functional and geometric domains. However, they fail in the presence of aberrant values, also called outliers, both in theory and practice. We propose a new algorithm that deals with outliers. We handle aberrant functional values with a method inspired from the k-nearest neighbors regression and the local median filtering, while the geometric outliers are handled using the distance to a measure. Combined with topological results on nested filtrations, our algorithm performs robust topological analysis of scalar fields in a wider range of noise models than handled by current methods. We provide theoretical guarantees on the quality of our approximation and some experimental results illustrating its behavior.

7.2.4. Zigzag Persistence via Reflections and Transpositions

Participants: Clément Maria, Steve Oudot.

We introduce [33] a simple algorithm for computing zigzag persistence, designed in the same spirit as the standard persistence algorithm. Our algorithm reduces a single matrix, maintains an explicit set of chains encoding the persistent homology of the current zigzag, and updates it under simplex insertions and removals. The total worst-case running time matches the usual cubic bound.

A noticeable difference with the standard persistence algorithm is that we do not insert or remove new simplices "at the end" of the zigzag, but rather "in the middle". To do so, we use arrow reflections and transpositions, in the same spirit as reflection functors in quiver theory. Our analysis introduces a new kind of reflection called the "weak-diamond", for which we are able to predict the changes in the interval decomposition and associated compatible bases. Arrow transpositions have been studied previously in the context of standard persistent homology, and we extend the study to the context of zigzag persistence. For both types of transformations, we provide simple procedures to update the interval decomposition and associated compatible homology basis.

7.2.5. Stable Topological Signatures for Points on 3D Shapes

Participants: Mathieu Carrière, Steve Oudot, Maksims Ovsjanikovs.

Comparing points on 3D shapes is among the fundamental operations in shape analysis. To facilitate this task, a great number of local point signatures or descriptors have been proposed in the past decades. However, the vast majority of these descriptors concentrate on the local geometry of the shape around the point, and thus are insensitive to its connectivity structure. By contrast, several *global* signatures have been proposed that successfully capture the overall topology of the shape and thus characterize the shape as a whole. We propose [29], [43] the first point descriptor that captures the topology structure of the shape as ‘seen’ from a single point, in a multiscale and provably stable way. We also demonstrate how a large class of topological signatures, including ours, can be mapped to vectors, opening the door to many classical analysis and learning methods. We illustrate the performance of this approach on the problems of supervised shape labeling and shape matching. We show that our signatures provide complementary information to existing ones and allow to achieve better performance with less training data in both applications.

7.2.6. Structure and Stability of the 1-Dimensional Mapper

Participants: Mathieu Carrière, Steve Oudot.

Given a continuous function $f : X \to \mathbb{R}$ and a cover I of its image by intervals, the Mapper is the nerve of a refinement of the pullback cover $f^{-1}(I)$. Despite its success in applications, little is known about the structure and stability of this construction from a theoretical point of view. As a pixelized version of the Reeb graph of f, it is expected to capture a subset of its features (branches, holes), depending on how the interval cover is positioned with respect to the critical values of the function. Its stability should also depend on this positioning. We propose [44] a theoretical framework that relates the structure of the Mapper to the one of the Reeb graph, making it possible to predict which features will be present and which will be absent in the Mapper given the function and the cover, and for each feature, to quantify its degree of unstability. Using this framework, we can derive guarantees on the structure of the Mapper, on its stability, and on its convergence to the Reeb graph as the granularity of the cover I goes to zero.

7.2.7. Persistence Theory: From Quiver Representations to Data Analysis

Participant: Steve Oudot.

Persistence theory emerged in the early 2000s as a new theory in the area of applied and computational topology. This book [35] provides a broad and modern view of the subject, including its algebraic, topological, and algorithmic aspects. It also elaborates on applications in data analysis. The level of detail of the exposition has been set so as to keep a survey style, while providing sufficient insights into the proofs so the reader can understand the mechanisms at work.

7.3. Data Structures and Robust Geometric Computation

7.3.1. A probabilistic approach to reducing the algebraic complexity of computing Delaunay triangulations

Participant: Jean-Daniel Boissonnat.

In collaboration with Ramsay Dyer (Johann Bernoulli Institute, University of Groningen, Netherlands) and Arijit Ghosh (Max-Planck-Institut für Informatik, Saarbrücken, Germany).

Computing Delaunay triangulations in $\mathbb{R}^d$ involves evaluating the so-called in_sphere predicate that determines if a point x lies inside, on or outside the sphere circumscribing $d+1$ points $p_0, ..., p_d$. This predicate reduces to evaluating the sign of a multivariate polynomial of degree $d+2$ in the coordinates of the points $x, p_0, ..., p_d$. Despite much progress on exact geometric computing, the fact that the degree of the polynomial increases with d makes the evaluation of the sign of such a polynomial problematic except in very low dimensions. In this paper, we propose a new approach that is based on the witness complex, a weak form of the Delaunay complex introduced by Carlsson and de Silva. The witness complex $\text{Wit}(L, W)$ is defined from two sets L and W in some metric space X: a finite set of points L on which the complex is built, and a set W of witnesses that serves as an approximation of X. A fundamental result of de Silva states that $\text{Wit}(L, W) = \text{Del}(L)$ if $W = X = \mathbb{R}^d$. In [25], [41], we give conditions on L that ensure that the witness complex and the Delaunay triangulation coincide when W is a finite set, and we introduce a new perturbation scheme to compute a perturbed set L' close to L such that $\text{Del}(L') = \text{Wit}(L', W)$. Our perturbation algorithm is a geometric application of the Moser-Tardos constructive proof of the Lovász local lemma. The only numerical operations we use are (squared) distance comparisons (i.e., predicates of degree 2). The time-complexity of the algorithm is sublinear in $|W|$. Interestingly, although the algorithm does not compute any measure of simplex quality, a lower bound on the thickness of the output simplices can be guaranteed.

7.3.2. Smoothed complexity of convex hulls

Participants: Marc Glisse, Rémy Thomasse.

In collaboration with O. Devillers (VEGAS Project-team) and X. Goaoc (Université Marne-la-Vallée)

We establish an upper bound on the smoothed complexity of convex hulls in $\mathbb{R}^d$ under uniform Euclidean (ℓ^2) noise. Specifically, let $\{p_1^*, p_2^*, ..., p_n^*\}$ be an arbitrary set of n points in the unit ball in $\mathbb{R}^d$ and let $p_i = p_i^* + x_i$, where $x_1, x_2, ..., x_n$ are chosen independently from the unit ball of radius δ. We show that the expected complexity, measured as the number of faces of all dimensions, of the convex hull of $\{p_1, p_2, ..., p_n\}$ is $O\left(n^{2-\frac{4}{d+1}}(1+1/\delta)^{d-1}\right)$; the magnitude δ of the noise may vary with n. For $d = 2$ this bound improves to $O\left(n^{\frac{2}{3}}(1+\delta^{-\frac{2}{3}})\right)$.

We also analyze the expected complexity of the convex hull of ℓ^2 and Gaussian perturbations of a nice sample of a sphere, giving a lower-bound for the smoothed complexity. We identify the different regimes in terms of the scale, as a function of n, and show that as the magnitude of the noise increases, that complexity varies monotonically for Gaussian noise but non-monotonically for ℓ^2 noise [31], [38].

7.3.3. Realization Spaces of Arrangements of Convex Bodies

Participant: Alfredo Hubard.

In collaboration with M. Dobbins (PosTech, South Korea) and A. Holmsen (KAIST, South Korea)

In [23], we introduce combinatorial types of arrangements of convex bodies, extending order types of point sets to arrangements of convex bodies, and study their realization spaces. Our main results witness a trade-off between the combinatorial complexity of the bodies and the topological complexity of their realization space. On one hand, we show that every combinatorial type can be realized by an arrangement of convex bodies and (under mild assumptions) its realization space is contractible. On the other hand, we prove a universality theorem that says that the restriction of the realization space to arrangements of convex polygons with a bounded number of vertices can have the homotopy type of any primary semialgebraic set.

7.3.4. Limits of order types

Participant: Alfredo Hubard.

In collaboration with X. Goaoc (Institut G. Monge), R. de Joannis de Verclos (CNRS-INPG), J-S. Sereni (LORIA), and J. Volec (ETH)

The notion of limits of dense graphs was invented, among other reasons, to attack problems in extremal graph theory. It is straightforward to define limits of order types in analogy with limits of graphs, and in [24] we examine how to adapt to this setting two approaches developed to study limits of dense graphs. We first consider flag algebras, which were used to open various questions on graphs to mechanical solving via semidefinite programming. We define flag algebras of order types, and use them to obtain, via the semidefinite method, new lower bounds on the density of 5- or 6-tuples in convex position in arbitrary point sets, as well as some inequalities expressing the difficulty of sampling order types uniformly. We next consider graphons, a representation of limits of dense graphs that enable their study by continuous probabilistic or analytic methods. We investigate how planar measures fare as a candidate analogue of graphons for limits of order types. We show that the map sending a measure to its associated limit is continuous and, if restricted to uniform measures on compact convex sets, a homeomorphism. We prove, however, that this map is not surjective. Finally, we examine a limit of order types similar to classical constructions in combinatorial geometry (Erdös-Szekeres, Horton...) and show that it cannot be represented by any somewhere regular measure; we analyze this example via an analogue of Sylvester's problem on the probability that k random points are in convex position.

GRAPHDECO Project-Team

7. New Results

7.1. Computer-Assisted Design with Heterogeneous Representations

7.1.1. BendFields: Regularized Curvature Fields from Rough Concept Sketches

Participants: Adrien Bousseau, Emmanuel Iarussi.

Designers frequently draw curvature lines to convey bending of smooth surfaces in concept sketches. We present a method to extrapolate curvature lines in a rough concept sketch, recovering the intended 3D curvature field and surface normal at each pixel of the sketch (Fig. 4). This 3D information allows us to enrich the sketch with 3D-looking shading and texturing. We first introduce the concept of *regularized curvature lines* that model the lines designers draw over curved surfaces, encompassing curvature lines and their extension as geodesics over flat or umbilical regions. We build on this concept to define the orthogonal cross field that assigns two regularized curvature lines to each point of a 3D surface. Our algorithm first estimates the projection of this cross field in the drawing, which is non-orthogonal due to foreshortening. We formulate this estimation as a scattered interpolation of the strokes drawn in the sketch, which makes our method robust to sketchy lines that are typical for design sketches. Our interpolation relies on a novel smoothness energy that we derive from our definition of regularized curvature lines. Optimizing this energy subject to the stroke constraints produces a dense non-orthogonal 2D cross field, which we then lift to 3D by imposing orthogonality. Thus, one central concept of our approach is the generalization of existing cross field algorithms to the non-orthogonal case. We demonstrate our algorithm on a variety of concept sketches with various levels of sketchiness. We also compare our approach with existing work that takes clean vector drawings as input.

This work is a collaboration with David Bommes from Titane project team at Inria Sophia-Antipolis, now at RWTH Aachen University. It has been published at ACM Transactions on Graphics (TOG) [7].

7.1.2. Line Drawing Interpretation in a Multi-View Context

Participant: Adrien Bousseau.

Many design tasks involve the creation of new objects in the context of an existing scene. Existing work in computer vision only provides partial support for such tasks. On the one hand, multi-view stereo algorithms allow the reconstruction of real-world scenes, while on the other hand algorithms for line-drawing interpretation do not take context into account. Our work combines the strength of these two domains to interpret line drawings of imaginary objects drawn over photographs of an existing scene (Fig. 5). The main challenge we face is to identify the existing 3D structure that correlates with the line drawing while also allowing the creation of new structure that is not present in the real world. We propose a labeling algorithm to tackle this problem, where some of the labels capture dominant orientations of the real scene while a free label allows the discovery of new orientations in the imaginary scene. We illustrate our algorithm by interpreting line drawings for urban planing, home remodeling, furniture design and cultural heritage.

This work is a collaboration with Jean-Dominique Favreau and Florent Lafarge from Titane project team, Inria Sophia-Antipolis. It has been published at the Conference on Computer Vision and Pattern Recognition (CVPR) [14].

7.1.3. WrapIt: Computer-Assisted Crafting of Wire Wrapped Jewelry

Participants: Adrien Bousseau, Emmanuel Iarussi.

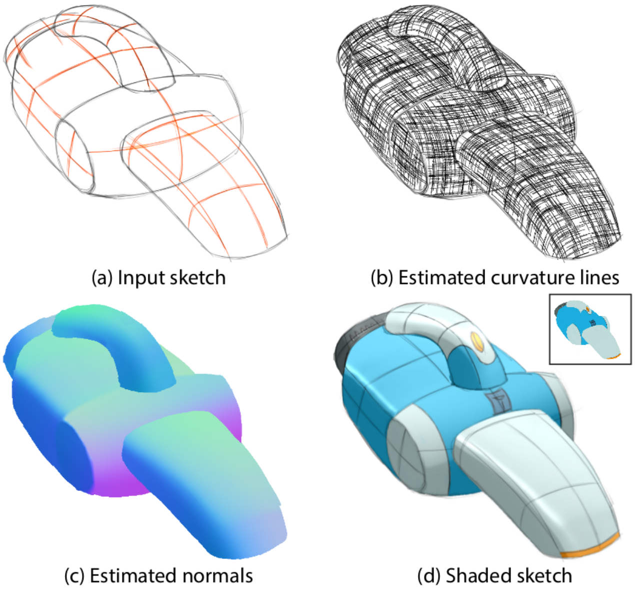


Figure 4. Our method [7] takes as input a rough design sketch with annotated curvature lines (a). We propose a novel smoothness energy to propagate the curvature information to all pixels (b), which allows us to recover surface normals (c) and compute shading (d).

Figure 5. Our method [14] takes as input several photographs of a scene, along with a line drawing of a new object (left). We exploit the dominant orientations of the existing scene to reconstruct the line drawing in 3D (right).

Wire wrapping is a traditional form of handmade jewelry that involves bending metal wire to create intricate shapes. The technique appeals to novices and casual crafters because of its low cost, accessibility and unique aesthetic. We present a computational design tool that addresses the two main challenges of creating 2D wire-wrapped jewelry: decomposing an input drawing into a set of wires, and bending the wires to give them shape (Fig. 6). Our main contribution is an automatic wire decomposition algorithm that segments a drawing into a small number of wires based on aesthetic and fabrication principles. We formulate the task as a constrained graph labeling problem and present a stochastic optimization approach that produces good results for a variety of inputs. Given a decomposition, our system generates a 3D-printed custom support structure, or *jig*, that helps users bend the wire into the appropriate shape. We validated our wire decomposition algorithm against existing wire-wrapped designs, and used our end-to-end system to create new jewelry from clipart drawings. We also evaluated our approach with novice users, who were able to create various pieces of jewelry in less than half an hour.

This work is a collaboration with Wilmot Li from Adobe, San Francisco. The project was initiated by a 3-months visit of Emmanuel Iarussi at Adobe. It has been published at ACM Transactions on Graphics (Proc. SIGGRAPH Asia) [8].

Figure 6. Our system [8] helps novices convert a line drawing (left) into a real piece of jewelry (right).

7.1.4. How Novices Sketch and Prototype Hand-Fabricated Objects

Participant: Adrien Bousseau.

We are interested in how to create digital tools to support informal sketching and prototyping of objects by *novices*. Achieving this goal first requires a deeper understanding of how novices currently generate, explore, and communicate design ideas with traditional tools, i.e., sketches on paper and hands-on prototyping materials. We describe a study framed around two all-day design charrettes where participants perform a complete design process including ideation sketching, concept development and presentation, fabrication planning documentation and collaborative fabrication of hand-crafted prototypes. This structure allows us to control key aspects of the design process while collecting rich data about creative tasks, including sketches on paper, physical models, and videos of collaboration discussions. We observed that while participants had no formal training in design, they made use of advanced visualization techniques to convey 3D concepts. Participants also extensively used physical materials (paper, foam, cardboard) both to support concept exploration and to communicate their ideas to collaborators. We deduce from these observations recommendations for the conception of design tools adapted to the needs and skills of novices.

This work is a collaboration with Wendy McKay, Theophanis Tsandilas and Lora Oehlberg from the InSitu project team - Inria Saclay, in the context of the ANR DRAO project. It is conditionally accepted to ACM CHI 2016.

7.1.5. Vectorizing Rough Line Drawings

Participant: Adrien Bousseau.

Our goal in this project is to convert rough, freehand bitmap sketches to clean vector drawings, keeping three main objectives in mind: (i) the vectorial curves should approximate well the input drawing, (ii) the drawing should be composed of a small number of curves with few control points to preserve the compactness and editability of vector graphics, and (iii) the algorithm should support user guidance to disambiguate the multiple interpretations inherent to artistic inputs. Unfortunately, existing vectorization algorithms only partly satisfy these requirements. In particular, while most methods employ curve fitting to satisfy the first objective of data fidelity, this fitting is performed locally and is often sub-optimal with respect to our second objective of *low complexity*. To achieve our objectives, we propose to cast line drawing vectorization as a global optimization that balances data fidelity with model complexity. We express data fidelity as the goodness of fit of Bézier curve segments, and we express model complexity as the number and degree of curve segments that compose the output drawing. Our algorithm produces clean, compact and editable vector drawings from bitmap sketches.

This ongoing work is a collaboration with Jean-Dominique Favreau and Florent Lafarge from Titane project team, Inria Sophia-Antipolis.

7.1.6. Exploring Design Spaces with Sketch-Based Rendering

Participant: Adrien Bousseau.

Designers often start product design by drawing many quick and imperfect sketches. These sketches typically capture shape variations of a concept from different viewpoints. We introduce *sketch-based rendering* as a way to help designers explore the design space induced by such sketches. Our interactive tool allows designers to interpolate between the sketches, providing a continuous, 3D-like visualization of the concept and its variations without requiring explicit 3D information.

We propose an iterative algorithm to match and warp between sketches using little user interaction. We designed this algorithm to address the specific challenges inherent to concept sketches, in particular the fact that they are dominated by contours rather than color or texture, and that these contours should not distort during interpolation. We also describe how to approximate the relative camera positions of different sketches from the magnitude of their 2D motion fields. This approach allows plausible 3D-like camera movements despite the presence of sketch distortions and variations that prevent standard camera calibration. Our tool, thus, fills a gap in the initial stage of the product design pipeline by allowing designers and their patrons to make better informed choices before proceeding to more expensive 3D modeling and prototyping.

This ongoing work is a collaboration with Ishan Darolia and Vinay Namboodiri from IIT Kampur and Rahul Arora and Karan Singh from University of Toronto.

7.1.7. Sketch-Based Inverse Procedural Modeling

Participant: Adrien Bousseau.

Designing and modeling 3D objects is a crucial skill in various areas of entertainment, science, and engineering. However, this task is notoriously hard and unintuitive, especially for novices. Prior work has addressed the modeling problem from many different directions. Sketch-based modeling exploits human intuition and experience in drawing objects. Nevertheless, the quality of the final 3D model depends on the sketching skills of the user, the amount of details added to the drawing, and ability to resolve inherent ambiguities of the sketching process. Another popular direction is procedural modeling, which has been successfully used to create detailed and complex cities, realistic and growing vegetation, and other man-made objects. But procedural modeling is difficult to control and thus hard to use as an exploratory design tool making it accessible only to experts. Our goal in this project is to leverage both the intuitiveness, freedom and flexibility of sketching and the precision, exactness, and detail amplification of procedural modeling. Users of our system begin to sketch a 3D model using a mouse or a digital pen on a tablet. After only a few strokes, our algorithm finds a compact 3D procedural representation that matches the sketch while augmented it with geometric details.

This ongoing work is a collaboration with Gen Nishida, Bedrich Benes, Ignacio Garcia Dorado and Daniel Aliaga from Purdue University.

7.1.8. A data-based approach to retrieve the viewpoint of a design sketch

Participants: Johanna Delanoy, Adrien Bousseau.

Designing objects requires frequent transitions from a 2D representation, the sketch, to a 3D one. Because 3D modeling is time consuming, it is made only during late phases of the design process. Our long term goal is to allow designers to automatically generate 3D models from their sketches. In this work, we address the preliminary step of recovering the viewpoint under which the object is drawn. We adopt a data-driven approach where we build correspondences between the sketch and 3D objects of the same class from a database. In particular, we relate the curvature lines and contours of the 3D objects to similar lines commonly drawn by designers using histograms of orientation. The 3D objects from the database are then used to vote for the viewpoints and the more probable ones are chosen. Our results on design sketches suggest that using both contours and curvature lines give higher precision than using either one. In particular, curvature information improves viewpoint retrieval when details of the objects are different from the sketch.

The work has been published in the journal Revue Française d'Informatique Graphique and presented at the 28th Journées de l'Association Française d'Informatique Graphique [5].

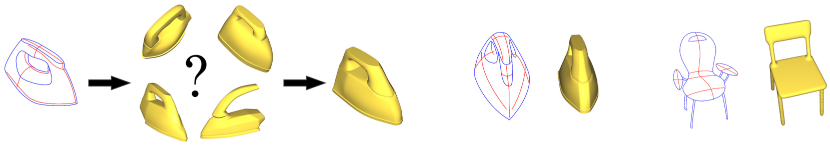


Figure 7. Our method [5] allows to retrieve the viewpoint of a design sketch, using a collection of 3D objects.

7.2. Graphics with Uncertainty and Heterogeneous Content

7.2.1. Multi-View Intrinsic Images for Outdoors Scenes with an Application to Relighting

Participants: Sylvain Duchêne, Clement Riant, Gaurav Chaurasia, Stefan Popov, Adrien Bousseau, George Drettakis.

We introduce a method to compute intrinsic images for a multi-view set of outdoor photos with cast shadows, taken under the same lighting (Fig. 8). We use an automatic 3D reconstruction from these photos and the sun direction as input and decompose each image into reflectance and shading layers, despite the inaccuracies and missing data of the 3D model. Our approach is based on two key ideas. First, we progressively improve the accuracy of the parameters of our image formation model by performing iterative estimation and combining 3D lighting simulation with 2D image optimization methods. Second we use the image formation model to express reflectance as a function of discrete visibility values for shadow and light, which allows us to introduce a robust visibility classifier for pairs of points in a scene. This classifier is used for shadow labelling, allowing us to compute high quality reflectance and shading layers. Our multi-view intrinsic decomposition is of sufficient quality to allow relighting of the input images. We create shadow-caster geometry which preserves shadow silhouettes and using the intrinsic layers, we can perform multi-view relighting with moving cast shadows. We present results on several multi-view datasets, and show how it is now possible to perform image-based rendering with changing illumination conditions.

This work was published in ACM Transactions on Graphics [2].


(a) Input image (b) Relighting +30 minutes

(c) Reflectance (d) Shading

Figure 8. Our algorithm [2] decomposes outdoor images (a) into reflectance (c) and shading (d). This decomposition enables relighting of photographs by rendering new shadows in the shading layer (b).

This work is part of an industrial partnership with Autodesk and has been published in ACM Transactions on Graphics [2].

7.2.2. A Bayesian Approach for Selective Image-Based Rendering using Superpixels

Participants: Rodrigo Ortiz Cayon, Abdelaziz Djelouah, George Drettakis.

Many recent Image-Based Rendering (IBR) algorithms have been proposed each having different strengths and weaknesses, depending on 3D reconstruction quality and scene content. Each algorithm operates with a set of hypotheses about the scene and the novel views, resulting in different quality/speed trade-offs in different image regions. We developed a principled approach to select the algorithm with the best quality/speed trade-off in each region. To do this, we propose a Bayesian approach, modeling the rendering quality, the rendering process and the validity of the assumptions of each algorithm. We then choose the algorithm to use with Maximum a Posteriori estimation. We demonstrate the utility of our approach on recent IBR algorithms which use oversegmentation and are based on planar reprojection and shape-preserving warps respectively. Our algorithm selects the best rendering algorithm for each superpixel in a preprocessing step; at runtime our selective IBR uses this choice to achieve significant speedup at equivalent or better quality compared to previous algorithms. The work has been published in the International Conference on 3D Vision (3DV) - 2015 [15].

7.2.3. Uncertainty Modeling for Principled Interactive Image-Based Rendering

Participants: Rodrigo Ortiz Cayon, George Drettakis.

Despite recent advances in IBR methods, they are limited in regions of the scene which are badly or completely unreconstructed. Such regions have varying degrees of uncertainty, which previous solutions treat with heuristic methods. Currently we attempt to develop a comprehensive model of uncertainty for interactive IBR. Regions with high uncertainty would feed an iterative multi-view depth synthesis algorithm. For the rendering we will formalize an unified IBR algorithm, which provides a good quality/speed tradeoff by combining the

Figure 9. In top-left, we use planes fronto-parallel to the input view which fail for trees and slanted planes. Using local plane estimation top-right the result is improved, especially for slanted planes (blue box). Using the shape preserving warp bottom-left of the warping method we previously developed, better results are achieved for the tree (red box), but the quality of the slanted planes is worse. Our algorithm [15] bottom-right makes the correct choice locally, giving the best solution in each case.

advantages of forward warping and depth-based backprojection algorithms and includes plausible stereoscopic rendering for unreconstructed volumetric regions.

7.2.4. Multi-view Inpainting

Participants: Theo Thonat, George Drettakis.

We are developing a new approach for removing objects in multi-view image datasets. For a given target image from which we remove objects, we use Image-Based Rendering for reprojecting the other images into the target and for regions not visible in any other image we use inpainting techniques. The difficulties reside in formalizing the unified algorithm and enforcing multi-view consistency. This is an ongoing project in collaboration with Adobe Research (E. Shechtman and S. Paris).

7.2.5. Beyond Gaussian Noise-Based Texture Synthesis

Participants: Kenneth Vanhoey, Georgios Kopanas, George Drettakis.

Texture synthesis methods based on noise functions have many nice properties: they are continuous (thus resolution-independent), infinite (can be evaluated at any point) and compact (only functional parameters need to be stored). A good method is also non-repetitive and aperiodic. Current techniques, like Gabor Noise, fail to produce structured content. They are limited to so-called “Gaussian textures”, characterized by second-order statistics like mean and variance only. This is suitable for noise-like patterns (e.g., marble, wood veins, sand) but not for structured ones (e.g., brick wall, mountain rocks, woven yarn). Other techniques, like Local Random-Phase noise, leverage some structure but as a trade-off with repetitiveness and periodicity.

In this project, we model higher-order statistics produced by noise functions. Then we define an algorithm for maximal-entropy sampling of the parameters of the noise functions so as to meet prescribed statistics to reproduce. This sampling ensures both the reproduction of higher-order visual features with high probability, like edges and ridges, and non-repetitiveness plus aperiodicity thanks to the stochastic sampling method. We are currently investigating a learning method so as to inject into the model the appropriate prescribed statistics deduced from an input exemplar image.

This ongoing work is a collaboration with Ian Jermyn from Durham University and will be submitted for publication in 2016.

7.2.6. Unifying Color and Texture Transfer for Predictive Appearance Manipulation

Participants: Fumio Okura, Kenneth Vanhoey, Adrien Bousseau, George Drettakis.

Recent color transfer methods use local information to learn the transformation from a Source to an Exemplar image, and then transfer this appearance change to a Target image (figure 10 (a) to (d)). These solutions achieve successful results for general mood changes, e.g., changing the appearance of an image from “sunny” to “overcast”. However, they fail to create new image content, such as leaves on a bare tree (figure 10 (d)). Texture transfer, on the other hand, can synthesize such content but tends to destroy image structure (figure 10 (e)). We propose the first algorithm that unifies color and texture transfer, outperforming both by automatically leveraging their respective strengths (figure 10 (f)). A key novelty in our approach resides in teasing apart appearance changes that can be modeled simply as changes in color versus those that require new image content to be generated. Our method starts with an analysis phase which evaluates the success of color transfer on the Source/Exemplar scene. To do so, color transfer parameters are learned on this pair, and applied on the Source. The color transferred Source image is then evaluated against the Exemplar which serves as a ground truth, using texture distance metrics (textons in our case). This provides information on the localization of success and failure of color transfer on this scene. This analysis then drives the synthesis: a selective, iterative texture transfer algorithm that simultaneously predicts the success of color transfer on the Target and synthesizes new content using texture transfer where needed. Synthesis exploits a dense pixel matching between the Source/Exemplar scene, on which information is learned, and the Target/Output scene, on which we want to synthesize. The algorithm iterates between synthesizing the new scene by locally using either color or texture transfer, and improving the dense matching on the scene being synthesized. As a result, it leverages the best of both techniques on a variety of scenes by transferring large temporal changes between photographs, such as change of season and flooding. We demonstrate this with seasonal changes on vegetation (e.g., trees) and snow, and on examples involving flooding.

This work is a collaboration with Alexei Efros from UC Berkeley in the context of the associate team CRISP. It has been published in Computer Graphics Forum [9] and was accepted and presented at the Eurographics symposium on Rendering.

Figure 10. Illustration of our appearance prediction method [9]. The future appearance of a target image (c) is predicted (f) based on the knowledge learned from a quasi-aligned source-exemplar pair ((a) and (b)) which characterizes an analogous transformation. The key insight is to selectively operate color transfer ((d): only operate rigid local color histogram transformations, i.e., change the background's overall mood) or texture transfer ((e): copying pixels or patches from the exemplar, i.e. synthesize the three's leaves) where suitable, so as to obtain an improved result (f).

7.2.7. Simplification of Triangle Meshes with Digitized Radiance

Participant: Kenneth Vanhoey.

Very accurate view-dependent surface color of virtual objects can be represented by outgoing radiance of the surface. This data forms a surface light field, which is inherently 4-dimensional, as the color is varying both spatially and directionally. Acquiring this data and reconstructing a surface light field of a real-world object can result in very large datasets, which are very realistic, but tedious to store and render. In this project, we consider the processing of outgoing radiance stored as a vertex attribute of triangle meshes, and especially propose a principled simplification technique. We show that when reducing the global memory footprint of such acquired objects, smartly reducing the spatial resolution, as opposed to the directional resolution, is an effective strategy for overall appearance preservation. To define such simplification, we define a new metric to guide an iterative edge collapse algorithm. Its purpose is to measure the visual damage introduced when operating a local simplification. Therefore, we first derive mathematical tools to calculate with radiance functions on the surface: interpolation, gradient computation and distance measurements. Then we derive a metric using these tools. We particularly ensure that the mathematical interpolation used in the metric is coherent with the non-linear interpolation we use for rendering, which makes the math coherent with the rendered object. As a result we show that both synthetic and acquired objects benefit from our radiance-aware simplification process: at equal memory footprint, visual quality is improved compared to state of the art alternatives.

This work is a collaboration with the ICube laboratory, Strasbourg, France. It was published in the Computer Graphics Journal [11] and was accepted and presented at Computer Graphics International 2015 in Strasbourg, France.

7.2.8. Video based rendering for vehicles

Participants: Abdelaziz Djelouah, Georgios Koulieris, George Drettakis.

The main objective of image based rendering methods is to provide high quality free-view point navigation in 3D scenes using only a limited set of pictures. Despite the good visual quality achieved by most recent methods, the results still look unrealistic because of the static nature of the rendered scenes. This project is in the general context of enriching image based rendering experience by adding dynamic elements and we are particularly interested by adding vehicles.

Vehicles represent an important proportion of the dynamic elements in any urban scene and adding an object with such a complex appearance model has many challenges. First, contrary to classic IBR the number of viewpoints is limited because all input videos must be recorded at the same time. Also, because of this limited number of viewpoints, using classic multi-view reconstruction methods does not produce good results. Instead we use 3D stock models as proxy for the cars. The first step is the registration of the 3D model with the input videos. Then, using the 3D model, the input videos are processed to extract the different visual layers (base color, reflections, transparency, etc.). Finally, the objective is to find the appropriate way to combine the 3D model and the extracted layers to provide the most realistic image from any viewpoint.

This ongoing work is a collaboration with Gabriel Brostow from University College London in the context of the CR-PLAY EU project and with Alexei Efros from UC Berkeley.

7.2.9. Finger-Based Manipulation in Immersive Spaces and the Real World

Immersive environments that approximate natural interaction with physical 3D objects are designed to increase the user’s sense of presence and improve performance by allowing users to transfer existing skills and expertise from real to virtual environments. However, limitations of current Virtual Reality technologies, e.g., low-fidelity real-time physics simulations and tracking problems, make it difficult to ascertain the full potential of finger-based 3D manipulation techniques. This project decomposes 3D object manipulation into the component movements, taking into account both physical constraints and mechanics. We fabricate five physical devices that simulate these movements in a measurable way under experimental conditions. We then implement the devices in an immersive environment and conduct an experiment to evaluate direct finger-based against ray-based object manipulation. The key contribution of this work is the careful design and creation of physical and virtual devices to study physics-based 3D object manipulation in a rigorous manner in both real and virtual setups.

This work was presented at IEEE Symposium on 3D User Interfaces [12], and is in collaboration with the EXSITU Inria group in Paris (T. Tsandilas, W. Mackay, L. Oehlberg).

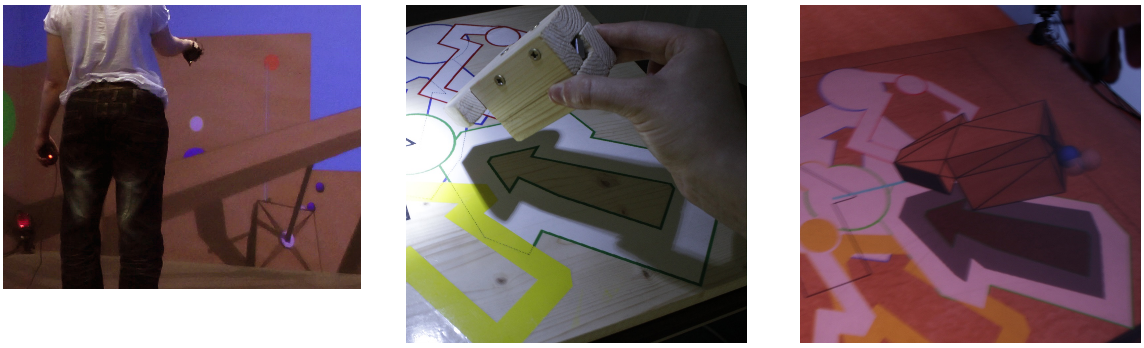


Figure 11. A user in our immersive environment (left) for finger-based manipulation [12]. Completing a 6 DoF manipulation task in real (center) and virtual (right) settings.

7.2.10. Gaze Prediction using Machine Learning for Dynamic Stereo Manipulation

Participants: Georgios Koulieris, George Drettakis.

Comfortable, high-quality 3D stereo viewing is becoming a requirement for interactive applications today. The main challenge of this project is to develop a gaze predictor in the demanding context of real-time, heavily task-oriented applications such as games. Our key observation is that player actions are highly correlated with the present state of a game, encoded by game variables. Based on this, we train a classifier to learn these correlations using an eye-tracker which provides the ground-truth object being looked at. The classifer is used at runtime to predict object category – and thus gaze – during game play, based on the current state of game variables. We use this prediction to propose a dynamic disparity manipulation method, which provides rich and comfortable depth. We evaluate the quality of our gaze predictor numerically and experimentally, showing that it predicts gaze more accurately than previous approaches. A subjective rating study demonstrates that our localized disparity manipulation is preferred over previous methods.

This is a collaboration with the Technical University of Crete (K. Mania) and Cottbuss University (D. Cunningham), and will be presented at IEEE VR 2016.

7.2.11. Compiling High Performance Recursive Filters

Infinite impulse response (IIR) or recursive filters, are essential for image processing because they turn expensive large-footprint convolutions into operations that have a constant cost per pixel regardless of kernel size. However, their recursive nature constrains the order in which pixels can be computed, severely limiting both parallelism within a filter and memory locality across multiple filters. Prior research has developed algorithms that can compute IIR filters with image tiles. Using a divide-and-recombine strategy inspired by parallel prefix sum, they expose greater parallelism and exploit producer-consumer locality in pipelines of IIR filters over multi-dimensional images. While the principles are simple, it is hard, given a recursive filter, to derive a corresponding tile-parallel algorithm, and even harder to implement and debug it. We show that parallel and locality-aware implementations of IIR filter pipelines can be obtained through program transformations , which we mechanize through a domain-specific compiler. We show that the composition of a small set of transformations suffices to cover the space of possible strategies. We also demonstrate that the tiled implementations can be automatically scheduled in hardware-specific manners using a small set of generic heuristics. The programmer specifies the basic recursive filters, and the choice of transformation requires only a few lines of code. Our compiler then generates high-performance implementations that are an order of magnitude faster than standard GPU implementations, and outperform hand tuned tiled implementations of specialized algorithms which require orders of magnitude more programming effort – a few lines of code instead of a few thousand lines per pipeline. This work was presented the High Performance Computing conference and is a collaboration with F. Durand, J. Ragan-Kelley and G. Chaurasia of MIT and S. Paris of Adobe [13].

7.2.12. Probabilistic Connections for Bidirectional Path Tracing

Participants: Sefan Popov, George Drettakis.

Bidirectional path tracing (BDPT) with Multiple Importance Sampling is one of the most versatile unbiased rendering algorithms today. BDPT repeatedly generates sub-paths from the eye and the lights, which are connected for each pixel and then discarded. Unfortunately, many such bidirectional connections turn out to have low contribution to the solution. The key observation in this projects is that we can importance sample connections to an eye sub-path by considering multiple light sub-paths at once and creating connections probabilistically. We do this by storing light paths, and estimating probability mass functions of the discrete set of possible connections to all light paths. This has two key advantages: we efficiently create connections with low variance by Monte Carlo sampling, and we reuse light paths across different eye paths. We also introduce a caching scheme by deriving an approximation to sub-path contribution which avoids high-dimensional path distance computations. Our approach builds on caching methods developed in the different context of VPLs. Our Probabilistic Connections for Bidirectional Path Tracing approach raises a major challenge, since reuse results in high variance due to correlation between paths. We analyze the problem of path correlation and derive a conservative upper bound of the variance, with computationally tractable sample weights. We present results of our method which shows significant improvement over previous unbiased global illumination methods, and evaluate our algorithmic choices.

This work was in collaboration with R. Ramamoorthi (UCSD) and F. Durand (MIT) and appeared in the Eurographics Symposium on Rendering [10].

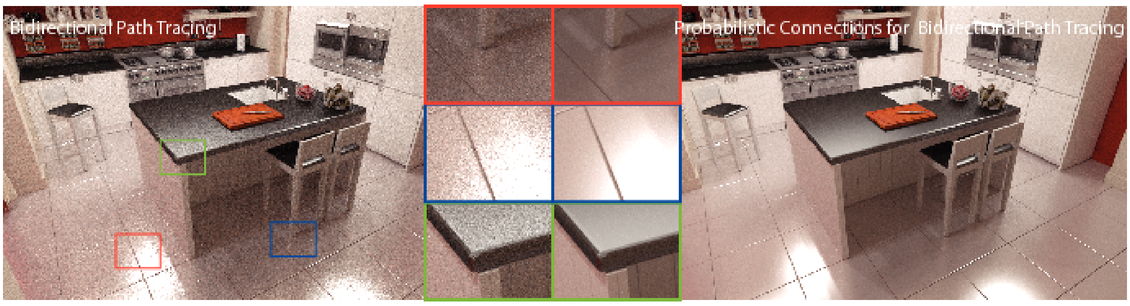


Figure 12. Our Probabilistic Connections for Bidirectional Path Tracing [10] approach importance samples connections to an eye sub-path, and greatly reduces variance, by considering and reusing multiple light sub-paths at once. Our approach (right) achieves much higher quality than bidirectional path-tracing on the left for the same computation time.

GRAPHIK Project-Team

7. New Results

7.1. Ontology-Based Query Answering with Existential Rules

Participants: Jean-François Baget, Meghyn Bienvenu, Fabien Garreau, Michel Leclère, Marie-Laure Mugnier, Swan Rocher, Federico Ulliana.

Since Meghyn Bienvenu joined the team very recently (September 2015), we only include here the work she did in collaboration with GraphIK members.

Ontolology-based query answering (and more generally *Ontology-Based Data Access*, *OBDA*) is a new paradigm in data management, which takes into account inferences enabled by an ontology when querying data. In other words, the notion of a database is replaced by that of a knowledge base, composed of data (also called facts) and of an ontology. In this context, *existential rules* (also called Datalog+) have been proposed to represent the ontological component [42], [41]. This expressive formalism generalizes both description logics used in OBDA (such as $\mathcal{EL}$ and DL-Lite), which form the cores of so-called tractable profiles of the Semantic Web ontological language OWL2) and Datalog, the language of deductive databases. Since about five years, we have been studying the theoretical foundations of this framework (mainly concerning decidability and complexity) and developing associated algorithmic techniques. We have started the development of a platform dedicated to OBDA with existential rules (see section 6.3).

Before presenting this year' results, we recall the two classical ways of processing rules, namely forward chaining and backward chaining, also known as "materialization" and "query rewriting" in the OBDA setting. In forward chaining, the rules are applied to enrich the initial data and query answering can then be solved solved by evaluating the query against the "saturate" database (as in a classical database system *i.e.*, with forgetting the rules). The backward chaining process can be divided into two steps: first, the initial query is rewritten using the rules into a first-order query (typically a union of conjunctive queries, UCQ); then the rewritten query is evaluated against the initial database (again, as in a classical database system). Since entailment is not decidable with general existential rules, both forward and backwards processes may not halt.

7.1.1. Embedding transitivity rules.

In recent years, many classes of existential rules have been exhibited for which CQ entailment is decidable. However, most of these classes cannot express transitivity of binary relations, a frequently used modelling construct. We began to investigate the issue of whether transitivity can be safely combined with decidable classes of existential rules. On the one hand, we obtained negative results, proving that transitivity is incompatible with many classes having finite chase, and with UCQ-reducible classes in general. Second, we showed that transitivity can be safely added to linear rules (a subclass of guarded rules, which generalizes the description logic DL-Lite$_R$) in the case of atomic CQs, and also for general CQs if we place a minor syntactic restriction on the rule set (only needed when predicate arity is strictly greater than 2). Finally, we pinpointed the combined and data complexities of query entailment over linear rules + transitivity.

▷ *IJCAI 2015* [22]

7.1.2. A generic algorithm for query reformulation.

We first designed and implemented a query reformulation algorithm that takes as input any set of existential rules and a UCQ q, and outputs a sound, minimal and complete UCQ-reformulation of q, whenever such a reformulation exists (i.e., when the set of existential rules is UCQ-reducible). The core operation, unification, relies on a special technique that we first developed for conceptual graphs ("piece-unification"). A noteworthy feature of the implemented unification is that it is able to process rules without decomposing their head into single atoms. Experiments showed that this feature has a very high impact on the efficiency of query reformulation in terms of running time.

This algorithm can be seen as an instantiation of a generic reformulation algorithm, parametrized by a reformulation operator. As a complementary work, we studied the properties that should be fulfilled by any reformulation operator to ensure the correctness and the termination of this generic algorithm and analyzed some known operators with respect to these properties.

▷ *Semantic Web Journal 2015* [15]

7.1.3. Optimization of query reformulation algorithms

Query reformulation techniques have the advantage of being independent from the data. However, a main bottleneck is that the size of the obtained query can be exponential in the size of the original query, hence the produced reformulation maybe not usable in practice (and the corresponding SQL query may not even be accepted by the RDBMS). To overcome this combinatorial explosion in practice, we made two proposals, which have in common to consider other forms of reformulation, while staying equivalent to UCQs in terms of expressivity.

We defined semi-conjunctive queries (SCQs), which are a syntactical extension of conjunctive queries allowing for internal disjunctions. Briefly, a union of SCQs can be encoded in a more compact form than a UCQ. We designed and implemented an algorithm called Compact, which computes a sound and complete reformulation of a UCQ in the form of a union of SCQs (USCQ). First experiments showed that USCQs are both very efficiently computable and (often) more efficiently evaluable than their equivalent UCQs.

We developed another solution, which starts from a simple observation: in practice, combinatorial explosion is mainly due to some very simple rules, which form the core of any ontology. These rules typically express concept and relation hierarchies, concept properties and relation signatures. We proposed a technique that consists in compiling these rules into a preorder on atoms and embedding this preorder into the reformulation process. This allows us to compute compact reformulations that can be considered as “pivotal” representations, in the sense that they can be easily translated into different kinds of queries that can be evaluated by different kinds of database systems (e.g., unfolded into a classical UCQ or a USCQ, processed as such on data saturated by the compilable rules, or transformed into a Datalog program). Experiments show that this technique leads to substantial gains in the query reformulation process, in terms of size and runtime, it scales on very large ontologies (several ten thousands of rules), and it is competitive w.r.t. other existing tools, including those tailored for more specific rules corresponding to DL-Lite ontologies. This technique has been implemented in the sofware platform Graal.

▷ *IJCAI 2015*[28], *RuleML 2015*[23]

7.1.4. Ontology-based query answering with Semantic Web languages

On the one hand, we proposed Deductive RDF Triplestores, which are RDF knowledge bases equipped with Datalog inference rules. This work was developed in the context of the tool MyCorporisFabrica http://www.mycorporisfabrica.org/, an ontology-based tool for querying complex anatomical models.

In particular, we studied how to extract modules from deductive RDF triplestores. Indeed, many ontologies are extremely large, while users often need to reuse only a small part of resources in their work. A module is a Deductive RDF Triplestore entailed from the reference knowledge base, which is defined upon a restricted vocabulary. We proposed a new semantics for bounded-level modules allowing one to control their size, and then presented extraction algorithms compliant with the novel semantics.

▷ *AAAI 2015*[30] and *Journal of Biomedical Semantics* [16]. In collaboration with Marie-Christine Rousset (U. of Grenoble) and MyCorporisFabrica’s team.

On the other hand, in the context of the Graal platform, we defined a translation from the Semantic Web Ontological Language OWL 2 to our existential rule format. This gave rise to the definition of the “existential rule” OWL 2 profile, which covers the so-called tractable profiles of OWL 2 (see Section 6.3).

▷ *RuleML challenge*[33] (this paper obtained the RuleML 2015 challenge award)

7.2. Reasoning with Imperfect Information and Priorities

Participants: Abdallah Arioua, Patrice Buche, Madalina Croitoru, Jérôme Fortin, Nouredine Tamani, Rallou Thomopoulos.

This year, we mainly explored the use of argumentation frameworks in practical applications. Indeed, we have been involved in three main projects that have employed argumentation techniques. The projects were all in the context of agronomy where the nature of the problem studies fits well the use of argumentation: (1) the knowledge bases considered to model the domain are inconsistent, (2) the reasoning / decision making has to take into account the inconsistency, (3) the end user is a non-computing expert thus explanation facilities are needed.

We enumerate below the three projects and explain our results:

- Bread Making project (financed by the Food and Bioproducts department at INRA) investigates the possibility of using wholemeal flour in bread as opposed to classic white flour. The main theoretical result that we exploited here was the instantiation work into the existential rule framework done with Srdjan Vesic [43]. We used reverse engineering and the subsequent logic-based argumentation in order to provide the experts with a cartography of possible pros and cons of using one type of flour vs the other.

  ▷ *Ecological Informatics 2015* [18]
- EcoBioCap (FP7 EU project led by INRA Montpellier, see Section 9.1) investigates the conception of biodegradable packaging for fruits and vegetables. The main theoretical result used here concerns the fuzzy aspects of argumentation but the modeling of the problem using argumentation and subsequent argument elicitation was also a very challenging process.

  ▷ *Computers and Electronics in Agriculture 2015* (2 papers) [14] [17]
- DURDUR (ANR project led by INRA Montpellier, see Section 9.1) investigates the technological itineraries to grow durum wheat for subsequent pasta and couscous making. This ongoing project investigates the use of argumentation for explanation facilities.

  ▷ Initial results have been published in *SUM 2015*[19] and *DEXA 2015* [20].

7.3. Quality and interoperability of large document catalogues

Participants: Michel Chein, Madalina Croitoru, Alain Gutierrez, Michel Leclère, Rallou Thomopoulos.

The work in this research line takes place in the ANR project Qualinca, devoted to methods and tools to repair linkage errors in bibliographical databases (see Qualinca in Section 9.1). Within this project, we specially work with our applicative partner ABES (French Agency for Academic Libraries, http://www.abes.fr/).

ABES manages several catalogues and authority bases, in particular the *Sudoc*, the collective catalogue of French academic libraries. ABES also provides services to libraries and end-users, as well as to other catalogue managers (e.g., OCLC for Worldcat and, in France, Adonis for the Isidore platform).

This year, we devoted most of our research effort to the following aspects in collaboration with ABES:

1. the finalization of a conceptual model of ABES librarian expertise in their linkage activity, and its formalization in our theoretical framework; the formalized model is both logical (the knowledge is expressed by facts, rules and constraints in first-order logic) and numerical (some predicates, which correspond to qualitative criteria, are computed by numerical functions, which themselves take as input the result of logical queries to the knowledge base).
2. the development of a diagnosis prototype, called SudoQual, which implements this model; in brief, SudoQual takes as input a given apellation (i.e., family name and first name), retrieves all references potentially associated with this appellation and outputs *sameAs* and *Different* links between these references. To develop SudoQual, we built an API on top of our tool Cogui.
3. first experiments with SudoQual on the Sudoc base, with the results being checked manually by ABES librarians.

▷ *Research report* [37]

This work required a tight collaboration with ABES (materialized by bimonthly meetings and numerous ponctual exchanges). The first experiments yield extremely satisfactory results, hence ABES is now considering turning SudoQual into a production tool used by librarians in their daily work to validate/correct autority links in the Sudoc catalogue. This requires to define a suitable user-interface, which is an issue we are currently discussing with ABES. We are also preparing experiments at a larger scale on a sample provided by ABES.

Besides, in collaboration with Qualinca partner LRI, we developed a method and a tool to fusion data linked by "same-as" links. More precisely, given an RDF dataset, our tool allows to merge "same as" data, which are often conflictual, into a unified and consistant representation using a multi-criteria decision method. The tool was evaluated on a dataset provided by INA and LIG, two other partners of Qualinca.

▷ *EGC 2015* [32]

Still with the LRI partner, who developed a logic-based decision tool that statuates on the validity of same-as links in RDF data, we investigated the use of argumentation techniques to explain why "same-as" links are invalidated by this tool.

▷ *SUM 2015* [25]

HEPHAISTOS Project-Team

7. New Results

7.1. Robotics

7.1.1. Cable-driven parallel robots (CDPR)

7.1.1.1. Analysis of Cable-driven parallel robots

Participants: Alessandro Berti, Laurent Blanchet, Houssein Lamine, Jean-Pierre Merlet [correspondant], Yves Papegay, Rémy Ramadour.

We have continued the analysis of suspended CDPRs for control and design purposes. For control it is essential to determine the current pose of the robot for given cable lengths (forward kinematics, FK) and to be able to calculate the cable lengths for a given pose of the platform (inverse kinematics, IK). If the cables are supposed to be non-deformable the IK problem is trivial and has a single solution but the FK is complex, admits several solutions and raises several issues. We have shown in the past that to get all FK solutions for a CDPR with m cables we have to consider not only the case where all cables are under tension but also have to solve the FK for all combinations of cables under tension with 1 to m cables. Surprisingly the FK is more difficult if the CDPR has less than 6 cables under tension. Our team, in collaboration with M. Carricato of Bologna University, is the first to have designed a solving algorithm that allow to compute in a guaranteed manner all FK solutions [21], [22]. The FK problem is different if it is intended to be used in a real-time context as in that case we have the extra information of the platform pose a short time before. After a small change in the cable lengths we may assume a small change in the pose platform but using Newton method with the previous pose cannot guarantee to provide the current pose. We have proposed an algorithm that is guaranteed to get the current pose and is also able to determine if the CDPR may be sufficiently close to a singularity so that multiple solutions are possible [11]. However the assumption of a small change in the platform pose may not always hold, a point that we have shown theoretically and experimentally. We have then proposed an algorithm that uses a model of the coiling process to determine if a drastic change in the pose may occur between two sampling time [11] and also allows one to better estimate the cable tensions on a trajectory. We have for example shown that sudden and important changes in these tensions may occur. Another issue arises for non-deformable cables and CDPR with more than 6 cables in a suspend configuration. In the past we have shown that there always will be at most 6 cables under tension whatever the number of cables. For a given pose there may be several possible set of cables under tension (called *cable configuration*), each of them having different characteristics in terms of maximal tension, sensitivity to disturbances, ... From a control viewpoint it makes sense to impose a given cable configuration at the pose by setting the lengths of slack cables to larger values than the one required for the pose. To determine the best cable configuration we have proposed several ranking index [12].

Even more complex kinematic problems are involved if we assume that the cable are deformable (e.g. are elastic or catenary-like). The cable model is included in the kinematic equations for getting a complete model. We have been interested in the catenary-like model that involves inverse hyperbolic functions and is valid for steel cable of relatively high length. As the IK has never been addressed with such a model we have proposed a solving algorithm [10] that has shown that the IK may have multiple solutions but also may have no solution for poses that are reachable with non-deformable cables. In the same way the DK has several solutions [13]. Finally efficient cables interference detection for sagging cables and the management of modular CDPR, whose geometry may be changed according to the task at hand, have been addressed [9]

7.1.1.2. Cable-Driven Parallel Robots for additive manufacturing in architecture

Participant: Yves Papegay.

Easy to deploy and to reconfigure, dynamically efficient in large workspaces even with payloads, cable-driven parallel robots are very attractive for solving displacement and positioning problems in architectural building at scale 1 and seems to be a good alternative to crane and industrial manipulators in this area.

In a collaboration with CNAM and Ecole Nationale Supérieure d'Architecture Paris-Malaquais, we worked on additive manufacturing of building based on ultra-high performance concrete and developed a CDPR as a proof of concept to power a large scale 3D-printer.

A real size industrial robot will be developed by the XtreeE start-up company.

7.1.2. Assistance

This is now the core of our activity and our work on CDPR is deeply connected to this field as they are an efficient solution for mobility assistance, a high priority for the elderly, helpers and medical community. We have presented our vision of assistance robotics in several occasions [24], [25], [23].

7.1.2.1. Assessment of elderly frailty

Participants: Karim Bakal, Jean-Pierre Merlet.

The assessment of elderly frailty is a difficult concept because it involves the physical capacities of a person and its environment (health-care services, families, funds...). We consider the assessment of upper limb capabilities by looking at the joint torques τ of the arm and the maximal force F that can be exerted by the hand, which are related by the equation

$$\tau = \mathbf{J^T} F \tag{1}$$

where $\mathbf{J}$ is a matrix which depends only upon the configuration of the arm. These equations constitute an underconstrained linear system. In biomechanics the torque τ is measured together with the configuration of the arm and the force F is evaluated by using the method of Chiacchio, that involves the pseudo-inverse of $\mathbf{J^T}$ to calculate F. But there are several uncertainties that are neglected when using this method: the measurement errors on τ and on the configuration of the arm together with uncertainties on the physical parameters of the arm (such as the length of the bones). The method of Chiacchio provides one of the possible solutions of equation (2) and not necessary the one corresponding to the force at the hand. We use another approach based on interval analysis. We assume that all uncertainties may be bounded (τ is an interval vector τ_m, $\mathbf{J^T}$ is an interval matrix) so that equation (2) become an interval linear system. Interval analysis then allows one to determine an approximation as accurate as wanted of the set F_s of all forces F that satisfy the equation and therefore this set includes the real force at the hand. Now assume that with the same arm configuration we measure the force at the hand, here again with some bounded uncertainties (i.e. F is an interval vector F_m). Here again we may use interval analysis applied on equation (2) in order to determine an interval vector τ_v for the τ that is guaranteed to include the real τ. Furthermore τ must be included in the intersection τ_i of τ_v and τ_m while F must be included in the intersection F_i of F_m and F_s. If τ_i is strictly included in τ_m, then we may compute a better approximation of F_s. Reciprocally if F_i is strictly included in F_m we will get a better τ_v. If one of these situation occurs we repeat the process until no significant improvement of F_s or τ_v is obtained. In a second step we consider that the uncertainties that lead to uncertainties in the matrix $\mathbf{J^T}$ are constrained as we have to satisfy $\tau_v = \mathbf{J^T} F_s$. Here again we use interval analysis to determine if this constraint does not allow to reduce the size of the interval on the physical parameters in which case we may obtain a new $\mathbf{J^T}$ that is included in the initial one. In turn this may allow to obtain better τ_v and F_s. The process stops when no improvement has been obtained for F_s, τ_v and the physical parameters.

To test this approach the right upper limb joint torque of 10 males and the force capacity at the right hand was measured by a dynamometer (Biodex III, Biodex Medical Systems) and respectively by a 6-axis load sensor during an experiment performed at HandiBio laboratory. The configuration of the upper limb was measured with a motion capture system (Qualisys, Sweden). The approach is currently being evaluated.

7.1.2.2. Walking analysis and Rehabilitation

Participants: Claire Maillard, Ting Wang, Jean-Pierre Merlet [correspondant].

The walkers of the `ANG` family allow one to determine accurately the trajectory of the walker and therefore to analyze the walking of the user. But these walkers may also be used to assess a rehabilitation process or the progress of an end-user involved in rehabilitation. For that purpose after having identified needs and requirements [17] we developed a new walker `ANG-med` that used infra-red distance sensors to measure the position of the subject during a rehabilitation exercise. Furthermore the software of this walker has been designed to support a message-passing scheme based on the HOP language of the INDES project team so that the walker may exchange information and control order with an external computer, together with allowing the download of new rehabilitation exercise through the robotics RAPP-store [26]. New exercises are designed as a set of such messages, that may include the calculation of exercise assessment indicators. `ANG-med` supports various modes: stand-alone (no external connection), passive mode (the walker only report indicator and status using a wifi connection) or full external control (an external computer fully control the walker except for emergency and real-time procedures).

`ANG-med` has been tested for one month in Centre Héliomarin de Vallauris and is now deployed in the rehabilitation center of MATIA in Spain, as part of the RAPP project. A start-up plan was proposed in November 2014 to transfer the walking analysis technology of HEPHAISTOS with the `ANG` walker in a company called Euthenia 9.2.1.3 .

7.1.2.3. Design and evaluation of assistive devices, ethics

Participants: Marc Beninati, Bernard Senach [correspondant], Jean-Pierre Merlet.

Providing appropriate support, services and information to the elderly, to their caregivers and to the medical profession, through a fleet of communicating devices must rely on a structured processes. A generic design and evaluation framework is being elaborated and will be validated through field experiments [20], [19], [18].

Assistance robotics raises many ethical questions. We started reflection about conducting experiments with frail and old people. A listing of questions to be addressed at each step of an experiment has been written (internal document). We have also hired a joint PhD student with University Bologna about the legal aspects of assistance robotics and we have initiated, together with Nathalie <Nevejans from University of Douai, a meeting with the OPECST at the French National Assembly to discuss legal and ethical aspects of robotics.

7.1.2.4. Smart Environment for Human Behaviour Recognition

Participants: Aurélien Massein, Yves Papegay, Odile Pourtallier.

Both economic motivations due to demographic evolution and willingness of people to live independently at home when aging, facing physical impairment or recovering from injuries has raised the need for activity monitoring at home, in rehabilitation center or in retirement home. Monitoring systems provide information that can range from a broad measure of the daily activity to a precise analysis of the ability of a person performing a task (cooking, dressing, ...) and its evolution.

The broad range of needs and contexts, together with the large variety of available sensors implies the necessity to carefully think the design of the monitoring system. An appropriate system should be inexpensive and forgettable for the monitored person, should respect privacy but collect necessary data, and should easily adapt to stick to new needs. We aim to provide an assisting tool for designing appropriate monitoring systems.

As part of a PhD work, optimal motion planning of a mobile robot with range sensors to locate targets in a room has been studied. Work in progress also include algorithms to deploy infra-red barriers in a large area with several interest places, to be able to locate people. An experimental set-up is in use in the lab and data analysis methods are developed to infer people behaviors.

7.2. Miscellaneous results

7.2.1. Symbolic tools for modeling and simulation

Participant: Yves Papegay.

This activity is the main part of a long-term ongoing collaboration with Airbus whose goal is to directly translate the conceptual work of aeronautics engineers into digital simulators to accelerate aircraft design.

An extensive modeling and simulation platform has been designed which includes a dedicated modeling language for the description of aircraft dynamics models in term of formulae and algorithms, and a symbolic compiler producing as target an efficient numerical simulation code ready to be plugged into a flight simulator, as well as a formatted documentation compliant with industrial requirements of corporate memory [14].

Technology demonstrated by our prototype has been transferred, final version of our modeling and simulation environment has been delivered to Airbus in November 2012. Developer level know-how has been transferred in 2013 to a software company in charge of industrialization and maintenance of the modeling and simulation environment.

Since 2014, we are working on several enhancements and extension of functionalities, namely to enhance the performances and the numerical quality of the generated C simulation code, and ease the integration of our environment into the airbus toolbox.

INDES Project-Team

6. New Results

6.1. Web programming

Participants: Yoann Couillec, Vincent Prunet, Manuel Serrano [correspondant].

6.1.1. Hop.js

Multitier programming languages unify within a single formalism and a single execution environment the programming of the different tiers of distributed applications. On the Web, this programming paradigm unifies the client tier, the server tier, and, when one is used, the database tier. This homogenization offers several advantages over traditional Web programming that rely on different languages and different environments for the two or three tiers of the Web application: programmers have only one language to learn, maintenance and evolution are simplified by the use of a single formalism, global static analyses are doable as a single semantics is involved, debugging and other runtime tools are more powerful as they access global informations about the execution.

The three first multitier platforms for the Web all appeared in 2006: GWT (a.k.a., Google Web Toolkit), Links, and Hop [6], [5]. Each relied on a different programming model and languages. GWT maps the Java programming model on the Web, as it allows, Java/Swing likes programs to be compiled and executed on the Web; Links is functional language with experimental features such as the storing of the whole execution context on the client; Hop is based on the Scheme programming language. These three pioneers have open the path for the other multitier languages such as, Ocsigen for Ocaml, UrWeb, js-scala, etc.

In spite of their interesting properties, multitier languages have not become that popular on the Web. Today, only GWT is widely used in industrial applications but arguably GWT is not a fully multitier language as developing applications with GWT requires explicit JavaScript and HTML programming. This lack of popularity of other systems is likely due to their core based languages than to the programming model itself.

JavaScript is the *defacto* standard on the Web. Since the mid 90's, it is the language of the client-side programming and more recently, with systems like Node.js, it is also a viable solution for the server-side programming. As we are convinced by the virtues of multitier programming we have started a new project consisting of enabling multitier programming JavaScript. We have created a new language called HopScript, which is a minimalist extension of JavaScript for multitier programming, and we have implemented a brand new runtime environment called Hop.js. This environment contains a builtin Web server, on-the-fly HopScript compilers, and many runtime libraries.

HopScript is a super set of JavaScript, *i.e.,* all JavaScript programs are legal HopScript programs. Hop.js is a compliant JavaScript execution environment as it succeeds at 99% of the Ecma 262 tests suite. The Hop.js environment also aims at Node.js compatibility. In its current version it supports about 70% of the Node.js runtime environment. In particular, it fully supports the Node.js modules, which lets Hop programs reuse existing Node.js modules as is.

After a full year of active development to enhance JavaScript and Node.js compatibility, to incorporate features of JavaScript 1.6, and to design new language constructs for machine-to-machine communication, we are now ready to release Hop.js. This will appear at the beginning of 2016.

6.1.2. Data source

During the past few years the volume of accumulated data has increased dramatically. New kinds of data stores have emerged as NoSQL family stores. Many modern applications now collect, analyze, and produce data from several heterogeneous sources. However implementing such applications is still difficult because of lack of appropriate tools and formalisms. We propose a solution to this problem in the context of the JavaScript programming language by extending array comprehensions. Our extension allows programmers to query data from usual stores, such as SQL databases, NoSQL databases, Semantic Web data repositories, Web pages, or even custom user defined data structures. The extension has been implemented in the Hop.js system. It has been described in the paper [10], which has been presented at the ACM DBPL'15 conference.

6.2. Distributed programming

Participants: Gérard Boudol, Johan Grande, Manuel Serrano [correspondant].

Shared-memory concurrency is a classic concurrency model which, among other things, makes it possible to take advantage of multicore processors that are now widespread in personal computers. Concurrent programs are prone to deadlocks which are notoriously hard to predict and debug. Programs using mutexes, a very popular synchronization mechanism, are no exception.

We have studied deadlock avoidance methods with the aim of making programming with mutexes easier. We first studied a method that uses a static analysis by means of a type and effect system, then a variation on this method in a dynamically typed language.

We developed more the second method. It mixes deadlock prevention and avoidance to provide an easy-to-use and expressive deadlock-free locking function. We implemented it as a Hop library. This lead us to develop a starvation-free algorithm to simultaneously acquire an arbitrary number of mutexes, and to identify the concept of asymptotic deadlock. While doing so, we also developed an optimization of exceptions (finally blocks).

Our performance tests seem to show that using our library has negligible impact on the performance of real-life applications. Most of our work could be applied to other structured programming languages such as Java.

This work has been presented at the 17th International Symposium on Principles and Practice of Declarative Programming (PPDP'15) [13]. More details can be found in Grande's PhD thesis [8].

6.3. Types

Participants: Ilaria Castellani, Bernard Serpette.

6.3.1. Behavioural Types

The survey paper https://hal.inria.fr/hal-01213201 presents a state-of-the-art of a recent trend of research on the use of behavioural types for specifying and analysing security properties of communication-centred systems. It is essentially an outcome of the working group on security of the BETTY COST Action, and it offers a unified overview of various proposals that have been put forward in the last few years, both within the BETTY community and outside it, to combine security analysis with behavioural types.

6.3.2. Abstract Rewriting Systems

We have formalised, with the Coq system, the beginning of Paul-André Melliès's thesis concerning abstract rewriting systems. Behind the interest of studying rewriting systems, which are the roots of all small step semantics of programming languages, this particular formalisation was attractive since it gives a concrete example where we have to manage dependant types.

This was done in collaboration with Eduardo Bonelli and Pablo Barenbaum of University of Quilmes, Argentina. The specification and the proofs of this work take 2200 lines of Coq.

6.4. Security

Participants: Ilaria Castellani, Francis Doliere Some, Nataliia Bielova, Bernard Serpette, Tamara Rezk [correspondant].

6.4.1. Hybrid Typing of Secure Information Flow in a JavaScript-like Language

We propose a novel type system for securing information flow in a core of JavaScript. This core takes into account the defining features of the language, such as prototypical inheritance, extensible objects, and constructs that check the existence of object properties. We design a hybrid version of the proposed type system. This version infers a set of assertions under which a program can be securely accepted and instruments it so as to dynamically check whether these assertions hold. By deferring rejection to runtime, the hybrid version can typecheck secure programs that purely static type systems cannot accept.

This work has been published at the 10th International Symposium on Trustworthy Global Computing [11].

6.4.2. Modular Monitor Extensions for Information Flow Security in JavaScript

Client-side JavaScript programs often interact with the web page into which they are included, as well as with the browser itself, through APIs such as the DOM API, the XMLHttpRequest API, and the W3C Geolocation API. Precise reasoning about JavaScript security must therefore take API invocation into account. However, the continuous emergence of new APIs, and the heterogeneity of their forms and features, renders API behavior a moving target that is particularly hard to capture. To tackle this problem, we propose a methodology for modularly extending sound JavaScript information flow monitors with a generic API. Hence, to verify whether an extended monitor complies with the proposed noninterference property, our methodology requires only to prove that the API satisfies a predefined set of conditions. In order to illustrate the practicality of our methodology, we show how an information flow monitor-inlining compiler can take into account the invocation of arbitrary APIs, without changing the code or the proofs of the original compiler. We provide an implementation of such a compiler with an extension for handling a fragment of the DOM Core Level 1 API. Furthermore, our implementation supports the addition of monitor extensions for new APIs at runtime. This work has been published at the 10th International Symposium on Trustworthy Global Computing [12].

6.4.3. Relaxed Noninterference

We have began a study concerning the use of gradual typing for down casting or declassification for information flow. The particularity of this work is to use a finite state machine to gradually accept the down casting process.

This work is done with Éric Tanter of University of Santiago de Chile, in the context of the project Conicyt Redes CEV Challenges on Electronic Voting.

6.4.4. Hybrid Monitoring of Attacker knowledge

Enforcement of non-interference requires to prove that an attacker's knowledge about the initial state remains the same after observing a programs public output. We define a powerful hybrid monitoring mechanism which evaluates dynamically the knowledge that is contained in program variables. To get a precise estimate of the knowledge, the monitor statically analyses non-executed branches. We show that our knowledge-based approach can be combined with existing dynamic monitors for non-interference. A distinguishing feature of such a combination is that the combined monitor is provably more powerful than each mechanism taken separately. We demonstrate this by proposing a knowledge-enhanced version of a dynamic monitor based on the no-sensitive-upgrade principle. We show how to use the knowledge computed by our hybrid monitor to quantify information leakage associated to the program output. The monitor and its static analysis has been formalized and proved correct within the Coq proof assistant.

6.4.5. A Taxonomy of Information Flow Monitors

We propose a rigorous comparison of information flow monitors with respect to two dimensions: soundness and transparency.

For soundness, we notice that the standard information flow security definition called *Termination-Insensitive Non-interference (TINI)* allows the presence of termination channels, however it does not describe whether the termination channel was present in the original program, or it was added by a monitor. We propose a stronger notion of noninterference, that we call *Termination-Aware Non-interference (TANI)*, that captures this fact, and thus allows us to better evaluate the security guarantees of different monitors. We further investigate TANI, and state its formal relations to other soundness guarantees of information flow monitors. For transparency, we identify different notions from the literature that aim at comparing the behaviour of monitors. We notice that one common notion used in the literature is not adequate since it identifies as better a monitor that accepts insecure executions, and hence may augment the knowledge of the attacker. To discriminate between monitors' behaviours on secure and insecure executions, we factorized two notions that we call true and false transparency. These notions allow us to compare monitors that were deemed to be incomparable in the past.

We analyse five widely explored information flow monitors: no-sensitive- upgrade (NSU), permissive-upgrade (PU), hybrid monitoring (HM), se- cure multi-execution (SME), and multiple facets (MF).

This work has been accepted for publication in the International Conference on Principles of Security and Trust (POST 2016).

6.4.6. A Study of JavaScript constructs used in Top Alexa Sites

Several works on JavaScript analysis have shown that including remote scripts can introduce severe security implications in the behavior of the whole web application. To deal with different kinds of attacks, a number of research groups are developing automatic tools to analyze JavaScript programs. However, most of these works rely on one assumption: the scripts are written in a subset of JavaScript language meaning that only certain constructs are used (that are easier to analyse automatically) and others are omitted (for example, eval is impossible to analyze statically). The goal of the internship was to account for the use of each JavaScript construct in real world programs. To achieve that, we first did a large-scale crawl of the top 10,000 Alexa sites, collecting both inlined scripts and remote scripts. Second, we established the popularity of remote scripts. Next, we accounted for the occurrence of JavaScript constructs in the collected programs. Finally, we use the occurrence of different constructs as basis to propose a subset of JavaScript language, which covers most of JavaScript programs found in the wild. One can rely on this evidence-based subset of JavaScript in future works on that language.

LAGADIC Project-Team

7. New Results

7.1. Visual tracking

7.1.1. Object detection

Participant: Eric Marchand.

We addressed the challenge of detecting and localizing a poorly textured known object, by initially estimating its complete 3D pose in a video sequence [45]. Our solution relies on the 3D model of the object and synthetic views. The full pose estimation process is then based on foreground/background segmentation and on an efficient probabilistic edge-based matching and alignment procedure with the set of synthetic views, classified through an unsupervised learning phase. Our study focuses on space robotics applications and the method has been tested on both synthetic and real images, showing its efficiency and convenience, with reasonable computational costs.

7.1.2. Registration of multimodal images

Participant: Eric Marchand.

This study has been realized in collaboration with Brahim Tamadazte and Nicolas Andreff from Femto-ST, Besançon. Following our developments in visual tracking and visual servoing from the mutual information [3], it concerned mutual information-based registration of white light images vs. fluorescence images for micro-robotic laser microphonosurgery of the vocal folds. Nelder-Mead Simplex for nonlinear optimization has been used to minimize the cost-function [43].

7.1.3. Pose estimation from RGB-D sensor

Participant: Eric Marchand.

RGB-D sensors have become in recent years a product of easy access to general users. They provide both a color image and a depth image of the scene and, besides being used for object modeling, they can also offer important cues for object detection and tracking in real-time. In this context, the work presented in this paper investigates the use of consumer RGB-D sensors for object detection and pose estimation from natural features. Two methods based on depth-assisted rectification are proposed, which transform features extracted from the color image to a canonical view using depth data in order to obtain a representation invariant to rotation, scale and perspective distortions. While one method is suitable for textured objects, either planar or non-planar, the other method focuses on texture-less planar objects [18]

7.1.4. 3D localization for airplane landing

Participants: Noël Mériaux, François Chaumette, Patrick Rives, Eric Marchand.

This study is realized in the scope of the ANR VisioLand project (see Section 9.2.2 . In a first step, we have considered and adapted our model-based tracker [2] to localize the aircraft with respect to the airport surroundings. Satisfactory results have been obtained from real image sequences provided by Airbus. In a second step, we have started to perform this localization from a set of keyframe images corresponding to the landing trajectory.

7.2. Visual servoing

7.2.1. Histogram-based visual servoing

Participants: Quentin Bateux, Eric Marchand.

Classically visual servoing considers the regulation in the image of a set of visual features (usually geometric features). Direct visual servoing schemes, such as photometric visual servoing, have been introduced in order to consider every pixel of the image as a primary source of information and thus avoid the extraction and the tracking of such geometric features. This year, we proposed a method to extend these works by using a global descriptor, namely intensity histograms, on the whole or multiple sub-sets of the images in order to achieve control of a 6 degrees of freedom (DoF) robot [30][53].

7.2.2. Photometric moment-based visual servoing

Participants: Manikandan Bakthavatchalam, François Chaumette.

This work also belongs to the class of direct visual servoing. Its goal was to use photometric moments as visual features in order to increase the convergence domain of this approach by reducing the non linearity of the control problem. In order to cope with appearance and disappearance of some parts of the environment during the camera motion, a spatial weight has been introduced in the definition of photometric moments. Thanks to a particular design of this weight, the analytical form of the interaction matrix has been obtained, from which it was possible to select a set of moment combinations to control all the six degrees of freedom of the system. Satisfactory experimental results have been obtained [29][8], even if the loss of invariance properties makes the optimal design of visual features still an open problem.

7.2.3. Model predictive visual servoing

Participants: Nicolas Cazy, Paolo Robuffo Giordano, François Chaumette.

The goal of this work is to exploit Model Predictive Control (MPC) techniques for dealing in a robust way with loss of features during a IBVS task. The work [31] provides an experimental validation of different correction schemes able to cope with loss of features due to occlusions of limited camera field of view. The reported results show the effectiveness of the proposed techniques during the servoing of four point features.

7.2.4. Nanomanipulation

Participants: Le Cui, Eric Marchand.

Following our work related to scanning electron micro- scope (SEM) calibration [12] we considered the control of a micro robot using a direct photometric visual servoing that uses only the pure image information as a visual feature, instead of using classic geometric features such as points or lines. However, in micro-scale, using only image intensity as a visual feature performs unsatisfactorily in cases where the photometric variation is low, such as motions along vision sensor's focal axis under a high magnification. In order to improve the performance and accuracy in those cases, an approach using hybrid visual features is proposed in this paper. Image gradient is employed as a visual feature on z axis while image intensity is used on the other 5 DoFs to control the motion. A 6-DoF micro-positioning task is accomplished by this hybrid visual servoing scheme [34].

We also considered a full scale autofocus approach for SEM [35]. The optimal focus (in-focus) position of the microscope is achieved by maximizing the image sharpness using a vision-based closed-loop control scheme. An iterative optimization algorithm has been designed using the sharpness score derived from image gradient information. The proposed method has been implemented and validated using a tungsten gun SEM at various experimental conditions like varying raster scan speed, magnification at real-time.

7.2.5. Audio-based control

Participants: Aly Magassouba, François Chaumette.

This study is not concerned with visual servoing, but to the application of the same principle of sensor-based control to audio sensors. It is made in collaboration with Nancy Bertin from Panama group at Irisa, Inria Rennes-Bretagne Atlantique. In a first step, we have determined the analytical form of the interaction matrix of audio features based on the time difference of arrival on two microphones. From this modeling step, we have determined the different virtual linkages that can be realized in function of the number and configuration of sources [41]. First experimental results using two microphones monunted on the Pioneer mobile robot (see Section 6.9) have been recently obtained.

7.3. Visual navigation of mobile robots

7.3.1. Visual navigation from straight lines

Participants: Suman Raj Bista, Paolo Robuffo Giordano, François Chaumette.

This study is concerned with visual autonomous navigation in indoor environments. As in our previous works concerning navigation outdoors [4], the approach is based on a topological localization of the current image with respect to a set of keyframe images, but the visual features used for this localisation as well as for the visual servoing is not based on points of interest, but straight lines that are more common indoors. Satisfactory experimental results have been obtained using the Pioneer mobile robot (see Section 6.9) [23].

7.3.2. Autonomous navigation of a wheelchair and social navigation

Participants: Vishnu Karakkat Narayanan, François Pasteau, Marie Babel.

Navigating within an unknown indoor environment using an electric wheelchair is a challenging task, especially if the user suffers from severe disabilities. We presented in [22] a framework for vision-based autonomous indoor navigation in an electric wheelchair capable of following corridors, and passing through open doorways using a single doorpost. The designed control schemes have been implemented onto a robotized wheelchair and experimental results show the robust behaviour of the designed system.

We then introduced in [40] a task-based control law which can serve as a low-level system for equitably joining interacting groups, while confirming to social conventions. The system uses the position and orientation of the participating humans with respect to a rigid sensor frame in order to control the translational and rotational velocity of a wheelchair so that the robot positions itself aptly at the meeting point

7.3.3. Semi-autonomous control of a wheelchair for navigation assistance

Participants: Vishnu Karakkat Narayanan, François Pasteau, Marie Babel.

To address the wheelchair driving assistance issue, we proposed in [56][28] a unified shared control framework able to smoothly correct the trajectory of the electrical wheelchair. The system integrates the manual control with sensor-based constraints by means of a dedicated optimization strategy. The resulting low-complex and low-cost embedded system is easily plugged onto on-the-shelf wheelchairs.

The robotic solution has been then validated through clinical trials that have been conducted within the Rehabilitation Center of Pôle Saint Hélier (France) with 25 volunteering patients presenting different disabling neuro-pathologies. This assistive tool is shown to be intuitive and robust as it respects the user intention, it does not alter perception while reducing the number of collisions in case of hazardous maneuvers or in crowded environment [27].

7.4. 3D Scene Mapping

7.4.1. Structure from motion

Participants: Riccardo Spica, Paolo Robuffo Giordano, François Chaumette.

Structure from motion (SfM) is a classical and well-studied problem in computer and robot vision, and many solutions have been proposed to treat it as a recursive filtering/estimation task. However, the issue of *actively* optimizing the transient response of the SfM estimation error has not received a comparable attention. In the work [50] we have addressed the active estimation of the 3D structure of an observed planar scene by comparing three different techniques: a homography decomposition (a well-established method taken as a baseline), a least-square fitting of a reconstructed 3D point cloud, and a direct estimation based on the observation of a set of discrete image moments made of a collection of image points belonging to the observed plane. The experimental results confirmed the importance of actively controlling the camera motion in order to obtained a faster convergence for the estimation error, as well as the superiority of the third method based on the machinery of image moments for what concerns robustness against noise and outliers. In [51] the active estimation scheme has been improved by considering a set of features invariant to camera rotations. This way, the dynamics of the structure estimation becomes independent of the camera angular velocity whose measurement is, thus, no longer required for implementing the active SfM scheme. Finally, in [46] the issue of determining online the ‘best’ combination of image moments for reconstructing the scene structure has been considered. By defining a new set of weighted moments as a weighted sum of traditional image moments, it is indeed possible to optimize for the weights online during the camera motion. The SfM scheme then automatically selects online the best combination of image moments to be used as measurements as a function of the current scene.

7.4.2. Scene Registration based on Planar Patches

Participants: Eduardo Fernandez Moral, Patrick Rives.

Scene registration consists of estimating the relative pose of a camera with respect to a scene previously observed. This problem is ubiquitous in robot localization and navigation. We propose a probabilistic framework to improve the accuracy and efficiency of a previous solution for structure registration based on planar representation. Our solution consists of matching graphs where the nodes represent planar patches and the edges describe geometric relationships. The maximum likelihood estimation of the registration is estimated by computing the graph similarity from a series of geometric properties (areas, angles, proximity, etc..) to maximize the global consistency of the graph. Our technique has been validated on different RGB-D sequences, both perspective and spherical [14].

7.4.3. Robust RGB-D Image Registration

Participants: Tawsif Gokhool, Renato José Martins, Patrick Rives.

Estimating dense 3D maps from stereo sequences remains a challenging task where building compact and accurate scene models is relevant for a number of tasks, from localization and mapping to scene rendering [20], [10] . In this context, this work deals with generating complete geometric and photometric “minimal” model of indoor/outdoor large-scale scenes, which are stored within a sparse set of spherical images to asset photo-geometric consistence of the scene from multiple points-of-views . To this end, a probabilistic data association framework for outlier rejection is formulated, enhanced with the notion of landmark stability over time. The approach was evaluated within the frameworks of image registration, localization and mapping, demonstrating higher accuracy and larger convergence domains over different datasets [39].

7.4.4. Accurate RGB-D Keyframe Representation of 3D Maps

Participants: Renato José Martins, Eduardo Fernandez Moral, Patrick Rives.

Keyframe-based maps are a standard solution to produce a compact map representation from a continuous sequence of images, with applications in robot localization, 3D reconstruction and place recognition. We have present a approach to improve keyframe-based maps of RGB-D images based on two main filtering stages: a regularization phase in which each depth image is corrected considering both geometric and photometric image constraints (planar and superpixel segmentation); and a fusion stage in which the information of nearby frames (temporal continuity of the sequence) is merged (using a probabilistic framework) to improve the accuracy and reduce the uncertainty of the resulting keyframes. As a result, more compact maps (with less keyframes) are created. We have validated our approach with different kind of RGB-D data including both indoor and outdoor sequences, and spherical and perspective sensors, demonstrating that our approach compares and outperforms the state-of-the-art [42].

7.4.5. Semantic Representation For Navigation In Large-Scale Environments

Participants: Romain Drouilly, Patrick Rives.

Autonomous navigation is one of the most challenging problem to address to allow robots to evolve in our everyday environments. Map-based navigation has been studied for a long time and researches have produced a great variety of approaches to model the world. However, semantic information has only recently been taken into account in those models to improve robot efficiency.

Mimicking human navigation is a challenging goal for autonomous robots. This requires to explicitly take into account not only geometric representation but also high-level interpretation of the environment [9]. We propose a novel approach demonstrating the capability to infer a route in a global map by using semantics. Our approach relies on an object-based representation of the world automatically built by robots from spherical images. In addition, we propose a new approach to specify paths in terms of high-level robot actions. This path description provides robots with the ability to interact with humans in an intuitive way. We perform experiments on simulated and real-world data, demonstrating the ability of our approach to deal with complex large-scale outdoor environments whilst dealing with labelling errors [37].

Mapping evolving environments requires an update mechanism to efficiently deal with dynamic objects. In this context, we propose a new approach to update maps pertaining to large-scale dynamic environments with semantics. While previous works mainly rely on large amount of observations, the proposed framework is able to build a stable representation with only two observations of the environment. To do this, scene understanding is used to detect dynamic objects and to recover the labels of the occluded parts of the scene through an inference process which takes into account both spatial context and a class occlusion model. Our method was evaluated on a database acquired at two different times with an interval of three years in a large dynamic outdoor environment. The results point out the ability to retrieve the hidden classes with a precision score of 0.98. The performances in term of localisation are also improved [36].

7.5. Control of single and multiple Unmanned Aerial Vehicles

7.5.1. Single UAV

Participant: Paolo Robuffo Giordano.

Over the last years the robotics community witnessed an increasing interest in the Unmanned Aerial Vehicle (UAV) field. In particular quadrotor UAVs have become more and more widespread in the community as experimental platform for, e.g., testing novel 3D planning, control and estimation schemes in real-world indoor and outdoor conditions. Indeed, in addition to being able to take-off and land vertically, quadrotors can reach high angular accelerations thanks to the relatively long lever arm between opposing motors. This makes them more agile than most standard helicopters or similar rotorcraft UAVs, and thus very suitable to realize complex tasks such as aerial mapping, air pollution monitoring, traffic management, inspection of damaged buildings and dangerous sites, as well as agricultural applications such as pesticide spraying.

Despite these clear advantages, a clear shortcoming of the quadrotor design lies in its inherent underactuation (only 4 actuated propellers for the 6 dofs of the quadrotor pose). This underactuation limits the quadrotor flying ability in free or cluttered space and, furthermore, it also degrades the possibility of interacting with the environment by exerting desired forces in arbitrary directions. In [24], a novel design for a quadrotor UAV with tilting propellers which is able to overcome these limitations has been presented and experimentally validated. Indeed, the additional set of 4 control inputs actuating the propeller tilting angles can be shown to yield full actuation to the quadrotor position/orientation in space, thus allowing it to behave as a fully-actuated flying vehicle and to overcome the aforementioned underactuation problem.

Furthermore, the issue of estimating online the UAV self-motion from vision has been considered. To this end, a novel nonlinear estimation scheme able to recover the metric UAV linear velocity from the *scaled* one obtained from the decomposition of the optical flow has been proposed in [15]. The observability conditions (in terms of persistency of excitation) needed to ensure a converging estimation have also been studied. The reported experimental results confirmed the effectiveness of the estimation scheme in recovering a reliable and accurate estimation of the UAV self-motion (linear and angular velocities) in realistic conditions.

This work has been realized in collaboration with the Max Planck Institute for Biological Cybernetics, Tübingen, Germany.

7.5.2. Collective control of multiple UAVs

Participants: Fabrizio Schiano, Paolo Robuffo Giordano.

The challenge of coordinating the actions of multiple robots is inspired by the idea that proper coordination of many simple robots can lead to the fulfilment of arbitrarily complex tasks in a robust (to single robot failures) and highly flexible way. Teams of multi-robots can take advantage of their number to perform, for example, complex manipulation and assembly tasks, or to obtain rich spatial awareness by suitably distributing themselves in the environment. Within the scope of robotics, autonomous search and rescue, firefighting, exploration and intervention in dangerous or inaccessible areas are the most promising applications.

In the context of multi-robot (and multi-UAV) coordinated control, *connectivity* of the underlying graph is perhaps the most fundamental requirement in order to allow a group of robots accomplishing common goals by means of *decentralized* solutions. In fact, graph connectivity ensures the needed continuity in the data flow among all the robots in the group which, over time, makes it possible to share and distribute the needed information. However, connectivity alone is not sufficient to perform certain tasks when only *relative sensing* is used. For these systems, the concept of *rigidity* provides the correct framework for defining an appropriate sensing and communication topology architecture. Rigidity is a combinatorial theory for characterizing the "stiffness" or "flexibility" of structures formed by rigid bodies connected by flexible linkages or hinges. In a broader context, rigidity turns out to be an important architectural property of many multi-agent systems when a common inertial reference frame is unavailable. Applications that rely on sensor fusion for localization, exploration, mapping and cooperative tracking of a target, all can benefit from notions in rigidity theory. The concept of rigidity, therefore, provides the theoretical foundation for approaching decentralized solutions to the aforementioned problems using distance measurement sensors, and thus establishing an appropriate framework for relating system level architectural requirements to the sensing and communication capabilities of the system.

In [26], a decentralized gradient-based rigidity maintenance action for a group of quadrotor UAVs has been proposed and tested in real experimental conditions. By starting in a rigid configuration, the group of UAVs is able to estimate their relative position from sole relative distance measurements, and then use these estimated relative positions in a control action able to preserve rigidity of the whole formation despite presence of sensor limitations (maximum range and line-of-sight occlusions), possible collisions with obstacles and inter-robot collisions. Furthermore, in [52] the novel case of *bearing rigidity* for directed graphs has been considered: here, rather than distances the measurements are the 3D bearing vectors expressed in the local body-frame of each agent. The theory has been developed for the case of planar agents in $SE(2)$ and a 'scale-free' bearing controller has been proposed, able to steer the robot group towards a desired bearing formation.

These works were realized in collaboration with the robotics group at the Max Planck Institute for Biological Cybernetics, Tübingen, Germany and with Technion, Israel.

7.5.3. Cooperative localization using interval analysis

Participants: Vincent Drevelle, Ide Flore Kenmogne Fokam.

In the context of multi-robot fleets, cooperative localization consists in gaining better position estimate through measurements and data exchange with neighboring robots. Positioning integrity (i.e., providing reliable position uncertainty information) is also a key point for mission-critical tasks, like collision avoidance. The goal of this work is to compute position uncertainty volumes for each robot of the fleet, using a decentralized method (i.e using only local communication with the neighbors). The problem is addressed in a bounder-error framework, with interval analysis and constraint propagation methods. These methods enable to provide guaranteed position error bounds, assuming bounded-error measurements. They are not affected by over-convergence due to data incest, which makes them a well sound framework for decentralized estimation. Encouraging results have already been obtained for multi-robot underwater positioning with acoustical range measurements. Ongoing work focuses on cooperative localization in a multi-UAV fleet with image-based measurements (bearings).

7.6. Medical robotics

7.6.1. Non-rigid target tracking in ultrasound images combining dense information and physically-based model

Participants: Lucas Royer, Alexandre Krupa.

This study concerns the real-time tracking of deformable targets within a sequence of ultrasound (US) images. The proposed approach combines dense information with a physically-based model and has therefore the advantage of not using any fiducial marker. The physical model is represented by a mass-spring damper system driven by external and internal forces. The external forces are obtained by maximizing an image similarity metric between a reference target and the deformed target along the time. The internal forces of the mass-spring damper system constrain the deformation to be physically plausible and therefore efficiently reduce the sensitivity to the speckle noise. This approach was first validated from simulated and real sequences of 2D US images [49]. It was then extended for deformable target tracking in a sequence of 3D ultrasound volumes and tested on a robotic setup used to apply deformation on an organic phantom [48]. The performance of this deformable 3D target tracking approach was evaluated with visual assessment combined with robotic odometry ground truth. This method was also tested and compared with respect to state-of-the-art techniques by using 3D image databases provided by MICCAI CLUST'14 and CLUST'15 challenges [47] (MICCAI Challenge on Liver Ultrasound Tracking). It was awarded by the organizers of the CLUST challenges as being the best method for accurate target tracking in 3D ultrasound sequences. We recently improved our approach in order to increase its robustness to the presence of ultrasound shadows, local illumination changes and image occlusions.

7.6.2. *3D steering of flexible needle by ultrasound visual servoing*

Participants: Pierre Chatelain, Jason Chevrie, Marie Babel, Alexandre Krupa.

The objective of this work is to provide robotic assistance during needle insertion procedures such as biopsy or ablation of localized tumor. In previous work, we designed a control approach based on a duty cycling technique for steering a beveled-tip flexible needle actuated by a robotic arm in such a way to control the needle curvature in 3D space and reach a desired target by visual servoing. In this preliminary work, the control approach was validated by using visual features extracted from 2 images provided by 2 orthogonal cameras observing a translucent gelatin phantom where the needle was inserted. This year, we have pursued our work towards this needle steering robotic assistance by developing a new algorithm able to track in real-time a flexible needle in a sequence of 3D ultrasound images (volumes). The flexible needle modeled as a polynomial curve is tracked during the automatic insertion using particle filtering. This new tracking algorithm enables real-time closed-loop needle control with 3D ultrasound feedback. The target to reach was manually defined by the user in the US image and can be on-line tracked thanks to the template tracking algorithm proposed in [21] based on ultrasound dense visual servoing [7]. Experimental results of an automatic needle tip positioning in a home-made gelatine phantom demonstrate the feasibility of 3D ultrasound-guided needle steering for reaching a desired target by ultrasound visual servoing [33]. Recently a new control law for needle steering that uses both direct manipulation of the needle base and the duty cycling method has been studied. It is based on a 3D model of a beveled tip needle using virtual springs that characterize the needle mechanical interaction with soft tissue. From this model, a measure of the controllability of the needle tip degrees of freedom was proposed in order to mix the control between the direct base manipulation and the duty cycling technique. Preliminary simulations show that this hybrid control allows better targeting capabilities in terms of larger needle workspace and reduced needle bending.

7.6.3. *Optimization of ultrasound image quality by visual servoing*

Participants: Pierre Chatelain, Alexandre Krupa.

This study focuses on a new ultrasound-based visual servoing approach that optimizes the positioning of an ultrasound probe manipulated by a robotic arm in order to improve the quality of the acquired ultrasound images. To this end, we use the recent framework of ultrasound confidence map, developed in the Chair for Computer Aided Medical Procedures and Augmented Reality of Prof. Nassir Navab, which aims at estimating the per-pixel quality of the ultrasound signal based on a model of sound propagation in soft tissues. More specifically, we treat the ultrasound confidence maps as a new modality and designed a visual servoing control law for image quality optimization. We illustrated our approach with the application of robotic tele-echography where the in-plane rotation of a 2D probe is visually servoed by the confidence map and the other degrees of freedom are teleoperated by the user. Experiments performed on both an ultrasound examination training phantom and ex vivo tissue samples validated this new concept [32]. Currently, we consider the confidence-driven servoing of other degrees of freedom, in particular out-of-plane motions that were controlled in our previous works from image moments [6], which could provide finer control of the image quality.

7.6.4. Visual servoing based on ultrasound elastography

Participants: Pedro Alfonso Patlan Rosales, Alexandre Krupa.

This study concerns the use of the ultrasound elastography as a new image modality for the control of the motion of an ultrasound probe actuated by a robotic manipulator. Elastography imaging is performed by applying continuous stress variation on soft tissues in order to estimate a strain map of the observed tissues. It is obtained by estimating, from the RF (radio-frequency) signal along each scan line of the probe transducer, the echo time delays between pre- and post-compressed tissue. Usually, this continuous stress variation is performed manually by the user who manipulates the US probe and it results therefore in a user-dependent quality of the elastography image. To improve the US elastography imaging, we recently developed an assistant robotic palpation system that automatically moves an ultrasound probe in such a way to optimize ultrasound elastography. The main originality of this preliminary work concerns the use of the elastography modality directly as input of the robot controller thanks to an innovative ultrasound elastography-based visual servoing approach.

7.6.5. Visual servoing using shearlet transform

Participants: Lesley-Ann Duflot, Alexandre Krupa.

Similar to wavelet transform, shearlet transform is usually used in the field of signal or image compression. At the best of our knowledge these image representations were never used directly as feedback of a closed-loop control scheme. The objective of this work is to study the feasibility of using the coefficients of shearlet transform of the observed ultrasound image directly as the visual features of an image-based visual servoing. In this study we estimated numerically the interaction matrix that links the time variation of the coarsest coefficients of the shearlet to the motion of the ultrasound probe. This shearlet-based visual servoing was experimentally tested for automatically positioning a 2D US probe, held by a robot, on a desired section of an abdominal phantom. The first results demonstrated promising performances.

LEMON Team

7. New Results

7.1. Hydrodynamics of the Tunquen lagoon, Chile

Participant: Antoine Rousseau.

In this internship co-advised with Céline Acary-Robert (Inria Chile), Loïc Dagnas developed a numerical hydrodynamic model for a specific lagoon of the chilean costline. This kind of lagoon is characterized by an intermittent connection to the sea and a regular fresh water input coming from the Andean mountains. The hydrodynamic model consists in a two-dimensional shallow water model, including tracer equations for the time evolution of temperature and salinity. The hydrodynamic circulation of the lagoon has been simulated taking into account various external forcings such as water exchanges with the atmosphere, wind effects and external pumping.

7.2. Upscaled modeling of Vaccares lake in Camargue

Participants: Carole Delenne, Antoine Rousseau, Vincent Guinot.

Sélim Cornet developed a numerical model for the hydrodynamics of Vaccares system in Camargue. The data and reference simulations (made with TELEMAC-2D) were provided by Tour du Valat (contact O. Boutron). Sélim's work consisted in the implementation and validation of the porosity shallow water model developped by Vincent GUINOT , in order to obtain accurate but inexpensive simulations of the Vaccares hydrosystem.

7.3. Numerical simulation of coastal flood made by Joanna storm (France, 2008)

Participant: Fabien Marche.

In collaboration with BRGM, a numerical platform based on Fabien Marche's numerical too WaveBox was developed in order to simulate coastal urban submersions associated with intense storms, see [22]. A nudging strategy is implemented with:

- a barotropic model at the regional scale,
- a spectral wave model with embeded meshes accounting for water level evolution and output from the large scale model,
- a free surface Shallow Water model (SURF2D, now called WaveBox) used at very high resolution for the submersion process.

7.4. Analysis of the inclusion of vorticity on fully nonlinear and weakly dispersive long wave models

Participant: Fabien Marche.

We study in [11] the propagation of long waves in the presence of vorticity. In the irrotational framework, the Green-Naghdi equations (also called Serre or fully nonlinear Boussinesq equations) are the standard model for the propagation of such waves. These equations couple the surface elevation to the vertically averaged horizontal velocity and are therefore independent of the vertical variable. In the presence of vorticity, the dependence on the vertical variable cannot be removed from the vorticity equation but it was however shown in [9] that the motion of the waves could be described using an extended Green-Naghdi system. In this paper we propose an analysis of these equations, and show that they can be used to get some new insight into wave-current interactions. We show in particular that solitary waves may have a drastically different behavior in the presence of vorticity and show the existence of solitary waves of maximal amplitude with a peak at their crest, whose angle depends on the vorticity. We also propose a robust and simple numerical scheme validated on several examples. Finally, we give some examples of wave-current interactions with a non trivial vorticity field and topography effects.

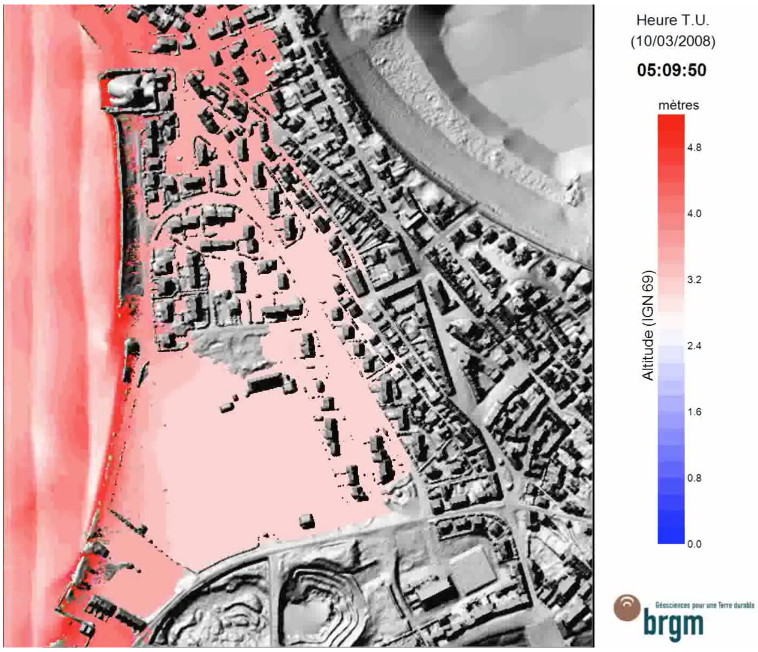


Figure 4. Simulation of Joanna storm (France, 2008). Snapshot of the movie made by BRGM, available on their Youtube page.

7.5. Multiscale aspects for confinement of coastal lagoons

Participant: Antoine Rousseau.

In [3] we expand a previous definition of paralic confinement (see [41]), and make it usable from the modeling slant, before implementing it in numerical tools. More specifically, we here deal with the multiscale aspect of the confinement. If a paralic environment is separated into two (or more) connected areas, we will show that is possible to split the confinement problem into two related problems, one for each area. We also focus on the importance of the interface length between the two subdomains.

7.6. Interface conditions for ocean models

Participant: Antoine Rousseau.

In [5] we are interested in the search of interface conditions to couple hydrostatic and nonhydrostatic ocean models. To this aim, we consider simplified systems and use a time discretization to handle linear equations. We recall the links between the two models (with the particular role of the aspect ratio $\delta = H/L$) and introduce an iterative method based on the Schwarz algorithm (widely used in domain decomposition methods). The convergence of this method depends strongly on the choice of interface conditions: this is why we look for exact absorbing conditions and their approximations in order to provide tractable and efficient coupling algorithms.

In [4] we present a study of optimized Schwarz domain decomposition methods for Navier-Stokes equations. Once discretized in time, optimal transparent boundary conditions are derived for the resulting Stokes equations, and a series of local approximations for these nonlocal conditions are proposed. Their convergence properties are studied, and numerical simulations are conducted on the test case of the driven cavity. It is shown that conditions involving one or two degrees of freedom can improve the convergence properties of the original algorithm.

7.7. Use of remote sensing data for hydraulic modelling

Participant: Carole Delenne.

Wetlands provide a vital resource to ecosystem services and associated rural livelihoods; but their extent, geomorphological heterogeneity and flat topography make the representation of their hydrological functioning complex. The main objective of this area of research is to assess the relevance of remote sensing data for the monitoring and hydraulic modelling of different hydrosystems. In [14], a semi automated method exploiting 526 MODIS 8-day 500 m resolution images was developed to study the spatial and temporal dynamics of the annual flood across the Niger Inner Delta over the period 2000?2011. The flooded area is detected using band ratio indexes. Results were evaluated against classified Landsat images, previous studies and field stage data for a range of hydrological units: river stretches, lakes, floodplains and irrigated areas. Depending on the study area, its extent, and the objective to be reached, different kinds of remote sensing data may be interesting: RADAR, multispectral, high/low spatial/temporal resolution, etc. Several paths for research are currently considered to upgrade the use of remote-sensing data in hydrodynamic modelling:

- use of detected flooded area for model validation and for the calibration of parameters such as friction coefficient.
- topography assessment using the detection of the flooded area of a given wetland at different times.
- characterization of statistical properties of the geometry of the urban medium (useful for large-scale models): statistical, subgrid-scale properties of the topography, and information regarding the flow connectivity properties of the urban medium.

7.8. Lumped hydrological models with infinite characteristic time transfer functions

Participant: Vincent Guinot.

Karst and mountainous catchments usually exhibit rainfall-runoff transfer functions involving multiple time scales. In most existing conceptual, hydrological models of such catchments, multiple time scale response is achieved by introducing several reservoirs and non-linear transfer functions. In [9], multiple time scales are introduced by proposing a transfer function with an infinite characteristic time. The heavy-tailed transfer function behaves asymptotically as an inverse power of time. In the limit of long time scales, the governing equation for the system obeys a fractional differential equation. With a single reservoir, the proposed approach is shown to perform satisfactorily compared to other models of similar or more complex structure. The fractional differential equation is shown to be useless for usual time scales and should not be used in practice.

7.9. Upscaled models for urban floods

Participant: Vincent Guinot.

Shallow water models with porosity have arisen over the last two decades as a promising alternative to refined flow models for the simulation of urban floods. Several porosity-based models have been proposed in the literature. In [10], the integral porosity formalism developed at the University of California Irvine is validated against scale model experiments. A sudden dike breaching near an idealized city layout is simulated in the scale model. Comparison with numerical simulations shows the superiority of the integral porosity model over the single porosity model in reproducing the effects of urban layout anisotropy on flood wave propagation properties. This research has initiated a collaboration between the LEMON team and UC Irvine for the development of a new porosity formalism.

7.10. Models for dispersion in porous media

Participants: Carole Delenne, Vincent Guinot.

Solute dispersion in porous media is usually modelled using Fick’s law or fractional variations of the solute dispersion equation. The Fickian model, however, is known to exhibit a number of drawbacks, such as poor scaling properties. This is also true for its fractional counterparts, that perform with limited success when compared to experimental data sets. In [13], a high-quality experimental device is built in the form of periodic heterogeneities of length 15 cm. Placing up to 10 periods in series allows the scaling properties of the dispersion model to be analyzed. Besides providing a high quality experimental database, the results in [13] indicate that (i) previously identified scaling trends for the dispersion coefficient may easily be explained by experiment variability, (ii) there exists a linear transport model that allows the experimental behaviour to be reproduces at all scales, (iii) this model is not the advection-dispersion model (even fractional).

7.11. Invasion in growth-fragmentation-death models

Participant: Fabien Campillo.

In collaboration with Nicolas Champagnat and Coralie Fritsch (Inria Nancy), we present in [20] two approaches to study invasion in growth-fragmentation-death models: one based on a stochastic individual based model and one based on an integro-differential model. The invasion of the population is described by the survival probability for the first model and by an eigenproblem for the second one. We study these two notions of invasion fitness, giving different characterizations of the growth of the population, and we make links between these two complementary points of view. We apply our work in the context of adaptive dynamics in a chemostat model.

7.12. Stochastic growth model with extinction

Participant: Fabien Campillo.

In collaboration with Marc Joannides and Irène Larramendy-Valverde (IMAG / Université de Montpellier), we consider in [6] a stochastic logistic growth model given by a stochastic differential equation featuring both birth and death rates in the drift and diffusion coefficients. Our aim is to infer these rates, based on discrete observations with possible extinction. Since extinction occurs eventually for the model, the density of the diffusion process is not absolutely continuous with respect to the Lebesgue measure; we established the associated Fokker-Planck equation together with appropriate numerical schemes. This formulation allows to design variants of the standard methods that can handle extinction.

MAESTRO Project-Team

7. New Results

7.1. Network Science

Participants: Eitan Altman, Konstantin Avrachenkov, Arun Kadavankandy, Jithin Kazhuthuveettil Sreedharan, Hlib Mykhailenko, Philippe Nain, Giovanni Neglia, Yonathan Portilla, Alexandre Reiffers-Masson.

7.1.1. Posting behavior in Social Networks and Content Active Filtering

In [57], Alexandre Reiffers-Masson and Eitan Altman in collaboration with Yezekael Hayel (UAPV), model the posting behavior in Social Networks in topics which have negative externalities, and propose content active filtering in order to increase content diversity. By negative externalities, it is meant that when the quantity of posted contents about some topic increases the popularity of posted contents decreases. They introduce a dynamical model to describe the posting behavior of users taking into account these externalities. Their model is based on stochastic approximations and sufficient conditions are provided to ensure its convergence to a unique rest point. They provide a closed form expression for this rest point. Content Active Filtering (CAF) are actions taken by the administrator of the Social Network in order to promote some objectives related to the quantity of contents posted in various topics. As objective of the CAF they consider maximizing the diversity of posted contents.

7.1.2. Network centrality measures

Recent papers studied the control of spectral centrality measures of a network by manipulating the topology of the network. In [56], Alexandre Reiffers-Masson, Eitan Altman and Yezekael Hayel (UAPV) extend these works by focusing on a specific spectral centrality measure, the Katz-Bonacich centrality. The optimization of the Katz-Bonacich centrality using a topological control is called the Katz-Bonacich optimization problem. The authors first prove that this problem is equivalent to a linear optimization problem. Thus, in the context of large graphs, one can use state-of-the-art algorithms. The authors provide a specific applications of the Katz-Bonacich centrality minimization problem based on the minimization of gossip propagation and make some experiments on real networks which validate the model assumptions.

Betweenness centrality is one of the basic concepts in the analysis of social networks. The initial definition for the betweenness of a node in a graph is based on the fraction of the number of geodesics (shortest paths) between any two nodes that this given node lies on, to the total number of the shortest paths connecting these nodes. This method has quadratic complexity and does not take into account indirect paths. In [45] K. Avrachenkov in collaboration with V. Mazalov (Korelian Institute of Applied Mathematical Research, Russia) and B. Tsynguev (Transbaikal State Univ., Russia) propose a new concept of betweenness centrality for weighted networks, called beta current flow centrality, based on Kirchhoff's law for electric circuits. In comparison with the original current flow centrality and alpha current flow centrality, this new measure can be computed for larger networks. The results of numerical experiments for some examples of networks, in particular, for the popular social network VKontakte as well as the comparison with PageRank method are presented.

PageRank has numerous applications in information retrieval, reputation systems, machine learning, and graph partitioning. In [44], K. Avrachenkov and A. Kadavankandy in collaboration with L.O. Prokhorenkova and A. Raigorodskii (both from Yandex Research) study PageRank in undirected random graphs with expansion property. The Chung-Lu random graph represents an example of such graphs. The authors show that in the limit, as the size of the graph goes to infinity, PageRank can be represented by a mixture of the restart distribution and the vertex degree distribution.

7.1.3. Mining social networks

Social Networks became a major actor in information propagation. Using the Twitter popular platform, mobile users post or relay messages from different locations. The tweet content, meaning and location show how an event-such as the bursty one "JeSuisCharlie" happened in France in January 2015 is comprehended in different countries. In [75], [76] researchers from UAPV and Inria (Mohamed Morchid, Yonathan Portilla, Didier Josselin, Richard Dufour, Eitan Altman, Marc El-Beze, Jean-Valère Cossu, Georges Linarès, Alexandre Reiffers-Masson), studied clustering of the tweets according to the co-occurrence of their terms, including the country, and forecasting the probable country of a non located tweet, knowing its content. First, they present the process of collecting a large quantity of data from the Twitter website. The dataset consists of 2.189 located tweets about "Charlie", from the 7th to the 14th of January. The authors then describe an original method adapted from the Author-Topic (AT) model based on the Latent Dirichlet Allocation method (LDA). They define a homogeneous space containing both lexical content (words) and spatial information (country). During a training process on a part of the sample, the authors provide a set of clusters (topics) based on statistical relations between lexical and spatial terms. During a clustering task, they evaluate the method effectiveness on the rest of the sample that reaches up to 95% of good assignments. It shows that the model is pertinent to foresee tweet location after a learning process.

7.1.4. Analysis of Internet Memes

Memes have been defined by R. Dawkins as cultural phenomena that propagate through non genetic ways. In [42], Eitan Altman and Yonathan Portilla examine three very popular Internet Memes and study their impact on the society in mediterranean countries. the authors use existing software tools (such as Google Trends) as well as tools that they develop in order to quantify the impact of the Memes on the mediterranean societies. The authors obtain quite different results with the different tools they use, which they explain based on some propagation characteristic of each one of the Memes. The analysis shows the extent to which these Memes cross borders and thus contribute to the creation of a globalized culture. The authors finally identify some of the impacts of the globalization of culture.

7.1.5. Trend detection in social networks using Hawkes processes

In [52], Julio Cesar Louzada Pinto and Tijani Chahed (Telecom SudParis) in collaboration with Eitan Altman propose a new trend detection algorithm, designed to find trendy topics being disseminated in a social network. The authors assume that the broadcasts of messages in the social network is governed by a self-exciting point process, namely a Hawkes process, which takes into consideration the real broadcasting times of messages and the interaction between users and topics. The authors formally define trendiness and derive trend indices for each topic being disseminated in the social network. These indices take into consideration the time between the detection and the message broadcasts, the distance between the real broadcast intensity and the maximum expected broadcast intensity, and the social network topology. The proposed trend detection algorithm is simple and uses stochastic control techniques in order to calculate the trend indices. It is also fast and aggregates all the information of the broadcasts into a simple one-dimensional process, thus reducing its complexity and the quantity of data necessary to the detection.

7.1.6. Study of the Youtube recommendation system

The Youtube recommendation system is one the most important view source of a video. In [54], Yonathan Portilla, Alexandre Reiffers-Masson, Eitan Altman in collaboration with Rachid El-Azouzi (UAPV) study the role of recommendation systems in boosting the popularity of videos. The authors first construct a graph that captures the recommendation system in Youtube and study empirically the relationship between the number of views of a video and the average number of views of the videos in its recommendation list. The authors then consider a random walker on the recommendation graph, i.e. a random user that browses through videos such that the video it chooses to watch is selected randomly among the videos in the recommendation list of the previous video it watched. The authors study the stability properties of this random process and show that the trajectory obtained does not contain cycles if the number of videos in the recommendation list is small (which is the case if the computer's screen is small).

7.1.7. Average consensus protocols

In [22] M. El Chamie (Univ. of Washington, USA), G. Neglia and K. Avrachenkov study the weight optimization problem for average consensus protocols by reformulating it as a Schatten norm minimization with parameter p. They show that as p approaches infinity, the optimal solution of the Schatten norm induced problem recovers the optimal solution of the original problem. Moreover, by tuning the parameter p in the proposed minimization, it is possible to trade-off the quality of the solution (i.e., the speed of convergence) for communication/computation requirements (in terms of number of messages exchanged and volume of data processed). They then propose a distributed algorithm to solve the Schatten norm minimization and show that it outperforms the other distributed weight selection methods.

7.1.8. Estimation techniques

The estimation of a large population's size by means of sampling procedures is a key issue in many networking scenarios. Their application domains span from RFID systems to peer-to-peer networks; from traffic analysis to wireless sensor networks; from multicast networks to WLANs. In [14], N. Accettura (Univ. of California Berkeley, USA), G. Neglia and L. A. Grieco (Politecnico di Bari, Italy) illustrate and classify in a coherent framework the main approaches proposed so far in the computer networks literature to deal with such a problem. In particular, starting from the methodologies proposed in ecological studies since the last century, they survey their counterparts in the computer network domain, finding that many lessons can be gained from this insightful investigation. Capture-Recapture techniques are deeply analyzed to allow the reader to exactly understand their pros, cons, and applicability bounds. Finally, they discuss some open issues that deserve further investigations and could be relevant to afford estimation problems in next generation Internet.

Online social networks (OSN) contain extensive amount of information about the underlying society that is yet to be explored. One of the most feasible technique to fetch information from OSN, crawling through Application Programming Interface (API) requests, poses serious concerns over the the guarantees of the estimates. In [70] J. Sreedharan and K. Avrachenkov in collaboration with B. Ribeiro (Purdue University, USA) focus on making reliable statistical inference with limited API crawls. Based on regenerative properties of the random walks, they propose an unbiased estimator for the aggregated sum of functions over edges and proved the connection between variance of the estimator and spectral gap. In order to facilitate Bayesian inference on the true value of the estimator, they derive the approximate posterior distribution of the estimate. Later the proposed ideas are validated with numerical experiments on inference problems in real-world networks.

7.1.9. Percolation in multilayer networks

In [79], P. Nain and his co-authors (S. Guha and P. Basu from Raytheon BB Technologies, D. Towsley from the Univ. of Massachusetts, C. Capar from Ericsson Research, A. Swami from the US Army Research Lab.) consider multiple networks formed by a common set of users connected via M different means of connectivity, where each user (node) is active, independently, in any given network with probability q. They show that when q exceeds a threshold $q_c(M)$, a giant connected component appears in the M-layer network—thereby enabling faraway users to connect using 'bridge' nodes that are active in multiple network layers, even though the individual layers may only have small disconnected islands of connectivity. They show that $q_c(M) \leq \sqrt{\log(1-p_c)}/\sqrt{M}$, where p_c is the bond percolation threshold of the underlying connectivity graph G, and $q_c(1) \equiv q_c$ is its site percolation threshold. The threshold $q_c(M)$ is found explicitly when G is a large random network with an arbitrary node-degree distribution and numerically for various regular lattices. Finally, an intriguingly close connection between this multilayer percolation model and the well-studied problem of site-bond percolation is revealed, in the sense that both models provide a smooth transition between the traditional site and bond percolation models. This connection is used to translate analytical approximations of the site-bond critical region developed in the 1990s, which are functions only of p_c and q_c of the respective lattice, to excellent general approximations of $q_c(M)$.

7.1.10. Extreme Value Theory for Complex Networks

In [20] J. Sreedharan and K. Avrachenkov in collaboration with N. Markovich (Institute of Control Sciences, Moscow) explore the dependence structure in the sampled sequence of complex networks. They consider randomized algorithms to sample the nodes and study extremal properties in any associated stationary sequence of characteristics of interest like node degrees, number of followers, or income of the nodes in online social networks, which satisfy two mixing conditions. Several useful extremes of the sampled sequence like the kth largest value, clusters of exceedances over a threshold, and first hitting time of a large value are investigated. The dependence and the statistics of extremes is abstracted into a single parameter that appears in extreme value theory, called the Extremal Index. The authors derive this parameter analytically and also estimate it empirically. They propose the use of the Extremal Index as a parameter to compare different sampling procedures. As a specific example, degree correlations between neighboring nodes are studied in detail with three prominent random walks as sampling techniques.

7.1.11. Random Matrix Theory for Complex Networks

In [68] A. Kadavankandy and K. Avrachenkov in collaboration with L. Cottatellucci (Eurecom) consider an extension of Erdős-Rényi graph known in the literature as the Stochastic Block Model (SBM). They analyze the limiting empirical distribution of the eigenvalues of the adjacency matrix of a SBM. They derive a fixed point equation for the Stieltjes transform of the limiting eigenvalue empirical distribution function (e.d.f.), concentration results on both the support of the limiting e.d.f. and the extremal eigenvalues outside the support of the limiting e.d.f. Additionally, they derive analogous results for the normalized Laplacian matrix and discuss potential applications of the general results in epidemics and random walks.

In [40], the same authors continue with the analysis of eigenvectors of a Stochastic Block Model. The eigenvalue spectrum of the adjacency matrix of a SBM consists of two parts: a finite discrete set of dominant eigenvalues and a continuous bulk of eigenvalues. They characterize analytically the eigenvectors corresponding to the continuous part: the bulk eigenvectors. For symmetric SBM adjacency matrices, the eigenvectors are shown to satisfy two key properties. A modified spectral function of the eigenvalues, depending on the eigenvectors, converges to the eigenvalue spectrum. Its fluctuations around this limit converge to a Gaussian process different from a Brownian bridge. This latter fact disproves that the bulk eigenvectors are Haar distributed.

7.2. Wireless Networks

Participants: Eitan Altman, Abdulhalim Dandoush.

7.2.1. A General SDN-based IoT Framework with NVF Implementation

The emerging technologies of IoT (Internet of Things), SDN (Software Defined Networking), and NFV (Network Function Virtualization) have a great potential for the information service innovation in the Cloud and big data era. In [26], Jie Li (Tsukuba Univ.) in cooperation with Eitan Altman and with Corinne Touati (Inria Grenoble-Rhône-Alpes) have studied architecture issues in Internet of Things based on SDN with NFV implementation. The contribution of the paper is in providing a view point for integrating these technologies based on their existing standards.

7.2.2. Self-Organizing Network (SON)

Self-Organizing Network (SON) technology aims at autonomously deploying, optimizing and repairing the Radio Access Networks. In [31], Abdoulaye Tall, Zwi Altman (Orange, Issy les Moulineaux) and Eitan Altman showed that in certain cases, it is essential to take into account the impact of the backhaul state in the design of the SON algorithm. They revisit the Base Station load definition taking into account the backhaul state. They provide an analytical formula for the load along with a simple estimator for both elastic and guaranteed bit-rate traffic. They incorporate the proposed load estimator in a self-optimized Load Balancing algorithm. Simulation results for a backhaul constrained heterogeneous network illustrate how the correct load definition can guarantee a proper operation of the SON algorithm.

SON is further studied by these authors in [58], [59] where the Vertical Sectorization (VS) is adapted. VS consists in creating vertically separated sectors in the original cell using an Active Antenna Systems (AAS) supporting two distinct beams with different downtilts. The total transmit power is split between the two sectors, while the frequency bandwidth can be reused by each sector, creating additional interference between the two sectors. For low traffic demand, VS may lead to performance degradation, while for high traffic demand in both sectors, VS is likely to bring about important capacity gains. Hence intelligent activation policy of VS is needed to fully benefit from this feature. The authors propose an approach taking advantage of the more focused downtilted beam. A dynamic alpha fair bandwidth sharing is proposed for low and medium load. It is autonomously replaced by full bandwidth reuse for high load scenarios using a threshold-based controller. A flow-level dynamic simulator is used to numerically validate the proposed mechanisms.

7.2.3. Automated Dynamic Offset for Network Selection in Heterogeneous Networks

Complementing traditional cellular networks with the option of integrated small cells and WiFi access points can be used to further boost the overall traffic capacity and service level. Small cells along with WiFi access points are projected to carry over 60% of all the global data traffic by 2015. With the integration of small cells on the radio access network levels, there is a focus on providing operators with more control over small cell selection while reducing the feedback burden. Altogether, these issues motivate the need for innovative distributed and autonomous association policies that operate on each user under the network operator's control, utilizing only partial information, yet achieving near-optimal solutions for the network. In [25], Majed Haddad (UAPV), Piotr Wiecek (Institute of Mathematics and Computer Science, Wroclaw), Saidi Habib (Inria project-team DYOGENE) and Eitan Altman propose a load-aware network selection approach applied to automated dynamic offset in heterogeneous networks. In particular, they investigate the properties of a hierarchical (Stackelberg) Bayesian game framework, in which the macro cell dynamically chooses the offset about the state of the channel in order to guide users to perform intelligent network selection decisions between macro cell and small cell networks. The authors effectively address the problem of how to intelligently configure a dynamic offset which optimizes network's global utility while users maximize their individual utilities.

7.2.4. Localization in ad-hoc wireless sensors networks

Range-based localization algorithms in wireless sensor networks are more accurate but also more computationally complex than the range-free algorithms. The work on this topic by M. S. Elgamel (Arab Academy for Science, Technology & Maritime Transport, Egypt) and A. Dandoush, previously reported, has been published in [23].

7.3. Network Engineering Games

Participants: Eitan Altman, Konstantin Avrachenkov, Giovanni Neglia.

7.3.1. Matching games and the association problem

In [33], Mikaël Touati, Jean-Marc Kélif (Orange Labs), Rachid El-Azouzi (UAPV), Marceau Coupechoux (Telecom ParisTech) and Eitan Altman propose two new algorithms for finding stable structures in ordinal coalition potential games. The first one is enumerative and it performs on a graph. The second one is a modified Deferred Acceptance Algorithm using counter-proposals. It finds a many-to-one matching. The authors illustrate with the example of video caching from a content creator's servers to a service provider's servers.

This is applied to the association of mobiles to IEEE 802.11-based WLANs in populated areas where many mobile terminals are covered by several Access Points (APs) [32]. These mobiles have the possibility to associate to the AP with the strongest signal (best-RSSI association scheme). This can lead to poor performances and overloaded APs. Moreover, the well-known anomaly in the protocol at the MAC layer may also lead to very unpredictable performances and affect the system throughput due to the presence of heterogeneous data rate nodes and the shared nature of the 802.11 medium. In [61], the same authors solve the joint resource allocation and mobile user association after modeling it as a matching game with complementarities, peer effects and selfish players.

7.3.2. Normalized Nash Equilibria for power control with correlated constraints

When correlated constraints are introduced to a game (i.e. the set of actions of a player depends on the policies of other players) there may exist infinitely many Nash equilibria. Assume one wishes to select a particular one u. According to the Karush Kuhn Tucker theorem, there exist Lagrange mutipliers such that the best response when all players use their equilibrium policy is the same as that obtained by optimizing the corresponding Lagrangian of that player. The Lagrange multipliers can be interpreted as marginal costs such that if they are imposed on the player as some tax to pay then this induces the player to use Nash equilibrium. The following question arises: does there exist an equilibrium u for which the corresponding Lagrange multipliers are player independent. If the answer is positive then this would make in many cases the billing scalable and simple to implement. An equilibrium u for which the corresponding Lagrange multipliers are player independent is called a normalized Nash equilibrium (NNE). In [39], [50] and [24], Arnob Ghosh (Univ. of Pennsylvania), Laura Cottatellucci (Eurecom) and Eitan Altman provide new conditions for existence and uniqueness of NNE and apply this for power control games arising in cognitive radio [24] and in heterogeneous networks [39], [50].

7.3.3. Admission control to an infinite server queue

In [36], Eitan Altman studies in collaboration with Piotr Wiecek (Wrocław Univ. of Technology) and Arnob Ghosh (Univ. of Pennsylvania) a mean field approximation of the M/M/∞ queueing system. The problem they consider is quite different from standard games of congestion as they consider the case in which higher congestion results in smaller costs per user. This is motivated by a situation in which some TV show is broadcast so that the same cost is needed no matter how many users follow the show. Using a mean-field approximation, they show that this results in multiple equilibria of threshold type which is explicitly computed. The authors further derive the social optimal policy and compute the price of anarchy, and show that the mean-field approximation becomes tight as the workload increases, thus the results obtained for the mean-field model well approximate the discrete one.

7.3.4. Posting Time of Content over a Temporally-Ordered Shared Medium

In [17], Eitan Altman in collaboration with Nahum Shimkin (Technion) consider a game of timing between a random number of content creators, who compete for position and exposure time over a shared medium such as an on-line classified list. Contents (such as ads, messages, multimedia items or comments) are ordered according to their submission times, with more recent submissions displayed at the top (and better) positions. The instantaneous effectiveness of each ad depends on its current display position, as well as on a time-dependent exposure function common to all. Each content creator may choose the submission time of her content within a finite time interval, with the goal of maximizing the total exposure of this content. The authors formulate the problem as a non-cooperative game, analyze its symmetric equilibrium, characterize it in terms of a differential boundary value problem and devise a numerical scheme for its computation.

7.3.5. Routing Games

A central question in routing games has been to establish conditions for the uniqueness of the equilibrium, either in terms of network topology or in terms of costs. This question is well understood in two classes of routing games. The first is the non-atomic routing introduced by Wardrop on 1952 in the context of road traffic in which each player (car) is infinitesimally small; a single car has a negligible impact on the congestion. Each car wishes to minimize its expected delay. Under arbitrary topology, such games are known to have a convex potential and thus a unique equilibrium. The second framework is splitable atomic games: there are finitely many players, each controlling the route of a population of individuals (let them be cars in road traffic or packets in the communication networks). In [64], Eitan Altman and Corinne Touati (Inria Grenoble-Rhône-Alpes) study two other frameworks of routing games in which each of several players has an integer number of connections (which are population of packets) to route and where there is a constraint that a connection cannot be split. Through a particular game with a simple three link topology, they identify various novel and surprising properties of games within these frameworks. The authors show in particular that equilibria are non unique even in the potential game setting of Rosenthal with strictly convex link costs. They further show that non-symmetric equilibria arise in symmetric networks.

7.3.6. Resilience of Routing in Parallel Link Networks

Aniruddha Singhal, Corinne Touati (both from Inria Grenoble-Rhône-Alpes) in collaboration with Eitan Altman and Jie Li (Univ. of Tsukuba) revisit in [63], the resilience problem of routing traffic in a parallel link network model with a malicious player using a game theoretic framework. Consider that there are two players in the network: the first player wishes to split its traffic so as to minimize its average delay, which the second player, i.e., the malicious player, tries to maximize. The first player has a demand constraint on the total traffic it routes. The second player controls the link capacities: it can decrease by some amount the capacity of each link under a constraint on the sum of capacity degradation. The authors first show that the average delay function is convex both in traffic and in capacity degradation over the parallel links and thus does not have a saddle point. They identify best responses strategies of each player and compute both the max-min and the min-max values of the game. One is especially interested in the min-max strategy as it guarantees the best performance under worst possible link capacity degradation. It thus allows to obtain routing strategies that are resilient and robust. The authors compare the results of the min-max to those obtained under the max-min strategies. They provide stable algorithms for computing both max-min and min-max strategies as well as for best responses.

7.3.7. The Social Medium Selection Game

In [72], Fabrice Lebeau (ENS Lyon) Corinne Touati and Nof Abuzainab (Inria Grenoble-Rhône-Alpes) in collaboration with Eitan Altman, consider competition of content creators in routing their content through various media. The routing decisions may correspond to the selection of a social network (e.g. twitter versus facebook or linkedin) or of a group within a given social network. The utility for a player to send its content to some medium is given as the difference between the dissemination utility at this medium and some transmission cost. The authors model this game as a congestion game and compute the pure potential of the game. In contrast to the continuous case, they show that there may be various equilibria. The authors show that the potential is M-concave which allows them to characterize the equilibria and to propose an algorithm for computing it. They then give a learning mechanism which leads to an efficient algorithm to determine an equilibrium. The authors finally determine the asymptotic form of the equilibrium and discuss the implications on the social medium selection problem.

7.3.8. Activation Games in Online Dating Platforms

In [41], Eitan Altman in collaboration with Francesco De Pellegrini (CREATE-NET, Trento) and Huijuan Wang (Delft Univ. of Technology) describe a model for the activation level of users in online dating platforms (ODPs). Such popular systems are conceived in order to match individuals from two groups of potential mates. The business of such platforms pivots around the customers' expectancy to get in contact with their future dates: upon the payment of a fee to the platform owner, ODPs provide specific tools to improve reach and visibility. However, ODPs require a critical number of active users in order to sustain their operations (and their business). Customers of the platform trade off on the price for being more visible and attract mates' contacts. A user becomes inactive if he or she is not contacted by others for some time: being contacted by potential mates acts as an activation signal. The aim of the analysis is to propose a game theoretical framework to capture such a complex activation problem in strategic form. The authors unveil the structure of Nash equilibria and further derive a Stackelberg formulation. The latter is a hierarchical game where the platform owner aims at maximizing profits while preserving the ODP activity level above a critical epidemic threshold.

7.3.9. Epidemics in Networks

Stojan Trajanovski, Huijuan Wang, Piet Van Mieghem (all from Delft Univ. of Technology), in collaboration with Yezekael Hayel (UAPV) and Eitan Altman have pursued their work in the Congas European project concerning malware attacks modeled as SIS (for Susceptible-Infected-Susceptible) epidemics in networks. In [34], the authors consider decentralized optimal protection strategies when a virus is propagating over a network. they assume that each node in the network can fully protect itself from infection at a constant cost, or the node can use recovery software, once it is infected. They model the system using a game theoretic framework and find pure, mixed equilibria, and the Price of Anarchy (PoA) in several network topologies.

Further, they propose both a decentralized algorithm and an iterative procedure to compute a pure equilibrium in the general case of a multiple communities network. Finally, the authors evaluate the algorithms and give numerical illustrations of all results.

They then considered the game-formation problem while balancing multiple, possibly conflicting objectives like cost, high performance, security and resiliency to viruses. In [60], Stojan Trajanovski, Fernando Antonio Kuiper and Piet Van Mieghem (all from Delft Univ. of Technology) in collaboration with Yezekael Hayel (UAPV) and Eitan Altman use a game-formation approach to network design where each player (node), aims to collectively minimize the cost of installing links, of protecting against viruses, and of assuring connectivity. In the game, minimizing virus risk as well as connectivity costs results in sparse graphs. They show that the Nash Equilibria are trees that, according to the Price of Anarchy (PoA), are close to the global optimum, while the worst-case Nash Equilibrium and the global optimum may significantly differ for small infection rate and link installation cost. Moreover, the types of trees, in both the Nash Equilibria and the optimal solution, depend on the virus infection rate, which provides new insights into how viruses spread: for a high infection rate, the path graph is the worst- and the star graph is the best-case Nash Equilibrium. However, for small and intermediate infection rates, trees different from the path and star graphs may be optimal.

7.3.10. Retrial games

In [46] K. Avrachenkov in collaboration with E. Morozov and R. Nekrasova (both from Petrozavodsk State Univ., Russia) consider a single-server retrial system with one and several classes of customers. In the case of several classes, each class has its own orbit for retrying customers. The retrials from the orbits are generated with constant retrial rates. In the single class case, the objective is finding an optimal retrial rate. Whereas in the multi-class case, a game theoretic framework is used and equilibrium retrial rates are found. The performance criteria balance the number of retrials per retrying customer with the number of unhappy customers.

7.3.11. Cooperative Network Design

The Network Design problem has received increasing attention in recent years. Previous works have addressed this problem considering almost exclusively networks designed by selfish users, which can be consistently suboptimal. In [18] K. Avrachenkov, J. Elias (Univ. Paris Descartes, France), F. Martignon (Univ. Paris Sud, France), G. Neglia and L. Petrosyan (St. Petersburg State Univ.) address the network design issue using cooperative game theory, which permits to study ways to enforce and sustain cooperation among users. Both the Nash bargaining solution and the Shapley value are widely applicable concepts for solving these games. However, the Shapley value presents several drawbacks in this context. For this reason, they solve the cooperative network design game using the Nash bargaining solution (NBS) concept. More specifically, they extend the NBS approach to the case of multiple players and give an explicit expression for users' cost allocations. They further provide a distributed algorithm for computing the Nash bargaining solution. Then, they compare the NBS to the Shapley value and the Nash equilibrium solution in several network scenarios, including real ISP topologies, showing its advantages and appealing properties in terms of cost allocation to users and computation time to obtain the solution.

Numerical results demonstrate that the proposed Nash bargaining solution approach permits to allocate costs fairly to users in a reasonable computation time, thus representing a very effective framework for the design of efficient and stable networks.

7.4. Green Networking and Smart Grids

Participants: Sara Alouf, Eitan Altman, Alberto Benegiamo, Alain Jean-Marie, Giovanni Neglia.

7.4.1. Energy efficiency and management in wireless networks

In [35], Rodrigo A. Vaca Ramirez and John S. Thompson (Univ. of New England), in collaboration with Eitan altman and Victor Ramos Ramos (UAM - Univ. Autonoma Metropolitana Unidad Iztapalapa) consider a low complexity virtual Multiple-input Multiple-output (MIMO) coalition formation algorithm. The goal is to obtain improvements in energy efficiency by forming multi-antenna virtual arrays for information transmission in the uplink. Virtual arrays are formed by finding a stable match between single antenna devices such as mobile station (MS) and relay stations (RS) by using a game theoretic approach derived from the concept of the college admissions problem. They focus on enhancing the MS performance by forming virtual coalitions with the RSs. Thus, power savings are obtained through multi-antenna arrays by implementing the concepts of spatial diversity and spatial multiplexing for uplink transmission. They focus on optimizing the overall consumed power rather than the transmitted power of the network devices. Furthermore, it is shown analytically and by simulations that when overall consumed power is considered as an optimization metric, the energy efficiency of the single antennas devices is not always improved by forming a virtual MIMO array. Hence, single antenna devices may prefer to transmit on their own when channel conditions are favorable. In addition, the simulation results show that the proposed framework provides comparable energy savings and a lower implementation complexity when compared to a centralized exhaustive search approach that is coordinated from the Base Station.

Sara Alouf, Ioannis Dimitriou (now at Univ. Patras, Greece) and Alain Jean-Marie had worked on the modeling of wireless communication base stations with autonomous energy supply (solar, wind). They had proposed a versatile 5-dimensional Markov model of the device, and shown that the Quasi Birth-Death framework is adequate for solving the model. This work has been completed with a companion product-form model based on E. Gelenbe's modeling of energy networks with signals [48].

7.4.2. Stochastic Geometric Models for Green Networking

In [16], Eitan Altman in collaboration with Cengis Hasan, Manjesh Kumar Hanawal (IIT Mumbai), Shlomo Shamai (Technion), Jean-Marie Gorce (Inria project-team SOCRATE), Rachid El-Azouzi (UAPV) and Laurent Roullet (Alcatel Lucent Bell Labs) study both the uplink and downlink energy efficiency based on the assumption that base stations are distributed according to an independent stationary Poisson point process. This type of modeling allows to make use of the property that the spatial distribution of the base stations after thinning (switching-off) is still a Poisson process. This implies that the probability of the SINR can be kept unchanged when switching-off base stations provided that one scales up the transmission power of the remaining base stations. The authors then solve the problem of optimally selecting the switch-off probabilities so as to minimize the energy consumptions while keeping unchanged the SINR probability distribution. They then study the trade-off in the uplink performance involved in switching-off base stations. These include energy consumption, the coverage and capacity, and the impact on amount of radiation absorbed by the transmitting user.

7.4.3. Direct Load Control

Energy demand and production need to be constantly matched in the power grid. The traditional paradigm to continuously adapt the production to the demand is challenged by the increasing penetration of more variable and less predictable energy sources, like solar photovoltaics and wind power. An alternative approach is the so called direct control of some inherently flexible electric loads to shape the demand. Direct control of deferrable loads presents analogies with flow admission control in telecommunication networks: a request for network resources (bandwidth or energy) can be delayed on the basis of the current network status in order to guarantee some performance metrics. In [53] G. Neglia, in collaboration with G. Di Bella (Telecom Italia, Italy), L. Giarré and I. Tinnirello (Univ. of Palermo, Italy) go beyond such an analogy, showing that usual teletraffic tools can be effectively used to control energy loads. In particular they propose a family of control schemes which can be easily tuned to achieve the desired trade-off among resources usage, control overhead and privacy leakage.

7.4.4. Charge of Electric Vehicles

The massive introduction of Electric Vehicles (EVs) will make fleet managers spend a significant amount of money to buy electric energy. If energy price changes over time, accurate scheduling of recharging times may result in significant savings. In [29] C. Rottondi (IDSIA Dalle Molle Institute for Artificial Intelligence, Switzerland), G. Neglia and G. Verticale (Politecnico di Milano, Italy) evaluate the complexity of the optimal scheduling problem considering a scenario with a fleet manager having full knowledge of the customers' traveling needs at the beginning of the scheduling horizon. They prove that the problem has polynomial complexity and provide complexity lower and upper bounds. Moreover, they propose an online sub-optimal scheduling heuristic that schedules the EVs' recharge based on historical travelling data. They compare the performance of the optimal and sub-optimal methods to a benchmark online approach that does not rely on any prior knowledge of the customers' requests, in order to evaluate whether the additional complexity required by the proposed strategies is worth the achieved economic advantages. Numerical results show up to of 35% cost savings with respect to the benchmark approach.

7.5. Content-Oriented Systems

Participants: Sara Alouf, Eitan Altman, Konstantin Avrachenkov, Alain Jean-Marie, Philippe Nain, Giovanni Neglia.

7.5.1. Modeling modern DNS caches

Sara Alouf and Nicaise Choungmo Fofack (former PhD student at MAESTRO, currently at Ingima) have thoroughly revised their study of the modern behavior of DNS caches. In particular the closure properties of the class of distributions called *diagonal matrix-exponential* are fully derived, hence the analytic models presented in [78] to study tree of caches with general caching durations are extended to the case of polytrees [15].

7.5.2. Data placement and retrieval in distributed/peer-to-peer systems

In previous years, Alain Jean-Marie and collaborators from the Univ. Montpellier have defined a family of combinatorial designs that minimize the variance in the availability of replicated documents in unreliable infrastructures. Then with Jean-Claude Bermond (CNRS, with the Inria project-team COATI), Dorian Mazauric (now with Inria project-team ABS) and Joseph Yu (UFV Vancouver), it was shown that *well-balanced families* solve the problem, and such families were constructed for small numbers of replicas. This work is now published in [21]. During the internship of Mikhail Grigorev, several methods for generating at random good solutions have been investigated.

7.5.3. Fairness in caching systems

Data offloading from the cellular network to lowcost WiFi has been the subject of several research works in the last years. In-network caching has also been studied as an efficient means to further reduce cellular network traffic. In [49] M. El Chamie (Univ. of Washington, USA), C. Barakat (Inria project-team DIANA) and G. Neglia consider a scenario where mobile users can download popular contents (e.g., maps of a city, shopping information, social media, etc.) from WiFi-enabled caches deployed in an urban area. They study the optimal distribution of contents among the caches (i.e., what contents to put in each cache) to minimize users' access cost in the whole network, and argue that this optimal distribution does not necessarily provide geographic fairness, i.e., users at different locations can experience highly variable performance. In order to mitigate this problem, they propose two different cache coordination algorithms based on gossiping. These algorithms achieve geographic fairness while preserving the minimum access cost for end users.

In [43] K. Avrachenkov in collaboration with V.S. Borkar (IIT Mumbai, India) consider the task of scheduling a crawler to retrieve from several sites their ephemeral content. This is content, such as news or posts at social network groups, for which a user typically loses interest after some days or hours. Thus development of a timely crawling policy for ephemeral information sources is very important. The authors first formulate this problem as an optimal control problem with average reward. The reward can be measured in terms of the number of clicks or relevant search requests. The problem in its exact formulation suffers from the curse of dimensionality and quickly becomes intractable even with moderate number of information sources. Fortunately, this problem admits a Whittle index, a celebrated heuristics which leads to problem decomposition and to a very simple and efficient crawling policy. The Whittle index is derived, together with its theoretical justification.

7.6. Advances in Methodological Tools

Participants: Eitan Altman, Konstantin Avrachenkov, Ilaria Brunetti.

7.6.1. Control theory

In [19] K. Avrachenkov in collaboration with O. Habachi (UAPV) and A. Piunovskiy and Y. Zhang (both from Univ. of Liverpool, UK) investigate infinite horizon deterministic optimal control problems with both gradual and impulsive controls, where any finitely many impulses are allowed simultaneously. Both discounted and long run time average criteria are considered. They establish very general and at the same time natural conditions, under which the dynamic programming approach results in an optimal feedback policy. The established theoretical results are applied to the Internet congestion control, and by solving analytically and nontrivially the underlying optimal control problems, they obtain a simple threshold-based active queue management scheme, which takes into account the main parameters of the transmission control protocols, and improves the fairness among the connections in a given network.

7.6.2. Game theory

7.6.2.1. Evolutionary games

Standard Evolutionary Game framework is a useful tool to study large interacting systems and to understand the strategic behavior of individuals in such complex systems. Adding an individual state to model a local feature of each player in this context, allows one to study a wider range of problems in various application areas as networking, biology, etc. In [47], Ilaria Brunetti and Eitan Altman in collaboration with Yezekael Hayel (UAPV) introduce such an extension of evolutionary game framework and particularly, focus on the dynamical aspects of this system. Precisely, the authors study the coupled dynamics of the strategies and the individual states inside a population of interacting individuals. They consider here a two-strategies evolutionary game. They first obtain a system of combined dynamics and they show that the rest-points of this system are equilibria of the evolutionary game with individual state. Second, by assuming two different time scales between states and strategy dynamics, one can compute explicitly the equilibria. Then, by transforming the evolutionary game with individual states into a standard evolutionary game, the authors obtains an equilibrium which is equivalent, in terms of occupation measure, to the previous one. Finally, they show a generalization of the model. All the results are illustrated with numerical results.

7.6.2.2. Stochastic Games

Motivated by uncertainty in the value of the interest rate, in [62] K. Avrachenkov in collaboration with A. Varava (KTH, Sweden) study discounted zero-sum stochastic games with and arbitrary discount factor. Their general goal is to obtain a power series expansion of the value of the game with respect to the discount factor around its nominal value. They consider a specific but important class of stochastic games – completely mixed stochastic games. As an illustrative example they take a tax evasion model.

MARELLE Project-Team

6. New Results

6.1. IDE for Coq

Participants: Enrico Tassi, Alexander Faithfull [ITU Copenhagen], Jesper Bengtson [ITU Copenhagen], Carst Tankink.

User interfaces for interactive proof assistants should rely on the advanced software available in integrated development environments. we collaborated with researchers from Copenhagen to build an Eclipse-based environment for the Coq system. This exploits the quick compilation chain that was developed for Coq 8.5. This work has been published in [15].

6.2. ELPI, Fast, Embeddable, λ-Prolog Interpreter

Participants: Enrico Tassi, Cvetan Dunchev [University of Bologna], Ferruccio Guidi [University of Bologna], Claudio Sacerdoti Coen [University of Bologna].

We developed a new interpreter that runs consistently faster than the other available implementations of λ-prolog. The key insight is the identification of a fragment of the language, which we call reduction-free fragment, that occurs quite naturally and that admits constant time reduction and unification rules. In the long run, we hope that this will contribute to developing elaborators that support a more efficient and adaptable usage of interactive proof tools. This work is published in [14].

6.3. Verified Proofs of Higher-Order Masking

Participants: Gilles Barthe [IMDEA Software, Madrid], Sonia Belaïd [Thales Communication], François Dupressoir [IMDEA Software, Madrid], Pierre-Alain Fouque [Université de Rennes, IUF], Benjamin Grégoire, Pierre-Yves Strub [IMDEA Software, Madrid].

We study the problem of automatically verifying higher-order masking countermeasures. We propose a method based on program verification techniques, to check the independence of sets of intermediate variables from secrets. This new language-based technique makes it possible to implement several algorithms that reduce the number of sets of variables that need consideration. The tool also has the capability to to give useful information when proofs fail, for instance by discovering possible attacks. This is based on EasyCrypt. This work has been published in [8].

6.4. Relational Reasoning via Probabilistic Coupling

Participants: Gilles Barthe [IMDEA Software, Madrid], Thomas Espitau [ENS Cachan], Benjamin Grégoire, Justin Hsu [University of Pennsylvania], Léo Stefanesco [ENS Lyon], Pierre-Yves Strub [IMDEA Software, Madrid].

Probabilistic coupling is a powerful tool for analyzing pairs of probabilistic processes. While the mathematical definition of coupling looks rather complex and cumbersome to manipulate, we show that the relational program logic pRHL—the logic underlying the EasyCrypt cryptographic proof assistant—already internalizes a generalization of probabilistic coupling. With this insight, constructing couplings is no harder than constructing logical proofs. We demonstrate how to express and verify classic examples of couplings in pRHL, and we mechanically verify several couplings in EasyCrypt. This work is described in [9].

6.5. Automated Proofs of Pairing-Based Cryptography

Participants: Gilles Barthe [IMDEA Software, Madrid], Benjamin Grégoire, /benedikt Schmidt [IMDEA Software, Madrid].

We implement a new tool, called AutoG&P, which supports extremely compact, and often fully automated, proofs of cryptographic constructions based on (bilinear or multilinear) Diffie-Hellman assumptions. For instance, we provide a 100-line proof of Waters' Dual System Encryption (CRYPTO'09), and fully automatic proofs of Boneh-Boyen Identity-Based Encryption (CRYPTO'04). Finally, we provide an automated tool that generates independently verifiable EasyCrypt proofs from AutoG&P proofs. This work has been published in [10].

6.6. Improvements on CBC MAC formalized in EasyCrypt

Participants: Benjamin Grégoire, Cécile Baritel-Ruet, Pierre-Alain Fouque.

In a paper of 2003, J. Black and P. Rogaway propose variations of cipher block chaining message authentication codes for the efficient authentication of arbitrary length messages. We formalize their work in EasyCrypt, resulting in formal proofs for CBC-MAC, EMAC, ECBC, FCBC and the most efficient of these variations, XCBC.

This work required the development of new EasyCrypt theories. A small flaw in the original paper was found and a fix has been proposed. This work was partially funded by the Brutus ANR project.

6.7. Buchberger's algorithm and advanced formalization of multinomials

Participant: Laurent Théry.

We studied how the Mathematical Components library could improve the formalization of of algorithms based on multivariate polynomials. In particular, we re-used Pierre-Yves Strub library of multivariate polynomials and re-did the proofs of correctnes for Buchberger's algorithm. This new piece of formalized algorithm is now available at the following address https://github.com/thery/grobner.

6.8. Proofs that e and π and transcendental

Participants: Sophie Bernard, Laurence Rideau, Yves Bertot, Pierre-Yves Strub [IMDEA Software, Madrid].

In the previous year, we developed formally verified proofs that e and π are transcendental. This year we cleaned up these proofs to obtain a common lemma that applies in both cases with simple hypotheses. In parallel, P.-Y. Strub streamlined the library on multivariate polynomials which plays a significant role in the case of π. This work has been published in [11].

In the future, we will probably extend this work to more general proofs of transcendance.

6.9. Algorithms for Real Algebraic Geometry

Participant: Cyril Cohen.

We formalized an efficient algorithm to count roots of a polynomial satisfying polynomial inequalities. This work was presented at the Types workshop in May and the Workshop on Algebra, Geometry, and Proofs in Symbolic computation.

6.10. Nominal sets in Coq

Participants: Cyril Cohen, Nicolas Tabareau, Matthieu Sozeau, Gabriel Lewertoski.

Cyril Cohen collaborated with members of the team $\pi.r^2$ on the implementation of nominal sets in Coq.

6.11. Formal Description of Dynamic Logic

Participants: Yves Bertot, Cyril Cohen, Jean-Yves Franceschi.

We developed a formal description of the language of dynamic logic in the Coq system.

6.12. Cubical Type Theory

Participants: Cyril Cohen, Thierry Coquand, Simon Huber, Anders Mörtberg.

We participate to the development of a software prototype, `cubicaltt`, https://github.com/mortberg/cubicaltt, that is expected to support an extension of type theory suited for homotopy type theory.

6.13. Finite set and finite maps

Participant: Cyril Cohen.

We extend the Math-Components library with a module concerning finite sets (in potentially infinite types) and finite maps. This module will play a crucial role in other experiments, like the experiments on dynamic logic, nominal sets, and cubical sets.

6.14. Formalization of a Newton Series Representation of Polynomials

Participants: Boris Djalal, Cyril Cohen.

We formalize an algorithm to change the representation of a polynomial to a Newton power series. This provides a way to compute efficiently polynomials whose roots are the sums or products of roots of other polynomials, and hence provides a base component of efficient computation for algebraic numbers. In order to achieve this, we formalize a notion of truncated power series and develop an abstract theory of poles of fractions. This work is described in [13].

6.15. Formal description of catalan numbers

Participant: José Grimm.

Catalan number can be defined by a recurrence, or by explicit formulas using binomial numbers. An important property is $C_{n+1} = \sum_{k \leq n} C_k C_{n-k}$. The easiest way to prove this formula is to use Dyck paths.

A Dyck path of size $2n$ is a sequence l of integers $+1$ and -1 so that the sum s_k of the k first terms is ≥ 0 for $k \leq 2n$ and $s_{2n} = 0$. The relation between Dyck paths and Catalan numbers is easy to prove and then properties of Dyck paths are quite simple to state and verify.

The proofs have been done with the Math-Components library.

6.16. Latex to XML translator

Participant: José Grimm.

This year, we released version 2.15.4 of Tralics, our LaTeX to XML translator. Array handling has been redesigned: for instance, an array preamble of the form `{>{$}c<{$}}` is now correctly interpreted; there is a possibility to add an attribute pair to any table, row or cell; for math environments like “align”, one label and one tag per row is allowed. Tralics is also able to read an XML file, and there are some primitives for inserting the result (or part of it) into the XML code under construction.

MCTAO Project-Team

6. New Results

6.1. Optimal control for quantum systems and applications to MRI

Participants: Bernard Bonnard, Thierry Combot [Université de Bourgogne, IMB], Alain Jacquemard [Université de Bourgogne, IMB], Dominique Sugny [Université de Bourgogne, LIC].

Important results have been obtained in this aera that we detail next :

- A complete solution to the time minimal control of a chain of three spins with Ising coupling which is a toy example applicable to quantum computing [42], [3].
- Optimal control of an ensemble of spins systems with application to MRI : this work is performed in the framework of the ANR project Explosys , based on our previous results in the contrast problem in Nuclear Magnetic Imaging. In relation with the laboratory Creatis (Insa Lyon) and TUM (S. Glaser) the objective is to design robust pulses control, with respect to the relaxation parameters and the B0 and B1 inhomogneities. The computations are intricated from both the numerical point of view and exact computaions. From this second point a systematic study of the controlled Bloch equation has been initiated using exact computer algebraic method in relation with the Inria project-team POLSYS .

6.2. Controllability and Optimal control at Low Reynolds number

Participants: Piernicola Bettiol [Université de Bretagne Occidentale (Brest)], Bernard Bonnard, Laetitia Giraldi, Pierre Martinon [project-team COMMANDS], Jean-Baptiste Pomet, Jérémy Rouot.

This new aera is somehow connected to the recent recruitment of L. Giraldi (CR2) in the Mc Tao team. The problem under study is to design strokes for swimmers at low Reynolds numbers, e .g. the Copepod swimmer (an abundant variety od zooplankton) or the Purcell swimmer. The problem was studied from the point of view of geometric optimal control [17], [18] combining theoretical and numerical computations or controllability techniques [19].

6.3. Averaging in control and application to space mechanics

Participants: Bernard Bonnard, Jean-Baptiste Caillau, Helen-Clare Henninger, Jana Němcová [Instritute of Chemical Tech, Prague, CZ], Jean-Baptiste Pomet, Jeremy Rouot.

We have obtained results on the structure of the average system for the planar minimum time problem without perturbation in [4], and the “double average” that takes the lunar perturbation into account in [13]. This is also the topic of Helen Henninger’s PhD [1].

The structure of the problem where the consumption (i.e. the L^1 norm of the control) is minimized is studied in [5].

The book [16] is a general reference opus edited by members of the team.

MODEMIC Project-Team

7. New Results

7.1. Mathematical models for microbial ecology

7.1.1. Differential equations models

Participants: Jérôme Harmand, Claude Lobry, Alain Rapaport, Yessmine Daoud, Sonia Hassam, Zeyneb Khedim, Alejandro Maximiliano Rojas.

Anaerobic digestion refers to the transformation of biodegradable material by micro-organisms in absence of oxygen (it can be found in waste-water treatments or industrial fermentation, and occurs naturally in soils). It receives an increasing consideration due to recent technological advances, but also because it is a source of renewable energy (bio-gas, fuel...). The anaerobic digestion is a complex set of bio-processes, for which there is a strong expectation of tractable models. We have proposed and studied new mathematical models that takes into account the following features:

- Microbial food chains are present in anaerobic digestion where the different reaction steps can be seen as such: the waste products of the organisms at one trophic level (i.e. one reaction step) are consumed by organisms at the next trophic level (i.e. the next reaction step). In [54] we study a model of a two-tiered microbial 'food chain' with feedback inhibition, which was recently presented as a reduced and simplified version of the anaerobic digestion model ADM1 of the International Water Association (IWA) (cf. [61]). It is known that in the absence of maintenance (or decay) the microbial 'food chain' is stable. In [61], using a purely numerical approach and ADM1 consensus parameter values, it was shown that the model remains stable when decay terms are added. In [54] we prove that introducing decay in the model preserves stability whatever its parameters values are and for a wide range of kinetics.
- In the thesis by Sonia Hassam [13], we have proposed a new procedure to easily and systematically obtain a simple model useful for control purposes of any process for which an ADM1 is available. The simplified model has two major characteristics : its states keep their physical meaning and it remains nonlinear. The technique is based on the state association technique proposed in [26].
- Zeyneb Khedim (University of Tlemcen, Algeria) has began her PhD in 2014. She is working on the modeling and control of anaerobic digestors. In particular, she works on the reduction of models using the state association approach proposed by Sonia Hassam but for substrates highly loaded in nitrogen such as algae. She has published this year a survey with Sonia Hassam [36].
- Yessmine Daoud (ENIT-LAMSIN, Tunis, Tunisia) continues her work on the analysis of a model of the literature to optimize anaerobic processes [35]. She is preparing a journal paper which should be submitted during 2016.

Formerly, the team has studied chemostat models where the bacterial compartment is split into "planktonic" and "attached" bacteria (such as in flocculation or biofilms formation), under the hypothesis that attachment and detachment are fast phenomena. Under certain mixing conditions, this last condition is no longer satisfied. We have studied on the non-reduced model the competition between a species that presents growth inhibition in planktonic form with a species that does not attach. This consideration leads to multiple positive equilibria but surprisingly it can also conduct to limit cycles [53] (paper under revision for Applied Math. Model.).

Spatial heterogeneity is often observed in non perfectly mixed bioprocesses or in populations in natural environments. The representation o spatial heterogeneity in population models with patches or interconnected models, rather than p.d.e., is one of the specialties of the team, that allows us to characterize non intuitive effects of spatialization :

- The very basic Rosenzweig-MacArthur model is subject to the "atto-fox" problem [2] when considered for homogeneous populations. Is it still true in case of heterogeneous populations? The idea is: the resource population being not small at the same time in different places is it possible that, thanks to dispersal, it will not disappear? One possible idealization of heterogeneous populations is to use reaction-diffusion equations. We do not take this direction for two reasons
- (i) Due to the presence of a limit cycle in the homogeneous system mathematics of such reaction diffusion are difficult.
- (ii) Idealization through reaction-diffusion is not the best one; patch-systems (or lattice differential equations in mathematical terms) are better in many cases.

Our ultimate objective is to provide mathematical results for systems with a large number of patches but, as a first step, in the paper [27] we consider two patches. It is proved that for some migration rates, stable periodic solutions avoiding "atto-fox" exist.

- The standard model for the dynamics of a fragmented density-dependent population is built from several local logistic models coupled by migrations.

First introduced in the 1970s and used in innumerable articles, this standard model applied to a two-patch situation has never been completely analyzed. The motivation for studying this problem came out from discussions at the Bernoulli semester organized in 2014 and 2015 by the team at the EPFL (see the 2014 activity report and Section 8.3.3.1). It addresses very fundamental issues in theoretical ecology. In the paper [15] written in collaboration with R. Arditi (U. Fribourg) an T. Sari (IRSTEA Montpellier), we complete this analysis and we delineate the conditions under which fragmentation associated to dispersal is either beneficial or detrimental to total population abundance. Therefore, this is a contribution to the SLOSS question. Importantly, we also show that, depending on the underlying mechanism, there is no unique way to generalize the logistic model to a patchy situation. In many cases, the standard model is not the correct generalization. We analyze several alternative models and compare their predictions. Finally, we emphasize the shortcomings of the logistic model when written in the r-K parameterization and we explain why Verhulst's original polynomial expression is to be preferred.

- We have carried on our former work on the role of particular interconnections patterns on the global stability of chemostat model with inhibition. While we focused formerly on the conditions for which a spatial structure ensures the global stability when the chemostat model is bi-stable in homogeneous environment, we have shown that at the opposite a spatial structure can make unstable the dynamics of the chemostat model with inhibition when it is stable in a homogeneous environment [30].
- In collaboration with Géosciences Rennes (Jean-Raynald de Dreuzy, Tristan Babey) and in the scope of the co-supervision of the PhD of Alejandro Rojas (also in the collaboration within the associated team with Chile), we have carried on the complete equivalence between several models used in Geosciences to characterize soil fractures : MINC (Multiple INteracting Continua), MRMT (Multi-Rate Mass Transfer) and SINC (Structured INteracting Continua). We have shown that the irreducibility of the network graph is not sufficient to obtain equivalence : a controllability assumption has also to be fulfilled [42]. Moreover, this kind of models has been used to fit experimental data of reconstituted soils at Inra Grignon and has shown the role of convection in the acquisition of pesticides by micro-organisms [46] (paper in preparation). This work will be continued in the framework of the new ANR project Soil$\mu 3D$ (see Section 8.2.1).

In resources/consumers models, heterogeneity can be also due to time varying inputs of resources (e.g. light in micro-algae populations). While, most of the literature studies periodic inputs, we have begun investigations of more general time varying inputs in chemostat like models, having in mind to characterize “pull-back attractors” (rather than forward attractors) [43].

7.1.2. Stochastic and hybrid discrete-continuous dynamical models

Participants: Bertrand Cloez, Claude Lobry.

7.1.3. Approximation of quasi-stationary distributions

The study of the long-time behavior of a stochastic process is one of the main questions of interest for modeling. In a standard Markov setting, this leads to the study of the convergence towards the invariant distribution. However, in many applications such as population dynamics for instance, the stochastic dynamics is killed in a finite (random) time so that the standard asymptotic regime is trivial. In this case, it can be interesting to focus on the behavior of the process conditionally to its non-extinction before a given time t Under appropriate assumptions, one can exhibit a convergence of this conditional distribution towards a law called *Quasi-Stationary Distribution*. Properties of this law is then fundamental. In [21], we study an algorithm to approximate this distribution and we provide proof of convergence as well as precise rates for convergence. This one is based on a reinforced random walk.

7.1.3.1. Lotka Volterra in fluctuating environment

In the paper [49], we consider two dimensional Lotka-Volterra systems in a fluctuating environment. Relying on recent results on stochastic persistence and piecewise deterministic Markov processes, we show that random switching between two environments that are both favorable to the same species can lead to the extinction of this species or coexistence of the two competing species. This work has been accepted in Journal of applied probabilities, provided major revisions. We submitted a new version with the new title: Lotka Volterra with randomly fluctuating environments or "how switching between beneficial environments can make survival harder".

7.2. Analysis and supervision of bioprocesses

7.2.1. Models development and identification

Participants: Yessmine Daoud, Jérôme Harmand, Nesrine Kalboussi, Guilherme Pimentel, Alain Rapaport.

Membrane bioreactors combine a filtration process (with a membrane) and a suspended growth rate bioreactor. This recent technology present many advantages compared to conventional ones, but is more sophisticated and requires refined control because of possible problems related to the risk of membrane fouling. After the PhD by Amine Charfi defensed in 2014 we continue to work on the modeling and control of membrane bioreactors.

- Within this framework, new results have been obtained and a new model including several fouling mechanisms has been proposed [22].
- In the scope of the PhD of Guilherme Pimentel (defensed in February 2015, [11]), we have proposed a simple three time scales model in view of the control of the cake formation [14], [33]. This model has been validated on real data from a pilot plant at Univ. Mons (Belgium).
- The PhD thesis by Nesrine Kalboussi (ENIT-LAMSIN, Tunis, Tunisia) has just begun. It is dedicated to the early detection and control of membrane fouling. At present time, Nesrine is working on the bibliography about modeling and control of membrane bioreactors.

In many bioprocesses models, the loss of nutrient used for the maintenance of bacteria is neglected compared to the important nutrient supply. In poor environment, such as natural one in the oceans, this is no longer verified. In our collaboration with the LOMIC lab (Banyuls-sur-Mer), we have shown that the consideration of a maintenance term in the chemostat model allows to fit the data observed in experimental chemostats, and moreover that the level of maintenance is correlated to the activities of bacteria under the presence of light [23]. This gives a possible explanation of the variable yield observed in the bacterial compartment of marine ecosystems.

7.2.2. Synthesis of control laws

Participants: Térence Bayen, Walid Bouhafs, Amel Ghouali, Jérôme Harmand, Claude Lobry, Guilherme Pimentel, Alain Rapaport, Victor Riqueleme.

We investigate two kinds of bioprocesses to be controlled, arising in industrial biotechnology (digesters, wastewater purification...) or in the bioremediation of natural environments (lakes, landfill...).

7.2.2.1. Industrial biotechnology

In the framework of the PhD of Guilherme Pimentel [11] and the pilot plant at Mons, a nonlinear predictive law based on the model exposed in 7.2.1 has been tested and validated for piloting the process[34].

Control of biological reactors are still of great interest, notably but not only with respect to anaerobic digestors that can be destabilized due to the accumulation of intermediate metabolites that can inhibate the growth of some bacteria.

- Amel Ghouali (Cotutelle Univ. Montpellier and Univ. of Tlemcen, Algeria) who has defensed her PhD in December has developed an optimal control strategy to optimize the production of biogas over a given period of time [12]. In particular, she has solved an original optimal control problem using the maximum principle of Pontryagin [25].
- Within the scope of the PhD thesis by Walid Bouhafs (ENIT-LAMSIN, Tunis, Tunisia), we have proposed a new optimal control strategy for systems in which two specific substrates are degraded by two different bacterial consortia, one being limited by the oxygen while the other is inhibited. Walid will defense his PhD in next February.

The minimal time criterion is of particular interest in biotechnology, as it leads to time-independent feedback controllers.

- The paper [19] is devoted to the study of the minimal time problem of a fed-batch reactor, under the presence of a saturation point on the singular locus (this typically occurs whenever the growth rate function is of Haldane type and when typically the maximum input flow rate is not high enough to maintain the substrate concentration constant). This brings non-intuitive issues for the optimal synthesis (existence of switching curve and point of prior saturation).
- In the work [47], we study the minimal time control to drive a chemostat model to a target point. Such a problem finds application typically in the case where the input substrate concentration changes yielding in a new steady state. Converging fast towards this new reference point is much desired in practice. One essential feature of the present work is that the system takes into account a recirculation of biomass (as it is more and more often the case in modern biotechnology). We depict an optimal synthesis and provide an optimal feedback control by using the Pontryagin Maximum Principle and geometric control theory for both Monod and Haldane kinetics.

7.2.2.2. Bioremediation of natural environments

In the scope of the associated team with Chile (see 8.3.1.1) and the co-supervision of the PhD of Victor Riquelme, we have carried on the study of optimal syntheses for the minimal time treatment of natural water reservoirs (such as lakes) [41]. We have proved that the minimal time strategy consists in a most-rapid approach to homogeneous concentrations, even though the optimal control problem is non convex. Moreover, we have shown that a large diffusion increases the treatment time when the resource is everywhere highly polluted, while it can at the opposite be beneficial when only part of the resource is polluted (paper under revision for SIAM J. Cont. & Optim.). This feature should serve the practitioners in the choice of pumps positioning in a originally clean water resource that is suddenly affected by a local pollution. Moreover, we have shown, in collaboration with A. Rousseau (EPI LEMON), how these analytic feedback laws obtained on a over-simplified representation of the spatial heterogeneity behave quite satisfactorily when simulated [17]. This year we have started to study to problem of treating two different pollutants, with a anaerobic/aerobic process in series.
Also in the scope of the associated team with Chile, we have characterized the optimal strategy to treat as fast as possible a landfill with the recirculation as a manipulated variable [40], [29], based on a model that we have proposed last year. In presence of singular arcs that are non-admissible (in the sense that the upper bound of the recirculation pump does not allow to stay on the singular arc), we have shown that a kind of *anticipation* law is necessary before operating optimally the switching. This analyses reveals several sub-domains for which the optimal policy requires different kind of measurements. Knowing in which sub-domain the initial stage

of landfill could be inform then the practitioners about which concentration (leachate or solubilized or both) should be ideally measured. This primarily work has led to the co-development of a software mock-up with Chilean partner (see 6.1), in order the study the consideration of spatial heterogeneity in landfill, with the approach exposed in 7.1.1 .

This year, again in the scope of the associated team with Chile and Inria Chile, we have begun a new investigation on modeling and control strategy for the regulation of a lagoon that communicates temporarily with the sea and whose water is exploited by pumping.

7.2.2.3. Theoretical development

The time crisis is an interesting criterion that measures the time spent by a system in a “bad” zone or in “danger”. Typically, when a desired species is under a given (low) threshold, one can consider that this defines a crisis domain. For controlled system, the minimal time crisis has already been proposed in the literature [56]. Nevertheless, only sufficient conditions (i.e. characterization of the solutions of the associated Hamilton-Bellman-Jacobi equation) have been given, and no necessary conditions have been yet proposed, due to the lack of continuity of the integrand cost. We have proposed a regularization of this problem by a family of optimal control problems for which the usual necessary conditions can be derived, and studied the convergence [20]. Practically, this allows to use classical software, such as Bocop, to approximate the optimal solutions. In the internship of C. Romero (U. Chile), this technique has been successfully applied on the Lotka-Volterra model with a control on the predator, and a threshold on the prey.

MORPHEME Project-Team

6. New Results

6.1. Exact continuous penalties for ℓ_2-ℓ_0 minimization

Participants: Emmanuel Soubies, Laure Blanc-Féraud, Gilles Aubert.

We consider the following ℓ_0-regularized least squares problem

$$\widehat{\mathrm{x}} \in \arg\min_{\mathrm{x}\in\mathrm{R}^N} G_{\ell_0}(x) := \frac{1}{2}\|A\mathrm{x}-d\|^2 + \lambda\|\mathrm{x}\|_0\,, \tag{2}$$

where $A \in \mathrm{R}^{M\times N}$, $d \in \mathrm{R}^M$ represents the data and $\lambda > 0$ is an hyperparameter characterizing the trade-off between data fidelity and sparsity. This problem finds a wide range of applications in signal/image processing, learning and coding areas among many others. We proposed a unified framework for exact continuous penalties approximating the ℓ_0-norm. In other words, we are concerned by the design of a class of continuous relaxations of G_{ℓ_0}, preserving all its global minimizers, and for which any local minimal point is also one of the initial functional. Hence, we highlight five *necessary and sufficient* conditions on the continuous penalty approximating the ℓ_0-norm ensuring that the minimizers of the underlying continuous relaxation of G_{ℓ_0} are consistent with those of G_{ℓ_0}. However, some local minimizer of the relaxed functional are not minimizer of G_{ℓ_0} which is an interesting point for such highly non-convex functional. This work offers a new way to compare penalties approximating the ℓ_0-norm. Finally, it is worth noting that the CEL0 penalty [1], [14], [17] is the inferior limit of the obtained class of penalties and seems to be the best choice to do in order to obtained an equivalent continuous reformulation of (1).

6.2. Application of the Continuous Exact ℓ_0 relaxation to Channel and DOA sparse estimation problems

Participants: Emmanuel Soubies, Laure Blanc-Féraud.

This work is made in collaboration with Adilson Chinatto, Cynthia Junqueira, João M. T. Romano (University of Campinas, Brazil) and Pascal Larzabal, Jean-Pierre Barbot (ENS Cachan, SATIE Lab).

This work is devoted to two classical sparse problems in array processing: Channel estimation and DOA (Direction Of Arrivals) estimation. We show how our results on ℓ_0 optimization [1], [14], [17] can be used, at the same computational cost, in order to obtain improvement in comparison with ℓ_1 optimization (usually used) for sparse estimation. Moreover, for the DOA case, we show that our analysis conducted in the Single Measurement Vector (SMV) case [1] can be generalized to the Multiple Measurement Vectors (MMV) case. In that case, the variable x is not a vector of $\mathbb{R}^N$ but a matrix of $\mathbb{R}^{N\times K}$ where N is the signal length and K the number of snapshots. Hence, one wants to apply sparsity to the rows of x, i.e. x must have a small number of nonzero rows, instead of applying the sparsity on all the components of x. This results in a row-structured sparsity penalty which is modelled using a mixed ℓ_2-ℓ_0 norm.

Finally, numerical experiments demonstrate the efficiency of the proposed approach compared to classical methods as ℓ_1 relaxation, Iterative Hard Thresholding or MUSIC algorithms and that it can reach the Cramer Rao Bound in some cases [4].

6.3. From TIRF microscope calibration to 3D biological reconstructions

Participants: Emmanuel Soubies, Laure Blanc-Féraud, Sébastien Schaub, Gilles Aubert.

This work is made in collaboration with Agata Radwanska, Ellen Van Obberghen-Schilling (iBV).

Total Internal Reflection Fluorescence microscopy (TIRF) is a method of choice to visualize membrane-substrate interactions. The principle of this device relies on the total internal reflection phenomenon generating an evanescent wave capable of producing a selective excitation of the dye molecules within a single layer of 100 to 500nm. The fast decay of the evanescent wave varies with respect to the incident angle of the light beam. Hence, intensity variations on TIRF images, occurring when changing the incident angle, are, in part, due to the axial positions of the observed structures. While a direct interpretation of Multi-Angle TIRF (MA-TIRF) images in terms of axial structure positions is not an easy task, reconstruction algorithms can be dedicated to compute a quantitative depth map with high axial resolution. However, the success of such reconstruction methods strongly depends on the system calibration.

We have proposed a pipeline for MA-TIRF calibration. Considering back focal plane (BFP) images of several solutions differing by their refractive indices, we validate the theoretical relation linking the tension applied to the galvanometric mirror (which controls the laser beam orientation) and the incident angle of the beam on the specimen. Then it is crucial to verify if the simple exponential decaying model of the evanescent wave is sufficient to describe our setup. To this end we propose to build a phantom sample (for which the geometry is known) using a large lens placed into a homogeneous fluorescent solution (Fig. 1 top). Based on a least square estimation, we showed a good agreement between the estimated slope of the lens (we assume the lens to be linear near the border) and the expected one up to 400nm depth (Fig. 1 bottom-left). To complete the validation procedure, we use a sample for which the structures of interest are labeled using two different fluorescent proteins sensitive to different wavelengths and emitting respectively green and red fluorescence. Then, using standard variational approaches, we obtain a co-localization of the reconstructed structures with a precision around 30-40nm (Fig. 1 bottom-middle) over at least 170nm depth showing the precision of the method. Finally, once this calibration step is achieved, we perform color-coded depth representation of 3D biological structures living in the vicinity of the cell membrane (Fig. 1 bottom-right). All these experiments have been made on an experimental TIRF system develloped at iBV lab in Valrose.

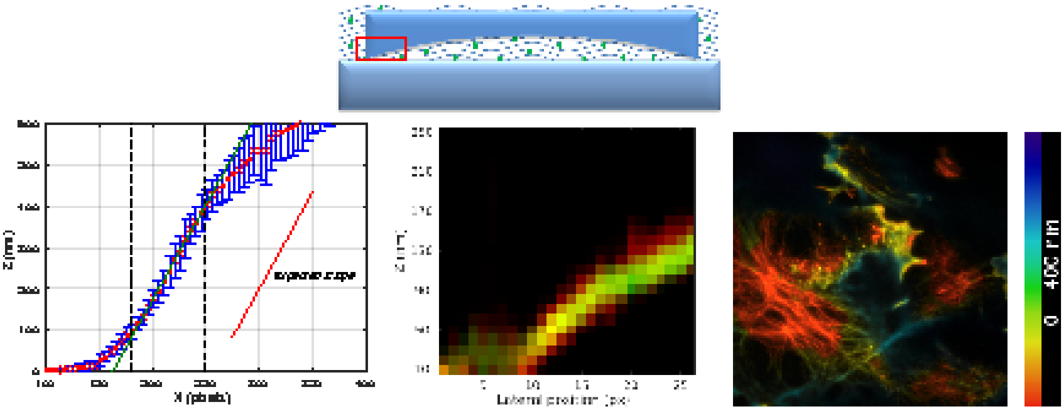


Figure 1. Top: Phantom sample constructed from a large lens and an homogeneous fluorescent solution. The red rectangle represents the observed region through the MA-TIRF setup. Bottom: estimated (green) and expected (red) slope of the lens within the red zone of the top figure (left), results of the co-localization experiment along a XY-line (middle) and a color-coded depth representation of a 3D biological reconstruction (right).

6.4. Phase estimation in Differential Interference Contrast (DIC) microscopy

Participants: Lola-Xiomara Bautista Rozo, Laure Blanc-Féraud.

We present a gradient-based optimization method for the estimation of a specimen phase function from polychromatic DIC images. The method minimizes the sum of a nonlinear least-squares discrepancy measure and a smooth approximation of the total variation. A new formulation of the gradient and a recent updating rule for the choice of the step size are both exploited to reduce computational time. Numerical simulations on two computer-generated objects show significant improvements, both in efficiency and accuracy, with respect to a more standard choice of the step size.

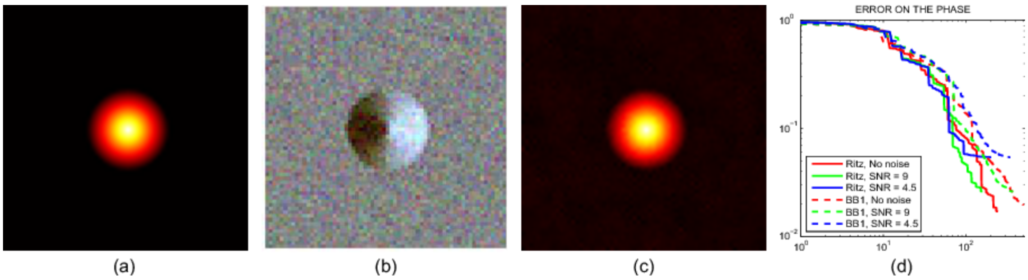


Figure 2. Data and results for the cone object. From left to right: true object, noisy DIC color image taken at angle $\tau_0 = 0°$ *and SNR = 4.5, reconstructed phase and the relative error versus the number of iterations.*

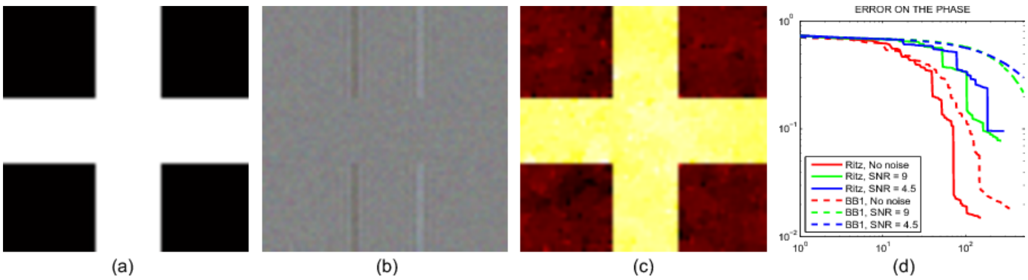


Figure 3. Data and results for the cross object. From left to right: true object, noisy DIC color image taken at angle $\tau_0 = 0°$ *and SNR = 4.5, reconstructed phase with and the relative error versus the number of iterations.*

6.5. Spatio-temporal registration of 3D microscopy image sequences of Arabidopsis floral meristems

Participants: Gaël Michelin, Grégoire Malandain.

This work is made in collaboration with Léo Guignard and Christophe Godin (Virtual Plants), within the Morphogenetics Inria Project Lab.

The shoot apical meristem (SAM) is at the origin of all the plant above-ground organs (including stems, leaves and flowers) and is a biological object of interest for the understanding of plant morphogenesis. The quantification of tissue growth at a cellular level requires the analysis of 3D microscopic image sequences of developing meristems. To address inter-individual variability, it is also required to compare individuals. This obviously implies the ability to process inter-individual registration, i.e. to compute spatial and temporal correspondences between sequences from different meristems.

In [8], we propose a spatial registration method dedicated to microscopy floral meristem (FM) images, based on the registration of both the outer and the inner surfaces of the L1 layer (the epidermal cell layer). A given meristem (one timepoint) can be compared to a sequence (several timepoints) of an other meristem: the goodness-of-fit criterion allows to identify the best corresponding time-point in this sequence of an other individual, achieving the temporal registration (see figure 4). Since the morphological deformations remain important between successive images of a sequence, images interpolation between time-points is also performed in order to refine the sequence temporal resolution and thus to ensure a precise temporal registration.

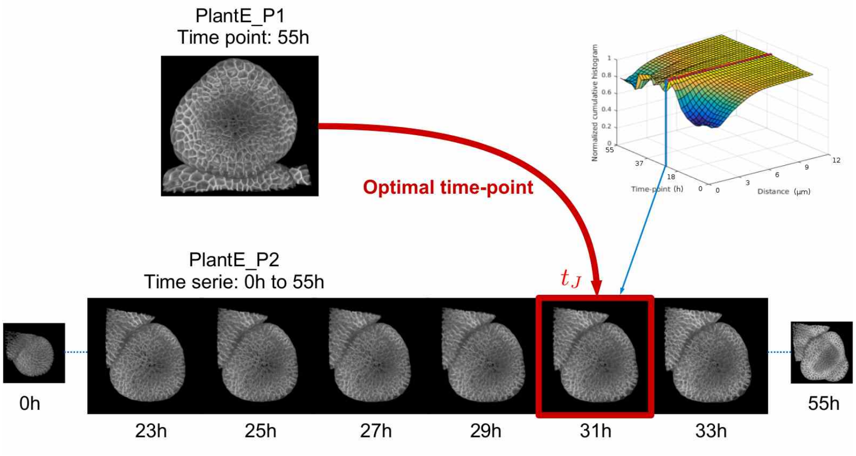


Figure 4. Inter-individual temporal registration result with 3D views of the registered meristem and the interpolated movie at several time-points.

6.6. Epidermal cell layer thickness variability in Arabidopsis floral meristems

Participants: Gaël Michelin, Grégoire Malandain.

This work is made in collaboration with Yassin Refahi (Sainsbury Lab., University of Cambridge) and Jan Traas (ENS Lyon), within the Morphogenetics Inria Project Lab.

Flowers from the same species display a great robustness in their global shape and their developing stage can be theoretically identifiable to their size. The cells in epidermal (L1) and sub-epidermal (L2) layers of the floral meristem divide anticlinally, i.e. in a sideway fashion that ensures that L1 and L2 remain distinct. Thus a goodness-of-fit criterion on L1 and L2 layers is considered as an adequate registration quality measure in the inter-individual spatio-temporal registration framework developed in [8].

The aim of the present work is to investigate the variability of L1 layer thickness over development stages of an individual and between individuals. The study results may impact the way we process the inter-individual spatial registration. Therefore we measured the thickness distribution (histogram) of the L1 cells and we plotted the distribution of cells thickness (see figure 5) on images provided from three distinct floral meristems at acquisition time-points. Our results tend towards showing that L1 thickness increases over time non-uniformly, with a higher L1 thickness on sepals for advanced developing stages. We also observed an inter-individual thickness variability of about 15% for developing floral meristems at close developing stages. Future investigations will consist in taking a larger set of data to assess our first observations, in providing a biological interpretation of these observations and in using this knowledge to propose a refined spatial registration method.

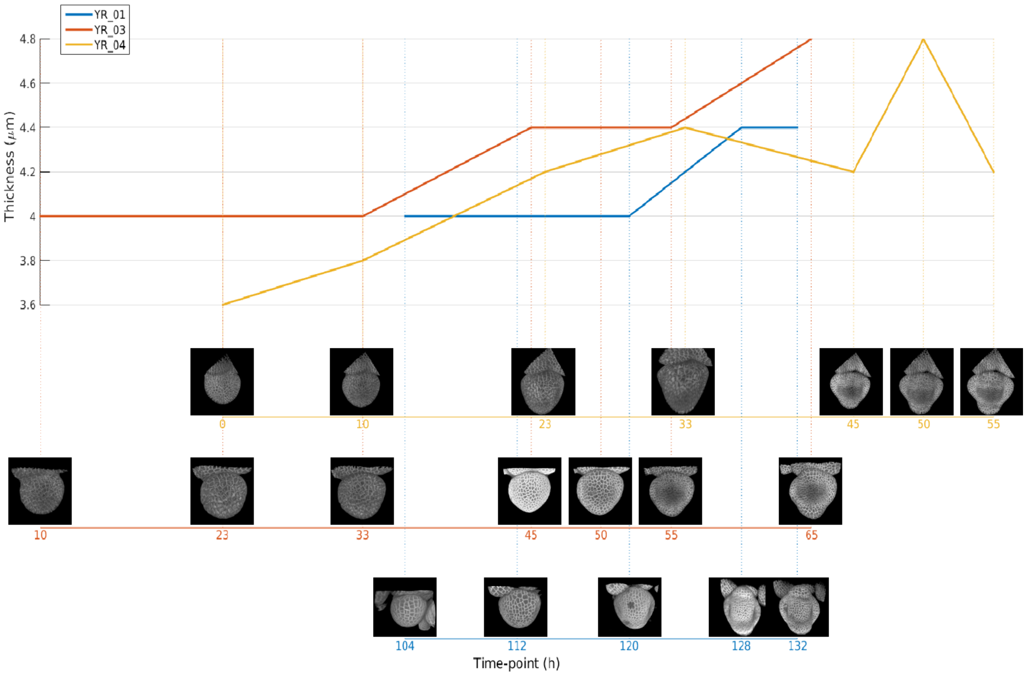


Figure 5. Epidermal cell layer thickness distribution over floral meristem development. Each row relates the measures at different developing times of a floral meristem.

6.7. Statistical Characterization, Modelling and Classification of Morphological Changes in imp Mutant Drosophila Gamma Neurons

Participants: Agustina Razetti, Caroline Medioni, Florence Besse, Xavier Descombes.

In Drosophila brain, gamma neurons in the mushroom body are involved in higher functions such as olfactory learning and memory. During metamorphosis, they undergo remodelling after which they adopt their adult shape. Some mutations alter remodelling and therefore neuronal final morphology, causing behavioural dysfunctions. The RNA binding protein Imp, for example, was shown to control this remodelling process at least partly by regulating profilin expression. This work aims at precisely characterizing the morphological changes observed upon imp knockdown in order to further understand the role of this protein. We developed a methodological framework that consists in the selection of relevant morphological features (axon length and shape and branch length distribution and density), their modelling and parameter estimation. We thus perform a statistical comparison and a likelihood analysis to quantify similarities and differences between wild type and mutated neurons. The data was a set of 3D images showing a single neuron taken with a confocal microscope and provided by F. Besse group, IBV. The workflow from raw data to the likelihood analysis is summarized on figure6 . We show that imp mutant neurons can be classified into two phenotypic groups (called Imp L and Imp Sh) that differ in several morphological aspects. We also demonstrate that, although Imp L and wild-type neurons show similarities, branch length distribution is discriminant between these populations. Finally, we study biological samples in which Profilin was reintroduced in imp mutant neurons, and show that defects in main axon and branch lengths are partially suppressed.

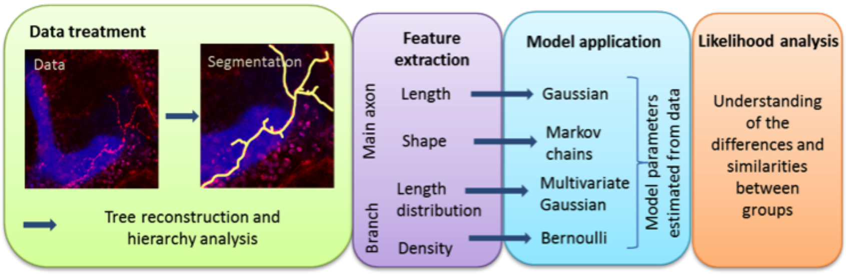


Figure 6. Summary of the workflow from raw data to the likelihood analysis.

6.8. Genome-wide search for factors that control the assembly of RNA granules

Participants: Wei Shen, Nicolas Cedilnik, Florence Besse, Xavier Descombes.

This work has been done in collaboration with Fabienne De Graeve from iBV

In vivo, mRNAs are packaged together with regulatory proteins into ribonucleoprotein particles (RNP) that control their fate and undergo extensive remodeling in response to developmental cues or environmental stresses. Cytoplasmic RNPs of different sizes, composition and regulatory properties have been described, including large macromolecular complexes such as P-bodies, stress granules, or germ cell granules. We aim at studying the different RNA granules distribution within the cytoplasm depending on genomic factors.

Before considering a spatial statistics analysis of the granules, it is necessary to detect them on confocal microscopy images of the cells. Therefore, we have studied a first pipeline for detecting these granules in confocal microscopy images. We have marked cells wih DAPI for detecting nuclei. These nuclei are then classified into "dead" or "alive" by a support vector machine (SVM) using intensity and shape criteria. In the second step we consider GFP marked images to segment the cytoplasm and detect the granules within the cytoplasm. The cytoplasm segmentation is performed using an active contour whereas the granule detection is based on a marked point process model optimized by the multiple births and cut algorithm.

The full pipeline has been validated on a few samples from a pilot study. The next step will consists of a validation on the full study before considering a genome-wide screening.

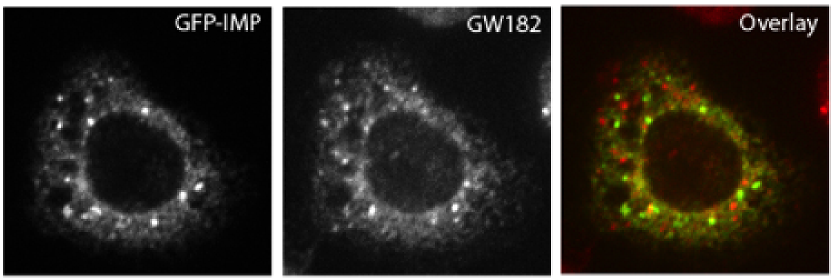


Figure 7. GFP-IMP particles are distinct from P-bodies. S2R+ cells expressing GFP-IMP fusions (left: green in the overlay) and stained with α-GW182 antibodies (middle: red in the overlay). GW182 is a well-described marker of P-bodies. Experiment performed in F. Besse lab at iBV (unpublished).

6.9. Cells detection using segmentation competition

Participants: Sen Wang, Emmanuel Soubies, Xavier Descombes.

Marked point processes have proved to be very efficient for segmenting a collection of objects from digital images. The multiple birth and death algorithm provides an optimization framework that allows reasonable time computation. This algorithm has to be embedded in a simluated annealing framework which involves parameters tuning (initial temperature and cooling scheme). This tedious task can be overcame considering a graph cut algorithm instead of the probabilistic death step. The algorithm then consists in successively adding new random objects in the configuration and selecting the most relevant using the graph cut algorithm. In the graph construction a node is associated to each object. In the orignal algorithm proposed by [21] the regularity condition imposed by the graph cut prevents to consider attractive interactions such as clustering or alignment constraints, which restricts the model to repulsive properties such as non overlap between objects. To overcome this restriction we have investigated new graph constructions by considering nodes defined by clusters of interacting objects. Different strategies have been compared to avoid being tracked in local minima defined by clusters while minimizing the number of required iterations. We have applied this new algorithm on different bioimagery problems such as axon extraction or cells detection (see figure 8).

6.10. Vesicles trajectory detection and analysis

Participant: Xavier Descombes.

This work has been done in collaboration with Maximilian Furthauer and Thomas Juan from iBV.

In many species, the left right asymetry of organs location is initiated in a ciliated cavity called Kupffer's vesicle in zebrafish. The cills beating induce a non symetrical flow in the cavity that can be studied by following the trajectory of exovesicles in the Kupffer's vesicle. The goal of this project is to automatically track these exovesicles and to perform a statistical analysis of theses trajectories in different conditions.

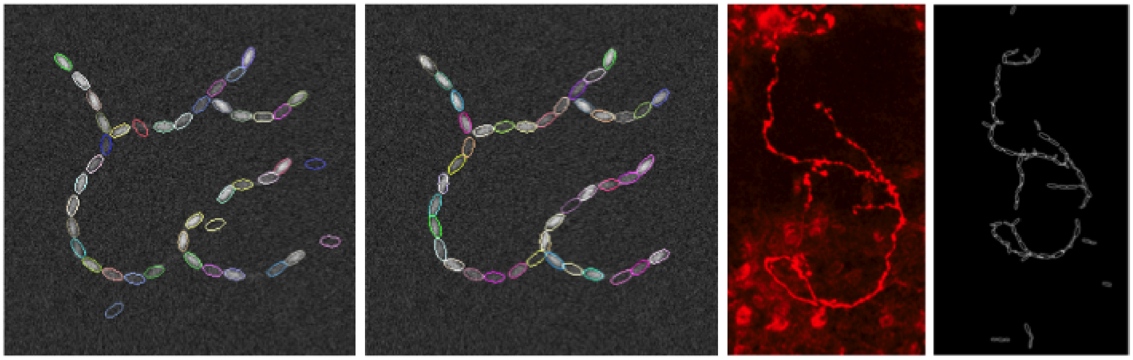


Figure 8. Axon detection on a synthetic image without (left) and with (middle left) attractive interactions and on real image (middle right) with attractive interactions (right).

We consider 2D time sequences of images. To extract the vesicles from the time sequence we first remove the background by substracting a local time mean. We then detect the cell border using an active contour computed on the spatial derivative of the images. The vesicles are then simply detected using a threshold followed by a morphological opening to remove the noise. The trajectory are finally obtained using a morphological closing in time. We aim at statistically comparing populations. In order to aggregate trajectories from several samples, we project the datasets into the same space using a continuous transformation of each cell into a reference disk. We thus project all the obtained trajectories from a given population into this disk. We compute the speed vector on each time point of each detected trajectory. To obtain a dense representation of the norm and the orientation of the vector speed in the reference disk, we extrapolate the obtained vector speed to a regular lattice with a Gaussian Markov random field. Finally, we obtain two spatial maps of respectively the norm and the orientation of the speed 9 .

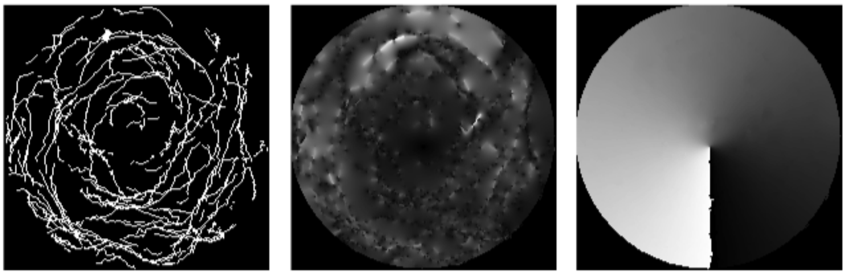


Figure 9. Exovesicles trajectories detected on the control population (left), the speed norm map (middle) and the angle map (left) showing an anticlockwise movement (black=0, white=2π).

6.11. Extraction and Analysis of the Vascular Network to Classify and Grade Kidney Tumors in Histological Imaging

Participants: Alexis Zubiolo, Eric Debreuve, Xavier Descombes.

This work is made in collaboration with Philippe Pognonec (Team TIRO, CEA/UNS), Damien Ambrosetti (Histopathology department, CHU Pasteur, Nice).

The renal carcinoma is the most frequent type of kidney cancer (between 90% and 95% of all cases). Twelve classes of carcinoma can be distinguished, among which the clear cell carcinoma (CCRCC) and the papillary carcinoma (PRCC) are the two most common (75% and 10% of the cases, respectively). After the carcinoma has been diagnosed, the tumor is ablated and prepared for histological examination (fixation, staining, slicing, observation with a microscope). Along with genetic tests and protein reactions, the histological study allows to classify and grade the tumor in order to make a prognosis and to take decisions for the subsequent patient treatment. Digital histology is a recent domain (routinely, histological slices are studied by MDs directly on the microscope). The pioneer works deal with the automatic analysis of cells. However, one crucial factor for carcinoma classification is the structure of the vascular network. Coarsely, CCRCC is characterized by a “fishnet” structure while the PRCC has a tree-like structure.

In this context, our goal was to extract the vascular network from a given histological slice, compute features of the underlying graph structure, and classify the tumor into CCRCC or PRCC based on these features. The histological images being huge (typically, 100k x 100k pixels), they must be split into tiles (with some overlap to ease the combination of results) and processed tile-wise. The first step is to combine the color channels so that the vessels are as highlighted as possible. Then, the vascular network is detected by a processing pipeline including tailored, Gabor-like filtering, thresholding, and extraction of the skeleton. Small gaps in the skeleton are filled and some pruning is performed. Finally, the skeleton is converted to a graph representation. Based on the medical interpretation procedure, we focused our analysis of the graph on the following elements: the number of terminal and junction nodes, and the terminal branches. We proposed to compute the ratio between the number of terminal nodes and the number of junctions (T/J ratio), and the average length of terminal branches. Both features seem to be adapted to classification, especially the T/J ratio which, on the fairly small database of cases we currently have, exhibits an average value 65% higher for PRCC.

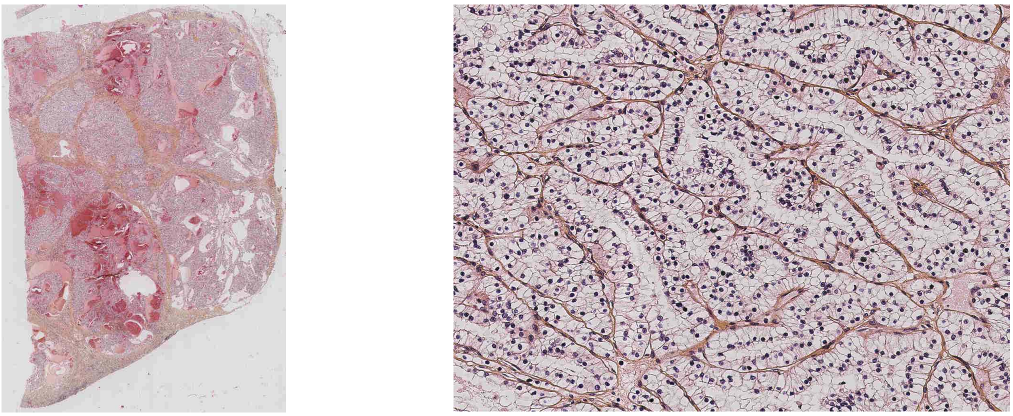


Figure 10. A histological slice through a kidney tumor: the whole slice (left) and a close-up (right).

NACHOS Project-Team

6. New Results

6.1. Electromagnetic wave propagation

6.1.1. Numerical study of the non-linear Maxwell equations for Kerr media

Participants: Loula Fezoui, Stéphane Lanteri.

The system of Maxwell equations describes the evolution of the interaction of an electromagnetic field with a propagation medium. The different properties of the medium, such as isotropy, homogeneity, linearity, among others, are introduced through *constitutive laws* linking fields and inductions. In the present study, we focus on non-linear effects and address non-linear Kerr materials specifically. In this model, any dielectric may become non-linear provided the electric field in the material is strong enough. As a first setp, we considered the one-dimensional case and study the numerical solution of the non-linear Maxwell equations thanks to DG methods. In particular, we make use of an upwind scheme and limitation techniques because they have a proven ability to capture shocks and other kinds of singularities in the fluid dynamics framework. The numerical results obtained in this preliminary study gave us confidence towards extending them to higher spatial dimensions. As a matter of fact, we recently started to work on the three-dimensional case and have initiated the development of a parallel simulation software based on our past contributions on DGTD methods for the case of linear propagation media.

6.1.2. High order geometry conforming DGTD method for nanophotonics

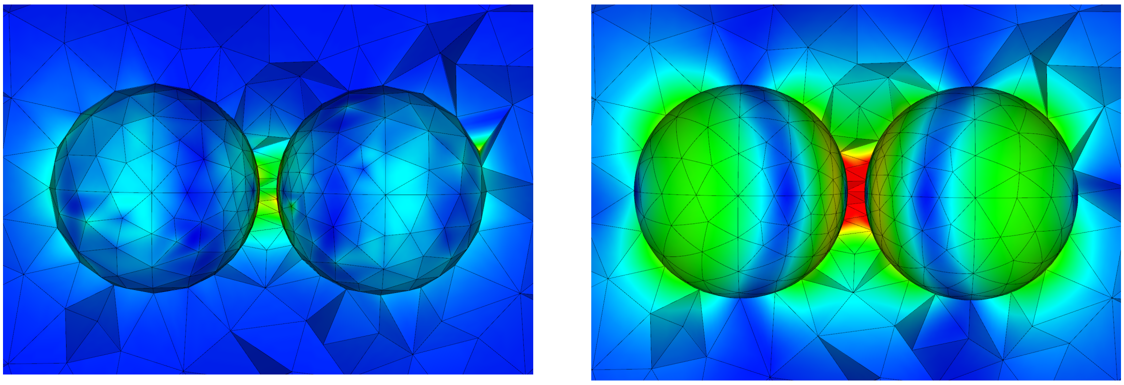
Participants: Stéphane Lanteri, Claire Scheid, Jonathan Viquerat.

Usually, unstructured mesh based methods rely on tessellations composed of straight-edged elements mapped linearly from a reference element, on domains which physical boundaries are indifferently straight or curved. Such meshes represent a serious hindrance for high order finite element (FE) methods since they limit the accuracy to second order in the spatial discretization. Thus, exploiting an enhanced representation of physical geometries is in agreement with the natural procedure of high order FE methods, such as the DG method. There are several ways to account for curved geometries. One could choose to incorporate the knowledge coming from CAD in the method to design the geometry and the approximation. These methods are called *isogeometric*, and have received a lot of attention recently. This naturally implies to have access to CAD models of the geometry. On the other hand, *isoparametric* usually rely on a polynomial approximation of both the boundary and the solution. This can be added fairly easily on top of existing implementations. In the present study we focus on the latter type of method, since our goal is first to envisage the benefit of curvilinear meshes for light/matter interaction with nanoscale structures.

6.1.3. Local approximation order DGTD method for nanophotonics

Participants: Stéphane Lanteri, Jonathan Viquerat.

High order DGTD methods for the numerical modeling of light/matter interactions on the nanoscale often assume a uniform distribution of the polynomial order to the cells of the underlying mesh. However, in the case of a mesh showing large variations in cell size, the time step imposed by the smallest cells can be a serious hindrance when trying to exploit high approximation orders. Indeed, a potentially large part of the CPU time is spent in the update of the physical field inside small cells where high polynomial orders might not be essential, while they are necessary in the larger cells. In this study, we consider the possibility of using a non-uniform distribution of the polynomial order in the framework of a global time step DGDT method. By imposing low orders in small cells and high orders in large cells, it is possible to significantly alleviate both the global number of degrees of freedom and the time step restriction with a minimal impact on the method accuracy. Strategies exploiting locally adaptive (LA) formulations usually combine both h- and p-adaptivity (where h denotes the discretization parameter in space and p the degree of the interpolation of the field components) in



Mesh with affine elements Mesh with curvilinear elements

Figure 4. Near-field visualization of the amplitude of the electric field Fourier transformed for a gold nanosphere dimer. Surface-to-surface distance is set to 4 nm. Calculations are based on a DGTD-$\mathbb{P}_4$ *method.*

order to concentrate the computational effort in the areas of high field variations. Here, the adopted point of view is quite different: starting from a given mesh and a uniform distribution of the polynomial order k, the LA strategy exploits all the polynomial orders p with $p \leq k$ to obtain a solution of similar accuracy with a reduced computational cost.

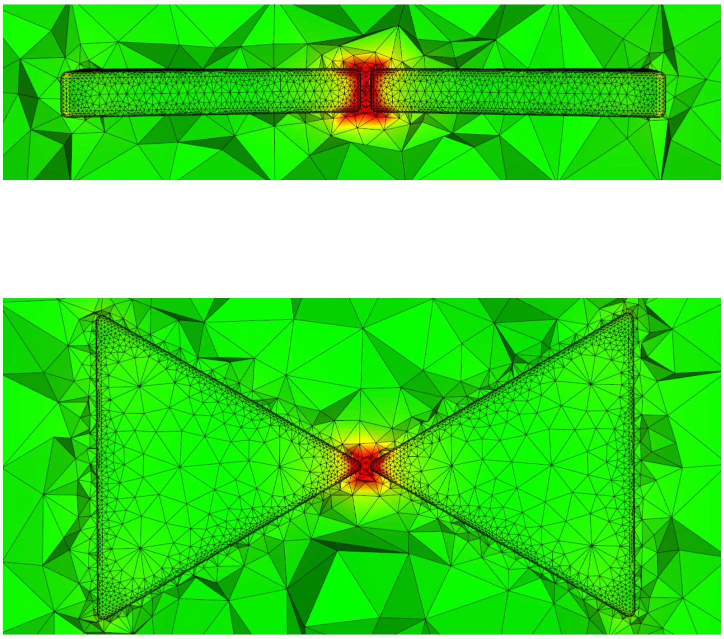


Figure 5. Near-field plasmonic interaction with a bowtie nanoantenns: contour line the amplitude of the discrete Fourier transform of the electric field.

6.1.4. Numerical treatment of non-local dispersion for nanoplasmonics

Participants: Stéphane Lanteri, Claire Scheid, Colin Vo Cong Tri.

When metallic nanostructures have sub-wavelength sizes and the illuminating frequencies are in the regime of metal's plasma frequency, electron interaction with the exciting fields have to be taken into account. Due to these interactions, plasmonic surface waves can be excited and cause extreme local field enhancements (surface plasmon polariton electromagnetic waves). Exploiting such field enhancements in applications of interest requires a detailed knowledge about the occurring fields which can generally not be obtained analytically. For the numerical modeling of light/matter interaction on the nanoscale, the choice of an appropriate model is a crucial point. Approaches that are adopted in a first instance are based on local (no interaction between electrons) dispersion models e.g. Drude or Drude-Lorentz. From the mathematical point of view, these models lead to an additional ordinary differential equation in time that is coupled to Maxwell's equations. When it comes to very small structures in a regime of 2 nm to 25 nm, non-local effects due to electron collisions have to be taken into account. Non-locality leads to additional, in general non-linear, partial differential equations and is significantly more difficult to treat, though. In this work, we study a DGTD method able to solve the system of Maxwell equations coupled to a linearized non-local dispersion model relevant to nanoplasmonics. While the method is presented in the general 3d case, in this preliminary stdudy, numerical results are given for 2d simulation settings.

6.1.5. Corner effects in nanoplasmonics

Participants: Camille Carvalho [ENSTA, POEMS project-team], Patrick Ciarlet [ENSTA, POEMS project-team], Claire Scheid.

The starting point of this ongoing work is the theoretical and numerical study of nanoplasmonic structures with corners. This is the central subject of the PhD thesis of Camille Carvalho. In the latter, the focus is made on a lossles Drude dispersion model with a frequency domain approach. Several well posedness problems arise due to the presence of corners and are addressed in the PhD thesis. A time domain approach in this context can also be investigated with the techniques developed in our project-team. Even if both approaches (time domain and frequency domain) represent similar physical phenomena, problems that arise are different and they appear as complementary; it is thus worth bridging the gap between the two frameworks. We propose to perform a thorough comparison in the case of structures with corners. Several extensions to other models are also envisaged, especially concerning the non local dispersion model.

6.1.6. DGTD method for nanoplasmonics based on generalized dispersion model

Participants: Stéphane Lanteri, Claire Scheid, Jonathan Viquerat.

In this work, we are concerned with the numerical modelling of the propagation of electromagnetic waves in dispersive materials for nanophotonics applications. We focus on a generalized model that allows for the description of a wide range of dispersive media. The underlying differential equations are recast into a generic form and we establish an existence and uniqueness result. We then turn to the numerical treatment and propose an appropriate DGTD framework. We obtain the semi-discrete convergence and prove the stability (and to a large extent, the convergence) in the fully discrete case when time integration is achieved with a 4 steps low storage Runge-Kutta scheme, *via* a technique relying on energy principles. Finally, we validate our results through the numerical simulation of two nanophotonics test cases.

6.1.7. Travelling waves for the non-linear Schrödinger equation in 2d

Participants: David Chiron [J.A. Dieudonné Laboratory, University of Nice-Sophia Antipolis], Claire Scheid.

We are interested in the numerical study of the two-dimensional travelling waves of the non-linear Schrödinger equation for a general non-linearity and with nonzero condition at infinity that can appear in optics. The equation has a variational structure that we propose to exploit to design a numerical method. We characterize the saddle points of the action as minimizers of another functional, allowing us to use a gradient flow. Combining this approach with a continuation method in the speed of the wave, we obtain the numerical solution for the full range of velocities. We plot the energy-momentum diagrams for different type of non-linearities. Through various examples, we show that even though the non-linearity has the same behaviour as the well-known Gross-Pitaevskii (GP) non-linearity, the qualitative properties of the travelling waves may be extremely different. For instance, we observe cusps, a modified Kadomtsev-Petviashvili I (KP-I) asymptotic in the transonic limit (as the speed of the wave approaches the speed of sound), various multiplicity results.

6.1.8. Multiscale DG methods for the time-domain Maxwell equations

Participants: Stéphane Lanteri, Raphaël Léger, Diego Paredes Concha [Instituto de Matemáticas, Universidad Católica de Valparaiso, Chile], Claire Scheid, Frédéric Valentin [LNCC, Petropolis, Brazil].

Although the DGTD method has already been successfully applied to complex electromagnetic wave propagation problems, its accuracy may seriously deteriorate on coarse meshes when the solution presents multiscale or high contrast features. In other physical contexts, such an issue has led to the concept of multiscale basis functions as a way to overcome such a drawback and allow numerical methods to be accurate on coarse meshes. The present work, which has been initiated in the context of the visit of Frédéric Valentin in the team, is concerned with the study of a particular family of multiscale methods, named Multiscale Hybrid-Mixed (MHM) methods. Initially proposed for fluid flow problems, MHM methods are a consequence of a hybridization procedure which caracterize the unknowns as a direct sum of a coarse (global) solution and the solutions to (local) problems with Neumann boundary conditions driven by the purposely introduced hybrid (dual) variable. As a result, the MHM method becomes a strategy that naturally incorporates multiple scales while providing solutions with high order accuracy for the primal and dual variables. The completely independent local problems are embedded in the upscaling procedure, and computational approximations may be naturally obtained in a parallel computing environment. In this study, a family of MHM methods is proposed for the solution of the time-domain Maxwell equations where the local problems are discretized either with a continuous FE method or a DG method (that can be viewed as a multiscale DGTD method). Preliminary results have been obtained in the 2d case for models problems.

6.1.9. HDG methods for the time-domain Maxwell equations

Participants: Alexandra Christophe-Argenvillier, Stéphane Descombes, Stéphane Lanteri.

This study is concerned with the development of accurate and efficient solution strategies for the system of 3d time-domain Maxwell equations coupled to local dispersion models (e.g. Debye, Drude or Drude-Lorentz models) in the presence of locally refined meshes. Such meshes impose a constraint on the allowable time step for explicit time integration schemes that can be very restrictive for the simulation of 3d problems. We consider here the possibility of using an unconditionally stable implicit time or a locally implicit time integration scheme combined to a HDG discretization method. In doing so, we extend our former study which was dealing with the 2d time-domain Maxwell equations for non-dispersive media.

6.1.10. HDG methods for the frequency-domain Maxwell equations

Participants: Thomas Frachon, Stéphane Lanteri, Liang Li [UESTC, Chengdu, China], Ludovic Moya, Ronan Perrussel [Laplace Laboratory, Toulouse].

In the context of the ANR TECSER project, we continue our efforts towards the development of scalable high order HDG methods for the solution of the system of 3d frequency-domain Maxwell equations. We aim at fully exploiting the flexibiity of the HDG discretization framework with regards to the adaptation of the interpolation order (p-adaptivity) and the mesh (h-adaptivity). In particular, we study the formulation of HDG methods on a locally refined non-conforming tetrahedral mesh and on a non-confoming hybrid cubic/tetrahedral mesh. We also investigate the coupling between the HDG formulation and a BEM (Boundary Element Method) discretization of an integral representation of the electromagnetic field in the case of propagation problems theoretically defined in unbounded domains.

6.2. Elastodynamic wave propagation

6.2.1. Sesimic wave interaction with viscoelastic media

Participants: Nathalie Glinsky, Stéphane Lanteri, Fabien Peyrusse [Department of Mathematics, Purdue University].

This work is concerned with the development of high order DGTD methods formulated on unstructured simplicial meshes for the numerical solution of the system of time-domain elastodynamic equations. These methods share some ingredients of the DGTD methods developed by the team for the time-domain Maxwell equations among which, the use of nodal polynomial (Lagrange type) basis functions, a second order leap-frog time integration scheme and a centered scheme for the evaluation of the numerical flux at the interface between neighboring elements. A recent novel contribution is the numerical treatment of viscoelastic attenuation. For this, the velocity-stress first order hyperbolic system is completed by additional equations for the anelastic functions including the strain history of the material. These additional equations result from the rheological model of the generalized Maxwell body and permit the incorporation of realistic attenuation properties of viscoelastic material accounting for the behaviour of elastic solids and viscous fluids. In practice, we need solving 3L additional equations in 2d (and 6L in 3d), where L is the number of relaxation mechanisms of the generalized Maxwell body. This method has been implemented in 2d and 3d.

6.2.2. DG method for arbitrary heterogeneous media

Participants: Nathalie Glinsky, Diego Mercerat [CETE Méditerranée].

We have recently devised an extension of the DGTD method for elastic wave propagation in arbitrary heterogeneous media. In realistic geological media (sedimentary basins for example), one has to include strong variations in the material properties. Then, the classical hypothesis that these properties are constant within each element of the mesh can be a severe limitation of the method, since we need to discretize the medium with very fine meshes resulting in very small time steps. For these reasons, we propose an improvement of the DGTD method allowing non-constant material properties within the mesh elements. A change of variables on the stress components allows writing the elastodynamic system in a pseudo-conservative form. Then, the introduction of non-constant material properties inside an element is simply treated by the calculation, via convenient quadrature formulae, of a modified local mass matrix depending on these properties. This new extension has been validated for a smoothly varying medium or a strong jump between two media, which can be accurately approximated by the method, independently of the mesh.

6.2.3. HDG method for the frequency-domain elastodynamic equations

Participants: Hélène Barucq [MAGIQUE-3D project-team, Inria Bordeaux - Sud-Ouest], Marie Bonnasse, Julien Diaz [MAGIQUE-3D project-team, Inria Bordeaux - Sud-Ouest], Stéphane Lanteri.

One of the most used seismic imaging methods is the full waveform inversion (FWI) method which is an iterative procedure whose algorithm is the following. Starting from an initial velocity model, (1) compute the solution of the wave equation for the N sources of the seismic acquisition campaign, (2) evaluate, for each source, a residual defined as the difference between the wavefields recorded at receivers on the top of the subsurface during the acquisition campaign and the numerical wavefields, (3) compute the solution of the wave equation using the residuals as sources, and (4) update the velocity model by cross correlation of images produced at steps (1) and (3). Steps (1)-(4) are repeated until convergence of the velocity model is achieved. We then have to solve $2N$ wave equations at each iteration. The number of sources, N, is usually large (about 1000) and the efficiency of the inverse solver is thus directly related to the efficiency of the numerical method used to solve the wave equation. Seismic imaging can be performed in the time-domain or in the frequency-domain regime. In this work which is conducted in the framework of the Depth Imaging Partnership (DIP) between Inria and TOTAL, we adopt the second setting. The main difficulty with frequency-domain inversion lies in the solution of large sparse linear systems which is a challenging task for realistic 3d elastic media, even with the progress of high performance computing. In this context, we study novel high order HDG methods formulated on unstructured meshes for the solution of the frency-domain elastodynamic equations. Instead of solving a linear system involving the degrees of freedom of all volumic cells of the mesh, the principle of a HDG formulation is to introduce a new unknown in the form of Lagrange multiplier representing the trace of the numerical solution on each face of the mesh. As a result, a HDG formulation yields a global linear system in terms of the new (surfacic) unknown while the volumic solution is recovered thanks to a local computation on each element.

6.2.4. Multiscale DG methods for the time-domain elastodynamic equations

Participants: Marie-Hélène Lallemand, Frédéric Valentin [LNCC, Petropolis, Brazil].

In the context of the visit of Frédéric Valentin in the team, we have initiated a study aiming at the design of novel multiscale methods for the solution of the time-domain elastodynamic equations, in the spirit of MHM (Multiscale Hybrid-Mixed) methods previously proposed for fluid flow problems. Motivation in that direction naturally came when dealing with non homogeneous anisotropic elastic media as those encountered in geodynamics related applications, since multiple scales are naturally present when high contrast elasticity parameters define the propagation medium. Instead of solving the usual system expressed in terms of displacement or displacement velocity, and stress tensor variables, a hybrid mixed-form is derived in which an additional variable, the Lagrange multiplier, is sought as representing the (opposite) of the surface tension defined at each face of the elements of a given discretization mesh. We consider the velocity/stress formulation of the elastodynamic equations, and study a MHM method defined for a heterogeneous medium where each elastic material is considered as isotropic to begin with. If the source term (the applied given force on the medium) is time independent, and if we are given an arbitrarily coarse conforming mesh (triangulation in 2d, tetrahedrization in 3d), the proposed MHM method consists in first solving a series of fully decoupled (therefore parallelizable) local (element-wise) problems defining parts of the full solution variables which are directly related to the source term, followed by the solution of a global (coarse) problem, which yields the degrees of freedom of both the Lagrange multiplier dependent part of the full solution variables and the Lagrange multiplier itself. Finally, the updating of the full solution variables is obtained by adding each splitted solution variables, before going on the next time step of a leap-frog time integration scheme. Theoretical analysis and implementation of this MHM method where the local problems are discretized with a DG method, are underway.

6.3. High performance numerical computing

6.3.1. Porting a DGTD solver for bioelectromagnetics to the DEEP-ER architecture

Participants: Alejandro Duran [Barcelona Supercomputing Center, Spain], Stéphane Lanteri, Raphaël Léger, Damian A. Mallón [Juelich Supercomputing Center, Germany].

We are concerned here with the porting of a Discontinuous Galerkin Time-Domain solver for computational bioelectromagnetics to the novel heterogeneous architecture proposed in the DEEP-ER european project on exascale computing. This architecture is based on a Cluster/Booster division concept (see Fig. 6). The Booster nodes are based on the Intel Many Integrated Core (MIC) architecture. Therfore, one objective of our efforts is the algorithmic adaptation of the DG kernels in order to leverage the vectorizing capabilities of the MIC processor. The other activities that are undertaken in the context of our contribution to this project aim at exploiting the software environments and tools proposed by DEEP-ER partners for implementing resiliency strategies and high performance I/O operations. In particular, the Cluster nodes are used for running some parts of the pre- and post-processing phases of the DGTD solver which do not lend themselves well to multithreading, as well as I/O intensive routines. One possibility to achieve this is to consider a model in which these less scalable and I/O phases are reverse-offloaded from Booster processes to Cluster processes in a one-to-one mapping. This is achieved by exploiting the OmpSs offload functionality, developed at Barcelona Supercomputing Center for the DEEP-ER platform. In future work, the OmpSs framework will also be leveraged to expose task-based parallelism and exploit task-based resilience.

6.3.2. Hybrid MIMD/SIMD high order DGTD solver for nanophotonics

Participants: Tristan Cabel, Gabriel Hautreux [CINES, Montpellier], Stéphane Lanteri, Raphaël Léger, Claire Scheid, Jonathan Viquerat.

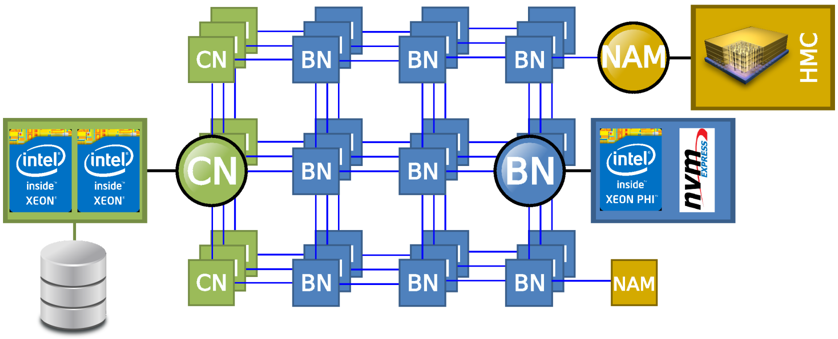


Figure 6. DEEP-ER hardware architecture sketch.

This work is concerned with the development of a scalable high order finite element type solver for the numerical modeling of light interaction with nanometer scale structures. From the mathematical modeling point of view, one has to deal with the differential system of Maxwell equations in the time domain, coupled to an appropriate differential model of the behavior of the underlying material (which can be a dielectric and/or a metal) at optical frequencies. For the numerical solution of the resulting system of differential equations, we adopt the high order DGTD (Discontinuous Galerkin Time-Domain) solver described in [21]. A hybrid MIMD/SIMD parallelization of this DGTD solver has been developed by combining the MPI and OpenMP parallel programming models. The performances of the resulting parallel DGTD solver have been assessed on the Curie system of the PRACE research infrastructure. For that purpose, we selected a use case typical of optical guiding applications. A Y-shaped waveguide is considered which consists in nanosphere embedded in vacuum (see Fig. 7). The constructed tetrahedral mesh consists of 520,704 vertices and 2,988,103 elements. The high order discontinuous finite element method designed for the solution of the system of time-domain Maxwell equations coupled to a Drude model for the dispersion of noble metals at optical frequencies is formulated on a tetrahedral mesh.

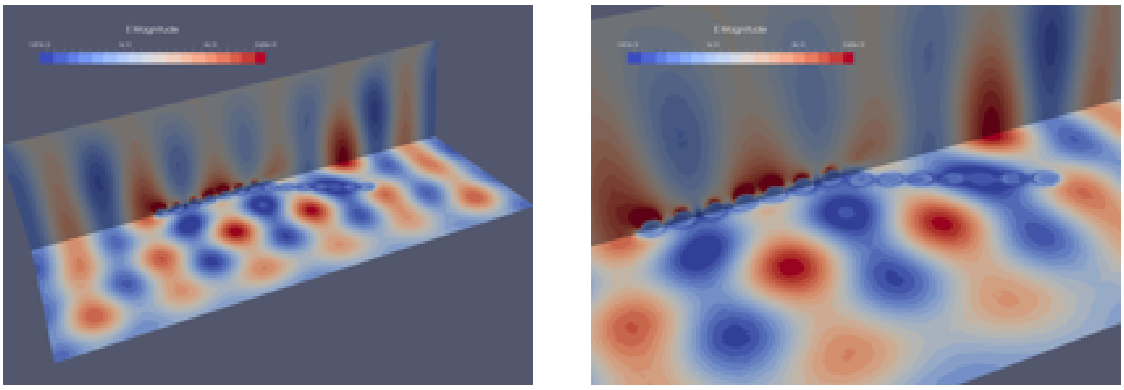


Figure 7. Y-shaped waveguide: contour lines of the amplitude of the discrete Fourier transform of the electric field.

6.4. Applications

6.4.1. Light diffusion in nanostructured optical fibers

Participants: Wilfried Blanc [Optical Fibers team, LPMC, Nice], Stéphane Lanteri, Paul Loriot, Claire Scheid.

Optical fibers are the basis for applications that have grown considerably in recent years (telecommunications, sensors, fiber lasers, etc.). Despite these undeniable successes, it is necessary to develop new generations of amplifying optical fibers that will overcome some limitations typical of silica. In this sense, the amplifying Transparent Glass Ceramics (TGC), and particularly the fibers based on this technology, open new perspectives that combine the mechanical and chemical properties of a glass host and the augmented spectroscopic properties of embedded nanoparticles, particularly rare earth-doped oxide nanoparticles. Such rare earth-doped silica-based optical fibers with transparent glass ceramic (TGC) core are fabricated by the Optical Fibers team of the Laboratory of Condensed Matter (LPMC) in Nice. The obejctive of this collaboration with Wilfried Blanc at LPMC is the study of optical transmission terms of loss due to scattering through the numerical simulation of light propagation in a nanostructured optical fiber core using a high order DGTD method developed in the team.

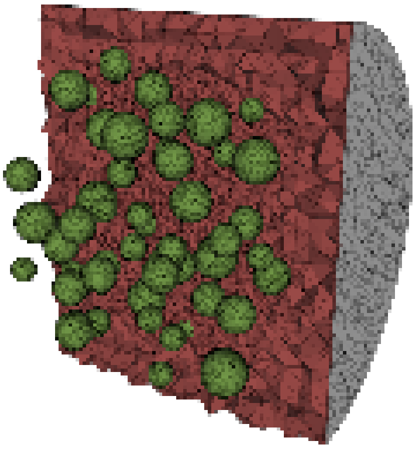


Figure 8. Unstructured tetrahdral mesh of a nanostructured optical fiber core.

6.4.2. Gap-plasmon confinement with gold nanocubes

Participants: Stéphane Lanteri, Antoine Moreau [Institut Pascal, Université Blaise Pascal], Claire Scheid, Jonathan Viquerat.

The propagation of light in a slit between metals is known to give rise to guided modes. When the slit is of nanometric size, plasmonic effects must be taken into account, since most of the mode propagates inside the metal. Indeed, light experiences an important slowing-down in the slit, the resulting mode being called *gap-plasmon*. Hence, a metallic structure presenting a nanometric slit can act as a light trap, i.e. light will accumulate in a reduced space and lead to very intense, localized fields. Recently, the chemical production of random arrangements of nanocubes on gold films at low cost was proved possible by Antoine Moreau and colleagues at Institut Pascal. Nanocubes are separated from the gold substrate by a dielectric spacer of variable thickness, thus forming a narrow slit under the cube. When excited from above, this configuration is able to support gap-plasmon modes which, once trapped, will keep bouncing back and forth inside the cavity. At visible frequencies, the lossy behavior of metals will cause the progressive absorption of the trapped electromagnetic field, turning the metallic nanocubes into efficient absorbers. The frequencies at which this absorption occurs can be tuned by adjusting the dimensions of the nanocube and the spacer. In collaboration with Antoine Moreau, we propose to study numerically the impact of the geometric parameters of the problem on the behaviour of a single nanocube placed over a metallic slab (see Fig. 9). The behavior of single nanocubes on metallic plates has been simulated, for lateral sizes c ranging from 50 to 80 nm, and spacer thicknesses δ from 3 to 22 nm. The absorption efficiency in the cube Q_{cube} at the resonance frequency is retrieved from the results of each computation (see Fig. 10).

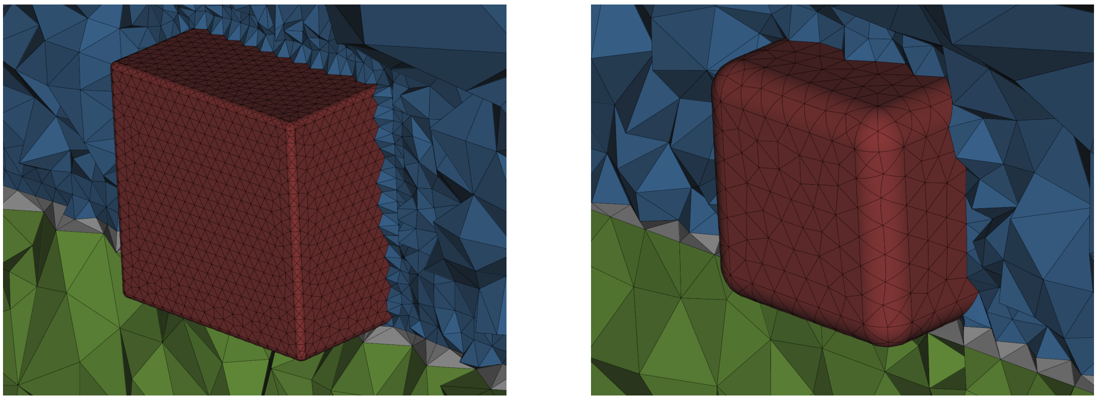
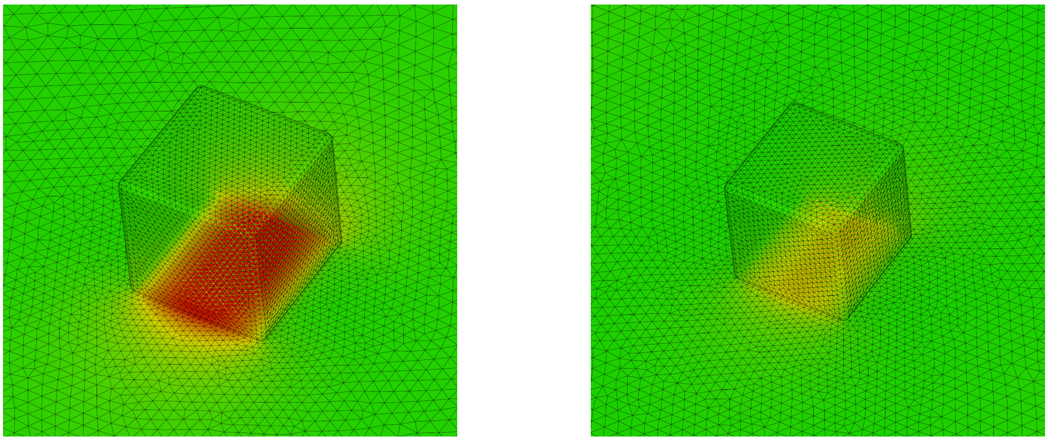


Figure 9. Meshes of rounded nanocubes with rounding radii ranging from 2 to 10 nm. Red cells correspond to the cube. The latter lies on the dielectric spacer (gray cells) and the metallic plate (green). Blue cells represent the air surrounding the device.



$c = 70$ nm, $\delta = 12$ nm $c = 60$ nm, $\delta = 18$ nm

Figure 10. Amplitude of the discrete Fourier transform of the magnetic field for different nanocube configurations. All field maps are scaled identically for better comparison. The obtained field is more intense for configurations that yield high Q_{cube} values.

6.4.3. Light propagation in power splitters

Participants: Julien Coulet, Carlos Henrique Da Silva Santos [Instituto Federal de Educação, Ciência e Tecnologia de Säo Paulo, Brazil], Hugo Enrique Hernandez Figueroa [Universidade Estadual de Campinas, Faculdade de Engenharia Elétrica e de Computaçäo, Säo Paulo, Brazil], Stéphane Lanteri, Frédéric Valentin [LNCC, Petropolis, Brazil].

Power splitters are passive devices widely used in signal processing, which splits an input signal into two or more output signals. The repartition of the input power over each output is specifc to the required usage. Even if power splitters are common in classical electronics, designing them at the micrometric scale is quietly recent and is an active field of research. The purpose of this study initiated in the framework of a collaboration with researchers at Unicamp in Säo Paulo is to study the electromagnetic wave propagation is such a power splitter geometry using a high order DGTD method developed in the team, see Fig. 11 .

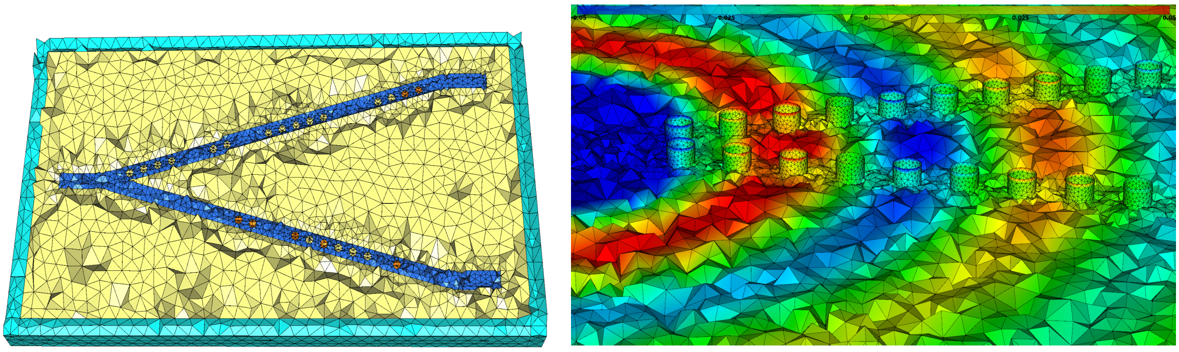
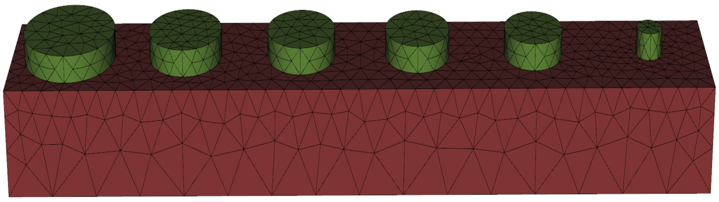


Figure 11. Geomtry of Y-shaped power splitter (left) and contour lines of the amplitude of the electric field (right).

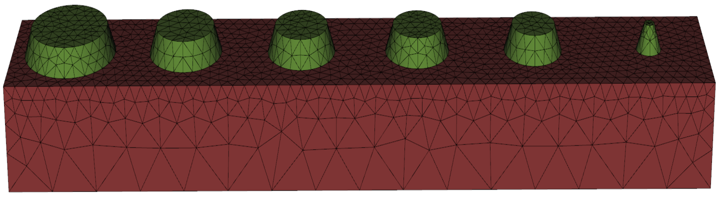
6.4.4. Dielectric reflectarrays

Participants: Maciej Klemm [Centre for Communications Research, University of Bristol], Stéphane Lanteri, Claire Scheid, Jonathan Viquerat.

In the past few years, important efforts have been deployed to find alternatives to on-chip, low-performance metal interconnects between devices. Because of the ever-increasing density of integrated components, intra- and inter-chip data communications have become a major bottleneck in the improvement of information processing. Given the compactness and the simple implantation of the devices, communications *via* free-space optics between nanoantenna-based arrays have recently drawn more attention. Here, we focus on a specific low-loss design of dielectric reflectarray (DRA), whose geometry is based on a periodic repartition of dielectric cylinders on a metallic plate. When illuminated in normal incidence, specific patterns of such resonators provide a constant phase gradient along the dielectric/metal interface, thus altering the phase of the incident wavefront. The gradient of phase shift generates an effective wavevector along the interface, which is able to deflect light from specular reflection. However, the flaws of the lithographic production process can lead to discrepancies between the ideal device and the actual resonator array. Here, we propose to exploit our DGTD solver to study the impact of the lithographic flaws on the performance of a 1D reflectarray (see Fig. 12). Efficient computations are obtained by combining high-order polynomial approximation with curvilinear meshing of the resonators, yielding accurate results on very coarse meshes (see Fig. 13). The study is continued with the computation of the reflection of a 2D reflectarray. This work constitutes the base of a wider study in collaboration with Maciej Klemm at the Centre for Communications Research, University of Bristol.

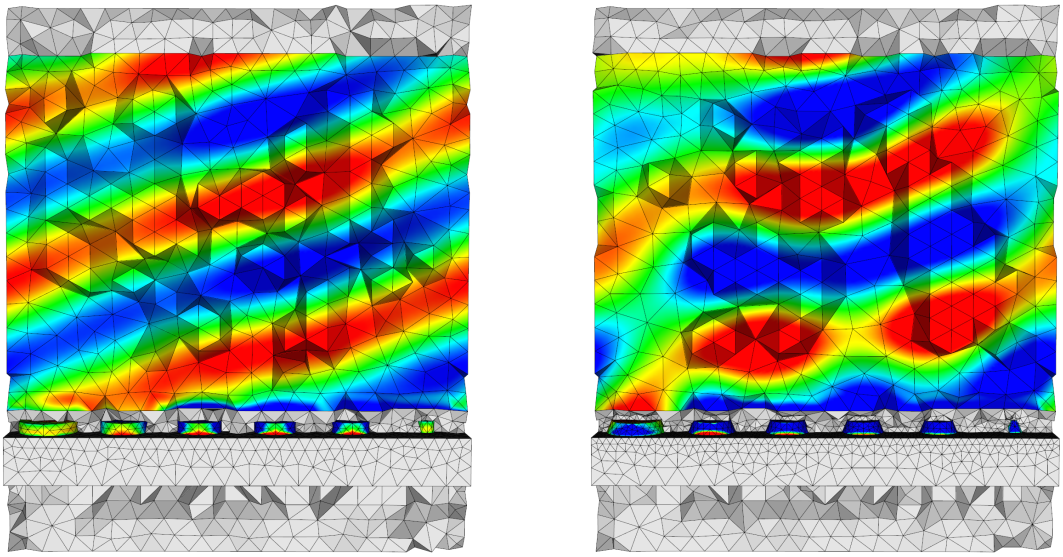


Ideal reflectarray



Realistic reflectarray

Figure 12. Ideal and realistic 1D dielectric reflectarray meshes. The red tetrahedra correspond to silver, while the green ones are made of an anisotropic dielectric material. The device is surrounded by air and terminated by a PML above and below, and by periodic boundary conditions on the lateral sides.



Ideal reflectarray Realistic reflectarray

Figure 13. Time-domain snapshot of E_y component for ideal and realistic 1D dielectric reflectarrays. Solution is obtained in established regime at $t = 0.1$ ps. Fields are scaled to $[-1, 1]$.

NEUROMATHCOMP Project-Team

6. New Results

6.1. Neural Networks as dynamical systems

6.1.1. Periodic forcing of stabilized E-I networks: Nonlinear resonance curves and dynamics

Participants: Romain Veltz, Terry Sejnowski [Salk Institute].

Inhibition stabilized networks (ISNs) are neural architectures with strong positive feedback among pyramidal neurons balanced by strong negative feedback from in-hibitory interneurons, a circuit element found in the hippocampus and the primary visual cortex. In their working regime, ISNs produce damped oscillations in the γ-range in response to inputs to the inhibitory population. In order to understand the proper-ties of interconnected ISNs, we investigated periodic forcing of ISNs. We show that ISNs can be excited over a range of frequencies and derive properties of the resonance peaks. In particular, we studied the phase-locked solutions, the torus solutions and the resonance peaks. More particular, periodically forced ISNs respond with (possibly multi-stable) phase-locked activity whereas networks with sustained intrinsic oscillations respond more dynamically to periodic inputs with tori. Hence, the dynamics are surprisingly rich and phase effects alone do not adequately describe the network response. This strengthens the importance of phase-amplitude coupling as opposed to phase-phase coupling in providing multiple frequencies for multiplexing and routing information.

This work has been published in Neural Computation and is available as [29].

6.1.2. A new twist for the simulation of hybrid systems using the true jump method

Participant: Romain Veltz.

The use of stochastic models, in effect piecewise deterministic Markov processes (PDMP), has become increasingly popular especially for the modeling of chemical reactions and cell biophysics. Yet, exact simulation methods, for the simulation of these models in evolving environments, are limited by the need to find the next jumping time at each recursion of the algorithm. We report on a new general method to find this jumping time for the True Jump Method. It is based on an expression in terms of ordinary differential equations for which efficient numerical methods are available. As such, our new result makes it possible to study numerically stochastic models for which analytical formulas are not available thereby providing a way to approximate the state distribution for example. We conclude that the wide use of event detection schemes for the simulation of PDMPs should be strongly reconsidered. The only relevant remaining question being the efficiency of our method compared to the Fictitious Jump Method, question which is strongly case dependent.

This work is available as [55].

6.1.3. On the effects on cortical spontaneous activity of the symmetries of the network of pinwheels in visual area V1

Participants: Romain Veltz, Pascal Chossat, Olivier Faugeras.

This work challenges and extends earlier seminal work. We consider the problem of describing mathematically the spontaneous activity of V1 by combining several important experimental observations including (1) the organization of the visual cortex into a spatially periodic network of hypercolumns structured around pinwheels, (2) the difference between short-range and long-range intracortical connections, the first ones being rather isotropic and producing naturally doubly periodic patterns by Turing mechanisms, the second one being patchy, and (3) the fact that the Turing patterns spontaneously produced by the short-range connections and the network of pinwheels have similar periods. By analyzing the preferred orientation (PO) maps, we are able to classify all possible singular points (the pinwheels) as having symmetries described by a small subset of the wallpaper groups. We then propose a description of the spontaneous activity of V1 using a classical voltage-based neural field model that features isotropic short-range connectivities modulated by non-isotropic long-range connectivities. A key observation is that, with only short-range connections and because the problem has full translational invariance in this case, a spontaneous doubly periodic pattern generates a 2-torus in a suitable functional space which persists as a flow-invariant manifold under small perturbations, for example when turning on the long-range connections. Through a complete analysis of the symmetries of the resulting neural field equation and motivated by a numerical investigation of the bifurcations of their solutions, we conclude that the branches of solutions which are stable over an extended range of parameters are those that correspond to patterns with an hexagonal (or nearly hexagonal) symmetry. The question of which patterns persist when turning on the long-range connections is answered by (1) analyzing the remaining symmetries on the perturbed torus and (2) combining this information with the Poincaré-Hopf theorem. We have developed a numerical implementation of the theory that has allowed us to produce the predicted patterns of activities, the planforms. In particular we generalize the contoured and non-contoured planforms predicted by previous authors.

This work has been published in Journal of Mathematical Neuroscience and is available as [27].

6.1.4. Biophysical reaction-diffusion model for stage II retinal waves and bifurcations analysis

Participants: Theodora Karvouniari, Bruno Cessac.

Retinal waves are spontaneous waves of spiking activity observed in the retina, during development only, playing a central role in shaping the visual system and retinal circuitry. Understanding how these waves are initiated and propagate in the retina could enable one to control, guide and predict them in the in vivo adult retina as inducing them is expected to reintroduce some plasticity in the retinal tissue and in the projections to the LGN. In this context, we propose a physiologically realistic reaction-diffusion model for the mechanisms of the emergence of stage II cholinergic retinal waves during development. We perform the bifurcation analysis when varying two biophysically relevant parameters, the conductances of calcium and potassium g_{Ca}, g_K respectively. The two main goals of our work are: firstly, reproduce the experimental recordings of developmental retinal waves by simulating our model and secondly, explore the different dynamical behaviours observed when varying these two parameters.

This work is available as [35].

6.1.5. Spatio-Temporal Linear Response of Spiking Neuronal Network Models

Participants: Rodrigo Cofré, Bruno Cessac.

We study the impact of a weak time-dependent external stimulus on the collective statistics of spiking responses in neuronal networks. We extend the current knowledge, assessing the impact over firing rates and cross correlations, to any higher order spatio-temporal correlation [1]. Our approach is based on Gibbs distributions (in a general setting considering non stationary dynamics and infinite memory) [2] and linear response theory. The linear response is written in terms of a correlation matrix, computed with respect to the spiking dynamics without stimulus. We give an example of application in a conductance based integrate-and fire model.

This work is available as [38].

6.1.6. Heteroclinic cycles in Hopfield networks

Participants: Pascal Chossat, Maciej Krupa.

It is widely believed that information is stored in the brain by means of the varying strength of synaptic connections between neurons. Stored patterns can be replayed upon the arrival of an appropriate stimulus. Hence, it is interesting to understand how an information pattern can be represented by the dynamics of the system. In this work, we consider a class of network neuron models, known as Hopfield networks, with a learning rule which consists of transforming an information string to a coupling pattern. Within this class of models, we study dynamic patterns, known as robust heteroclinic cycles, and establish a tight connection between their existence and the structure of the coupling.

This work has been published in Journal of Nonlinear Science and is available as [20].

6.2. Mean field approaches

6.2.1. Confronting mean-field theories to measurements: a perspective from neuroscience

Participant: Bruno Cessac.

Mean-field theories in neuroscience are usually understood as ways to bridge spatial and temporal scales by lumping together the activities of many single neurons, and then explaining or predicting the spatio-temporal variations of mesoscopic or macroscopic quantities measurable with current technologies: EEG, MEG, fMRI, optical imaging, etc. This is very much alike the situation in statistical physics where macroscopic quantities such as pressure, conductivity and so on are explained by the interactions between ”microscopic” entities like atoms or molecules.

The situation in neuroscience is different however: the laws governing the microscopic dynamics in physics do not have the same structure as the laws governing neuronal dynamics; for example, interactions between neurons are not symmetric. Moreover, it is yet unclear what the relevant macroscopic quantities are in order to account for, say, visual perception. At the present stage of research, these quantities are considered to be what is measurable with currently available technologies, whereas better theories could reveal new types of phenomenological observables with a higher explanatory power.

We review mean-field methods coming from physics and their consequences on neuronal dynamics predictions.

This work is available as [30], [31], [32].

6.2.2. A Formalism for Evaluating Analytically the Cross-Correlation Structure of a Firing-Rate Network Model

Participants: Diego Fasoli, Olivier Faugeras, Stefano Panzeri.

We introduce a new formalism for evaluating analytically the cross-correlation structure of a finite size firing-rate network with recurrent connections. The analysis performs a first-order perturbative expansion of neural activity equations that include three different sources of randomness: the background noise of the membrane potentials, their initial conditions, and the distribution of the recurrent synaptic weights. This allows the analytical quantification of the relationship between anatomical and functional connectivity, i.e. of how the synaptic connections determine the statistical dependencies at any order among different neurons. The technique we develop is general, but for simplicity and clarity we demonstrate its efficacy by applying it to the case of synaptic connections described by regular graphs. The analytical equations obtained in this way reveal previously unknown behaviors of recurrent firing-rate networks, especially on how correlations are modified by the external input, by the finite size of the network, by the density of the anatomical connections and by correlation in sources of randomness. In particular, we show that a strong input can make the neurons almost independent, suggesting that functional connectivity does not depend only on the static anatomical connectivity, but also on the external inputs. Moreover we prove that in general it is not possible to find a mean-field description à la Sznitman of the network, if the anatomical connections are too sparse or our three sources of variability are correlated. To conclude, we show a very counterintuitive phenomenon, which we call stochastic synchronization, through which neurons become almost perfectly correlated even if the sources of randomness are independent. Due to its ability to quantify how activity of individual neurons and the correlation among them depends upon external inputs, the formalism introduced here can serve as a basis for exploring analytically the computational capability of population codes expressed by recurrent neural networks.

This work is available as [22].

6.2.3. Asymptotic Description of Neural Networks with Correlated Synaptic Weights

Participants: Olivier Faugeras, James Maclaurin.

We study the asymptotic law of a network of interacting neurons when the number of neurons becomes infinite. Given a completely connected network of neurons in which the synaptic weights are Gaussian correlated random variables, we describe the asymptotic law of the network when the number of neurons goes to infinity. We introduce the process-level empirical measure of the trajectories of the solutions to the equations of the finite network of neurons and the averaged law (with respect to the synaptic weights) of the trajectories of the solutions to the equations of the network of neurons. The main result of this article is that the image law through the empirical measure satisfies a large deviation principle with a good rate function which is shown to have a unique global minimum. Our analysis of the rate function allows us also to characterize the limit measure as the image of a stationary Gaussian measure defined on a transformed set of trajectories.

This work is available as [23].

6.2.4. *Clarification and Complement to "Mean-Field Description and Propagation of Chaos in Networks of Hodgkin-Huxley and FitzHugh-Nagumo Neurons"*

Participants: Mireille Bossy, Olivier Faugeras, Denis Talay.

In this work, we clarify the well-posedness of the limit equations to the mean-field N-neuron models proposed in [1] and we prove the associated propagation of chaos property. We also complete the modeling issue in [1] by discussing the well-posedness of the stochastic differential equations which govern the behavior of the ion channels and the amount of available neurotransmitters.

This work is available as [18].

6.3. Neural fields theory

6.3.1. *ERRATUM: A Center Manifold Result for Delayed Neural Fields Equations*

Participants: Romain Veltz, Olivier Faugeras.

Lemma C.1 in [95] is wrong. This lemma is used in the proof of the existence of a smooth center manifold, Theorem 4.4 in that paper. An additional assumption is required to prove this existence. We spell out this assumption, correct the proofs, and show that the assumption is satisfied for a large class of delay functions τ. We also weaken the general assumptions on τ.

This work has been published in SIAM J. Math. Anal. and is available as [28].

6.3.2. *A general framework for stochastic traveling waves and patterns, with application to neural field equations*

Participants: James Inglis, James Maclaurin.

In this work we present a general framework in which to rigorously study the effect of spatio- temporal noise on traveling waves and stationary patterns. In particular, the framework can incorporate versions of the stochastic neural field equation that may exhibit traveling fronts, pulses or stationary patterns. To do this, we first formulate a local SDE that describes the position of the stochastic wave up until a discontinuity time, at which point the position of the wave may jump. We then study the local stability of this stochastic front, obtaining a result that recovers a well-known deterministic result in the small-noise limit. We finish with a study of the long-time behavior of the stochastic wave.

This work has appeared in SIAM J. on Applied Dynamical Systems (SIADS) [49].

6.4. Slow-Fast Dynamics in Neural Models

6.4.1. *From Canards of Folded Singularities to Torus Canards in a Forced van der Pol Equation*

Participants: John Burke [Boston University, USA], Mathieu Desroches, Albert Granados [Technical University of Denmark, Lyngby, Denmark], Tasso Kaper [Boston University, USA], Maciej Krupa, Theodore Vo [Boston University, USA].

In this work, we study canard solutions of the forced van der Pol equation in the relaxation limit for low-, intermediate-, and high-frequency periodic forcing. A central numerical observation made herein is that there are two branches of canards in parameter space which extend across all positive forcing frequencies. In the low-frequency forcing regime, we demonstrate the existence of primary maximal canards induced by folded saddle nodes of type I and establish explicit formulas for the parameter values at which the primary maximal canards and their folds exist. Then, we turn to the intermediate- and high-frequency forcing regimes and show that the forced van der Pol possesses torus canards instead. These torus canards consist of long segments near families of attracting and repelling limit cycles of the fast system, in alternation. We also derive explicit formulas for the parameter values at which the maximal torus canards and their folds exist. Primary maximal canards and maximal torus canards correspond geometrically to the situation in which the persistent manifolds near the family of attracting limit cycles coincide to all orders with the persistent manifolds that lie near the family of repelling limit cycles. The formulas derived for the folds of maximal canards in all three frequency regimes turn out to be representations of a single formula in the appropriate parameter regimes, and this unification confirms the central numerical observation that the folds of the maximal canards created in the low-frequency regime continue directly into the folds of the maximal torus canards that exist in the intermediate- and high-frequency regimes. In addition, we study the secondary canards induced by the folded singularities in the low-frequency regime and find that the fold curves of the secondary canards turn around in the intermediate-frequency regime, instead of continuing into the high-frequency regime. Also, we identify the mechanism responsible for this turning. Finally, we show that the forced van der Pol equation is a normal form-type equation for a class of single-frequency periodically driven slow/fast systems with two fast variables and one slow variable which possess a non-degenerate fold of limit cycles. The analytic techniques used herein rely on geometric desingularisation, invariant manifold theory, Melnikov theory, and normal form methods. The numerical methods used herein were developed in Desroches et al. (SIAM J Appl Dyn Syst 7:1131–1162, 2008, Nonlinearity 23:739–765 2010).

This work has been published in J. Nonlinear Sci. and is available as [19].

6.4.2. Extending the zero-derivative principle for slow-fast dynamical systems

Participants: Eric Benoît [Université de La Rochelle, France], Morten Brøns [Technical University of Denmark, Lyngby, Denmark], Mathieu Desroches, Maciej Krupa.

Slow-fast systems often possess slow manifolds, that is invariant or locally invariant sub-manifolds on which the dynamics evolves on the slow time scale. For systems with explicit timescale separation, the existence of slow manifolds is due to Fenichel theory, and asymptotic expansions of such manifolds are easily obtained. In this work, we discuss methods of approximating slow manifolds using the so-called zero-derivative principle. We demonstrate several test functions that work for systems with explicit time scale separation including ones that can be generalized to systems without explicit timescale separation. We also discuss the possible spurious solutions, known as ghosts, as well as treat the Templator system as an example.

This work has been published in ZAMP and is available as [17].

6.4.3. Canards, folded nodes and mixed-mode oscillations in piecewise-linear slow-fast systems

Participants: Mathieu Desroches, Antoni Guillamon [Polytechnic University of Catalunya, Barcelona, Spain], Enrique Ponce [University of Sevilla, Spain], Rafel Prohens [University of the Balearic Islands, Palma, Spain], Serafim Rodrigues [Plymouth University, UK], Antonio Teruel [University of the Balearic Islands, Palma, Spain].

Canard-induced phenomena have been extensively studied in the last three decades, both from the mathematical and from the application viewpoints. Canards in slow-fast systems with (at least) two slow variables, especially near folded-node singularities, give an essential generating mechanism for Mixed-Mode oscillations (MMOs) in the framework of smooth multiple timescale systems. There is a wealth of literature on such slow-fast dynamical systems and many models displaying canard-induced MMOs, in particular in neuroscience. In parallel, since the late 1990s several papers have shown that the canard phenomenon can be faithfully reproduced with piecewise-linear (PWL) systems in two dimensions although very few results are available in the three-dimensional case. This work aims to bridge this gap by analyzing canonical PWL systems that display folded singularities, primary and secondary canards, with a similar control of the maximal winding number as in the smooth case. We also show that the singular phase portraits are compatible in both frameworks. Finally, we show on an example how to construct a (linear) global return and obtain robust PWL MMOs.

This work has been accepted for publication in SIAM Review and is available as [46].

6.4.4. Canard solutions in planar piecewise linear systems with three zones

Participants: Soledad Fernández-García [Inria Paris-Rocquencourt, France], Mathieu Desroches, Maciej Krupa, Antonio Teruel [University of the Balearic Islands, Palma, Spain].

In this work, we analyze the existence and stability of canard solutions in a class of planar piecewise linear systems with three zones, using a singular perturbation theory approach. To this aim, we follow the analysis of the classical canard phenomenon in smooth planar slow-fast systems and adapt it to the piecewise-linear framework. We first prove the existence of an intersection between repelling and attracting slow manifolds, which defines a maximal canard, in a non-generic system of the class having a continuum of periodic orbits. Then, we perturb this situation and prove the persistence of the maximal canard solution, as well as the existence of a family of canard limit cycles in this class of systems. Similarities and differences between the piecewise linear case and the smooth one are highlighted.

This work has been published Dynam. Syst. and is available as [24].

6.4.5. Spike-adding mechanism in parabolic bursters: the role of folded-saddle canards

Participants: Mathieu Desroches, Maciej Krupa, Serafim Rodrigues [Plymouth University, UK].

The present work develops a new approach to studying parabolic bursting, and also proposes a novel four-dimensional canonical and polynomial-based parabolic burster. In addition to this new polynomial system, we also consider the conductance-based model of the Aplysia R15 neuron known as Plant's model, and a reduction of this prototypical biophysical parabolic burster to three variables, including one phase variable, namely Rinzel's theta model. Revisiting these models from the perspective of slow-fast dynamics reveals that the number of spikes per burst may vary upon parameter changes, however the spike-adding process occurs in a brutal (explosive) fashion that involves special solutions called canards. This spike-adding canard explosion phenomenon is analysed by using tools from geometric singular perturbation theory in tandem with numerical bifurcation techniques. We find that the bifurcation structure persists across both parabolic bursters, that is, spikes within the burst are incremented via the crossing of an excitability threshold given by a particular type of canard orbit, namely the strong canard of a folded-saddle singularity. Using these findings, we construct a new polynomial approximation of Plant's model, which retains all the key elements for parabolic bursting, including the canard mediated spike-adding transitions. Finally, we briefly investigate the presence of spike-adding via canards in planar phase models of parabolic bursting, namely the theta model by Ermentrout and Kopell.

This work has been submitted for publication and is available as [47].

6.4.6. Canards and spike-adding transitions in a minimal piecewise-linear Hindmarsh-Rose square-wave burster

Participants: Mathieu Desroches, Soledad Fernández-García [Inria Paris-Rocquencourt, France], Maciej Krupa.

We construct a piecewise-linear (PWL) approximation of the Hindmarsh-Rose (HR) neurone model that is minimal, in the sense that the vector field has the least number of pieces, in order to reproduce all the dynamics present in the original HR model with the classical parameter values. This includes spiking, square-wave bursting, and also special trajectories called canards, which possess long repelling segments and organise the transition between stable bursting patterns with n and $n+1$ spikes. This is the spike-adding canard explosion. We propose a first approximation of the smooth bursting model, using a continuous PWL system, and show that its fast subsystem cannot possess a homoclinic bifurcation, which is necessary to obtain proper square-wave bursting. We then relax the assumption of continuity of the vector field across all zones and show that we can obtain a homoclinic bifurcation in the fast subsystem. We use the recently developed canard theory for PWL systems in order to reproduce the spike-adding canard explosion feature of the HR model as studied, e.g., in [66].

This work has been submitted for publication and is available as [45].

6.4.7. Ducks in space

Participants: Daniele Avitabile [University of Nottingham, UK], Mathieu Desroches, Edgar Knobloch [University of California at Berkeley, USA], Maciej Krupa.

A subcritical pattern-forming system with nonlinear advection in a bounded domain is recast as a slow-fast system in space and studied using a combination of geometric singular perturbation theory and numerical continuation. Two types of solutions describing the possible location of stationary fronts are identified, one of which is present for all values of the bifurcation parameter while the other is present for zero or sufficiently small inlet boundary conditions but only when the bifurcation parameter is large enough. For slightly larger inlet boundary condition a continuous transition from one type to the other takes place as the bifurcation parameter increases. The origin of the two solution types is traced to the onset of convective and absolute instability on the real line. The role of canard trajectories in the transitions between these states is clarified and the stability properties of the resulting spatial structures are determined. Front location in the convective regime is highly sensitive to the upstream boundary condition and its dependence on this boundary condition is studied using a combination of numerical continuation and Monte Carlo simulations of the partial differential equation. Statistical properties of the system subjected to random or stochastic boundary conditions are interpreted using the deterministic slow-fast spatial-dynamical system.

This work has been submitted for publication and is available as [43].

6.5. Spike Train statistics

6.5.1. Statistical models for spike trains analysis in the retina.

Participant: Bruno Cessac.

Recent advances in multi-electrodes array acquisition have made it possible to record the activity of up to several hundreds of neurons at the same time and to register their collective activity (spike trains). For the retina, this opens up new perspectives in understanding how retinal structure and ganglion cells encode information about a visual scene and what is transmitted to the brain. Especially, two paradigms can be confronted: in the first one, ganglion cells encode information independently of each others; in the second one non linear dynamics and connectivity contribute to produce a population coding where spatio-temporal correlations, although weak, play a significant role in spike coding. Confronting these two paradigms can be done at an experimental and at a theoretical level. On experimental grounds, new methods to analyse the role of weak correlations in spike train statistics are required. On theoretical grounds, mathematical results have been established, in neuronal models, showing how non linear dynamics and connectivity contribute to produce a correlated spike response to stimuli. In the context of the ANR KEOPS project, we have been working on these two aspects and we present our main results.

This work is available as [33].

6.5.2. Spectral dimension reduction on parametric models for spike train statistics

Participants: Cesar Ravello, Ruben Herzog, Bruno Cessac, Maria-Jose Escobar, Adrian Palacios.

It has been shown that the neurons of visual system present correlated activity in response to different stimuli. The role of these correlations is an unresolved subject. These correlations vary according to the stimulus, specially with natural images. To uncover the role of these correlation and characterize the population code, it is necessary to measure the simultaneous activity of large neural populations. This has been achieved thanks to the advent of Multi-Electrode Array technology, opening up a way to better characterize how the brain encodes information in the concerted activity of neurons. In parallel, powerful statistical tools have been developed to accurately characterize spatio-temporal correlations between neurons. Methods based on *Maximum Entropy Principle*, where statistical entropy is maximized under a set of constraints corresponding to specific assumptions on the relevant statistical quantities, have been proved successfully, specially when they consider *spatiotemporal* correlations. They are although limited by (i) **the assumption of stationarity**, (ii) **the many possible choice of constraints**, and (iii) **the huge number of free parameters**. We present our results on these two aspects obtained in the context of ANR KEOPS.

This work is available as [54].

6.6. Visual Neuroscience

6.6.1. Shifting stimulus for faster receptive fields estimation of ensembles of neurons

Participants: Bruno Cessac, Matthias Hennig [University of Edinburg, UK], Gerrit Hilgen [Institute of Neuroscience, Medical School, Newcastle University, Newcastle, UK], Pierre Kornprobst, Daniela Pamplona, Sahar Pirmoradian [University of Edinburg, UK], Evelyne Sernagor [Institute of Neuroscience, Medical School, Newcastle University, Newcastle UK].

The Spike Triggered Average (STA) is a classical technique to find a discrete approximation of the Receptive Fields (RFs) of sensory neurons [63], a required analysis in most experimental studies. One important parameter of the STA is the spatial resolution of the estimation, corresponding to the size of the blocks of the checkerboard stimulus images. In general, it is experimentally fixed to reach a compromise: If too small, neuronal responses might be too weak thus leading to RF with low Signal-to-Noise-Ratio; on the contrary, if too large, small RF will be lost, or not described with enough details, because of the coarse approximation. Other solutions were proposed consisting in starting from a small block size and updating it following the neuron response in a closed-loop to increase its response [70], [78], [77]. However, these solutions were designed for single cells and cannot be applied to simultaneous recordings of ensembles of neurons (since each RF has its own size and preferred stimulus).

To solve this problem, we introduced a modified checkerboard stimulus where blocks are shifted randomly in space at fixed time steps. This idea is inspired from super-resolution techniques developed in image processing [84]. The main interest is that the block size can be large, enabling strong responses, while the resolution can be finer since it depends on the shift minimum size. In [52], we show that the STA remains an unbiased RF estimator and, using simulated spike trains from an ensemble of Linear Nonlinear Poisson cascade neurons, it was predicted that this approach improves RF estimation over the neuron ensemble, in terms of resolution and convergence. In [53], we test these predictions experimentally on the RFs estimation of 8460 ganglion cells from two mouse retinas, using recordings performed with a large scale high-density multielectrode array. We compare RFs obtained using (i) the classical checkerboard stimulus with block size of 160μm and (ii) our checkerboard stimulus with block size of 160μm and arbitrary shifts of 40μm in $x-$ and $y-$directions. Results show how spatial resolution can be improved and that our approach allows to recover 51$\%$ of the mapped RFs at a resolution of 40μm, while in the classical case, 41$\%$ of the RFs could be found at a resolution of only 160μm. Thus, our approach improves not only the quality of the RF estimation but also the amount of successfully mapped RFs in neural ensembles.

This work was presented in [52], [53] and it is being used in current experimental protocols by E. Sernagor (Newcastle University), partner of the EC IP project FP7-ICT-2011-9 no. 600847 (RENVISION).

6.6.2. Using neural mechanisms underlying motion analysis for optical flow estimation

Participants: Manuela Chessa [University of Genoa, DIBRIS, Italy], Pierre Kornprobst, Guillaume S. Masson [Institut de Neurosciences de la Timone, Team InVibe], Kartheek Medathati, Fabio Solari [University of Genoa, DIBRIS, Italy].

We explore how motion information, also called optical flow, is estimated from natural moving sequences. Owing to application potential, optical flow estimation has been studied extensively by computer vision. On the other hand the neural mechanisms underlying motion analysis in the visual cortex have been extensively studied almost with little interaction with computer vision community resulting in few mathematical models. Even though there was some early interaction among the two communities for example, methods by Heeger et.al, Sejnowski et. al, comparatively little work has been done in terms of examining or extending the mathematical models proposed in biology in terms of their engineering efficacy on modern optical flow estimation datasets.

Pursuing this idea, in [26], we proposed a neural model inspired from the ones presented in [87], [86] which are popular models of primate velocity encoding. We started from a classical V1-MT feedforward architecture. We modeled V1 cells by motion energy (based on spatio-temporal filtering), and MT pattern cells (by pooling V1 cell responses). The efficacy of this architecture and its inherent limitations in the case of real videos were not known. To answer this question, we proposed a velocity space sampling of MT neurones (using a decoding scheme to obtain the local velocity from their activity) coupled with a multi-scale approach. After this, we explored the performance of our model on the Middlebury dataset. To the best of our knowledge, this is the only neural model in this dataset. The results were promising and suggested several possible improvements, in particular to better deal with discontinuities. An extension was proposed in [40].

We also focused on the decoding the motion energies which is of natural interest for developing biologically inspired computer vision algorithms for dense optical flow estimation. In [37], we addressed this problem by evaluating four strategies for motion decoding: intersection of constraints, maximum likelihood, linear regression on MT responses and neural network based regression using multi scale-features. We characterized the performances and the current limitations of the different strategies, in terms of recovering dense flow estimation using Middlebury benchmark dataset widely used in computer vision, and we highlight key aspects for future developments.

This work was partially funded by the EC IP project FP7-ICT-388 2011-8 no. 318723 (MatheMACS).

6.6.3. Bio-Inspired Computer Vision: Towards a Synergistic Approach of Artificial and Biological Vision

Participants: Pierre Kornprobst, Guillaume S. Masson [Institut de Neurosciences de la Timone, Team InVibe], Kartheek Medathati, Heiko Neumann [Ulm University, Germany].

Studies in biological vision have always been a great source of inspiration for design of computer vision algorithms. In the past, several successful methods were designed with varying degrees of correspondence with biological vision studies, ranging from purely functional inspiration to methods that utilise models that were primarily developed for explaining biological observations. Even though it seems well recognised that computational models of biological vision can help in design of computer vision algorithms, it is a non-trivial exercise for a computer vision researcher to mine relevant information from biological vision literature as very few studies in biology are organised at a task level.

In [42], we aim to bridge this gap by providing a computer vision task centric presentation of models primarily originating in biological vision studies. Not only we revisit some of the main features of biological vision and discuss the foundations of existing computational studies modelling biological vision, but also consider three classical computer vision tasks from a biological perspective: image sensing, segmentation and optical flow. Using this task-centric approach, we discuss well-known biological functional principles and compare them with approaches taken by computer vision. Based on this comparative analysis of computer and biological vision, we present some recent models in biological vision and highlight a few models that we think are promising for future investigations in computer vision. To this extent, this paper provides new insights and a starting point for investigators interested in the design of biology-based computer vision algorithms and pave a way for much needed interaction between the two communities leading to the development of synergistic models of artificial and biological vision.

[42] is under review. This work was partially funded by the EC IP project FP7-ICT-388 2011-8 no. 318723 (MatheMACS).

SCALE Team

7. New Results

7.1. Programming Languages for Distributed Systems

7.1.1. Multi-active Objects

Participants: Ludovic Henrio, Justine Rochas, Vincenzo Mastandrea.

The active object programming model is particularly adapted to easily program distributed objects: it separates objects into several *activities*, each manipulated by a single thread, preventing data races. However, this programming model has its limitations in terms of expressiveness – risk of deadlocks – and of efficiency on multicore machines. We proposed to extend active objects with *local multi-threading*. We rely on declarative *annotations* for expressing potential concurrency between requests, allowing easy and high-level expression of concurrency. This year we realized the following:

- We proved the correctness of our compiler from ABS language into ProActive multi-active objects. This translation can be generalised to many other active object languages. This work has been published as a research report, and is under submission to a conference. The proof brought us very deep and interesting understanding on the differences between the languages.
- We started to work on static detection of deadlocks for multi-active object. This is the work of Vincenzo Mastandrea who is starting a Labex PhD in collaboration with the FOCUS EPI (Univ of Bologna). An article is currently submitted to a conference on this subject.
- We are formalising in Isabelle/HOL a first version of the semantics of multiactive objects. This work was done in collaboration with Florian Kammuller
- We organised a workshop on active object languages with the main teams in Europe involved in the development of active-object languages. A journal survey paper on the subject is currently being written.
- We implemented a debugger for multi active object programs.

We plan to continue to improve the model, especially about compile-time checking of annotations and about fault tolerance of multiactive objects.

7.1.2. Behavioural Semantics

Participants: Ludovic Henrio, Eric Madelaine, Min Zhang, Siqi Li.

We are conducting a large study on Parameterised Networks of Automata (pNets) from a theoretical perspective. We started last year with some 'pragmatic' expressiveness of the pNets formalism, showing how to express a wide range of classical constructs of (value-passing) process calculi, but also complex interaction patterns used in modern distributed systems. After publishing those results [13], we focused on open systems and our formalism is able to represent operators of composition of processes, they are represented as hierarchically composed automata with holes and parameters. We defined a semantics for open pNets and a bisimulation theory for them. This study was driven by several usecase examples including a hierarchical broadcast algorithm and several operators of concurrent processes. A short presentation is accepted for publication in the journal "Science China: Information Sciences -". A full paper on the subject of the semantics and bisimulation for open pNets is under submission to a conference.

In parallel we have started the study of a denotational semantics for open pNets, based on the Universal Theory of Processes (UTP). The idea in the long term would be to draw links between the operational, denotational, and algebraic models of the pNet formalism. A short presentation of our preliminary results will be presented at the conference PDP'16 (work in progress session).

7.1.3. GPU-based High Performance Computing for finance

Participants: Michael Benguigui, Françoise Baude.

We have pursued our work on pricing American multi-dimensional (so very computation intensive) options in finance and we hav been able to extend this to the computation of Value At Risk (consists in repeating the American option pricing, but we have found a financial grounded optimization that avoids us to replicate the most time consuming phase).

Moreover, the balancing of work is taking in consideration the heterogeneous nature of the involved GPUs, and is capable to harness the computing power of multi-core CPUs that also support running OpenCL codes. As our scheduling solution is capable to get a reasonable prediction of the workload of each slave computation, we have leveraged this to run the whole pricing and VaR computations onhybrid and heterogeneous clusters. These last results have been incorporated in the PhD thesis of M. Benguigui.

7.1.4. Scalable and robust Middleware for distributed event based computing

Participants: Maeva Antoine, Fabrice Huet, Françoise Baude.

In the context of the FP7 STREP PLAY and French SocEDA ANR research projects terminated late 2013, we initiated and pursued the design and development of the Event Cloud.

As a distributed system handling huge amount of information, this middleware can suffer from data imbalances. In a journal extension of a previous workshop paper [6], we have enlarged our litterature review of structured peer to peer systems regarding the way they handle load imbalance to the case of distributed big data systems. We have generalized those popular approaches by proposing a core API that we have proved to be indeed also applicable to the Event Cloud middleware way of implementing a load balancing policy.

7.1.5. Vercors: Integrated environment for verifying and running distributed components

Participants: Ludovic Henrio, Oleksandra Kulankhina, Eric Madelaine.

It is the general prurpose of the Vercors platform to target the generation of distributed applications with safety guarantees. In Vercors, the approach starts from graphical specification formalisms allowing the architectural and behavioral description of component systems. From this point, the user can automatically verify application properties using model-checking techniques. Finally, the specified and verified component model can be translated into executable Java code. The Vercors tool suite is distributed as an Eclipse plugin. This year

- we implemented a first reliable version of the whole tool chain including generation of verifiable models and executable Java code.
- We applied the approach to several examples including Peterson's leader election algorithm, a workflow executor, and the control and management of service composition [7].
- A paper accepted at FASE'2016 presents an overview of this work; a research report provides the full version of the paper [20]. The theoretical background was published as a research report and an improve version is being submitted as a journal paper.

The practical implementation allowed us to improve the presentation of the theory and better evaluate it.

7.2. Run-time/middle-ware level

7.2.1. Virtual Machines Scheduling

Participants: Fabien Hermenier, Vincent Kherbache.

In [19], we present BtrPlace as an application of the dynamic bin packing problem with a focus on its dynamic and heterogeneous nature. We advocate flexibility to answer these issues and present the theoretical aspects of BtrPlace and its modeling using Constraint Programming.

We also continued our work on scheduling VM migrations. In [14], [17], we propose a model for VM migration that consider their memory workload and the network topology. This model was then implemented in place of the previous migration scheduler in BtrPlace. Experiments on a real testbed show the new scheduler outperforms state-of-the-art approaches that cap the migration parallelism by a constant to reduce the completion time. Besides an optimal capping, it reduces the migration duration by 20.4% on average and the completion time by 28.1%. In a maintenance operation involving 96 VMs to migrate between 72 servers, it saves 21.5% Joules against the native BtrPlace. Finally, its current library of 6 constraints allows administrators to address temporal and energy concerns, for example to adapt the schedule and fit a power budget.

Finally, in [10] we transfer the principles of using Constraint Programming to propose a multi-objective job placement algorithm devoted to High Performance Computing (HPC). One of the key decisions made by both MapReduce and HPC cluster management frameworks is the placement of jobs within a cluster. To make this decision, they consider factors like resource constraints within a node or the proximity of data to a process. However, they fail to account for the degree of collocation on the cluster's nodes. A tight process placement can create contention for the intra-node shared resources, such as shared caches, memory, disk, or network bandwidth. A loose placement would create less contention, but exacerbate network delays and increase cluster-wide power consumption. Finding the best job placement is challenging, because among many possible placements, we need to find one that gives us an acceptable trade-off between performance and power consumption. We then propose to tackle the problem via multi-objective optimization. Our solution is able to balance conflicting objectives specified by the user and efficiently find a suitable job placement.

7.3. Application level

7.3.1. DEVS-based Modeling & Simulation

Participants: Olivier Dalle, Damian Vicino.

DEVS is a formalism for the specification of discrete-event simulation models, proposed by Zeigler in the 70's, that is still the subject of many research in the simulation community. Surprisingly, the problem of representing the time in this formalism has always been somehow neglected, and most DEVS simulators keep using Floating Point numbers for their arithmetics on time values, which leads to a range of systematic errors, including severe ones such as breaking the causal relations in the model.

In [15] we propose simulation algorithms, based on the Discrete Event System Specification (DEVS) formalism, that can be used to simulate and obtain every possible output and state trajecto-ries of simulations that receive input values with uncertainty quantification. Then, we present a subclass of DEVS models, called Finite Forkable DEVS (FF-DEVS), that can be simulated by the proposed algorithms. This subclass ensures that the simulation is forking only a finite number of processes for each simulation step. Finally, we discuss the simulation of a traffic light model and show the trajectories obtained when it is subject to input uncertainty.

We have also worked on improving the simulation of DEVS models in some particular situations[16]. Parallel Discrete Event System Specification (PDEVS), for example, is a well-known formalism used to model and simulate Discrete Event Systems. This formalism uses an abstract simulator that defines a set of abstract algorithms that are parallel by nature. To implement simulators using these abstract algorithms , several architectures were proposed. Most of these architectures follow distributed approaches that may not be appropriate for single core processors or microcontrollers. In order to reuse efficiently PDEVS models in this type of systems, we define a new architecture that provides a single threaded execution by passing messages in a call/return fashion to simplify the execution time analysis.

This work has also been presented and defended in the PhD Thesis of D. Vicino[5].

7.3.2. Simulation of Software-Defined Networks

Participants: Olivier Dalle, Damian Vicino.

Software Defined Networks (SDN) is a new technology that has gained a lot of attention recently. It introduces programmatic ways to reorganize the network logical topology. To achieve this, the network interacts with a set of controllers, that can dynamically update the configuration of the network routing equipments based on the received events. As often with new network technologies, discrete-event simulation proves to be an invaluable tool for understanding and analzing the performance and behavior of the new systems. In [8], we use such smulations for evaluating the impact of Software-Defined Networks' Reactive Routing on BitTorrent performance. Indeed, BitTorrent uses choking algorithms that continuously open and close connections to different peers. Software Defined Networks implementing Reactive Routing may be negatively affecting the performances of the system under specific conditions because of it lack of knowledge of BitTorrent strategies.

STARS Project-Team

7. New Results

7.1. Introduction

This year Stars has proposed new results related to its three main research axes : perception for activity recognition, semantic activity recognition and software engineering for activity recognition.

7.1.1. Perception for Activity Recognition

Participants: Julien Badie, Slawomir Bak, Piotr Bilinski, François Brémond, Duc Phu Chau, Etienne Corvée, Antitza Dancheva, Kanishka Nithin Dhandapani, Carolina Garate, Furqan Muhammad Khan, Michal Koperski, Thi Lan Anh Nguyen, Javier Ortiz, Ujjwal Ujjval.

The new results for perception for activity recognition are:

- Pedestrian Detection using Convolutional Neural Networks (see 7.2)
- Head detection for eye tracking application (see 7.3)
- Minimizing hallucination in Histogram of Oriented Gradients (see 7.4)
- Hybrid approaches for Gender estimation (see 7.5)
- Automated Healthcare: Facial-expression-analysis for Alzheimer's patients in musical mnemotherapy (see 7.6)
- Robust Global Tracker based on an Online Estimation of Tracklet Descriptor Reliability (see 7.7)
- Optimizing people tracking for a video-camera network (see 7.8)
- Multi-camera Multi-object Tracking and Trajectory Fusion (see 7.9)
- Person Re-Identification in Real-World Surveillance Systems (see 7.10)
- Human Action Recognition in Videos (see 7.11)

7.1.2. Semantic Activity Recognition

Participants: Vasanth Bathrinarayanan, François Brémond, Duc Phu Chau, Serhan Cosar, Alvaro Gomez Uria Covella, Carlos Fernando Crispim Junior, Ramiro Leandro Diaz, Giuseppe Donatielo, Baptiste Fosty, Carolina Garate, Alexandra Koenig, Michal Koperski, Farhood Negin, Thanh Hung Nguyen, Min Kue Phan Tran, Philippe Robert.

For this research axis, the contributions are :

- Evaluation of Event Recognition without using Ground Truth (see 7.12)
- Semantic Event Fusion of Different Visual Modality Concepts for Activity Recognition (see 7.13)
- Semi-supervised activity recognition using high-order temporal-composite patterns of visual concepts (see 7.14)
- From activity recognition to the assessment of seniors' autonomy (see 7.15)
- Serious Games Interfaces using an RGB-D camera (see 7.16)
- Assistance for Older Adults in Serious Game using an Interactive System (see 7.17)
- Generating Unsupervised Models for Online Long-Term Daily Living Activity Recognition (see 7.18)

7.1.3. Software Engineering for Activity Recognition

Participants: Sabine Moisan, Annie Ressouche, Jean-Paul Rigault, Ines Sarray, Imane Khalis, Nazli Temur, Daniel Gaffé, Rachid Guerchouche, Matias Marin, Etienne Corvée, Carolina Da Silva Gomes Crispim, Anais Ducoffe, Jean Yves Tigli, François Brémond.

The contributions for this research axis are:

- Run-time Adaptation of Video Systems (see 7.19)
- Scenario Description Language (see 7.20)
- Scenario Recognition (see 7.21)
- The Clem Workflow (see 7.22)
- Safe Composition in WComp Middleware for Internet of Things (see 7.23)
- Design of UHD panoramic video camera (see 7.24)
- Brick & Mortar Cookies (see 7.25)
- Monitoring Older People Experiments (see 7.26)

7.2. Pedestrian Detection using Convolutional Neural Networks

Participants: Ujjwal Ujjwal, François Brémond.

Keywords: Pedestrian detection, CNN

The objective of the work was to perform pedestrian detection in different settings. The settings corresponded to different types of camera-views as well as different types of camera settings (e.g- moving camera vs. static camera). The work followed a wide range of experiments using different public implementations of convolutional neural networks and on different types of datasets. We detail the experiments one by one in the following subsections :

Experiments on CNN architectures

We started with an evaluation of different CNN architectures for pedestrian detection. Towards this end, we implemented three important and famous architectures - LeNet [72], AlexNet [69] and CifarNet [68]. For the purpose of training and validation we extracted patches from the public datasets of Inria [55], Daimler [58], TUD-Brussels [92], Caltech [57], ViPer [62], USC [93] and MIT [78]. The breakup of the dataset used for training was as shown in table 1 .

Table 1. Training, testing and validation patch details for CNN training

	Pedestrian Patches	Non-Pedestrian Patches
Training	131,183	61,500
Validation	65,591	30,700
Testing	65,591	30,600

Implementation of all the three models for pedestrian detection which gave a very high accuracy (94.2% (LeNet), 98% (AlexNet) and 98.2% (CifarNet)) for classification at patch level. Though these results were good at the patch level, more thorough understanding was needed to determine the effect of network architecture on classification. This was important because the three architectures vary greatly in terms of number of layers and other parameters. Moreover the practical problem in pedestrian detection chiefly involves detecting pedestrians in an image (i.e. when full-scale images instead of pre-defined patches are available).

The first set of experiments was done using sliding windows. This had to be abandoned soon, since for each image this was taking an impractical time ($> 3 minutes/image$). Moreover sliding window is less suited in its naive setting due to the fact that each candidate window had to be rescaled to meet the network input size and tested individually by extracting features over it. This was followed by efforts to understand and implement a wide range of other techniques for full-scale detection using CNN. This is still an open problem though some encouraging advancements through R-CNN [61] and OverFeat [83] have been made. A major difficulty lies in lack of robust implementations of CNN which allow for integrated training and testing with object localization. Moreover existing implementations are less flexible and often make it difficult to carry out modifications required to implement new techniques independently.

We settled with the R-CNN which uses region proposals extracted using selective search [87] to extract object proposals and then train a CNN using those proposals and subsequently classifying using a SoftMax classifier or a SVM.

The evaluation was done on both moving cameras and static cameras and the evaluation showed that the network was performing a little satisfactorily, though below the state-of-art performance standards. The performance metric was Average Miss Rate (AMR) Vs. False Positives Per Image (FP/I). A good detector must exhibit a very low AMR alongwith very low FP/I. Table 2 summarizes the detection results, with table 3 summarizing state-of-art results on different pedestrian detection datasets.

Table 2. Our R-CNN results on different pedestrian datasets.

Dataset	#Images	AMR	FP/I
Inria	741	0.27	0.36
DAIMLER	15K	0.38	0.40
Caltech	16K	0.46	0.43
USC	584	0.02	0
PETS 2009 S2.L1	5565	0.42	0.29
PETS 2009 S2.L2	1744	0.35	0.19

Table 3. State-of-art results on different pedestrian datasets.

Dataset	#Images	AMR	FP/I
Inria	741	0.14	0.1
DAIMLER	15K	0.29	0.21
Caltech	16K	0.12	0.1
PETS 2009 S2.L1	5565	0.22	0.1
PETS 2009 S2.L2	1744		

It was felt that more work is needed to organize a CNN library and subsequently work to improve the above results.

Conclusion

Pedestrian detection finds its applications in different settings. It is also highly influenced by a wide variety of variations which have many practical ramifications in areas such as surveillance. It is important to develop a robust and high-performance system for pedestrian detection that is able to take into account a very wide range of such variations such as occlusion and poor visibility. CNNs have shown great promise in object detection and recognition lately and this inspires it growing applications in pedestrian detection. While the current results of our R-CNN experiments do not match the state-of-art it has shown some promise by providing consistent numbers across datasets which shows that CNNs are a good way to transcend a system beyond dataset-specific restrictions. An important factor is the instance of moving cameras vis-a-vis static cameras. While the present experiments show that decent performance is obtained on moving camera databases, consistent and similar performance is also obtained in the context of static camera databases such as PETS. This shows that with better training and improved practices of dataset handling such as augmentation and dataset structuring by clustering based methods can help in pushing the performance to acceptable levels for applications in automated surveillance and driving applications.

Further Work We intend to take this study forward, by looking into novel approaches to gather more information about a pedestrian dataset from CNN, while further increasing the detection results.

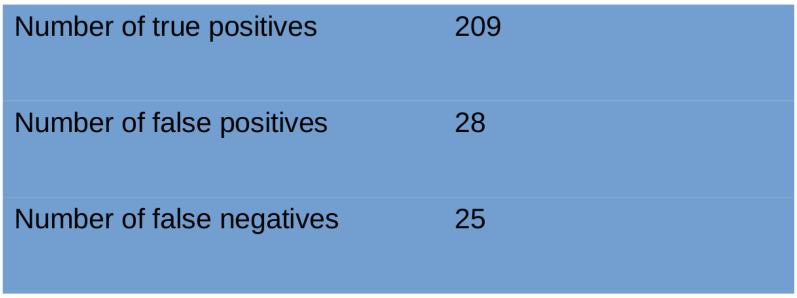
7.3. Head Detection for Eye Tracking Application

Participants: Thanh Hung Nguyen, Antitza Dantcheva, François Brémond.

Keywords: computer vision, head detection, eye tracking

Head detection [77] uses RGBD sensor (Kinect 2 sensor) which is supported by SUP platform of STARS team. For the eye tracking, we use the open-source library (OpenBR) which has good performance in our test.

Until now, the head detector was working well when people were standing but not good enough when people were lying or bending as you can see on Figure 5 and Figure 6 . This experiment was realised on simple datasets where mostly people was close to camera and walking. The main reason is the lack of samples in the learning process for the challenges cases (lying, bending). So at this moment we are collecting the head samples for it.



Number of true positives	209
Number of false positives	28
Number of false negatives	25

Figure 5. Performance of head detection on simple dataset

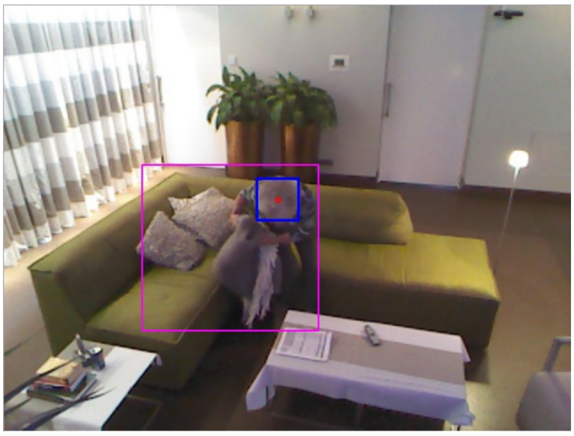


Figure 6. An example of experiment

7.4. Minimizing Hallucination in Histogram of Oriented Gradients

Participants: Javier Ortiz, Slawomir Bak, Michal Koperski, François Brémond.

keywords: computer vision, action recognition, re-identification,

Challenges in histogram representation

In computer vision very often histogram of values is used as a feature representation. For instance HOG descriptor is in fact histogram of gradient orientations. Also histogram of codewords in Bag of Words representation is a very popular action recognition representation. In such a case we are not interested in absolute values for given bin, but rather in shape of histogram to find the patterns. To make histogram representation independent from absolute values we use L1 or L2 normalization. In the normalization process we treat histogram as a vector and we transform it in the way that it should have unit length according to given norm (L1, L2). The drawback of such approach is that normalization process may amplify the noise for histograms where absolute values are very small (no pattern or only noise). Such histogram after normalization can be very similar to histogram with strong absolute values. This situation is showed in figure 7 . Although original histogram in second row of figure 7 contains almost no information after normalization the values are amplified and the result is exactly the same as for histogram in the first row. Such behavior is called hallucinations.

Proposed normalization method

We propose to add an artificial bin with given value which would prevent small noisy values from being amplified during normalization process. In figure 7 in first row we show artificial bin in pink color. Thanks to that histograms after normalization (last column) are different. If we analyze cumulative sum of histogram values across data (we sum values of all bins in whole histogram and we draw distribution), we can find that in some data-sets we obtain bi-nomial distribution see figure 8 . The gap between two Gaussians indicates convenient border between noise and meaningful data. On the other hand many data-sets do not have this bi-nomial feature and for that we do not have formalization to find the value of artificial bin. This problem would be a subject of further studies.

The following method was successfully applied to person re-identification and action recognition problem. Further details can be found in the paper [42]

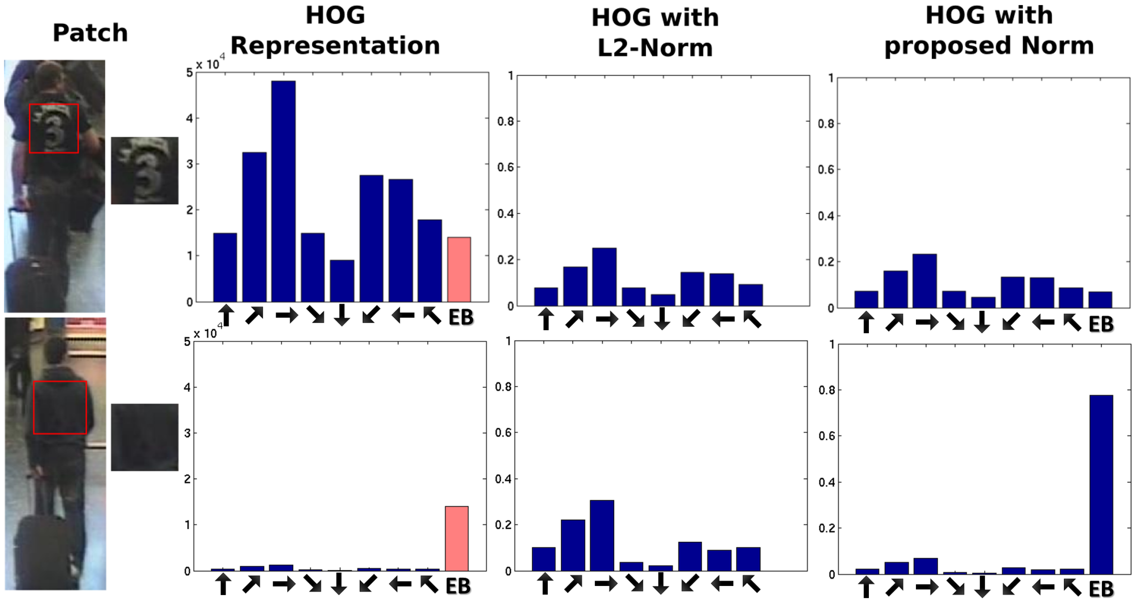


Figure 7. HOG representations of two patches with different amount of texture. Each bin in blue represents the sum of magnitudes of edges for a particular orientation. The bin marked as EB represents the proposed extra bin

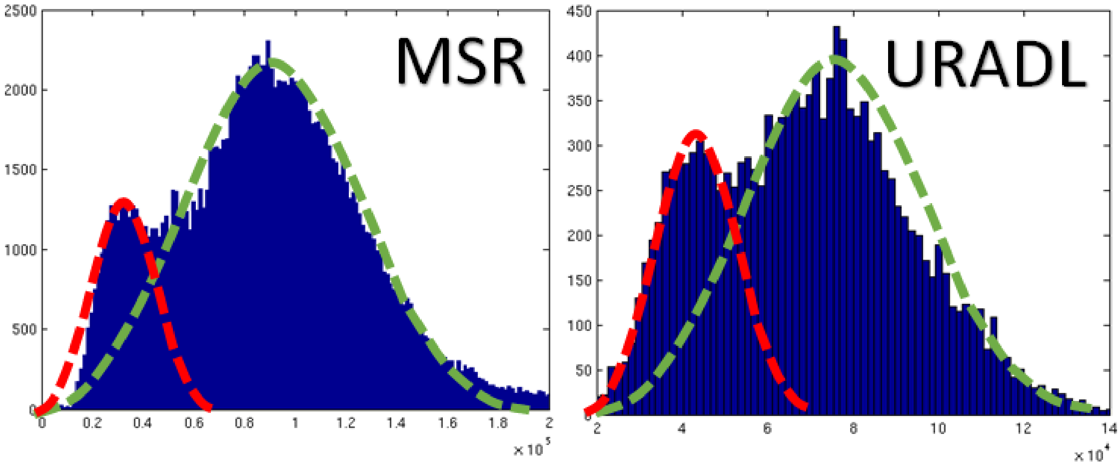


Figure 8. Distribution of the sum of bins for randomly sampled patches for URADL and MSRDailActivity3D data-set.

7.5. Hybrid Approaches for Gender estimation

Participants: Antitza Dantcheva, François Brémond, Philippe Robert.

keywords: gender estimation, soft biometrics, biometrics, visual attributes

Automated gender estimation has numerous applications including video surveillance, human computer-interaction, anonymous customized advertisement and image retrieval. Gender estimation remains a challenging task, which is inherently associated with different biometric modalities including fingerprint, face, iris, voice, body shape, gait, as well as clothing, hair, jewelry and even body temperature [31]. Recent work has sought to further the gains of single-modal approaches by combining them, resulting into hybrid cues that offer a more comprehensive gender analysis, as well as higher resilience to degradation of any of the single sources.

Can a smile reveal your gender?

In this work we propose a novel method for gender estimation, namely the use of dynamic features gleaned from smiles and show that (a) facial dynamics can be used to improve appearance-based gender-estimation, (b) that while for adults appearance features are more accurate than dynamic features, for subjects under 18 years old facial dynamics outperform appearance features. While it is known that sexual dimorphism concerning facial appearance is not pronounced in infants and teenagers, it is interesting to see that facial dynamics provide already related clues. The proposed system, fusing a state-of-the-art appearance and dynamics-based features (see Figure 9), improves the appearance based algorithm from 78.0% to 80.8% in video-sequences of spontaneous smiles and from 80% to 83.1% in video-sequences of posed smiles for subjects above 18 years old (see Table 4). These results show that smile-dynamics include pertinent and complementary information to appearance gender information.

While this work studies video sequences capturing frontal faces expressing human smiles, we can envision that additional dynamics, such as other facial expressions or head and body movements carry gender information as well.

Distance-based gender prediction: What works in different surveillance scenarios?

Table 4. True gender classification rates. Age given in years.

Age	< 10	$10 - 19$	$20 - 29$	$30 - 39$	$40 - 49$	> 49
Subj. amount	48	95	60	49	72	33
OpenBR	58.33%	50.53%	81.67%	75.51%	75%	81.82%
Combined Age-Groups	< 20		> 19			
Subj. amount	143		214			
OpenBR	52.45%		78.04%			
Dynamics (SVM+PCA)	60.1%		69.2%			
Dynamics (AdaBoost)	59.4%		61.7%			
OpenBR + Dynamics (Bagged Trees)	60.8%		80.8%			

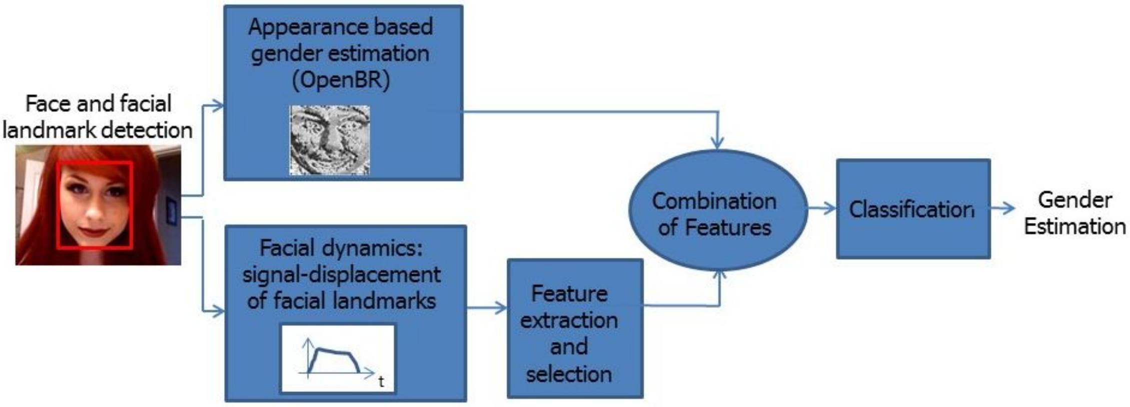


Figure 9. General Scheme of the facial appearance and dynamics framework.

In this work we fuse features extracted from face, as well as from body silhouette towards gender estimation. Specifically, for face, a set of facial features from the OpenBR library, including histograms of local binary pattern (LBP) and scale-invariant feature transform (SIFT) are computed on a dense grid of patches. Subsequently, the histograms from each patch are projected onto a subspace generated using PCA in order to obtain a feature vector, followed by a Support Vector Machine (SVM) used for the final face-based-gender decision. Body based features include geometric and color based features, extracted from body silhouettes, obtained by background subtraction, height normalization and SVM-classification for the final body-based-gender-decision. We present experiments on images extracted from video-sequences, emphasizing on three distance-based settings: close, medium and far from the TunnelDB dataset (see Figure 10). As expected, while face-based gender estimation performs best in the close-distance-scenario, body-based gender estimation performs best when the full body is visible - in the far-distance-scenario (see Table 5). A decision-level-fusion of face and body-based features channels the benefits of both approaches into a hybrid approach, providing results that demonstrate that our hybrid approach outperforms the individual modalities for the settings medium and close.

Table 5. True gender classification rates. Age given in years.

textbfSystem Distance	**FAR**	**MEDIUM**	**CLOSE**
FBGE	57.14	79.29	89.29
BBGE 89.3 85 79.3			
Fusion BBGE & FGBE	82.9	88.6	95

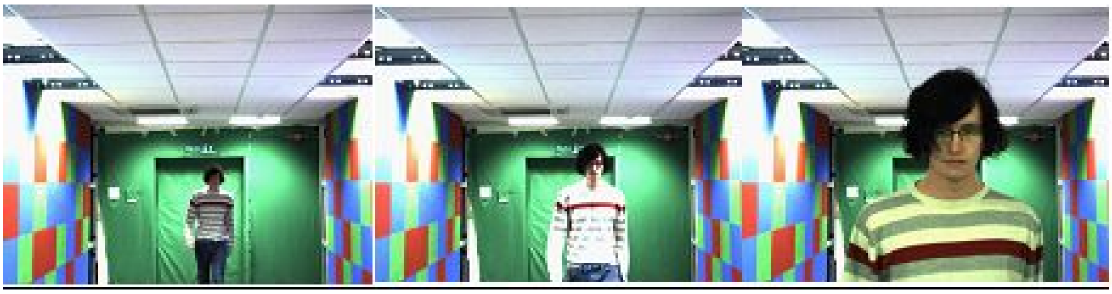


Figure 10. Three distance-based settings in the TunnelDB dataset: far, medium and close.

While the dataset is relatively unconstrained in terms of illumination, body and face are captured facing relatively frontally towards the camera. Future work will involve less constraints also towards the pose of humans.

7.6. Automated Healthcare: Facial-expression-analysis for Alzheimer's Patients in Musical Mnemotherapy

Participants: Antitza Dantcheva, François Brémond, Philippe Robert.

keywords: automated healthcare, healthcare monitoring, expression recognition

In this work we seek to apply computer vision towards increasing the life quality of patients with Alzheimer's disease (AD), and particularly in applying computer vision towards interventions to delay functional decline and to decrease the burden of the most disturbing behavioral symptoms. Towards this we design a smart interaction tool, that "reads" emotions of AD patients. This approach is becoming necessary now, because the increasing prevalence of chronic disorders and its impact on functional decline is challenging the sustainability of healthcare systems. Firstly, we have assembled a dataset of video-sequences acquired in the Alzheimer's Disease - clinique Fondation GSF Noisiez. Multiple patients and sessions have been captured during musical mnemotherapy. We then have annotated several sequences per one of four facial expressions, that occur in the recorded dataset including: neutral, talking, smile and sad. We then proceed to classify these expressions for 10 patients based on two approaches, that we study individually, as well as fused. The first approach contains face detection, facial landmark localization and signal displacement analysis for different facial landmarks, which are ranked based on categorization-pertinence, fused and classified into one of the four expression-categories (see Figure 11). In the second approach, we use face detection, eyes-detection, face normalization and HOG-features, which we classify into one of the four expression-categories.

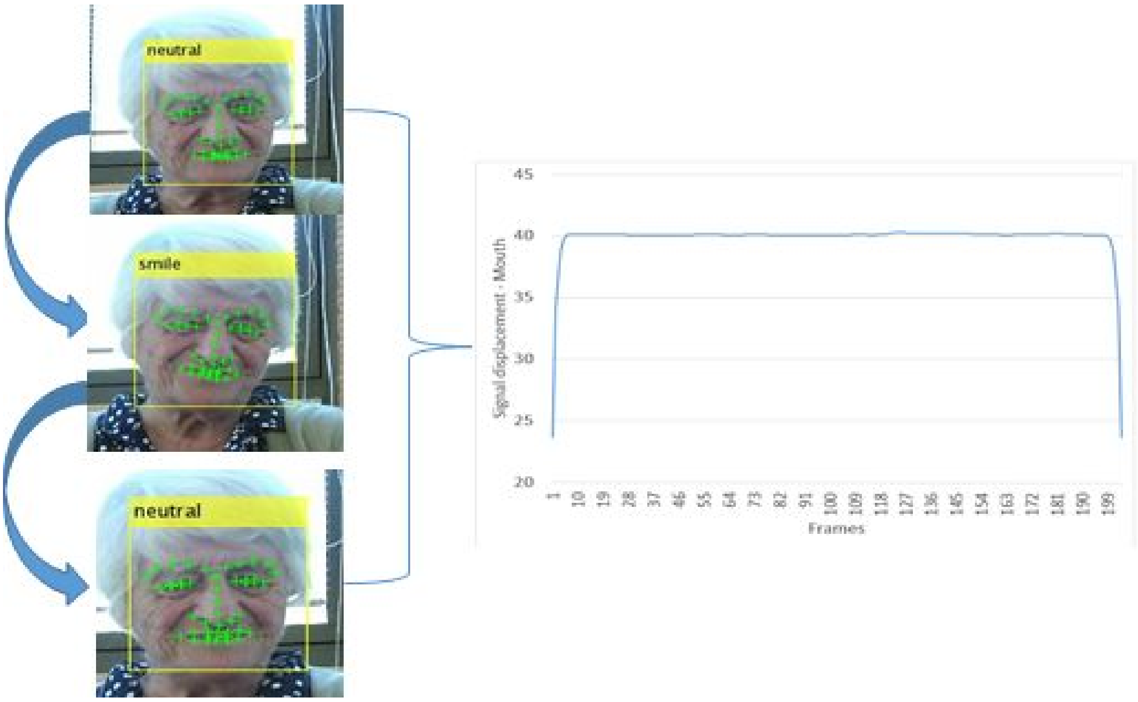


Figure 11. Smile detection based on signal displacement in the mouth region in the dataset collected at the Fondation GSF Noisiez in Biot, France.

The here used real-world-data challenges, as expected, all utilized computer vision algorithms (from face detection - due to no constraints of pose and illumination, to classifiers - due to a large intra-class-variation of facial expressions). Nevertheless, we obtain promising results that we envision improving by analyzing 2D and depth, as well as infrared data.

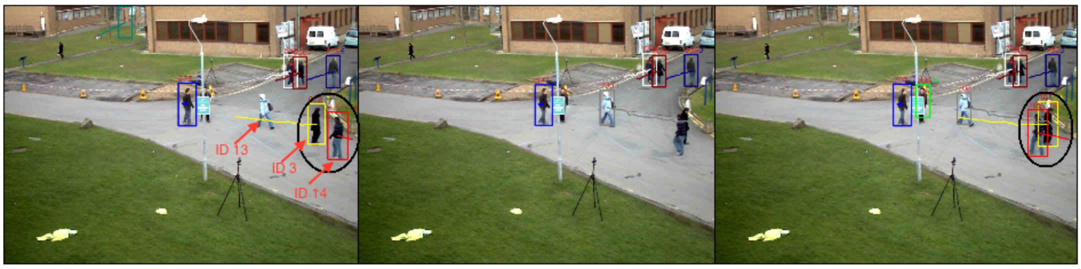
7.7. Robust Global Tracker Based on an Online Estimation of Tracklet Descriptor Reliability

Participants: Thi Lan Anh Nugyen, Chau Duc Phu, François Brémond.

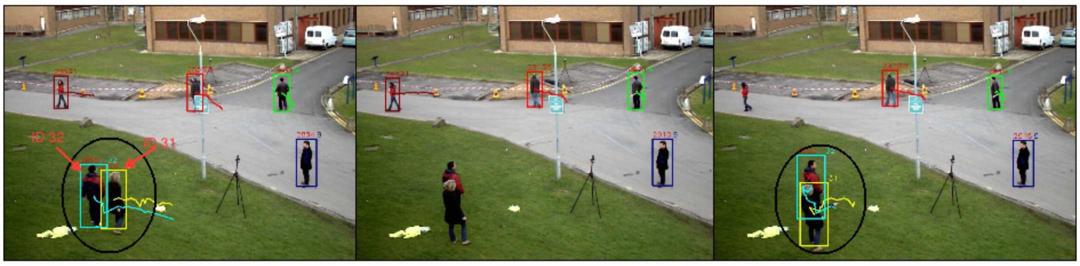
Keywords: Tracklet fusion, Multi-object tracking

Object tracking - the process of locating a moving object (or multiple objects) in one camera or in a camera network over time - is an important part in surveillance video processing. However, the video context variation requires trackers to face plenty of challenges. For example, objects change their movement direction or their appearances, poses; objects are occluded by other objects or background; illumination is changed... In order to overcome above challenges, calculating the object appearance model overtime to adapt tracker to context variation is a necessary work.

In the state of the art, some online learning approaches [52], [48] have been proposed to track objects in various video scenes in each frame. These approaches learn online discriminative object descriptors to the current background as in [52] or learn an object appearance model which discriminates objects overtime as in [48]. However, the limitation of these approaches is that the reliability of object descriptor computed on only current frame is sensitive. False positives can reduce tracking quality. Furthermore, these algorithms try to find the discriminative descriptors or signatures of one object compared to its neighborhood but not considering to the correlation of this object with its can-match candidates. Meanwhile, global tracking methods [91], [98] show their dominant ability over previous methods in noisy filtering. The approach in [98] proposes an algorithm that recovers fragmentation of object trajectories by using enhanced covariance-based signatures and an online threshold learning. The approach in [91] proposes a hierarchical relation hypergraph based tracker. These global tracking algorithms have significant results in matching short trajectories and filtering some noise. However, object descriptor weights are fixed for the whole video. Therefore, their tracking performances can be reduced if the scene changes.



(a) frame 137 (b) frame 140 (c) frame 161



(d) frame 553 (e) frame 563 (f) frame 865

Figure 12. PETs2009 dataset: The online computation of discriminative descriptor weights depending on each video scene.

In this work, we propose a new approach to improve the tracking quality by a global tracker which merges all tracklets belonging to an object in the whole video. Particularly, we compute online descriptor reliability over time based on their discrimination. Based on the computed discriminative descriptor weights, the global matching score over descriptors of 2 tracklets is given. Then, we apply Hungarian algorithm to optimize

Table 6. Tracking performance.The best values are printed in bold, the second best values are printed in italic.

Dataset	Method	MOTA	MOTP	GT	MT	PT	ML	FG
PETS2015	Chau *et Al.* [53]	–	–	2	0.0	100.0	**0.0**	2
	Ours (Proposed approach + [53])	–	–	2	**100.0**	0.0	**0.0**	**1**
PETS2009	Chau *et Al.* [53]	0.62	0.63	21	–	–	–	8
	Bae *et Al.with all* [48]	0.83	0.69	23	**100**	0	**0.0**	**4**
	Zamir **et Al.** [95]	**0.90**	0.69	21	–	–	–	–
	Bae **et Al.**-*global association* [48]	0.73	0.69	23	**100**	0	**0.0**	12
	Badie **et Al.** [47]	**0.90**	**0.74**	21	–	–	–	–
	Badie **et Al.** [47] + [53]	0.85	0.71	21	66.6	23.9	9.5	*6*
	Ours (Proposed approach + [53])	*0.86*	*0.72*	21	*76.2*	14.3	9.5	**4**
TUD-Stadtmitte	Milan *et Al.* [74]	**0.71**	**0.65**	9	**70.0**	20.0	**0.0**	–
	Yan *et Al.* [94]	–	–	10	**70.0**	30.0	**0.0**	–
	Chau *et Al.* [53]	0.45	0.62	10	60.0	40.0	**0.0**	13
	Ours (Proposed approach + [53])	0.47	**0.65**	10	**70.0**	30.0	**0.0**	**7**
TUD-Crossing	Tang *et Al.* [84]	–	–	–	**53.8**	38.4	7.8	–
	Chau *et Al.* [53]	0.69	0.65	11	46.2	53.8	**0.0**	14
	Ours (Proposed approach + [53])	**0.72**	**0.67**	11	**53.8**	46.2	**0.0**	**8**

tracklet matching. On the other hand, a motion model is also combined with appearance descriptors in a flexible way to improve the tracking quality. Figure12 shows the visual explanation . In frame 137, two objects have similar appearance but move with different direction. In this case, motion descriptor is more reliable. Inversely, in frame 553, two objects go consistently together but their coat and hair's colors are different. Therefore, the appearance descriptors are more reliable than motion one.

The proposed approach gets results of tracker in [53] as input and is tested on challenge datasets. The comparable results of this tracker with other trackers from the state of the art are shown in Table 6 . This paper is accepted in PETs workshop [41].

7.8. Optimizing People Tracking for a Video-camera Network

Participants: Julien Badie, François Brémond.

Keywords: tracking quality estimation, error recovering, tracklet matching

This work addresses the problem of improving tracking quality during runtime. Most state-of-the-art tracking or high-level algorithms such as event recognition have difficulties to handle erroneous inputs. This framework detects and repairs detection or tracking errors. It works in an online situation and even in the case where prior knowledge of the scene (such as contextual information or training data) is not available.
The Global Tracker (figure 13) uses tracking results (tracklets) as input and produces corrected tracklets as output.

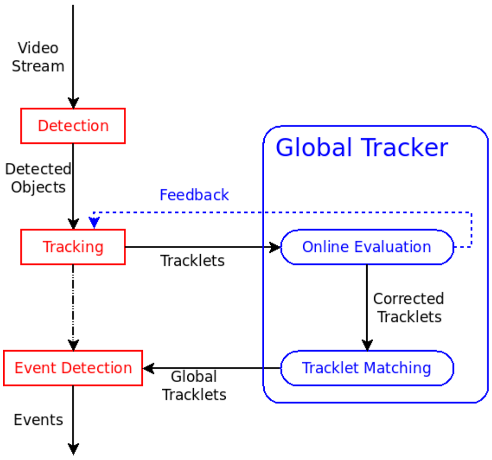


Figure 13. The Global Tracker framework, combining online evaluation and tracklet matching to improve tracking results.

The Global Tracker framework is divided into two main modules:

- **Online evaluation of tracking results**: the quality of the tracking results is computed by analyzing the variation of each tracklet feature. A significant feature variation is considered as a potential error, an anomaly. To determine if this anomaly is a real error or a natural phenomenon, we use information given by the object neighborhood and the context. Finally the errors are corrected either by removing the erroneous nodes (basic approach) or by sending a signal to the tracking algorithm in order to tune its parameters for the next frames (feedback approach).

- **Tracklets matching over time**: tracklets representing the same object are merged together in a four-step algorithm. First we select key frames (frames that are close to the mean value of the features) for each tracklet. Then a visual signature is computed based on these key frames. The distance between each pair of signature is then computed. Finally the tracklet merging is done using unsupervised learning and a constrained clustering algorithm where all tracklets representing the same object are put in the same cluster.

This approach has been tested on several datasets such as PETS2009 (table 7), CAVIAR (table 8), TUD, I-LIDS and VANAHEIM and with different kinds of scenarios (tracking associated with a controller, 3D camera, camera network with overlapping or distant cameras). In each case, we are able to reach or outperform the state-of-the-art results.

Table 7. Tracking results on sequence S2.L1.View1 of the PETS2009 dataset

Methods	MOTA	MOTP
Zamir et al. [95]	**0.90**	0.69
Milan et al. [75]	**0.90**	**0.74**
Online evaluation	0.90	0.74
Tracklet matching	0.83	0.68
Global Tracker	***0.92***	***0.76***

Table 8. Tracking results on the Caviar dataset

Methods	MT (%)	PT (%)	ML (%)
Li et al. [73]	84.6	14.0	1.4
Kuo et al. [70]	84.6	14.7	***0.7***
Online Evaluation	82.6	11.7	5.7
Tracklet matching	84.6	9.5	5.9
Global Tracker	***86.4***	***8.3***	5.3

This approach is described more in detail in the PhD manuscript [27].

7.9. Multi-camera Multi-object Tracking and Trajectory Fusion

Participants: Kanishka Nithin Dhandapani, Thi Lan Anh Nguyen, Julien Badie, François Brémond.

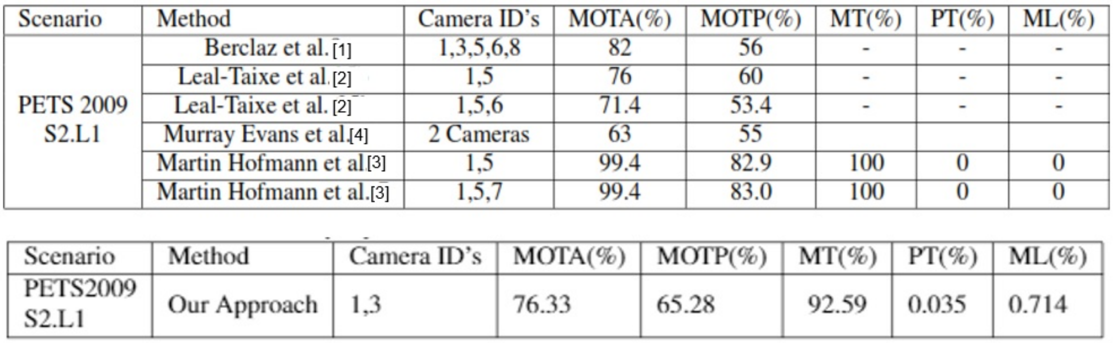
Keywords: Multicamera, Tracklet association, Trajectory fusion, Object Tracking.

In spite of number of solutions that exist for multi-object tracking, it is still considered most challenging and unsolved computer vision problems, mainly due to inter and intra-occlusions, inferior visibility in crowed scenes, object re-entry, abrupt movement of object, placement of cameras and other detection inaccuracies that occur in single camera. Such drawbacks in single-camera multi-target tracking can be solved to an extent by obtaining more visual information on the same scene (more cameras). Few works done in the past years are [50], [79], [65], [59]. However they have their own problems such as not real run time performance, complex optimizations, hypothesizing 3D reconstruction and data association together might lead to suboptimal solutions.

We present a multi-camera tracking approach that associates and performs late fusion of trajectories in a centralized manner from distributed cameras. We use multiple views of the same scene to recover information that might be missing in a particular view. For detection we use background subtraction followed by discriminatively Trained Part Based Models . For object tracking, we use an object appearance-based tracking algorithm introduced by Chau et al [54] that combines a large set of appearance features such as 2D size, 3D displacement, colour histogram, and dominant colour to increase the robustness of the tracker to manage occlusion cases. Each camera in the network runs the detection and tracking chain independent of each other in a distributed manner. After a batch of frames, the data from each camera is gathered to a central node by projecting the trajectories of people to the camera with the most inclusive view through a planar homography technique and then global association and fusion are performed. Unlike the temporally local (frame to frame) data-association method, global data association has ability to deal with challenges posed by noisy detections. Global association also increases the temporal stride under optimization, therefore more stable and discriminative properties of targets can be used. Trajectory similarities are calculated as heuristically weighted combination of individual features based on geometry, appearance and motion. Association is modeled as a complete K-partite graph (all pairwise relationships inside the temporal window are taken into account) K corresponds to number of cameras in network. For simplicity purpose, we use K=2. Since we use complete K-partite graph, we have an optimal solution. Whereas methods that model association as complex multivariate optimization, upon scaling, face the problem of being stuck at local minima and may provide sub-optimal solutions. Fusion is performed using adaptive weighting method. Where the weights are derived from reliability attribute of each tracker. This enables correct and consistent trajectories after fusion even if the individual trajectories have inherent noises, occlusion and false positives

Our approach is evaluated on the publicly available PETS2009 dataset. PETS2009 is a challenging dataset due to its low FPS and interobject occlusions . We choose View1,View3 and View5 in S2.L1 scenario to evaluate. The results can be seen in Figure 14

The results are encouraging and are very raw and preliminary with lot of scope and room for improvement. With more fine tuning, error rate can be improved. However too significant errors in people detection to build on top of it. Thus, we need training detector on specific datasets to improve the approach. As future work, we will study if we can improve the optimization stage with a more complex optimization using minimal graph flow would improve the results drastically.



Scenario	Method	Camera ID's	MOTA(%)	MOTP(%)	MT(%)	PT(%)	ML(%)
PETS 2009 S2.L1	Berclaz et al.[1]	1,3,5,6,8	82	56	-	-	-
	Leal-Taixe et al.[2]	1,5	76	60	-	-	-
	Leal-Taixe et al. [2]	1,5,6	71.4	53.4	-	-	-
	Murray Evans et al[4]	2 Cameras	63	55			
	Martin Hofmann et al[3]	1,5	99.4	82.9	100	0	0
	Martin Hofmann et al.[3]	1,5,7	99.4	83.0	100	0	0

Scenario	Method	Camera ID's	MOTA(%)	MOTP(%)	MT(%)	PT(%)	ML(%)
PETS2009 S2.L1	Our Approach	1,3	76.33	65.28	92.59	0.035	0.714

Figure 14. Result of our approach on PET2009 dataset.

7.10. Person Re-identification in Real-World Surveillance Systems

Participants: Furqan Mohammad Khan, François Brémond.

Keywords: re-identification, long term visual tracking, signature modeling

Cost of supervised metric learning Person re-identification problem has recently received a lot of attention and the recent focus is to use supervised model training to learn cross camera appearance transformation. In general, $O(n^2)$ models are trained in a surveillance network with n cameras, one for each camera pair. $2p$ tracks are required to train one model with p person identities. In a real-world surveillance network with non-overlapping fields of view, a person appears only in a subset of cameras (see figure 15 , courtesy of [51]). This puts the requirement of number of tracks to train all models at $O(n^2p)$, or more precisely, $n(n-1)p$. That is, to train each model with 100 people in a 10 camera network we need 9000 tracks. For supervised training, these tracks need to be given consistent identities, and worse, have their bounding boxes marked. This is a significant burden on human annotators for deployment in real-world. Further, the annotation cost has to be repaid at a significant fraction if only one new camera is added to the system (may be due to failure of an existing camera), or if the lighting changes significantly (in case of outdoor surveillance). In our opinion, this is a significant bottleneck for supervised metric learning based re-identification in real-world.

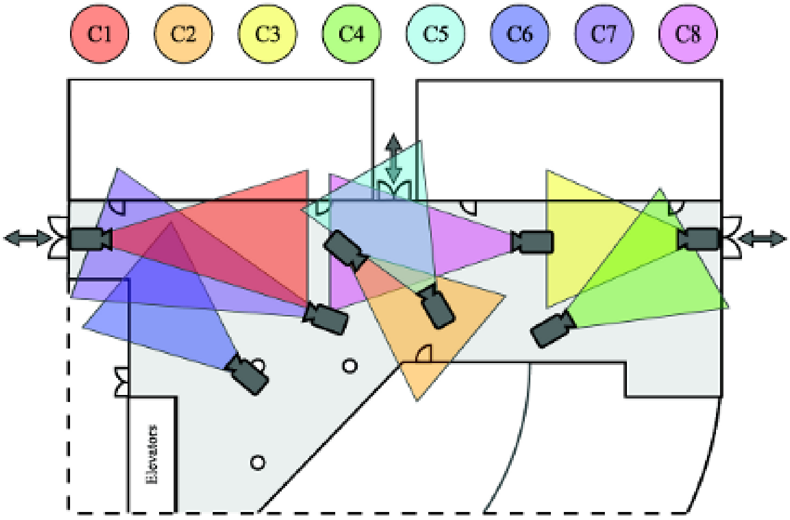


Figure 15. Camera arrangement in multi-camera surveillance scenario of SAIVT-SoftBio dataset [51]

Improved re-identification through signature modeling Re-identification is challenging because variance is intra-class appearance in often higher than inter-class appearance due to varying lighting conditions and viewpoints, and non-uniqueness of clothing. More importantly, in real-world when re-identification is fed by automated human detectors and trackers, significant mis-alignment or partial visibility of the person within proposed bounding box make it difficult to extract relevant features. Our work focuses on improving signature construction from low level features for multi-shot re-identification. We explicitly model multi-modality of person appearance using a feature mixture (corresponding publication is under review at this moment). This improves state-of-the-art re-identification performance on SAIVT-SoftBio [51] dataset and performs equally well as state-of-the-art metric learning methods on iLIDS-VID [88] and PRID2011 [64] datasets. The performance comparison of our method with state-of-the-art is presented using CMC in figure 16 (our results are denoted by MCAM).

7.11. Human Action Recognition in Videos

Participants: Piotr Bilinski, François Brémond.

Keywords: Action Recognition, Video Covariance Matrix Logarithm, VCML, Descriptor

Video Covariance Matrix Logarithm for Human Action Recognition in Videos

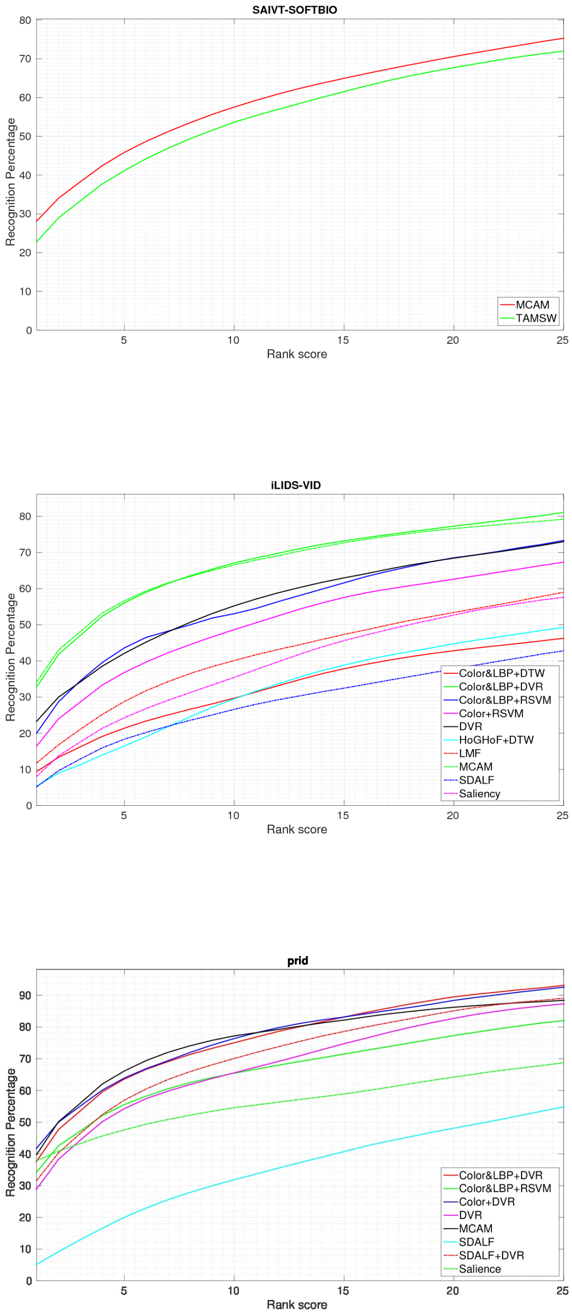


Figure 16. Performance comparison of our MCAM approach using CMC curves on different datasets. ***Top****: Comparison with TAMSW [49] on SAIVT-SoftBio dataset;* ***middle****:Comparison with Color+RSVM [88], Color&LBP+DTW [88], Color&LBP+DVR [88], Color&LBP+RSVM [88], DVR [88], HoG-HoF+DTW [88], LMF [97], Salience [96], and SDALF [60] on iLIDS-VID dataset;* ***bottom****: Comparison with Color+DVR [88], Color&LBP+DVR [88], Color&LBP+RSVM [88], DVR [88], Salience [96], and SDALF [60] on PRID2011 dataset.*

In this work, we propose a new local spatio-temporal descriptor for videos and we propose a new approach for action recognition in videos based on the introduced descriptor. Overview of the proposed action recognition approach based on the introduced descriptor is presented in Figure 17 . The new descriptor is called the Video Covariance Matrix Logarithm (VCML). The VCML descriptor is based on a covariance matrix representation, and it models relationships between different low-level features, such as intensity and gradient. We apply the VCML descriptor to encode appearance information of local spatio-temporal video volumes, which are extracted by the (Improved) Dense Trajectories. Then, we present an extensive evaluation of the proposed VCML descriptor with the (Improved) Fisher Vector encoding and the Support Vector Machines on four challenging action recognition datasets (*i.e.* URADL, MSR Daily Activity 3D, UCF50, and HMDB51 datasets). We show that the VCML descriptor achieves better results than the state-of-the-art appearance descriptors. In comparison with the most popular visual appearance descriptor, *i.e.*the HOG descriptor, the VCML achieves superior results. Moreover, we present that the VCML descriptor carries complementary information to the HOG descriptor and their fusion gives a significant improvement in action recognition accuracy (*e.g.* the VCML improves the HOG by 15% on the HMDB51 dataset). Finally, we show that the VCML descriptor improves action recognition accuracy in comparison to the state-of-the-art (Improved) Dense Trajectories, and that the proposed approach achieves superior performance to the state-of-the-art methods. The proposed VCML based technique achieves 94.7% accuracy on the URADL dataset, 85.9% on the MSR Daily Activity 3D dataset, 92.1% on the UCF50 dataset, and 58.6% on the HMDB51 dataset. More results and comparisons with the state-of-the-art are presented in Table 9 and Table 10 . To the best of our knowledge, this is the first time covariance based features are used to represent the trajectories. Moreover, this is the first time they encode the structural information and they are applied with the (Improved) Fisher Vector encoding for human action recognition in videos. This work has been published in [40].

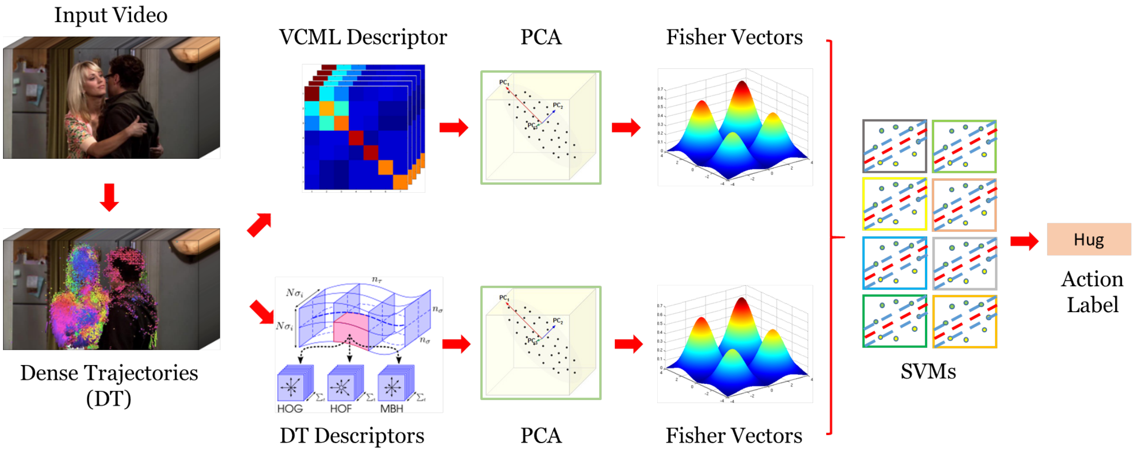


Figure 17. Overview of our action recognition approach based on the introduced VCML descriptor.

7.12. Evaluation of Event Recognition without Using Ground Truth

Participants: Ramiro Diaz, Carlos Fernando Crispim Junior, François Brémond.

Keywords: Computer Vision, Event Recognition, Video Summarization.

The main goal of the work is to improve the Event Recognition process and to improve the way we build the event models as well. The work concerns the Valrose Nursing Home, it consists in monitoring older people with health issues like Dementia and who are in need of care and stimulation.

Table 9. Comparison with the state-of-the-art on the URADL and MSR Daily Activity 3D datasets. The table presents the accuracy of our approach using Dense Trajectories (DT) and Improved Dense Trajectories (IDT).

URADL		MSR Daily Activity 3D	
Benabbas *et Al.*, 2010	81.0	Koperski *et Al.*, 2014	72.0
Raptis and Soatto, 2010	82.7	JPF – Wang *et Al.*, 2012	78.0
Messing *et Al.*, 2009	89.0	Oreifej and Liu, 2013	80.0
Bilinski and Bremond, 2012	93.3	AE – Wang *et Al.*, 2012	85.7
Dense Trajectories	94.0	Dense Trajectories	76.2
Our Approach (DT)	**94.0**	**Our Approach (DT)**	**78.1**
Our Approach (IDT)	**94.7**	**Our Approach (IDT)**	**85.9**

Table 10. Comparison with the state-of-the-art on the UCF50 and HMDB51 datasets. The table presents the accuracy of our approach using Dense Trajectories (DT) and Improved Dense Trajectories (IDT).

UCF50		HMDB51	
Kantorov and Laptev, 2014	82.2	Kantorov and Laptev, 2014	46.7
Shi *et Al.*, 2013	83.3	Jain *et Al.*, 2013	52.1
Oneata *et Al.*, 2013	90.0	Oneata *et Al.*, 2013	54.8
Wang and Schmid, 2013	91.2	Wang and Schmid, 2013	57.2
Dense Trajectories	84.2	Dense Trajectories	47.0
Our Approach (DT)	**88.1**	**Our Approach (DT)**	**52.9**
Our Approach (IDT)	**92.1**	**Our Approach (IDT)**	**58.6**

Since the video dataset contains data for about 8 months, a new evaluation method is required to properly analyze the whole dataset and gain a deeper understanding of it. Our approach consists in displaying the data in a way that can be useful either for doctors, as well for engineers to detect failures and to improve the event recognition process in an efficient way. Because of this need, a new evaluation tool has been developed and named Event Plotter.

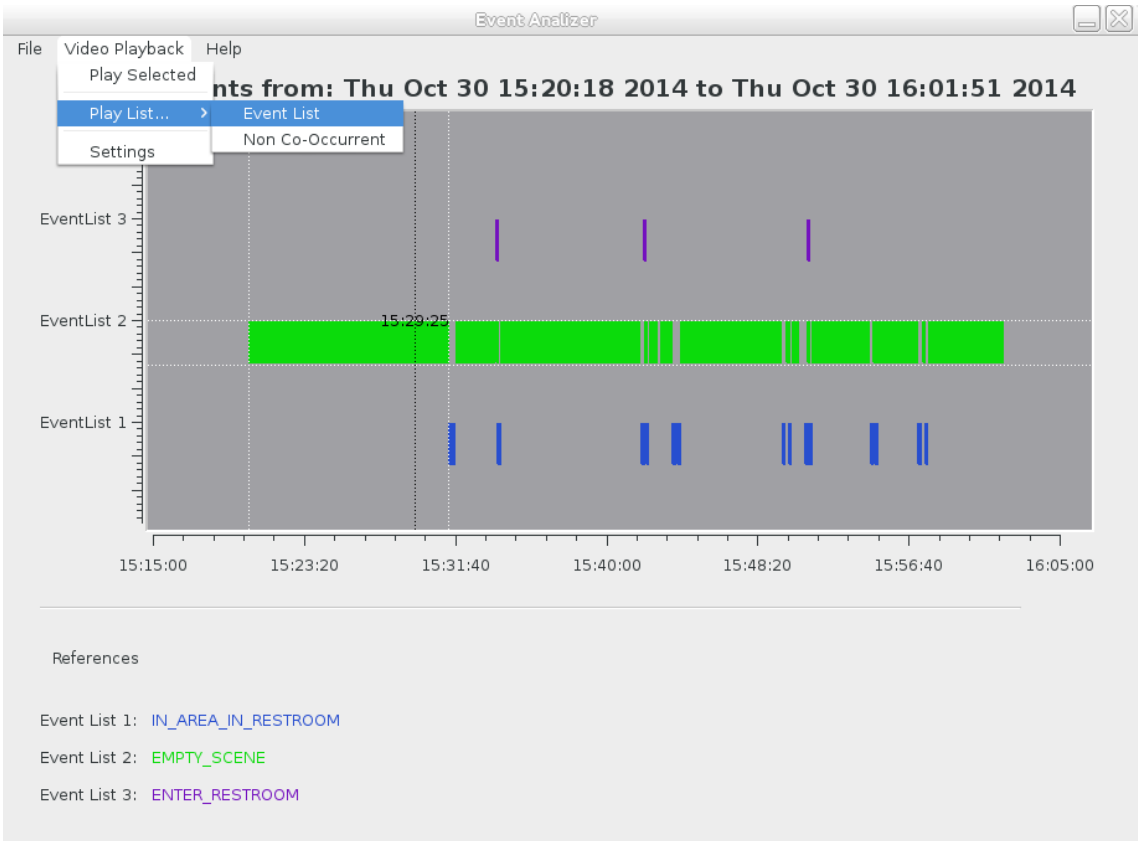
This tool provides a new method for event evaluation. First of all, as we do not have ground truth information for the total duration of the 8 months, but just for one week, so another method is needed to check the event model efficiency. To address this issue, the tool displays all the events in the desired time period (as clusters on a timeline) and single events (or time intervals) can be selected to quickly check the video and visualize the results of the event recognition working on the fly -see Figure 18 -. The goal of this work is to understand how event recognition works, change the models on the fly, import them, and see how the recognition changes in real time.

Also, to compare new event models with the old ones, video summarization is implemented as well. Event based video summarization is utilized here to check how the recognition of one particular event type changes globally on the whole video and to display the recognition results. Also video summarization can be useful for doctors to check the way patients behave, for example playing all the videos of event ”Get-up from bed”, trying to predict patterns.

The data processed to address this issue was 1 week, because it was the time corresponding to the ground truth data. With these processed data, we have tested the efficiency of the Event Plotter tool, and we are currently improving the event recognition process by changing event models, adding new zones, and testing them on the fly.

7.13. Semantic Event Fusion of Different Visual Modality Concepts for Activity Recognition

Participants: Carlos Fernando Crispim-Junior, François Brémond.



References

Event List 1: IN_AREA_IN_RESTROOM

Event List 2: EMPTY_SCENE

Event List 3: ENTER_RESTROOM

Figure 18. GUI of Event Plotter with 3 loaded Event Lists.

Keywords: Knowledge representation formalism and methods, Uncertainty and probabilistic reasoning, Concept synchronization, Activity recognition, Vision and scene understanding, Multimedia Perceptual System,

Combining multimodal concept streams from heterogeneous sensors is a problem superficially explored for activity recognition. Most studies explore simple sensors in nearly perfect conditions, where temporal synchronization is guaranteed. Sophisticated fusion schemes adopt problem-specific graphical representations of events that are generally deeply linked with their training data and focus on a single sensor. In this work we have proposed a hybrid framework between knowledge-driven and probabilistic-driven methods for event representation and recognition. It separates semantic modeling from raw sensor data by using an intermediate semantic representation, namely concepts. It introduces an algorithm for sensor alignment that uses concept similarity as a surrogate for the inaccurate temporal information of real life scenarios (Fig. 20). Finally, it proposes the combined use of an ontology language, to overcome the rigidity of previous approaches at model definition, and a probabilistic interpretation for ontological models, which equips the framework with a mechanism to handle noisy and ambiguous concept observations, an ability that most knowledge-driven methods lack (Fig. 19). We evaluate our contributions in multimodal recordings of elderly people carrying out instrumental activities of daily living (Table 11). Results demonstrated that the proposed framework outperforms baseline methods both in event recognition performance and in delimiting the temporal boundaries of event instances

This work has been developed as a collaboration between different teams in Dem@care consortium (Inria, University of Bordeaux, and CERTH). We thank the other co-authors for their contributions and support in the development of this work up to its submission for publication.

Table 11. Comparison to baseline methods in the test set

mean F_1-score	**Fusion approach**		
	Baselines		**Ours**
IADL	SVM	OSF	
S. bus line	44.19	31.36	73.10
M.finances	43.99	0.00	43.73
P. pill box	45.86	49.11	65.02
P. drink	20.02	24.29	64.03
Read	90.18	91.82	95.22
T.telephone	72.12	0.00	75.58
W. TV	2.32	0.00	35.80
W. Plant	0.00	0.00	100.00
Average	39.83	24.57	69.06

OSF: Ontology-based Semantic Fusion

7.14. Semi-supervised Activity Recognition Using High-order Temporal-composite Patterns of Visual Concepts

Participants: Carlos Fernando Crispim-Junior, Michal Koperski, Serhan Cosar, François Brémond.

Keywords: visual concepts, semi-supervised activity recognition, complex activities, cooking composite activities

Methods for action recognition have evolved considerably over the past years and can now automatically learn and recognize short term actions with satisfactory accuracy. Nonetheless, the recognition of activities - a composition of actions and scene objects - is still an open problem due to the complex temporal, composite structure of this category of events. Existing methods either still focus on simple activities or oversimplify the modeling of complex activities by only targeting whole-part relations between activity components, like actions. In this work, we have investigated a hierarchical, semi-supervised approach that unsupervisely learns actions from the composite patterns of atomic concepts (*e.g.*, slice, tomato), and complex activities from the

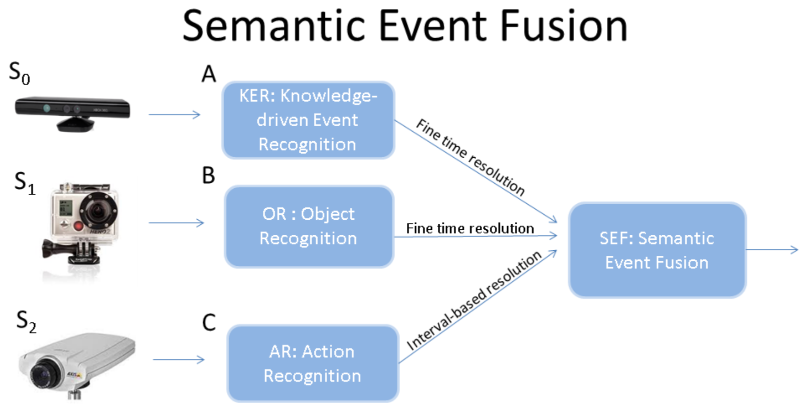


Figure 19. Semantic event fusion framework: detector modules (A-C) process data from their respective sensors (S0-S2) and output concepts (objects and low-level events). Semantic Event Fusion uses the ontological representation to initialize concepts to event models and then infer complex, composite activities. Concept fusion is performed on millisecond temporal resolution to cope with instantaneous errors of concept recognition.

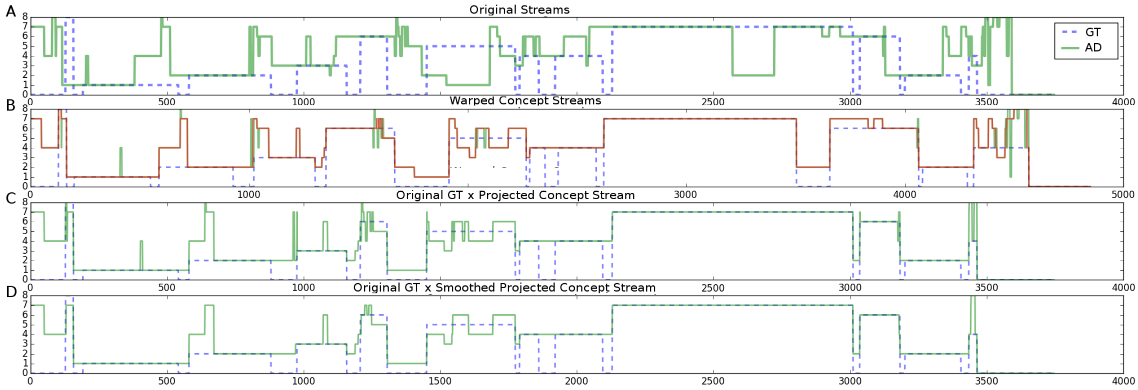


*Figure 20. Semantic alignment between the concept stream of the action recognition detector (AR) and a concept stream (GT) generated from events manually annotated by domain experts using the time axis of the color-depth camera. X-axis denotes time in frames, and Y-axis denotes activity code (1-8), respectively, search bus line on the map, establish bank account balance, prepare pill box, prepare a drink, read, talk on the telephone, watch TV, and water the plant. From top to bottom, images denote: (**A**) original GT and AR streams, (**B**) GT and AR streams warped, AR stream warped and smoothed (in red), (**C**) original GT and AR stream warped and then backprojected onto GT temporal axis, (**D**) original GT and AR warped, backprojected, and then smoothed with median filtering.*

temporal patterns of concept compositions (actions). On a first step, an unsupervised, inductive approach iteratively builds a multi-scale, temporal-composite model of the concept occurrences during the activity taking place (Fig. 22). Then, activity recognition is performed by comparing the similarity of the generated model of a given video and *a priori* learned and labeled unsupervised models. We have evaluated the proposed method in the MPII Cooking Composite Activities dataset (Fig. 21), a video collection where people perform a set of complex activities related to cooking recipes. To tackle this dataset it is necessary to recognize a large variety of visual concepts (*e.g.*, from actions, such as cutting and stirring, to objects, such as tomato and cutting board). Moreover, the detection of cooking activities is a very challenging problem since we observe a low inter-class variance between activity classes, and a high intra-class variance within an activity due to person to person differences in performing them. The proposed approach presents a mean average precision (mAP) of 56.36% ± 5.1%, and then outperforms previous methods ([81], mean AP 53.90%). This improvement is devoted to the modeling of deeper composite and temporal relations between visual concepts (from 2nd to 5th order compositions). The performance of the proposed method is mostly limited by the performance of low-level concept detectors. Future work will investigate ways to extend the current probabilistic model to handle more efficiently the differences in concept detector performance.

Figure 21. Illustation of one of the cooking recipes of Cooking Composite dataset [81]

7.15. From Activity Recognition to the Assessment of Seniors' Autonomy

Participants: Carlos Fernando Crispim-Junior, Carola Strumia, Alvaro Gomez Uria Covella, Alexandra Konig, François Brémond.

Activity recognition plays a fundamental role in several research fields as a way to extract semantic meaning from images and videos, to find more accurate matches for textual queries in video search engines, and to analyze long-term activity patterns in assisted living scenarios, such as seniors living at home. In this sense, we have continued our work on activity monitoring by proposing a novel knowledge-based event monitoring system that combines the observations of a vision system with expert knowledge and scene semantics, to recognize daily living activities in assisted living scenarios.

The approach's novelty lies in the combination of a flexible constraint-based ontology language for event modeling with efficient and robust algorithms to detect, track and re-identify people using color-depth sensing (low-level vision). The robust low-level vision promotes the modeling of longer and more complex events,

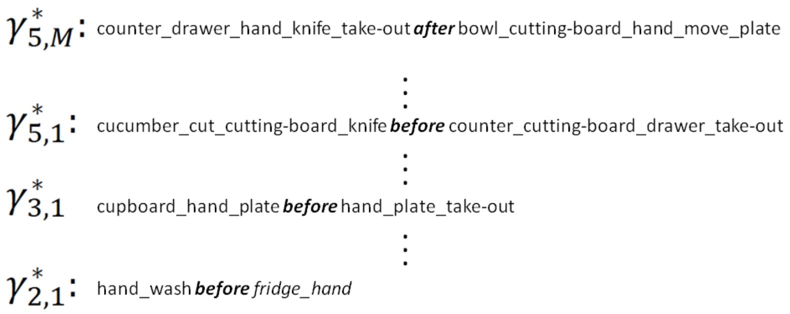


Figure 22. Example of temporal concept composite patterns of different arities extracted from an unlabeled activity.

while the ontology language provides a flexible way to describe event and incorporate domain knowledge, and ease knowledge transfer across different scenes. The proposed approach has been investigated for two assisted living scenarios: a) the monitoring of physical tasks and daily living activities in observation rooms of hospital and clinics, and b) daily and nightly activities of seniors living in nursing home apartments. To evaluate our approach performance compared to state of art methods, we have computed its results for GAADRD dataset. This is public dataset, which is composed of videos of seniors performing physical tasks and activities of daily living. Evaluation results (Table 12) have demonstrated that our approach achieves an average $F_1 - score$ 20 % higher than the baseline method [89].

Table 12. Recognition of IADLs - GAADRD data set - F 1 -score

Event	DT-HOG	DT-HOF	DT-MBHy	Proposed
Account Balance	44.96	34.71	42.98	66.67
Prepare Drink	81.66	44.87	52.00	100.00
Prepare Drug Box	14.19	0.00	0.00	57.14
Read Article	52.10	42.86	33.91	63.64
Talk on telephone	82.35	0.00	33.76	100.00
Turn on radio	85.71	42.52	58.16	94.74
Water Plant	0.00	0.00	0.00	52.63
Average ± SD	51.8 ± 34.4	23.6 ± 22.3	31.5 ± 23.3	76.4 ± 21.0

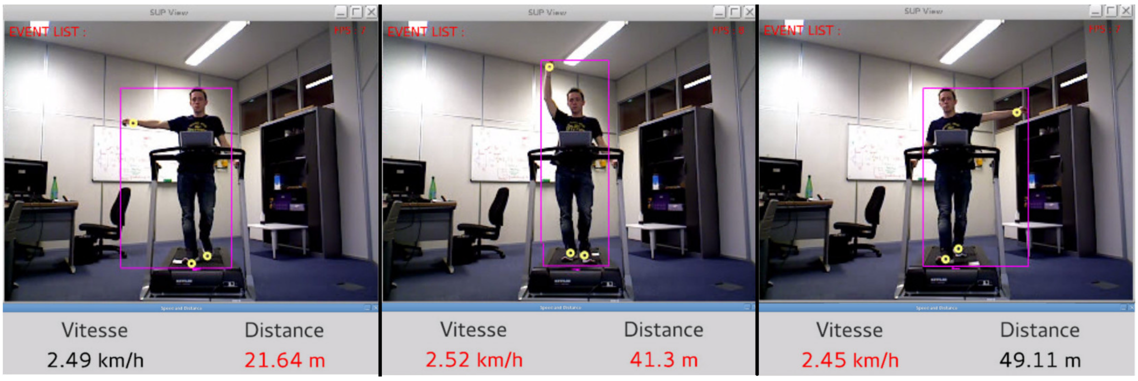
Given the satisfactory performance of the proposed activity recognition framework we have also investigated it as a method to automatically measure a seniors' autonomy in quantitative and objective fashion. To do so, we have developed a probabilistic model that takes as input the recognized activities and gait-patterns from the period of time the person performs physical tasks. The proposed autonomy model has presented an average performance of 83.67 %, which suggests that the use of such technologies may provide clinicians with diagnostic relevant information, and decrease observer's biases when compared to clinical scales. The results of this investigation have been published in [33].

7.16. Serious Games Interfaces Using an RGB-D Camera : Results and Perspectives

Participants: Baptiste Fosty, François Brémond.

Keywords: RGB-D camera analysis, walking speed, serious games, startup project

Within the context of the development of serious games for people suffering from Alzheimer disease (Az@Game project), we have developed algorithms to interact with the virtual environment through simple gesture recognition and walking speed computation. We have shown in previous work that the walking speed measured by our system is accurate enough within this context and reproducible. A paper has been submitted in Gait and Posture journal (now in reviewing process).
Concerning the gesture recognition algorithm, it consists in recognizing three basic gestures (right arm left on the side, left arm left on the side, right or left arm left on top). We performed a small experimentation to test the robustness of the system in detecting these gestures where participants (10 in total) had to perform 10 times each gesture while walking at 2.5km/h on the treadmill (see Figure 23). The results are shown in Table 13 .



Vitesse 2.49 km/h
Distance 21.64 m

Vitesse 2.52 km/h
Distance 41.3 m

Vitesse 2.45 km/h
Distance 49.11 m

Figure 23. Display of gesture recognition and speed computation. Distance is red when right arm is left, speed is red when left arm is left, both are red when one arm is on top. Yellow dots correspond to feet and arm detection.

Table 13. Gesture recognition results.

	Right	Left	Top	TOTAL
Recall	97.0	95.0	100.0	97.3
Precision	100.0	97.9	94.3	97.3
F-Score	**98.5**	**96.4**	**97.1**	**97.3**

Following that, we decided to study whether this system would be useful in rehabilitation. Some experts of this domain in a rehabilitation center (Centre Hélio Marin in Vallauris, France) have been interviewed and they were very enthusiasts about using this type of system to get objective gait parameters. To go further on the market opportunity and evaluate the feasability of a technology transfer, we studied the concurrent products and contacted more than thirty other rehabilitation centers to have a deeper understanding of the needs and validate our idea. This investigation lead to the proposition of a startup project(BOMOTECH) to Inria which has been accepted and funded for the next 7 months, during what the goal is to get closer from a marketable tool.

7.17. Assistance for Older Adults in Serious Game Using an Interactive System

Participants: Minh Khue Phan Tran, François Brémond, Philippe Robert.

Keywords: Older Adults, Assistance, Serious Games

Serious Games offer a new way to older adults to improve their abilities such as vision, balance or memory. However, cognitive impairment causes a lot of difficulties to them when actively practising these games. Their engagement and motivation are reduced rapidly when encountering successive problems without any help. Our hypothesis is that this problematic situation can be handled if they are assisted regularly. We propose then an interactive system which can determine dynamically the situations and provide different helps in real-time. We focuse on two main problems that the older players encounter regularly :

- they forget how to continue to play the game.
- they make a lot of errors.

The system determines the above problems by computing various characteristics of the player (skeleton, postures, gestures ...) along the game states. This process is presented in Figure 24 . The characteristics of the player, which are collected by the Recognition Module thanks to Kinect Camera and the related SDK, are sent to the Interaction Module. This module associates these data with the game states provided by the game in order to recognize the problem and interacts with the player through a 3D-animated avatar.

The system is tested with 3 groups of patients described by 3 different cognitive states : mnesic plaints, MCI and Alzheimer. The patients are invited to play a concentration-based game with a Kinect camera. Each patient plays 3 phases : playing with therapist, playing alone and playing with the avatar. The playing time and the final score of each game phase are recorded. Here, the system takes into account the player's gestures and the game states for recognizing two situations :

- the player reacts too late and too slowly to the current game task.
- the player makes many mistakes.

The experimental results confirmed our hypothesis. Most of the patients have the best performance in phase "playing with the avatar". Their playing time is shorter and their final score is higher in phase "playing with assistance" than in phase "playing alone". The results are presented in the publication [36] accepted by the Games and Learning Alliance Conference in December 2015. Future work aims at improving the system and compare its efficiency with the one of "humans assistances".

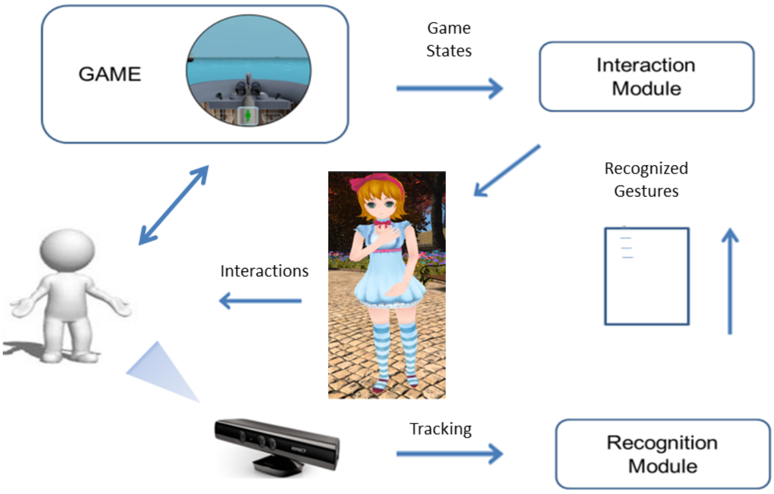


Figure 24. Assisting older adults in serious game playing

7.18. Generating Unsupervised Models for Online Long-Term Daily Living Activity Recognition

Participants: Farhood Negin, Serhan Coşar, Michal Koperski, François Brémond.

Keywords: Unsupervised Activity Recognition

Generating Unsupervised Models for Online Long-Term Daily Living Activity Recognition

In this work, we propose an unsupervised approach that offers a comprehensive representation of activities by modeling both global and body motion of people. Compared to existing supervised approaches, our approach automatically learns and recognizes activities in videos without user interaction. First, the system learns important regions in the scene by clustering trajectory points. Then, a sequence of primitive events is constructed by checking whether people are inside a region or moving between regions. This enables to represent the global movement of people and automatically split the video into clips. After that, using action descriptors [90], we represent the actions occurring inside each region. Combining action descriptors with global motion statistics of primitive events, such as time duration, an activity model that represents both global and local action information is constructed. Since the video is automatically clipped , our approach performs online recognition of activities. The contributions of this work are twofolds: (i) generating unsupervised human activity models that obtains a comprehensive representation by combining global and body motion information, (ii) recognizing activities online and without requiring user interaction. Experimental results show that our approach increases the level of accuracy compared to existing approaches. Figure 25 illustrates the flow of the system.

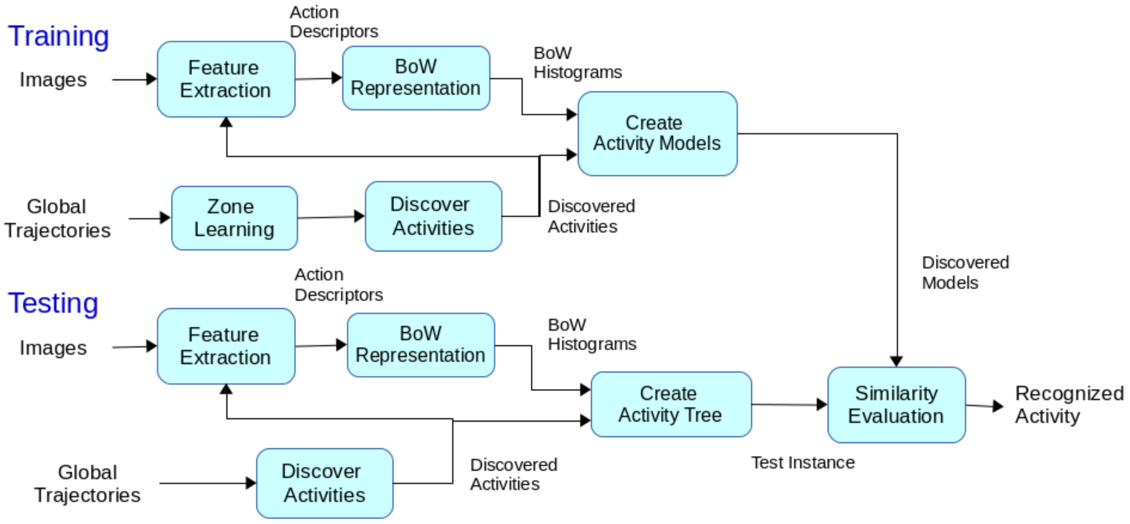


Figure 25. Architecture of the framework: Training and Testing phases

The performance of the proposed approach has been tested on the public GAADRD dataset [67] and CHU dataset (http://www.demcare.eu/results/datasets) that are recorded under EU FP7 Dem@Care Project1 in a clinic in Thessaloniki, Greece and in Nice, France, respectively. The datasets contain people performing everyday activities in a hospital room. The activities considered in the datasets are listed in Table 1 and Table 2. Each person is recorded using RGBD camera of 640x480 pixels of resolution. The GAADRD dataset contains 25 videos and the CHU dataset contains 27 videos. Each video lasts approximately 10-15 minutes.

We have compared our approach with the results of the supervised approach in [90]. We did also a comparison with an online supervised approach that follows [90]. For doing this, we train the classifier on clipped videos and perform the testing using sliding window. In the online approach, a SVM is trained using the action descriptors extracted from groundtruth intervals. We have also tested different versions of our approach that i) only uses global motion features and ii) which only uses body motion features. We have randomly selected 3/5 of the videos in both datasets for learning the activity models. The codebook size is set to 4000 visual words for all the methods.

The performance of the online supervised approach and our approach in GAADRD dataset are presented in Table 1. In all approaches that use body motion features, HoG descriptors are selected since they give the best results. It can be clearly seen that, using models that represent both global and body motion features, our unsupervised approach enables to obtain high sensitivity and precision rates. Compared to the online version of [90], thanks to the learned zones from positions and discovered activities, we obtain better activity localization, thereby better precision. However, since the online version of [90] utilizes only dense trajectories (not global motion), it fails to localize activities. Hence, it detects the intervals that does not include an activity (e.g. walking from radio desk to phone desk) and for ”prepare drug box“, ”watering plant“, and ”reading“ activities, it cannot detect the correct intervals of the activities. Compared to the unsupervised approach that either use global motion features or body motion features, we can see that, by combining both features, our approach achieves more discriminative and precise models, thereby improves both sensitivity and precision rates. By combining global and body motion features, our approach benefits from discriminative properties of both feature types. Table 1 also presents the results of the supervised approach in [90]. Although the supervised approach uses groundtruth intervals in test videos in an offline recognition scheme, it fails to achieve accurate recognition. As our approach learns the zones of activities, we discover the places where the activities occur, thereby we achieve precise and accurate recognition results. Since this information is missing in the supervised approach, it detects ”turning on radio“ while the person is inside drink zone preparing drink.

Table 2 shows the results of the online supervised approach and our approach in CHU dataset. MBH descriptor along y axis and HoG descriptor gives the best results for our approach and the online supervised approach, respectively. In this dataset, since people tend to perform activities in different places (e.g. preparing drink at phone desk), it is not easy to obtain high precision rates. However, compared to the online version of [90], our approach detects all activities and achieves a much better precision rate. The online version of [90] again fails to detect activities accurately, thereby misses some of the ”preparing drink“ and ”reading“ activities and gives many false positives for all activities.

Thanks to the activity models learned in unsupervised way, we accurately perform online recognition. In addition, the zones learned in an unsupervised way help to model activities accurately, thereby most of the times our approach achieves more accurate recognition compared to supervised approaches. This paper has been published in third Asian Conference on Pattern Recognition (ACPR 2015) [35].

7.19. Run-time Adaptation of Video Systems

Participants: Sabine Moisan, Jean-Paul Rigault, François Brémond.

In the framework of our research on model engineering techniques for video-surveillance systems, we have focused this year on run-time reconfiguration of such systems. The goal is to follow the “model at run-time” approach and to obtain context-aware self-adaptive video systems. In this approach models are kept and used at run-time. In our case, these models describe all the possible run-time configurations. They are specified using *Feature Models*.

Run-time reconfiguration means to react to context changes by tuning, adding, removing, or replacing components of the video chain, and possibly changing the chain itself.

So far, we have defined a run-time architecture consisting of three layers. The lower level describes the video analysis components and the context events. The upper one handles feature model adaptation. The middle layer is an adapter: it transforms lower level context event occurrences into upper level feature reconfiguration; in the other direction, it transforms the corresponding feature reconfigurations into video components reconfigurations.

Table 14. The activity recognition results for GAADRD dataset. Bold values represent the best sensitivity and precision results for each class.

	Supervised Approach [90]		Online Version of [90]		Unsupervised (Only Global Motion)		Unsupervised (Only Body Motion)		Proposed Approach	
ADLs	Sens. (%)	Prec. (%)	Sens. (%)	Prec. (%)	Sens. (%)	Prec. (%)	Sens. (%)	Prec. (%)	Sens. (%)	Prec. (%)
Answering Phone	**100**	88	**100**	70	**100**	**100**	57	**100**	**100**	**100**
Establish Acc. Bal.	67	**100**	**100**	29	**100**	86	50	**100**	**100**	86.67
Preparing Drink	**100**	69	**100**	69	78	**100**	**100**	**100**	**100**	**100**
Prepare Drug Box	**58.33**	**100**	11	20	33.34	**100**	33.34	**100**	33.34	**100**
Watering Plant	**54.54**	**100**	0	0	44.45	57	33	**100**	44.45	**100**
Reading	**100**	**100**	88	37	**100**	**100**	38	**100**	**100**	**100**
Turn On Radio	60	86	**100**	75	89	89	44	**100**	89	**100**
AVERAGE	77.12	91.85	71.29	42.86	77.71	90.29	50.71	**100**	**80.97**	98.10

Table 15. The activity recognition results for CHU dataset. Bold values represent the best sensitivity and precision results for each class.

	Supervised Approach [90]		Online Version of [90]		Proposed Approach	
ADLs	Sens. (%)	Prec. (%)	Sens. (%)	Prec. (%)	Sens. (%)	Prec. (%)
Answering Phone	57	78	**100**	**86**	**100**	65
Preparing Drink	78	**73**	92	43	**100**	58
Prepare Drug Box	**100**	83	**100**	43	**100**	**100**
Reading	35	**100**	92	36	**100**	**78**
Using Bus Map	90	**90**	**100**	50	**100**	47
AVERAGE	72.0	**84.80**	90.95	48.76	**100**	70.00

This year we focused on the upper layer. We first formalized the run-time feature reconfiguration rules. First, any reconfiguration should respect the semantics of feature models and their attached constraints. Second, the reconfiguration should satisfy the requests from the middle layer, essentially selections and deselections of features. From there two requirements, we identified three possible outcomes: successful reconfiguration, impossible reconfiguration (selection/deselection conflicts), and “undefined” reconfiguration (not enough information to get through the process). We also determined the actions to take in these cases. In particular, in the last two cases, we decided to let the component configuration unchanged.

To implement this upper layer, we first attempted to rely on an existing feature model manipulation framework, namely FAMILIAR [45]. However, this approach suffers from a number of drawbacks. First, FAMILIAR is a standalone Java program, whereas the rest of the system is written in C++, for performance reasons and library availability. Hence, using FAMILIAR implies superfluous back and forth inter-module communications and data transformations. Second, we confirmed that FAMILIAR is more a system deployment tool than a run-time reconfiguration one. In particular it cannot fulfill all the reconfiguration rules that we have formalized. Therefore, we are completing a full re-implementation of the upper layer.

The programming language homogeneity permits a more efficient integration of the three layers. It particular, it becomes easier to incorporate our extensions to feature models such as quality metrics [34].

7.20. Scenario Description Language

Participants: Sabine Moisan, Annie Ressouche, Jean-Paul Rigault, Nazli Temur, François Brémond.

Last year, we developed a scenario recognition engine based on the Synchronous Model of reactive systems. We now need a scenario description language friendly to our end users who are not computer scientists in general. In fact, Stars has already defined such a language. However, it is a declarative language based on (temporal) constraints. This is certainly not the most natural and the simplest way for end users to express their domain specific scenarios.

Consequently, we started this year a comparative study of different means to express scenarios in various domains (video understanding but also games, movies, music, criminology, military strategy...). We investigated 16 formalisms covering these domains. We defined a comparison grid based on criteria relevant for our video understanding goals. We retained 9 such criteria: application domain scope, ease of use, representation of scenario basic elements (background, scene, roles...), modularity (possibility of scenario hierarchy), time representation (absolute, logical, multi-clocks, no clocks...), expression of temporal constraints, representation of repetitive patterns, support for concurrency and parallelism, and finally formal foundations.

To complete the study, we conducted an experiment: describing a case study scenario using some of these formalisms to concretely estimate their advantages and drawbacks, especially their ease of use.

At this time, none of the studied languages fulfills completely our needs. Many languages are graphical ones. While this may appear as user friendly, scalability and automatic analysis become an issue. Some languages lack formal semantics, which is not acceptable in our case; others are merely extensions of computer languages, hence dedicated to specialists.

We plan to define our own version, which will rely on solid semantic foundations. (see section 7.21). To enforce user-friendliness, we started to collaborate with ergonomists.

7.21. Scenario Recognition

Participants: Annie Ressouche, Sabine Moisan, Jean-Paul Rigault, Ines Sarray, Daniel Gaffé.

Keywords: Synchronous Modeling, Model checking, Mealy machine, Cognitive systems.

For a long time, Stars strategy has been to favor the easy generation of activity recognition systems. These systems correspond to a succession of pattern matching and clustering algorithms, combined with adequate knowledge representation (e.g. scene topology, temporal constraints) at different abstraction levels (from raw signal to semantics). Due to the large range of application domains (surveillance, safety, health care, ...), we propose a generic approach to design activity recognition engines. Moreover, such domains require high dependability due to possible safety issues. Thus, our approach should also rely on formal methods to describe, analyze, verify, and generate effective recognition engines. We consider activity recognition engines as reactive systems that react to input events from their environment and produce output events in the form of alarms or notifications. Such engines are intrinsically real time, reactive and they evolve in discrete time. As a consequence, to recognize scenarios, we adapt the usual techniques of synchronous modeling approach to express scenario behaviors. This approach facilitates scenario validation and allows us to generate a recognizer for each scenario.

Our previous developments, on top of existing synchronous languages as Lustre and LE (see secrion 7.22), were convenient for rapid prototyping. However, even if LE is not a closed environment, it appeared as difficult as Lustre to customize, for efficiency reasons. This year, in the framework of Ines Sarray PhD thesis, we began to define a synchronous semantics for the future scenario language (see section 7.20). The idea is to generate automatically recognition engines at compilation time. The compilation itself is totally based on the semantics. To complete this approach we will rely on both our experiment with the LE language last year and on the LE compilation process.

7.22. The Clem Workflow

Participants: Annie Ressouche, Daniel Gaffé, Imane Khalis.

Keywords: Synchronous languages, Synchronous Modeling, Model checking, Mealy machine.

This research axis concerns the theoretical study of a synchronous language LE –with modular compilation– and the development of a toolkit (see Figure 26) around the language to design, simulate, verify, and generate code for programs. The novelty of the approach is the ability to manage both modularity and causality.

This year, we continued to focus on the improvement of both LE language and compiler concerning data handling and the generation of back-ends, required by other research axis of the team (see 7.21 and 7.23). We also improved the design of a new simulator for LE programs which integrates our new approach. In CLEM we generate an independent intermediate code (LEC) before specific target generations. This code represents the semantics of programs with 4-valued equation systems. In our design flow, we need to simulate programs at this level. Last year, we begun to develop such a simulator in order to integrate the data part of the language. The simulator GUI has been designed in Qt and the simulator takes into account the values carried by signals. This year, during her internship, Imane Khalis has completed the simulator to allow an external computation of data values and a communication with the simulator through a socket mechanism. With this last development, the LEC simulator is complete and is integrated in the CLEM toolkit.

7.23. Safe Composition in WCOMP Middleware for Internet of Things

Participants: Annie Ressouche, Daniel Gaffé, Ines Sarray, Jean-Yves Tigli.

Keywords: Synchronous Modeling, Ubiquitous Computing, middleware, internet of things

The aim of this research axis is to federate the inherent constraints of an activity recognition platform like SUP (see section 6.7) with a service-oriented middleware approach dealing with dynamic evolutions of system infrastructure in ubuquitous computing, and particularly in the Internet of Things (IoT). The Rainbow team (Nice-Sophia Antipolis University) proposes a component-based adaptive middleware (WComp [86], [85], [66]) to dynamically adapt and recompose assemblies of components.

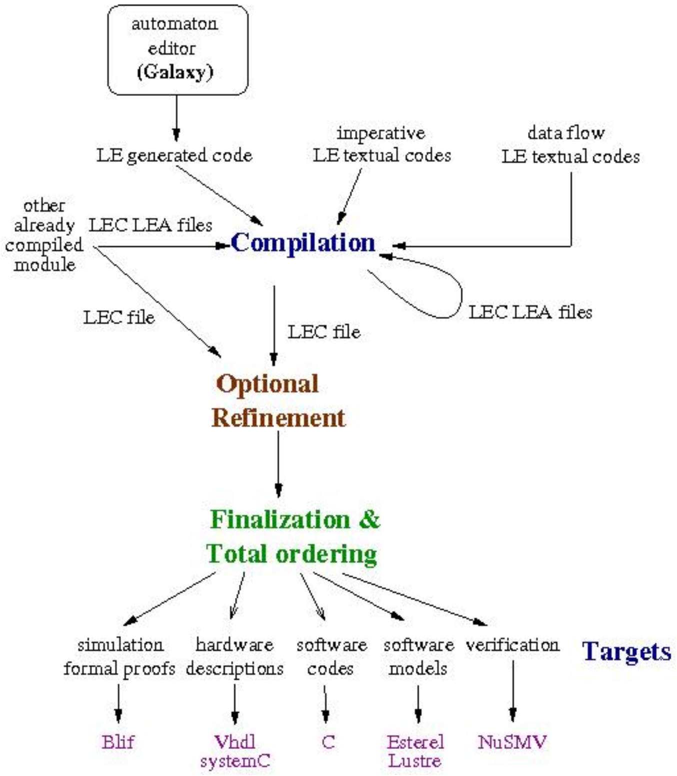


Figure 26. The Clem Toolkit

IoT is a way to combine computation and communication capabilities, sometimes in large scale information systems, with a huge number of complex devices connected to the physical world. Such infrastructures are often dedicated to the deployment of multiple applications, running concurrently. These applications are using shared devices from a common environment through different network middleware and numerous IoT protocols. Indeed, “Things”, also called the *Entities of Interest* [63], are the part of the real world in which devices are interacting and which must not be neglected. We aimed to model and validate concurrent accesses to shared devices without neglecting their associated *Entity of Interest*, their common physical context. One of the main challenge is then how to guarantee and validate some safety and integrity properties throughout the system’s evolution. In WComp middleware, we use synchronous models to facilitate the study and the validation of new composition mechanisms between applications at runtime. Then key problems to solve are: (1) how to specify and respect the "Thing" behavior? (2) how to ensure a safe combination of these multiple accesses when several services accesses a same entry of an *Entity of Interest* ? (3) how to manage multiple uses when applications simultaneously use a same service ?

This year, we addressed these problems by relying on formal method to model device behaviors as synchronous automata, taking into consideration their impact on the Entity Of Interest. Such an approach allows applying model-checking techniques to verify safety properties of applications. The main contribution is the definition of a sound way to compose models allowing context change adaptation. This composition relies on synchronous parallel composition paradigm. We proved that this operation preserves safety properties. However, it is not sufficient to obtain a global model of this composition because some devices may interact with the same Entity Of Interest. Moreover, several applications may use the same device services and then they can have concurrent accesses to their entries, so it can have an unexpected impact on our Entity Of Interest. Therefore, we added constraints to the device models composition and to applications level. We defined a generic way to express these constraints, independently of the knowledge about the devices and the applications, only their type is sufficient. We proposed the Description Constraint Language (DCL) to express these generic constraints and their compilation into LE Mealy machines. Thus we rely on CLEM model-checking facilities (see section7.22) to validate the constraints. As a consequence, this approach ensures the adaptation to a context change and offers a means to formally perform validation.

These results have been published in [43]

7.24. Design of UHD Panoramic Video Camera

Participants: Carolina Da Silva Gomes Crispim, Rachid Guerchouche, Daniel Gaffé, François Brémond.

The goal of this work is to investigate the possibilities of designing a new camera-based system for retail. This work was carried in the context of a collaboration between STARS and Neosensys. The system is composed of several high definition cameras placed in a configuration such as it makes it possible to obtain a panoramic vision with 360° of field of view. The work was divided into 2 parts: theoretical part and practical part.

In the theoretical part, the different characteristics of the desired system were studied, such as: number of cameras, resolution of each camera, the different characteristics of the sensors (WDR, HDR, exposure) etc. Depending on these characteristics, data transmission through an IP network was addressed. In addition to the hardware characteristics, the possibility of embedding stitching capabilities was studied. After spending some time understanding the background behind such techniques, an existing implementation of the stitching was adopted. Simulations were then made in order to estimate the characteristics of an FPGA capable of handling 5 cameras with 12MP resolution each. An existing FPGA architecture extensively used in the industry was chosen and a mathematical model was provided in order to estimate the characteristics of such FPGA according to the different parameters of the camera-based system.

In the practical part, an implementation of the two first steps of the stitching algorithm (homograph estimation and warping) was performed on a FPGA using 2 cameras. The problems of code optimisation were addressed in order to achieve a functioning implementation with respect to the memory and computation capabilities of the FPGA.

7.25. Brick & Mortar Cookies

Participant: Anaïs Ducoffe.

The objective of the BMC project is to create a software that aims to present attendance and attractiveness of the customer in stores, based on automatic video analysis. This final system is designed to be used without changing current camera network of the customer store, dedicated to security purpose. Analysis should be given at different time and space resolutions. For instance, day attendance can be as interesting as year attendance. Moreover, shop owners want to be able to compare two given years or months, etc... As space resolution is concerned, the software should be able to give information about the global attractiveness of the store but should also analyze some specific zones.

IVA embedded on Bosch cameras

Intelligence Video Analysis (IVA) is embedded in some models of Bosch cameras. The algorithms are composed of human detection and tracking. They can be configured directly on the camera interface via *tasks*. Following Bosch *tasks* were selected and studied:

- **Loitering and idle object in a field tasks** enable to detect stop actions in a zone, when they happened and the stop positions.
- **Entering and leaving field tasks** enable to know when a person enters or leaves a zone.
- **Detect people in a field task** enables counting people in a zone.
- **Crossing lines tasks** for counting people entering or leaving shop. We are able to know when the line was crossed and in what sense.

It is not possible to get people trajectory when metadata from Bosch cameras are acquired in offline mode. Then we studied live connection to get metadata directly from the camera stream using a RTSP connection. Metadata information is saved in XML format.

The previously enumerated tasks use algorithms to detect people and get their trajectories. STARS team has developed similar algorithms and has adapted their parameters values to the specific needs of this software. Moreover these algorithms can be run on any type of video cameras (live and offline modes) whereas Bosch IVA can only be run in live mode on compatible Bosch cameras. Stars algorithms can also combine several cameras at the same time in order to track people across the camera network. We need those algorithms to sell a system that doesn't need a new camera network but can be used with existing ones. They will be integrated in the final product.

Tests in real conditions

A system for testing cameras and our software was installed in partner store (Super U). Cameras were installed and configured to process all our use cases and test our mechanism to extract the metadata. We used only Bosch camera with embedded IVA. We successfully acquire 2 hours of the desired metadata. The results of embedded algorithms are reliable on realistic data : we get good results in counting people and trajectories are accurate.

Metadata storage in database

Metadata have to be stored in a hierarchical way as request of the metadata by the application should be easy and quick. We choose to store metadata in a database. This database design was constraint by data storage speed and a quick access for live computation. Different parts of database (store information, devices description etc...) were designed to be as much independent as possible.

Web interface (GUI)

The graphic interface design is in progress. The interface will be a web based one to narrow compatibility problems: the application should be used as well with a computer as a tablet.

7.26. Monitoring Older People Experiments

Participants: Matias Marin, Etienne Corvée, François Brémond.

This year we have conducted many experiments, especially in Nice and partially in Thessaloniki, Lulea, Taiwan and Dublin, to validate our studies on monitoring older people suffering from various behavioral disorders in the framework of several projects.

DEM@CARE PROJECT

For the project Dem@care (see section 9.2.1.3), we use PCs with ASUS cameras, for monitoring and collecting a video dataset associated with metadata [0]. The software CAR is installed to automatically annotate the videos [0]. Data is recorded locally, and backups are made automatically and remotely: one on the server (LAB at the nursery home) and another backup at Inria. These data can be reached locally at the nursery home, thanks to the server located in the lab; also, they are all accessible from Inria network by ssh.

SafEE PROJECT

SafEE project (see section 9.1.1.2) experiments in the nursing home have started at the end of 2015, and are made up with different technologies (wifi, wired network, smart phones, Kinect, RFID, tablet...). In the nursing room, 2 PCs with KINECT2 are connected for monitoring the residents and are stored in a database. Another PC with Windows software is configured for SafEE serious game (cognitive games and music box, 1.7 version). Moreover, a Wifi access point will be used by medical staff at the nursery home to connect to a Graphical User Interface through a website designed by INDES team [0], to consult patient data (their daily activities), from the activity history stored in the database or in real-time. Another device called AromaCare is installed in the rooms, which is a connected aroma diffuser by Wi-Fi. With the app Aroma Therapeutics (smartphone or tablet) we can manage several diffusers, by scheduling different programs each day and change the intensity of the diffusion.

In patients' home the same devices than Nursing Home have to be configured. Today, only recordings are done and stored at Inria.

OTHER PROJECTS at ICP (Institut Claude Pompidou)

ICP has now a remote access by using rdesktop, which is safer than team-viewer screen sharing session. The installation of new experimentations (e.g. praxis, relaxation, serious games) is now in progress: the configuration expected includes PCs with KINECT2 connected at ICP network and accessible from Inria. Some experimentations will use wireless sensors (e.g. accelerometer, pressure), controlled by the app wireless tag (on smartphone or tablet) to measure fine patient activities: motion, kettle utilization, etc.

---

[0] https://team.inria.fr/stars/demcare-chu-dataset/

[0] https://team.inria.fr/stars/software/car-complex-activity-recognition-component-installation/

[0] (http://webrobotics.inria.fr:8081/hop/events

TITANE Project-Team

7. New Results

7.1. Analysis

7.1.1. Planar Shape Detection and Regularization in Tandem

Participants: Sven Oesau, Florent Lafarge, Pierre Alliez.

In collaboration with EADS ASTRIUM

We contributed a method for planar shape detection and regularization from raw point sets. The geometric modeling and processing of man-made environments from measurement data often relies upon robust detection of planar primitive shapes. In addition, the detection and reinforcement of regularities between planar parts is a means to increase resilience to missing or defect-laden data as well as to reduce the complexity of models and algorithms down the modeling pipeline. The main novelty behind our method is to perform detection and regularization in tandem. We first sample a sparse set of seeds uniformly on the input point set, then perform in parallel shape detection through region growing, interleaved with regularization through detection and reinforcement of regular relationships (coplanar, parallel and orthogonal). In addition to addressing the end goal of regularization, such reinforcement also improves data fitting and provides guidance for clustering small parts into larger planar parts (Figure 1). We evaluate our approach against a wide range of inputs and under four criteria: geometric fidelity, coverage, regularity and running times. Our approach compares well with available implementations such as the efficient RANSAC-based approach proposed by Schnabel and co-authors in 2007 [8]. This work has been published in the Computer Graphics Forum journal.

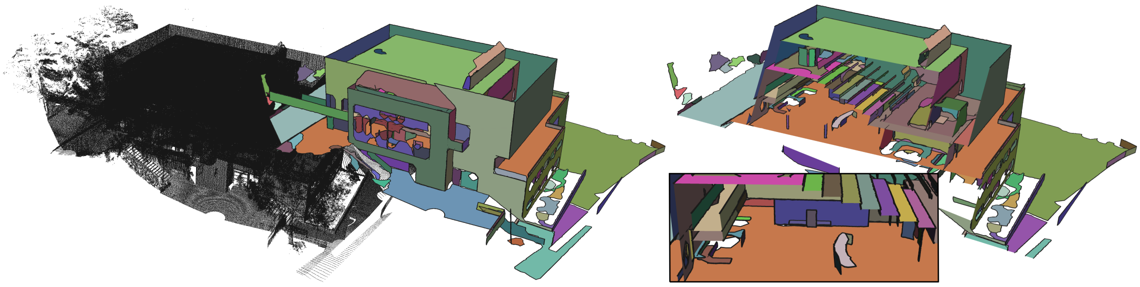


Figure 1. Shape detection and regularization. The input point set (5.2M points) has been acquired via a LIDAR scanner, from the inside and outside of a physical building. 200 shapes have been detected, aligned with 12 different directions in 179 different planes. The cross section depicts the auditorium in the upper floor and the entrance hall in the lower floor. The closeup highlights the steps of the auditorium which are made up of perfectly parallel and orthogonal planes.

7.1.2. Image partitioning into convex polygons

Participants: Liuyun Duan, Florent Lafarge.

In collaboration with Geoimage

The over-segmentation of images into atomic regions has become a standard and powerful tool in Vision. Traditional superpixel methods, that operate at the pixel level, cannot directly capture the geometric information disseminated into the images. We propose an alternative to these methods by operating at the level of geometric shapes. Our algorithm partitions images into convex polygons. It presents several interesting properties in terms of geometric guarantees, region compactness and scalability. The overall strategy consists in building a Voronoi diagram that conforms to preliminarily detected line-segments, before homogenizing the partition by spatial point process distributed over the image gradient. Our method is particularly adapted to images with strong geometric signatures, typically man-made objects and environments (Figure 2). We show the potential of our approach with experiments on large-scale images and comparisons with state-of-the-art superpixel methods [17]. This work has been published in the Computer Graphics Forum journal. Published in the proceedings of CVPR (IEEE conference on Computer Vision and Pattern Recognition).

Figure 2. Image partitioning into convex polygons.

7.1.3. Object Classification via Planar Abstraction

Participants: Sven Oesau, Florent Lafarge, Pierre Alliez.

In collaboration with EADS ASTRIUM.

We contributed a supervised machine learning approach for classification of objects from sampled point data. The main idea consists in first abstracting the input object into planar parts at several scales, then discriminate between the different classes of objects solely through features derived from these planar shapes. Abstracting into planar shapes provides a means to both reduce the computational complexity and improve robustness to defects inherent to the acquisition process. Measuring statistical properties and relationships between planar shapes offers invariance to scale and orientation. A random forest is then used for solving the multiclass classification problem. We demonstrate the potential of our approach on a set of indoor objects from the Princeton shape benchmark and on objects acquired from indoor scenes and compare the performance of our method with other point-based shape descriptors [22] (see Figure 3).

7.1.4. Optimizing partition trees for multi-object segmentation with shape prior

Participants: Emmanuel Maggiori, Yuliya Tarabalka.

This work has been done in collaboration with Dr. Guillaume Charpiat (TAO team, Inria Saclay).

Partition trees, multi-class segmentation, shape priors, graph cut.

A partition tree is a hierarchical representation of an image. Once constructed, it can be repeatedly processed to extract information. Multi-object multi-class image segmentation with shape priors is one of the tasks that can be efficiently done upon an available tree. The traditional construction approach is a greedy clustering based on color similarities. However, not considering higher level cues during the construction phase leads to trees that might not accurately represent the underlying objects in the scene, inducing mistakes in the later segmentation. We proposed a method to optimize a tree based both on color distributions and shape priors [15]. It consists in pruning and regrafting tree branches in order to minimize the energy of the best segmentation that can be extracted from the tree. Theoretical guarantees help reduce the search space and make the optimization efficient. Our experiments (see Figure 4) show that we succeed in incorporating shape information to restructure a tree, which in turn enables to extract from it good quality multi-object segmentations with shape priors. Published in the proceedings of BMVC (British Machine Vision Conference).

7.2. Reconstruction

7.2.1. LOD Generation for Urban Scenes

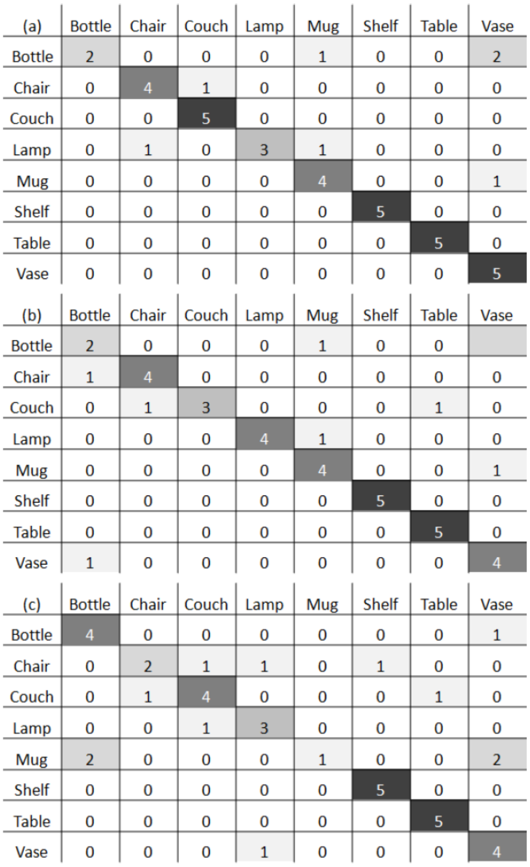
Participants: Florent Lafarge, Pierre Alliez.

We contributed a novel approach that reconstructs 3D urban scenes in the form of levels of detail (LODs). Starting from raw data sets such as surface meshes generated by multi-view stereo systems, our algorithm proceeds in three main steps: classification, abstraction and reconstruction (Figure 5). From geometric attributes and a set of semantic rules combined with a Markov random field, we classify the scene into four meaningful classes. The abstraction step detects and regularizes planar structures on buildings, fits icons on trees, roofs and facades, and performs filtering and simplification for LOD generation. The abstracted data are then provided as input to the reconstruction step which generates watertight buildings through a min-cut formulation on a set of 3D arrangements. Our experiments on complex buildings and large scale urban scenes show that our approach generates meaningful LODs while being robust and scalable. By combining semantic segmentation and abstraction it also outperforms general mesh approximation approaches at preserving urban structures [10]. Published in the ACM Transactions on Graphics journal.

7.2.2. A Surface Reconstruction Method for In-Detail Underwater 3D Optical Mapping

Participant: Pierre Alliez.

In collaboration with Ricard Campos and Rafael Garcia from the Computer Vision and Robotics Group from University of Girona, and Mariette Yvinec from the GEOMETRICA Inria project-team.



(a)	Bottle	Chair	Couch	Lamp	Mug	Shelf	Table	Vase
Bottle	2	0	0	0	1	0	0	2
Chair	0	4	1	0	0	0	0	0
Couch	0	0	5	0	0	0	0	0
Lamp	0	1	0	3	1	0	0	0
Mug	0	0	0	0	4	0	0	1
Shelf	0	0	0	0	0	5	0	0
Table	0	0	0	0	0	0	5	0
Vase	0	0	0	0	0	0	0	5

(b)	Bottle	Chair	Couch	Lamp	Mug	Shelf	Table	Vase
Bottle	2	0	0	0	1	0	0	
Chair	1	4	0	0	0	0	0	0
Couch	0	1	3	0	0	0	1	0
Lamp	0	0	0	4	1	0	0	0
Mug	0	0	0	0	4	0	0	1
Shelf	0	0	0	0	0	5	0	0
Table	0	0	0	0	0	0	5	0
Vase	1	0	0	0	0	0	0	4

(c)	Bottle	Chair	Couch	Lamp	Mug	Shelf	Table	Vase
Bottle	4	0	0	0	0	0	0	1
Chair	0	2	1	1	0	1	0	0
Couch	0	1	4	0	0	0	1	0
Lamp	0	0	1	3	0	0	0	0
Mug	2	0	0	0	1	0	0	2
Shelf	0	0	0	0	0	5	0	0
Table	0	0	0	0	0	0	5	0
Vase	0	0	0	1	0	0	0	4

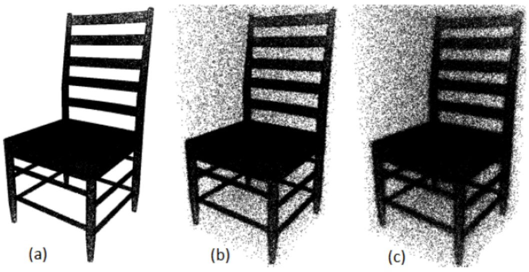
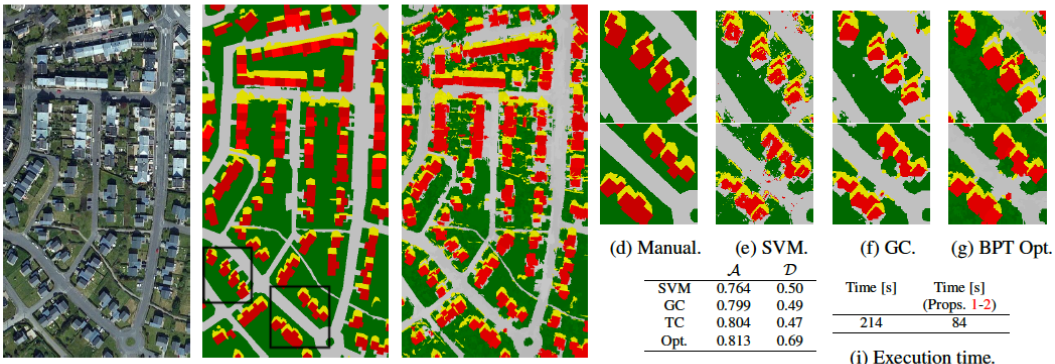


Figure 3. Classification. Left: We used four tabletop object classes from the Princeton Shape Benchmark: Bottle, Lamp, Mug and Vase. We also select four furniture object classes common to indoor scenes: Chair, Couch, Shelf and Table. Right: We evaluate our approach through computing a confusion matrix, for an increasing amount of noise and outliers. (a): Without noise and outliers. The precision of the class prediction is 82, 5%. The classifier is not reliable for classifying the bottles, which get mislabeled as vases. (b): Added 10% outliers and 0.5% noise. Compared to the noise-free version the precision slightly dropped to 77.5%. (c): Added 20% outliers and 1% noise. The method maintains a precision of 70% for this level of noise.



(a) Color image. (b) Manual segment. (c) BPT optimization. (d) Manual. (e) SVM. (f) GC. (g) BPT Opt.

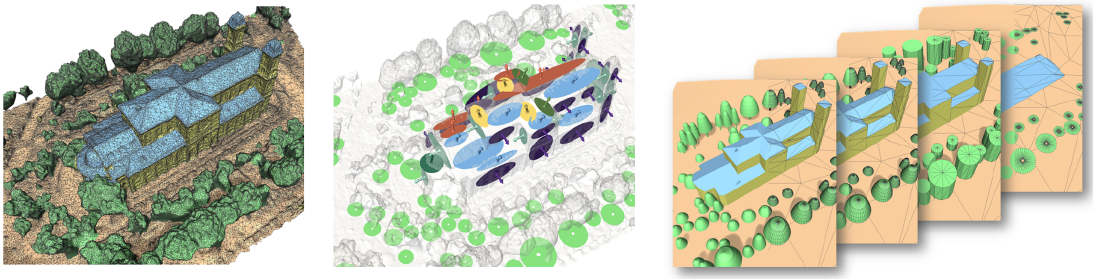
	$\mathcal{A}$	$\mathcal{D}$
SVM	0.764	0.50
GC	0.799	0.49
TC	0.804	0.47
Opt.	0.813	0.69

(h) Performance.

Time [s]	Time [s] (Props. 1-2)
214	84

(i) Execution time.

Figure 4. Classification results for the satellite image over Brest. $\mathcal{A}$ denotes overall classification accuracy, and $\mathcal{D}$ denotes average buildings overlap. The performance of the proposed binary partition tree (BPT) optimization method is compared with the following methods: 1) support vector machines (SVM) classification; 2) graph cut (GC) with α-expansion; 3) cut on the BPT, regularized by the number of regions without using shape priors (TC).



Classification Abstraction Reconstruction

Figure 5. LOD Generation for Urban Scenes. Main steps of our algorithm.

Underwater range scanning techniques are starting to gain interest in underwater exploration, providing new tools to represent the seafloor. These scans (often) acquired by underwater robots usually result in an unstructured point cloud, but given the common downward-looking or forward-looking configuration of these sensors with respect to the scene, the problem of recovering a piecewise linear approximation representing the scene is normally solved by approximating these 3D points using a heightmap (2.5D). Nevertheless, this representation is not able to correctly represent complex structures, especially those presenting arbitrary concavities normally exhibited in underwater objects. We present a method devoted to full 3D surface reconstruction that does not assume any specific sensor configuration. The method presented is robust to common defects in raw scanned data such as outliers and noise often present in extreme environments such as underwater, both for sonar and optical surveys (Figure 6). Moreover, the proposed method does not need a manual preprocessing step. It is also generic as it does not need any information other than the points themselves to work. This property leads to its wide application to any kind of range scanning technologies and we demonstrate its versatility by using it on synthetic data, controlled laser-scans, and multibeam sonar surveys. Finally, and given the unbeatable level of detail that optical methods can provide, we analyze the application of this method on optical datasets related to biology, geology and archeology [4]. Published in the International Journal of Robotics Research.

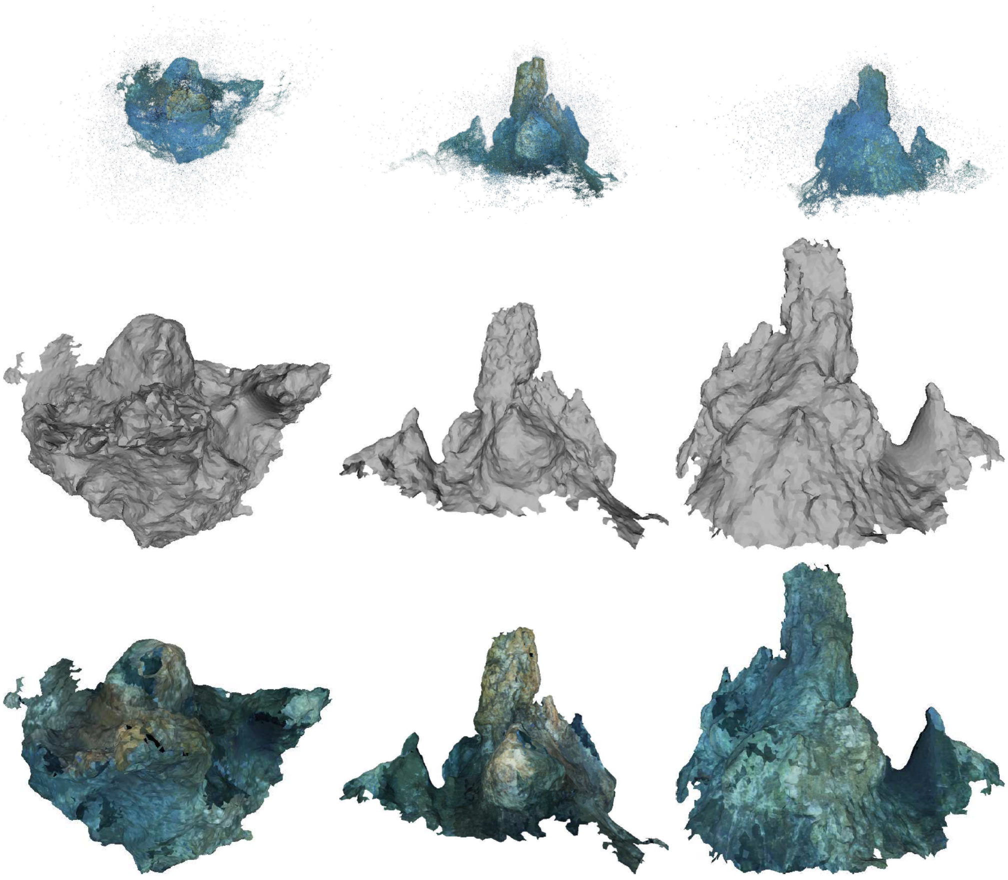


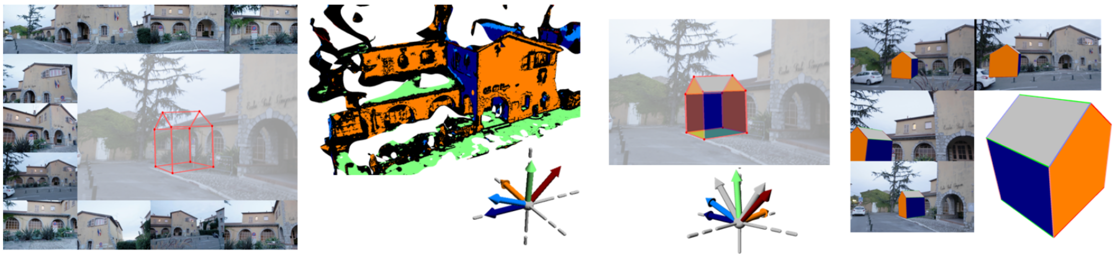
Figure 6. Underwater reconstruction.

7.2.3. Line Drawing Interpretation in a Multi-View Context

Participants: Jean-Dominique Favreau, Florent Lafarge.

In collaboration with Adrien Bousseau from the Inria project-team GraphDeco.

Many design tasks involve the creation of new objects in the context of an existing scene. Existing work in computer vision only provides partial support for such tasks. On the one hand, multi-view stereo algorithms allow the reconstruction of real-world scenes, while on the other hand algorithms for line-drawing interpretation do not take context into account. Our work combines the strength of these two domains to interpret line drawings of imaginary objects drawn over photographs of an existing scene. The main challenge we face is to identify the existing 3D structure that correlates with the line drawing while also allowing the creation of new structure that is not present in the real world. We propose a labeling algorithm to tackle this problem, where some of the labels capture dominant orientations of the real scene while a free label allows the discovery of new orientations in the imaginary scene (Figure 7). We illustrate our algorithm by interpreting line drawings for urban planing, home remodeling, furniture design and cultural heritage [18]. Published in the proceedings of CVPR (IEEE conference on Computer Vision and Pattern Recognition).



(a) Input (b) Dominant 3D orientations (c) Orientation labeling (d) Output

Figure 7. Line Drawing Interpretation. (a) Our algorithm takes as input multiple images of a scene along with a line-drawing traced over one of these images. (b) We first detect the dominant orientations of the existing scene from its multi-view stereo reconstruction. (c) Our labeling algorithm estimates the orientation of each facet of the drawing, favoring orientations already present in the scene. We visualize each dominant orientation with a random color, gray denotes new orientations. (d) We finally solve for the 3D model corresponding to the estimated orientations.

7.2.4. Marked point process model for curvilinear structures extraction

Participant: Yuliya Tarabalka [contact].

In collaboration with Seong-Gyun Jeong and Dr. Josiane Zerubia (AYIN team, Inria-SAM).

In this work, we proposed a new marked point process (MPP) model and the associated optimization technique to extract curvilinear structures [12]. Given an image, we compute the intensity variance and rotated gradient magnitude along the line segment. We constrain high level shape priors of the line segments to obtain smoothly connected line configuration. The optimization technique consists of two steps to reduce the significance of the parameter selection in our MPP model. We employ a Monte Carlo sampler with delayed rejection to collect line hypotheses over different parameter spaces. Then, we maximize the consensus among line detection results to reconstruct the most plausible curvilinear structures without parameter estimation process. Experimental results (see Figure 8) show that the algorithm effectively localizes curvilinear structures on a wide range of datasets.

7.2.5. Inference of curvilinear structure based on learning a ranking function and graph theory

Participant: Yuliya Tarabalka [contact].

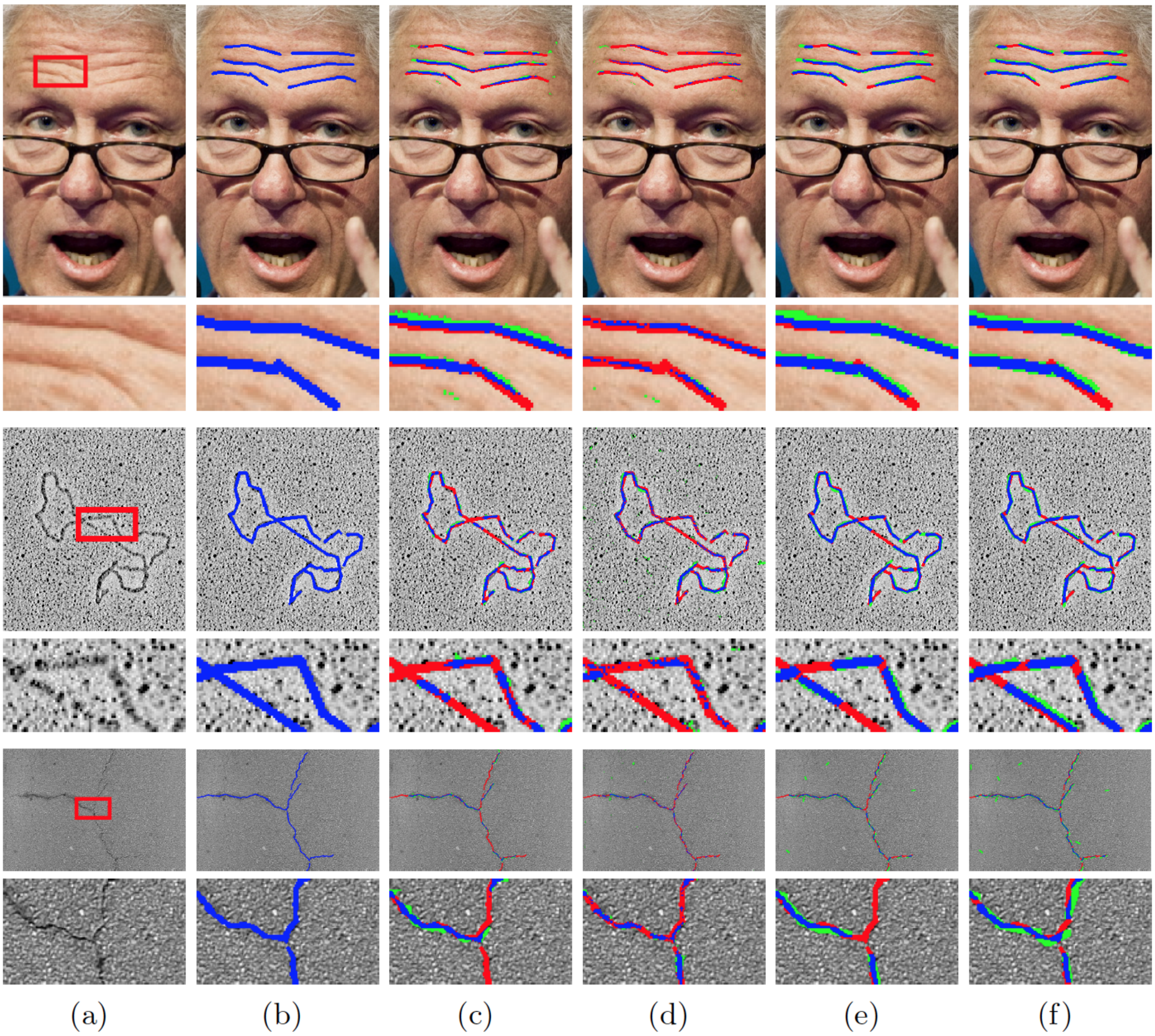
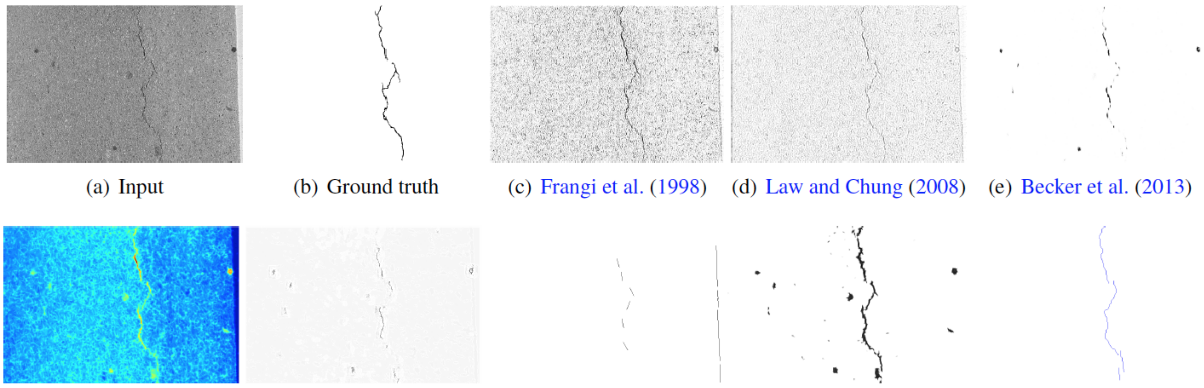


Figure 8. We visualize the localization of the curvilinear structures on input images (a). We compare with the results of a manually labeled image by a human expert (b), morphological filtering [Talbot 2007] (c), supervised feature learning [Becker 2013] (d), baseline MPP (e), and the proposed algorithm (f). Threshold values of (c) and (d) are chosen to achieve the closest recall scores to the proposed method. We use blue pixels to indicate areas which are completely corresponding to (b). Green and red pixels denote over-detected and under-detected areas, respectively, as compared with ground-truth. The name of the test images is from top to bottom: WRINKLE, DNA, *and* CRACK.

In collaboration with Seong-Gyun Jeong and Dr. Josiane Zerubia from the AYIN team and Dr. Nicolas Nisse from the COATI project-team.

Curvilinear structure extraction, inference of structured data, ranking learning, graphical model, shape simplification.

To detect curvilinear structures in natural images, we proposed a novel ranking learning system and an abstract curvilinear shape inference algorithm based on graph theory. We analyze the curvilinear structures as a set of small line segments. In this work, the rankings of the line segments are exploited to systematize the topological feature of the curvilinear structures. A Structured Support Vector Machine is employed to learn the ranking function that predicts the correspondence of the given line segments and the latent curvilinear structures. We first extract curvilinear features using morphological profiles and steerable filter responses. Also, we propose an orientation-aware feature descriptor and a feature grouping operator to improve the structural integrity during the learning process. To infer the curvilinear structure, we build a graph based on the output rankings of the line segments. We progressively reconstruct the curvilinear structure by looking for paths between remote vertices in the graph. Experimental results (see Figure 9 for an example of the experimental results' comparison on the CRACK dataset) show that the proposed algorithm faithfully detects the curvilinear structures within various datasets.


(a) Input (b) Ground truth (c) Frangi et al. (1998) (d) Law and Chung (2008) (e) Becker et al. (2013)

(f) Output rankings (g) Sironi et al. (2014) (h) Jeong et al. (2015) (i) Dissimilarity scores (j) Structure inference

Figure 9. Inference of curvilinear structure on the CRACK *dataset. Our approach is depicted by (j).*

7.3. Approximation

7.3.1. Isotopic approximation within a tolerance volume

Participants: Manish Mandad, Pierre Alliez.

In collaboration with David Cohen-Steiner from the GEOMETRICA project-team.

We introduce an algorithm that generates a surface triangle mesh given an input tolerance volume. The mesh is guaranteed to be within the tolerance, intersection free and topologically correct. A pliant meshing algorithm is used to capture the topology and discover the anisotropy in the input tolerance volume in order to generate a concise output. We first refine a 3D Delaunay triangulation over the tolerance volume while maintaining a piecewise-linear function on this triangulation, until an isosurface of this function matches the topology sought after. We then embed the isosurface into the 3D triangulation via mutual tessellation, and simplify it while preserving the topology. Our approach extends to surfaces with boundaries and to non-manifold surfaces. We demonstrate the versatility of our approach on a variety of data sets and tolerance volumes [7]. Figure 10 illustrates the robustness of our approach on defect-laden inputs.

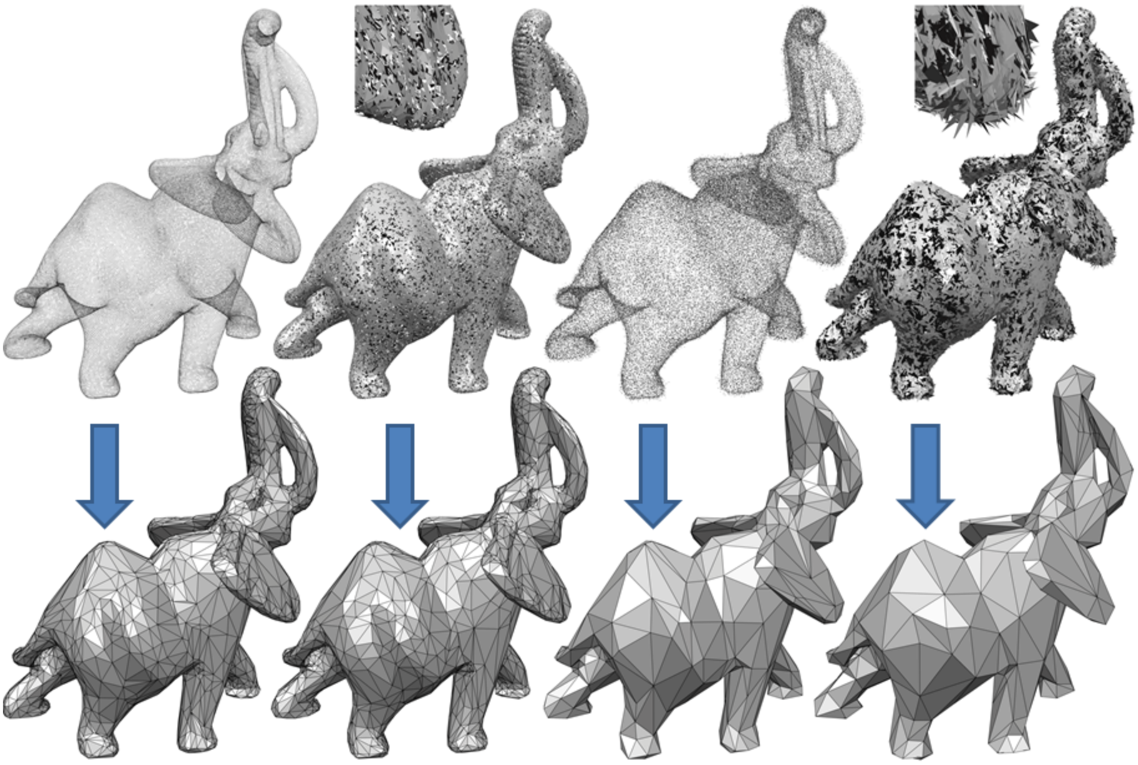


Figure 10. Isotopic approximation within a tolerance volume.

7.3.2. Structure-Aware Mesh Decimation

Participants: David Salinas, Florent Lafarge, Pierre Alliez.

We contributed to a novel approach for the decimation of triangle surface meshes. Our algorithm takes as input a triangle surface mesh and a set of planar proxies detected in a pre-processing analysis step, and structured via an adjacency graph. It then performs greedy mesh decimation through a series of edge collapse operators, designed to approximate the local mesh geometry as well as the geometry and structure of proxies (Figure 11). Such structure-preserving approach is well suited to planar abstraction, i.e., extreme decimation approximating well the planar parts while filtering out the others. Our experiments on a variety of inputs illustrate the potential of our approach in terms of improved accuracy and preservation of structure [9].

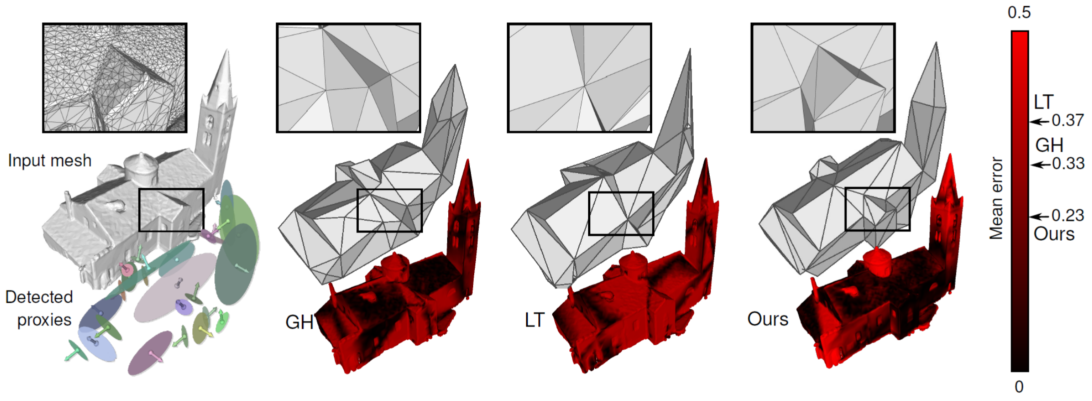


Figure 11. Structure-aware mesh decimation. Our algorithm simplifies dense triangle surface meshes via a structured set of planar proxies (left) that guides the decimation process while preserving the structure. At coarse complexity (here 50 vertices), common mesh decimation approaches (middle, quadric error metrics from Garland-Heckbert, and volume-preserving from Lindstrom-Turk) fail to reach low approximation error (see colored meshes) while preserving structure (see closeups).

7.3.3. CGALmesh: a Generic Framework for Delaunay Mesh Generation

Participants: Clément Jamin, Pierre Alliez.

In collaboration with Mariette Yvinec and Jean-Daniel Boissonnat from the GEOMETRICA project-team.

CGALmesh is the mesh generation software package of the Computational Geometry Algorithm Library (CGAL). It generates isotropic simplicial meshes – surface triangular meshes or volume tetrahedral meshes – from input surfaces, 3D domains as well as 3D multi-domains, with or without sharp features (see Figure 12). The underlying meshing algorithm relies on restricted Delaunay triangulations to approximate domains and surfaces, and on Delaunay refinement to ensure both approximation accuracy and mesh quality. CGALmesh provides guarantees on approximation quality as well as on the size and shape of the mesh elements. It provides four optional mesh optimization algorithms to further improve the mesh quality. A distinctive property of CGALmesh is its high flexibility with respect to the input domain representation. Such a flexibility is achieved through a careful software design, gathering into a single abstract concept, denoted by the oracle, all required interface features between the meshing engine and the input domain. We already provide oracles for domains defined by polyhedral and implicit surfaces [6].

7.4. Watermarking

7.4.1. Anti-Cropping Blind Resynchronization for 3D Watermarking

Participants: Xavier Rolland-Nevière, Pierre Alliez, Gwenaël Doërr.

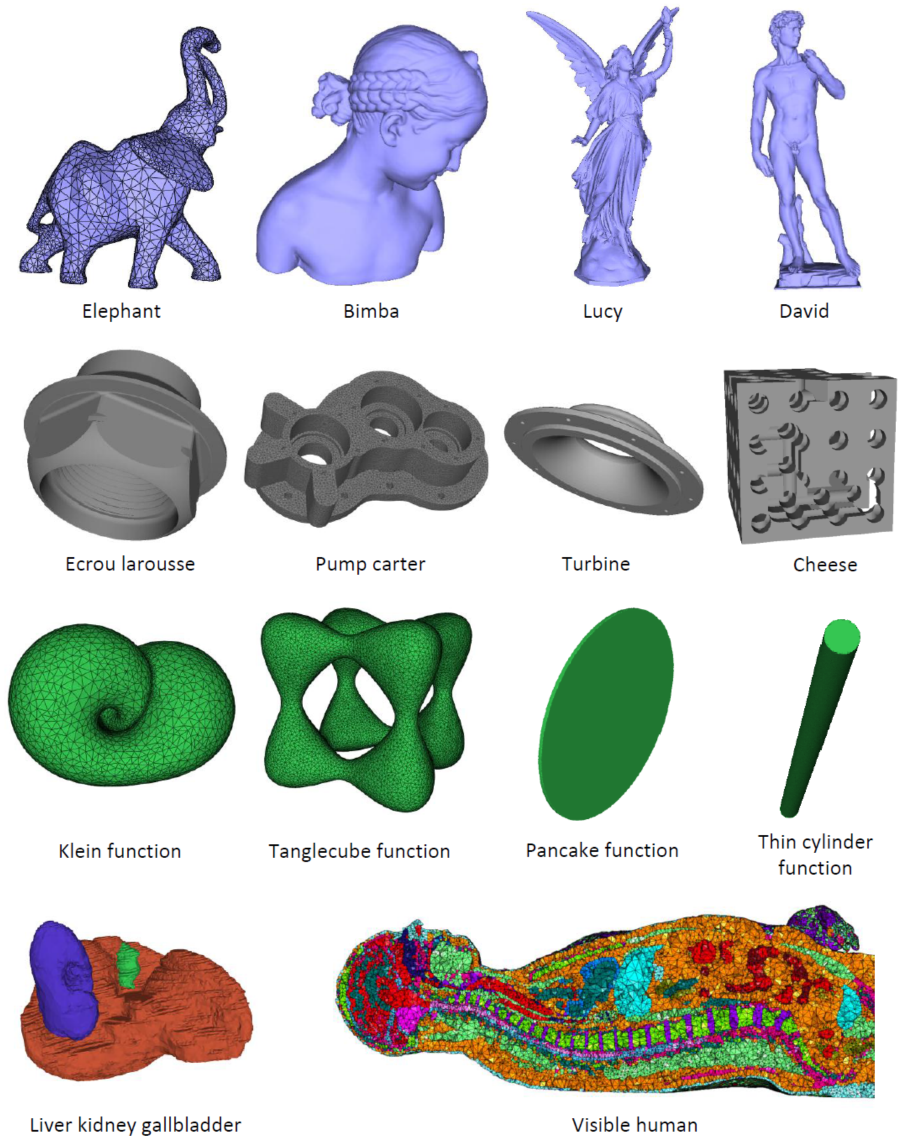


Figure 12. Input domains: domains bounded by smooth surfaces (top row, in blue), CAD models with sharp features (second row, in grey), implicit functions (third row, in green), 3D images (bottom row, multicolored).

Radial-based 3D watermarking alters the distances between the center of mass of the 3D mesh and its vertices. These watermarking systems are inherently sensitive to cropping. To address this limitation, this paper introduces a complementary blind resynchronization module to transmit critical synchronization information to the watermark decoder. Spherical patterns formed by several secret landmark vertices are embedded alongside the payload and blindly retrieved by the decoder, thereby conveying the synchronization information needed. Experimental results showcase significant improvement against cropping, while preserving performances against valumetric attacks thanks to a control parameter that automatically switches between alternate resynchronization modes [16].

TOSCA Project-Team

7. New Results

7.1. Probabilistic numerical methods, stochastic modelling and applications

Participants: Mireille Bossy, Nicolas Champagnat, Madalina Deaconu, Coralie Fritsch, Benoît Henry, James Inglis, Antoine Lejay, Oana-Valeria Lupascu, Sylvain Maire, Paolo Pigato, Alexandre Richard, Denis Talay, Etienne Tanré, Denis Villemonais.

7.1.1. Published works and preprints

- M. Bossy with H. Quinteros (UChile) submitted a paper [36] on the strong convergence of the symmetrized Milstein scheme for some CEV-like SDEs.
- M. Bossy and J.-F. Jabir (University of Valparaíso) submitted a paper [35] on the particle approximation for Lagrangian stochastic models with specular boundary condition.
- M. Bossy with N. Maizi (Mines ParisTech) and O. Pourtallier (Inria) published a book chapter [31] on game theory analysis for carbon auction market through electricity market coupling hypothesis.
- M. Bossy, O. Faugeras (Inria Sophia, EPI NEUROMATHCOMP), and D. Talay published a clarification on the well-posedness of the limit equations to the mean-field N-neuron models proposed in [58] and proven the associated propagation of chaos property. They also have completed the modeling issue in [58] by discussing the well-posedness of the stochastic differential equations which govern the behavior of the ion channels and the amount of available neurotransmitters. See [15].
- M. Bossy, N. Champagnat, S. Maire and L. Violeau worked with H. Leman (CMAP, Ecole Polytechnique) and M. Yvinec (Inria Sophia, GEOMETRICA team) on Monte Carlo methods for the linear and non-linear Poisson-Boltzmann equations [14]. These methods are based on walk on spheres algorithm, simulation of diffusion processes driven by their local time, and branching Brownian motion to deal with the nonlinear case.
- Together with M. Baar and A. Bovier (Univ. Bonn), N. Champagnat studied the adaptive dynamics of populations under the assumptions of large population, rare and small mutations [34]. In this work, the three limits are taken simultaneously, contrary to the classical approach, where the limits of large population and rare mutations are taken first, and next the limit of small mutations [59]. We therefore obtain the precise range of assumptions under which these limits can be taken, and provide explicit biological conditions for which our approximation is valid.
- N. Champagnat and C. Fritsch worked with F. Campillo (Inria Sophia-Antipolis, LEMON team) on the links between a branching process and an integro-differential equation of a growth-fragmentation-death model [37]. They proved that the two representations of the model lead to the same criteria of invasion of a population in a given environment.
- Using a new method to compute the expectation of an integral with respect to a random measure, N. Champagnat and B. Henry obtained explicit formulas for the moments of the frequency spectrum in the general branching processes known as Splitting Trees, with neutral mutations and under the infinitely-many alleles model [40]. This allows them to obtain a law of large numbers for the frequency spectrum in the limit of large time.
- N. Champagnat and P.-E. Jabin (Univ. Maryland) improved significantly the description of the functional spaces in the preprint [41], devoted to the study of strong existence and pathwise uniqueness for stochastic differential equations (SDE) with rough coefficients, typically in Sobolev spaces.

- N. Champagnat and D. Villemonais obtained criteria for existence and uniqueness of quasi-stationary distributions (QSD) and Q-processes for general absorbed Markov processes [17]. A QSD is a stationary distribution conditionally on non-absorbtion, and the Q-process is defined as the original Markov process conditioned to never be absorbed. The criteria ensure exponential convergence of the t-marginal of the process conditioned not to be absorbed at time t, to the QSD and also the exponential ergodicity of the Q-process.
- N. Champagnat and D. Villemonais obtained criteria for existence, uniqueness and exponential convergence in total variation to QSD for general absorbed and killed diffusion processes [43], [42]. For diffusions without killing [43], the criterion obtained is equivalent to the property that a diffusion on natural scale coming down from infinity has uniformly (w.r.t. the initial condition) bounded expectation at a fixed time t. The criteria obtained for diffusion processes with killing on $[0, \infty)$ [42] combine the last criteria and conditions on the killing time only close to 0, provided ∞ is an entrance boundary.
- N. Champagnat and D. Villemonais obtained criteria for existence, uniqueness and exponential convergence in total variation to QSD for general multi-dimensional birth and death processes in $\mathbb{Z}_+^d$ absorbed at the boundary $\mathbb{Z}_+^d \setminus \mathbb{N}^d$ [44]. These birth and death models are motivated by population dynamics and the criteria obtained assume stronger intra-specific competition than inter-specific competition. These resuls are the first one for such processes, except for the particular case of branching processes, which can be studied using very specific methods.
- M. Deaconu, S. Herrmann and S. Maire introduced a new method for the simulation of the exit time and position of a δ-dimensional Brownian motion from a domain. This method is based on the connexion between the δ-dimensional Bessel process and the δ-dimensional Brownian motion thanks to an explicit Bessel hitting time distribution associated with a particular curved boundary. This allows to build a fast and accurate numerical scheme for approximating the brownian hitting time [19].
- M. Deaconu and O. Lupaşcu worked with L. Beznea (Bucharest, Romania) on the probabilistic interpretation of fragmentation phenomena. They constructed a continuous time branching process and characterized its behavior by using new potential theoretical tools [12].
- M. Deaconu, O. Lupaşcu and L. Beznea (Bucharest, Romania) started a new challenging work on the description of rupture phenomena like avalanches, by using fragmentation models. The physical properties of the model are deeply involved in this study. The first results concern a stochastic equation of fragmentation and branching processes related to avalanches [13].
- M. Deaconu, B. Dumortier and E. Vincent are working with the Venathec SAS on the acoustic control of wind farms. They constructed a new approach to control wind farms with a control model based on real-time source separation. They first designed a deterministic algorithm in order to maximize the electric production of the wind farms under the legal acoustic constraints. They showed that it is a non linear knapsack optimization problem and they proposed an efficient solution in that context using a branch and bound algorithm based on continuous relaxation. This work was published at the EWEA 2015 [30].
- In [49], B. Henry showed a central limit theorem for the population counting process of a supercritical Splitting Tree in the limit of large time. Thanks to the results of [40], he also obtained a central limit theorem for the frequency spectrum of Splitting Trees with neutral mutations and under the infinitely-many alleles model.
- S. Herrmann and E. Tanré have proposed a new very efficient algorithm to simulate the first-passage-time of a one-dimensional Brownian motion over a continuous curved boundary [23].
- J. Inglis and E. Tanré together with F. Delarue and S. Rubenthaler (Univ. Nice – Sophia Antipolis) completed their study of the mean-field convergence of a highly discontinuous particle system modeling the behavior of a spiking network of neurons [21].

- In collaboration with J. Maclaurin (Inria Sophia, EPI NEUROMATHCOMP) J. Inglis has presented a general framework to rigorously study the effect of spatio-temporal noise on traveling waves and stationary patterns. In particular the framework can incorporate versions of the stochastic neural field equation that may exhibit traveling fronts, pulses or stationary patterns. They have formulated a local SDE that describes the position of the stochastic wave up until a discontinuity time, at which point the position of the wave may jump and studied the local stability of this stochastic front and the long-time behavior of the stochastic wave [50].
- A. Lejay has continued his work on the Snapping Out Brownian motion, especially with regard to the simulation issues, with potential application to brain imaging techniques [33], [53].
- A. Lejay has continued his work on the simulation of processes with either discontinuous drift (with Arturo Kohatsu-Higa, Ritsumeikan Universitey and Kazuhiro Yasuda, Hosei University, Japan) [52] or with discontinuous coefficients (with Lionel Lenêtre and Géraldine Pichot, EPI SAGE, Irisa) [54].
- A. Lejay has continued his work on the theory of rough paths, notably with the sensitivity aspects with Laure Coutin (Univ. Toulouse III) [47].
- In collaboration with Ivan Dimov and Jean-Michel Sellier (BAS), S. Maire developed a new Monte Carlo method, called the walk on equations, to solve linear systems of equations [22].
- In collaboration with Xuan Vu, Caroline Chaux-Moulin and Nadege Thirion-Moreau, S. Maire developed a stochastic algorithm to decompose large non-negative tensors with applications in spectroscopy [28].
- In collaboration with Martin Simon, Sylvain Maire developed a variant of the walk on spheres method to deal with diffusion equations appearing in electrical impedance tomography.
- With Giang Nguyen, Sylvain Maire worked on finite differences techniques to deal with many kinds of boundary conditions that are met during the Monte Carlo simulation of diffusions [25].
- A. Richard submitted a paper [56] on the spectral representation of L^2-indexed increment-stationary processes. The main result states that any random field (i.e. process indexed by a multidimensional parameter of a function in L^2) with stationary increments can be written as an integral against a random measure satisfying certain properties. Applications to sample path properties of a multiparameter fractional Brownian motion are exhibited.
- D. Villemonais worked with P. Del Moral (Univ. Sydney) on the conditional ergodicity of time inhomogeneous diffusion processes [48]. They proved that, conditionally on non extinction, an elliptic time-inhomogeneous diffusion process forgets its initial distribution exponentially fast. An interacting particle scheme to numerically approximate the conditional distribution is also provided.
- D. Villemonais proved a Foster-Lyapunov type criterion which ensures the α-positive recurrence of birth and death processes. This criterion also provides a non-trivial subset of the domain of attraction for quasi-stationary distributions. Finally, this study leads to a Foster-Lyapunov type criterion which ensures the exponential ergodicity of a Fleming-Viot type particle system whose particles evolve as birth and death processes. The criterion also ensures the tightness of the sequence of empirical stationary distributions considered as a family of random measures. A numerical study of the speed of convergence of the particle system is also obtained under various settings [29].
- J. Inglis and D. Talay ended their work on mean-field limits of a stochastic particle system smoothly interacting through threshold hitting-times and applications to neural networks with dendritic component [51].

7.1.2. Other works in progress

- Together with M. Andrade (Univ. Paris 7) and R. Ferrière (ENS Paris and Univ. Arizona), N. Champagnat is working on the phenomenon of clustering in populations structured by space and traits for which local adaptation favors different trait values at different spatial locations. Two methods are used and numerically validated: a Turing instablity method and a Hamilton-Jacobi approximation of the population density. This work is currently being written.

- N. Champagnat and J. Claisse (Ecole Polytechnique) are currently working on the ergodic and infinite horizon controls of discrete population dynamics with almost sure extinction in finite time. This can either correspond to control problems in favor of survival or of extinction, depending on the cost function. They have proved that these two problems are related to the QSD of the processes controled by Markov controls. This work is currently being written.
- N. Champagnat and C. Fritsch worked with F. Campillo (Inria Sophia-Antipolis, LEMON team) on the variations of the principal eigenvalue (resp. the survival probability) of an integro-differential equation (resp. branching process) of growth-fragmentation-death models with respect to an environmental parameter. This work is currently being written.
- N. Champagnat, K. Coulibaly-Pasquier (Univ. Lorraine) and D. Villemonais are currently working on general criteria for existence, uniqueness and exponential convergence in total variation to QSD for multi-dimensional diffusions in a domain absorbed at its boundary. These results both improve and simplify the existing results and methods. This work is currently being written.
- N. Champagnat and D. Villemonais are currently working on extensions of their work [17] to general penalized processes, including time-inhomogeneous Markov processes with absorption. Their method allows to improve significantly the former results of [60], [61]. This work is currently being written.
- N. Champagnat and D. Villemonais are also working on extensions of the criteria of [17] in the form of Foster-Lyapunov criteria allowing to deal with cases where the convergence of conditional distribution to the QSD is not uniform with respect to the initial distribution. This work is currently being written.
- M. Deaconu and S. Herrmann are working on the numerical approach of the time-space Dirichlet problem.
- M. Deaconu, O. Lupaşcu and L. Beznea (Bucharest, Romania) worked on the numerical scheme for the simulation of an avalanche by using the fragmentation model. This work is currently being written.
- M. Deaconu, B. Dumortier and E. Vincent are working with the Venathec SAS on the acoustic control of wind farms. They plan to submit another article to IEEE transaction on sustainable energy soon. Currently they work on handling uncertainties in the model in order to design a stochastic algorithm.
- C. Fritsch worked with F. Campillo (Inria Sophia-Antipolis, LEMON team) and O. Ovaskainen (Univ. Helsinki) about the numerical analysis of the invasion of mutant populations in a chemostat, using branching processes and integro-differential models.
- C. Fritsch started a collaboration with B. Cloez (INRA, Montpellier) on a central limit theorem of mass-structured individual-based chemostat model.
- With P. Pigato, A. Lejay has continued his work on the estimation of parameters of skew diffusions.
- Within the ANESTOC Associate Team, R. Rebolledo (Pontificia Universidad Católica de Chile) and A. Richard initiated a work on the long-term behavior of a class of non-Markovian stochastic differential equations. These equations of Volterra type can be used to model the motion of a particle subject to friction forces in a heat bath, which could also be interesting in neuroscience for ion channels.
- A. Richard and E. Tanré are working with P. Orio (CINV, Chile) on the measurement of long-range dependence in series of neuronal spikes, and are providing a leaky integrate-and-fire model with fractional noise to include this effect. So far, we produced numerical experiments that confirm the existence of memory in our model, and A. Richard and E. Tanré now work on the convergence of the statistical estimator that measures this phenomenon.

- A. Richard, E. Tanré and S. Torres (Universidad de Valparaíso, Chile) are working on the definition of a skew fractional Brownian motion. The skew Brownian motion (sBm) is a process which is partly reflected when it reaches the horizontal line, making it a natural model for the motion of a particle crossing media with different diffusion properties. The fractional sBm is a modification of this process to incorporate long-range dependences. So far, we constructed a reflected fractional Brownian motion, and we are now investigating its approximation by a discrete-time process.
- During her internship supervised by E. Tanré and Romain Veltz (NEUROMATHCOMP team), Roberta Evangelista worked on "A stochastic model of gamma phase modulated orientation selectivity". Neurons in primary visual cortex (V1) are known to be highly selective for stimulus orientation. Recent experimental evidence has shown that, in awake monkeys, the orientation selectivity of V1 neurons is modulated by gamma oscillations. In particular, neurons' firing rate in response to the preferred orientation changes as a function of the gamma phase of spiking. The effect is drastically reduced for non-preferred orientations. We have introduced a stochastic model of a network of orientation-dependent excitatory and inhibitory spiking neurons. We have found conditions on the parameters such that the solutions of the mathematical model reproduce the experimental behavior.
- During his internship supervised by E. Tanré and Romain Veltz (NEUROMATHCOMP team), Quentin Cormier studies numerically and theoretically a model of spiking neuron in interaction with plasticity. The synaptic weights evolve according to biological law of plasticity. We study the existence of separable time scales. During his internship, Quentin Cormier also develop a numerical code to simulate large networks of neurons evolving according to this dynamics.
- C. Graham (Ecole Polytechnique) and D. Talay have written a large part of the second volume of their series on Mathematical Foundation of Stochastic Simulation.

7.2. Financial Mathematics

Participants: Mireille Bossy, Madalina Deaconu, Antoine Lejay, Sylvain Maire, Khaled Salhi, Denis Talay, Etienne Tanré.

7.2.1. Published works and preprints

- In collaboration with Jerome Lelong and Christophe Deluigi, Sylvain Maire built a new algorithm for the automatic integration and approximation of irregular functions [18]. This algorithm is tested numerically on the pricing of multidimensional exotic options.
- In collaboration with V. Reutenauer and C. Michel (CA-CIB), D. Talay and E. Tanré worked on a model in financial mathematics including bid-ask spread cost. They study the optimal strategy to hedge an interest rate swap that pays a fixed rate against a floating rate. They present a methodology using a stochastic gradient algorithm to optimize strategies. A paper is in revision [55].

7.2.2. Other works in progress

- K. Salhi works on partial hedging of options in an incomplete market, under constraints on the initial capital of the investor and assuming that the stock price is described by a Lévy process. In this case, perfect hedging is no more possible and we talk about partial hedging and minimization of risk. K. Salhi focuses on the Conditional Value-at-Risk minimization. He tries to give a numerical approximation to the solution in this context.
- In collaboration with J. Bion-Nadal (Ecole Polytechnique and CNRS), D. Talay pursued the study of a new calibration methodology based on dynamical risk measures and stochastic control PDEs.

VIRTUAL PLANTS Project-Team

6. New Results

6.1. Analysis of structures resulting from meristem activity

6.1.1. Acquisition and design of plant geometry

Participants: Frédéric Boudon, Christophe Pradal, Christophe Godin, Christian Fournier, Ibrahim Chedaddi, Mathilde Balduzzi, Julien Diener.

Virtual 3D model of plants are required in many areas of plant modeling. They can be used for instance to simulate physical interaction of real plant structures with their environment (ligth, rain, wind, pests, ...), to set up initial conditions of growth models or to assess their output against real data. In the past decade, methods have been developed to digitize plant architectures in 3D [76], [63]. These methods are based on direct measurements of position and shape of every plant organ in space. Although they provide accurate results, they are particularly time consuming. More rapid and automated methods are now required in order to collect plant architecture data of various types and sizes in a systematic way. In this aim, we explore the use of laser scanner and pictures.

- *Reconstruction of tree structures from 3D laser scanner data.* (Chakkrit Preuksakarn, Mathilde Balduzzi, Frédéric Boudon, Christophe Godin, Pascal Ferraro [Labri, Bordeaux], Yassin Refahi)

  We investigate the possibility to use 3D laser scanners to automate plant digitizing. We are developing algorithms to reconstruct branching systems without leaves or foliage from scanner data or from scan simulated on plant mock-up obtained using different digitizing method.

  For the branching systems, we previously proposed a reconstruction method to reconstruct plausible branching structures from laser scanner data based on the concept of space colonization [73]. Additionally, a number of automatic methods were proposed in the literature. The question of their comparison and relative accuracy is however critical for further exploitation in biological applications. To address such problem, we developed an evaluation pipeline that takes two plant structures as input and compares their organization using two indices of geometrical and structural similarities [55]. A first comparative evaluation of the different methods of the literature has been designed and conducted. A graphical editor has been developed and makes it possible to test the different methods and correct manually the reconstruction. A procedure to automatically determine phyllotactic angles from scans of small plants has been added to the reconstruction pipeline and has been tested on database of 150 scans of arabidopsis thaliana with different genotypes. The editor has also been tested on apple trees and large African trees.

  In the context of the PhD of M. Balduzzi, we also investigated the reconstruction of tree foliage from 3D scans. Such elements are crucial to study the interaction of the plant with its environment. However, laser scans contain outliers in the silhouette of the scans that make the meshing of the pointset extremely difficult. New generation of laser scanners provide intensity of the laser reflected on the surface of scanned objects. This intensity depends on the distance to the object, its optical property and the incidence angle. A first work on this topic showed that after correcting the distance effect, the incidence angle can be deduced from the intensity. From this result, we developed a reconstruction pipeline using the scan intensities and based on Shape-From-Shading. Outliers being along the edge of the surface point cloud, we chose to develop a propagation SFS method initialized with points of the scans with high quality. We proved that surface with constant intensity are necessarily surfaces of constant slope or sand-pile surfaces. Using this result, a propagation method along iso-intensity regions was developed. These surfaces can then be sampled to provide a smooth point set without outliers.

- *Reconstruction of annual plants from multi-view images.* (Simon Artzet, Jerome Chopard, Christian Fournier, Christophe Pradal, Christophe Godin, Xavier Sirault [CSIRO-HRPPC, Canberra])

Image-based phenotyping platforms in semi-controlled conditions offer large possibilities to perform genetic analyses of plant growth, architecture, light interception, and biomass accumulation over large time series for thousands of plants. However, methods for image analyses currently available are still very crude and need improvement and robustness to process huge amount of data. We are developping an integrated pipeline allowing assessment of growths of individual organs, of plant geometry, and of derived variables such as light interception. The pipeline currently consists of 2D image analysis workflows built with standard image libraries (OpenCV, Scikit.Image), algorithms for 3D reconstruction, segmentation and tracking of plant organs for maize (under development), and workflows for estimation of light interception by plants during their growth.A 3D FSPM model for maize architectural development, is used to help segmenting plant images and to automate the mapping between segmented 3D objects and plant organs defined in the model.

- *Reconstruction of root structures.* (Julien Diener, Fredéric Boudon, Christophe Pradal, Christophe Godin, Philippe Nacry [BPMP, INRA], Christophe Périn [AGAP, CIRAD], Anne Dievart [AGAP, CIRAD], Xavier Draye [UCL, Belgium])

  This research theme is supported by the Agropolis through the Rhizopolis project and by NUMEV.

  Similarly to aerial part of plants, new needs for automatic digitizing of root systems emerge. Most existing methods focus only on semi-automatic approaches. This does not support the high-throughput capabilities of acquisition systems. In the context of the RhizoScan project, we previously designed a prototype of an automatic image analysis pipeline to extract root system architecture of branching systems grown in Petri boxes. This pipeline provides i) a set of model based image segmentation method, ii) the extraction of a graph representation of the root system, and iii) a method to identify the root axes organization. This year, we improved and extended the pipeline in the following way:

  1. We integrated a validation step in the workflow based on the comparison method presented in [55].
  2. We developed a standard file format for root architecture (RSML) described in [19] during an international collaboration with the niversité Catholique de Louvain (Belgium), the CPIB of the University of Notthingham (UK), the University of Vienna (Austria), the Jülich research center (Germany) and INRA.

- *Reconstruction of virtual fruits from pictures.* (Ibrahim Chedaddi, Mik Cieslak, Nadia Bertin [Inra, Avignon], Frédéric Boudon, Christophe Godin, Michel Genard [Inra, Avignon], Christophe Goz-Bac [Université Montpellier 2])

  This research theme is supported by the Agropolis project MecaFruit3D.

  The aim of this work is to provide methods for generating fruit structure that can be integrated with models of fruit function. To this end, a modeling pipeline has been developed in the OpenAlea platform. It involves two steps: (1) generating a 3D volumetric mesh representation of the entire fruit, and (2) generating a complex vascular network that is embedded within this mesh using the concept of space colonization [75]. Previous studies demonstrated the possibility to create species-specific models of fruit structure with relatively low effort [57]. We focus now on validating the vascular networks by comparing them to experimental data from the literature. This work has been presented at the ISHS symposium in Montpellier [38]

  Using these fruit virtual structures, a mechanical model of fruit growth is also developed (see section 6.3.2) taking into account the distribution of water fluxes in the fruit.

6.1.2. Modeling the plant ontogenic programme

Participants: Christophe Godin, Yann Guédon, Jean-Baptiste Durand, Pierre Fernique, Marc Labadie, Christophe Pradal, Jean Peyhardi.

This research theme is supported by two PhD programmes.

The remarkable organization of plants at macroscopic scales may be used to infer particular aspects of meristem functioning. The fact that plants are made up of the repetition of many similar components at different scales, and the presence of morphological gradients, e.g. [52], [65], [66], [62], provides macroscopic evidence for the existence of regularities and identities in processes that drive meristem activity at microscopic scales. Different concepts have been proposed to explain these specific organizations such as "morphogenetic programme" [71], "age state" [61] or "physiological age" [54]. All these concepts state that meristem fate changes according to position within the plant structure and during its development. Even though these changes in meristem fate are specific to each species and lead to the differentiation of axes, general rules can be highlighted [61], [54]. Here we develop computational methods to decipher these rules.

- *Relating branching structure to the shoot properties* (Jean Peyhardi, Yann Guédon, Evelyne Coste [AGAP, AFEF team], Catherine Trottier [I3M], Yves Caraglio [AMAP], Pierre-Eric Lauri [AGAP, AFEF team])

  Shoot branching structures often take the form of a succession of homogeneous branching zones and have been analyzed using segmentation models such as hidden semi-Markov chains. Axillary meristem fates are influenced by local properties of the parent shoot such as for instance its growth rate or local curvature. The objective of this work, which was part of the PhD subject of Jean Peyhardi, is to develop statistical models that generalize hidden semi-Markov chains with the capability to incorporate explanatory variables that vary along the parent shoot (e.g. leaf growth rate, leaf surface, internode length, local curvature of the parent shoot). More precisely, the simple multinomial distributions that represent the axillary productions observed in the different branching zones are replaced by multinomial generalized linear models (GLMs). Since the two classical categories of multinomial GLMs that correspond either to nominal or ordinal categorical response variables were not appropriate, we chose to develop a new family of multinomial GLMs called partitioned conditional GLMs [72] that enable to tackle hierarchically-structured categorical response variables. Typically, we need to distinguish different timing of branching events (e.g. immediate shoot, one-year-delayed shoot and latent bud), different categories of offspring shoots (e.g. among one-year-delayed shoots, vegetative short shoot, vegetative long shoot and flowering shoot) and to specialize the explanatory variables for certain categories of offspring shoots (e.g. the growth of the parent shoot influence the immediate offspring shoots but not the one-year-delayed offspring shoots). The resulting integrative models are called semi-Markov switching partitioned conditional GLMs and have been applied to apple and pear tree branching structures.

- *Genetic determinisms of the alternation of flowering in apple tree progenies.* (Jean-Baptiste Durand, Alix Allard [AGAP, AFEF team], Jean Peyhardi, Baptiste Guitton [AGAP, AFEF team], Yan Holtz [AGAP, AFEF team] Catherine Trottier, Evelyne Costes [AGAP, AFEF team], Yann Guédon)

  A first study was published to characterize genetic determinisms of the alternation of flowering in apple tree progenies [58]. Data were collected at two scales: at whole tree scale (with annual time step) and a local scale (annual shoot, which corresponds to portions of stem that were grown during the same year). Two replications of each genotype were available.

  Indices were proposed for early detection of alternation during the juvenile phase. They were based on a trend model and a quantification of the deviation amplitudes and dependency, with respect to the trend. This allowed early quantification of alternation from the yearly numbers of inflorescences at tree scale. Some quantitative trait loci (QTL) were found in relation with this indices.

  For better interpretation of the relationships of alternation at both scales, new models and indices were developped for sequences of flowering events at axis scale. New data sets where collected in other F1 progenies. Ancestral relationships between parents of different progenies were taken into account to enhance the power of QTL detection, and other QTL were found using these new indices.

- *Identifying and characterizing patterns in tree-structured data* (Pierre Fernique, Jean-Baptiste Durand, Yann Guédon).

  In the context of Pierre Fernique's PhD (Montpellier 2 University and CIRAD), two complementary approaches were developed for analyzing patterns in tree-structured data:

  - multitype branching processes relying on local dependency properties for analyzing motifs.
  - multiple change-point models relying on long-term dependencies for segmenting trees in homogeneous zones.

  In multitype branching processes, the plant development is viewed as a demographic process, a parent entity of a given type generating child entities of different types (e.g. vegetative and flowering entities). Formally, the botanical entity properties are summarized as a categorical state variable. The number of child entities in each state is modeled through discrete multivariate distributions. Model selection procedures are necessary to specify parsimonious generation distributions. We developed an approach based on probabilistic graphical models to identify and exploit properties of conditional independence between numbers of children in different states, so as to simplify the specification of their joint distribution. The graph building stage was based on exploring the space of possible chain graph models, which required defining a notion of neighbourhood of these graphs [59]. To relax the strong constraints regarding dependencies induced by parametric distributions, mixture of graphical models were also considered [60]. Multitype branching processes were applied to the analysis of the patchiness pattern (consisting of canopies made of clumps of either vegetative or flowering growth units) in mango trees. To identify the clumps, a novel approach based on tree-segmentation was developed [35].

- *Simulating fruit tree phenology* (A.S. Briand, Frédéric Boudon, Frédéric Normand [CIRAD, HortSys, Réunion Island], Anaëlle Dambreville, Jean-Baptiste Durand, Pierre Fernique, Yann Guédon, Christophe Pradal, Pierre-Eric Lauri [AFEF team, AGAP])

  Mango is a tropical tree characterized by strong asynchronisms within and between trees. To study more precisely the interplay between the plant structural components, we built an integrative model to simulate the plant development based on the L-system formalism and GLM to model the dependencies between events. With such model, we showed the importance of architectural and temporal factors in the development of the units of the trees. The model also simulates the phenology of shoots and inflorescences. For this, the sizes of the different organs is modelled by statistical laws estimated from measurements that depends on their locations in the architecture. The growth speed of organs is modulated by the temperature. This structural and phenological model has been presented at the ISHS symposium on Montpelier [37].

  This year, the model has been extended, during the intership of S. Persello to take into account fruiting probabilities and coupled with an ecophysiological model of fruit growth [68], [69]. The global aim is to have a crop simulation model to predict fruit yield and quality on mango tree. An overview of this global model based on the coupling of different structural or ecophysiological submodels has been also presented in different ISHS symposia [40], [50]

- *Integrative developmental growth stages of shoots* (Anaëlle Dambreville, Yann Guédon, Pierre-Eric Lauri [AFEF team, AGAP], Frédéric Normand [CIRAD, HortSys, Réunion Island])

  Plant growth, i.e. the increase of organ dimensions over time, and development, i.e. the change in plant structure, are often studied as two separate processes. However, there is structural and functional evidence that these two processes are strongly related. Our aim was to investigate the co-ordination between growth and development using mango trees, which have well-defined developmental stages. Developmental stages, determined in an expert way, and organ sizes, determined from objective measurements, were collected during the vegetative growth and flowering phases of two cultivars of mango. For a given cultivar and growth unit type (either vegetative or flowering), a multi-stage model based on absolute growth rate (AGR) sequences deduced from the measurements was first built, and then growth stages deduced from the model were compared with developmental stages. Strong matches were obtained between growth stages and developmental stages, leading to a consistent definition of integrative developmental growth stages [14]. The growth stages highlighted growth asynchronisms between two topologically connected organs, namely the vegetative axis and its leaves. Integrative developmental growth stages emphasize that developmental stages are closely related to organ growth rates and can be interpreted in terms of the possible physiological processes (hydraulics, biomechanics and carbohydrate partitioning) underlying these stages. We also explore growth stages deduced from relative growth rate (RGR) sequences applying the same methodology. AGR and RGR have different meanings regarding plant metabolism since AGR represents net sink strength whereas RGR represents net sink activity. For vegetative growth units, the match rates between RGR-based stages and developmental stages were rather similar to the match rates between AGR-based stages and developmental stages, because of the rich information provided by the four organs modeled (the axis and three selected leaves). The match rates were far lower for the inflorescences where only the main axis was modeled. This is related to the fact that, compared to AGRs, RGRs amplify the variations at the beginning of growth of an organ while damping the variations at the end of growth.

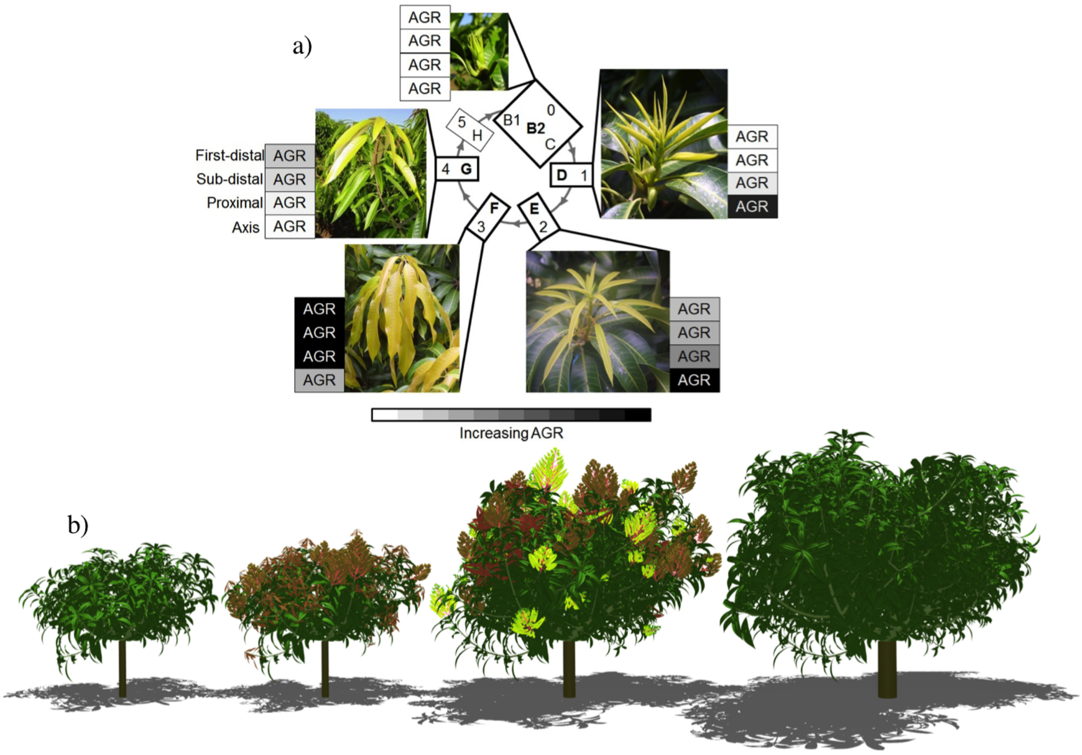


Figure 1. a) Correspondences between developmental stages of mango growth units determined from morphological observations in an expert way (inside the central circle) and growth stages obtained using segmentation models (outside the circle) [14]. Main developmental stages in bold are illustrated by the photographs. Variations in absolute growth rate for axis and leaves are illustrated using a white to black scale. b) Simulation of the development of a mango tree over two cycles [37]. The first and last image corresponds to the end of the vegetative period of the 3rd and 5th growing cycle (June), respectively while the second and third images correspond to the flowering phase (August) of the 3rd and 4th cycles, respectively. The different colours of the inflorescences of the 3rd image show different developmental stages and the flowering asynchronism over the tree.

- *Characterizing the successive flowering phases of strawberry in relation to genetic determinants* (Yann Guédon, Marc Labadie, Béatrice Denoyes [INRA, UMR BFP, Villenave d'Ornon], Justine Perrotte)

Our aim was to characterize the successive flowering phases of perpetual flowering strawberry genotypes, which is of particular importance for better predicting fruit production. We applied multiple change-point models for the synchronous segmentation of the individuals of a given genotype in successive flowering phases. We identified two groups of genotypes that differ by the intensity of the flowering at the end of the flowering period. Using a genetic approach, we identified a locus controlling the flowering intensity at the end of the flowering period that likely explain these two groups of genotypes. A multivariate generalization of the synchronous segmentation approach is developed in the context of Marc Labadie's PhD, the idea being to characterize not only the flowering pattern as in our first study but more generally the developmental pattern combining vegetative development, branching and flowering.

- *Self-nested structure of plants.* (Christophe Godin, Romain Azaïs, Farah Ben Naoum, Jean-Baptiste Durand, Alain Jean-Marie)

  In a previous work [7], we designed a method to compress tree structures and to quantify their degree of self-nestedness. This method is based on the detection of isomorphic subtrees in a given tree and on the construction of a DAG (Directed Acyclic Graph, equivalent to the original tree, where a given subtree class is represented only once (compression is based on the suppression of structural redundancies in the original tree). In the compressed graph, every node representing a particular subtree in the original tree has exactly the same height as its corresponding node in the original tree.

  The method proposed in [7] thus compresses a tree in width, but not in height. In a new work, we designed an extension of this compression method in which a tree is compressed in both width and height. The method is based on the detection of so-called *quasi-isomorphic paths* in a tree and on the compression of these paths in height. A paper describing the corresponding algorithms has been recently accepted in the Journal of Theoretical Biology (To appear).

6.1.3. Analyzing the influence of the environment on the plant ontogenic programme

Participants: Jean-Baptiste Durand, Christian Fournier, Christophe Godin, Yann Guédon, Christophe Pradal, Jean Peyhardi, Pierre Fernique, Guillaume Garin.

This research theme is supported by three PhD programs.

The ontogenetic programme of a plant is actually sensitive to environmental changes. If, in particular cases, we can make the assumption that the environment is a fixed control variable (see section 6.1.2), in general the structure produced by meristem results from a tight interaction between the plant and its environment, throughout its lifetime. Based on observations, we thus aim to trace back to the different components of the growth (ontogenetic development and its modulation by the environment). This is made using two types of approaches. On the one hand, we develop a statistical approach in which stochastic models are augmented with additional time-varying explanatory variables that represent the environment variations. The design of estimation procedures for these models make it possible to separate the plant ontogenetic programme from its modulation by the environment. On the other hand, we build reactive models that make it possible to simulate in a mechanistic way the interaction between the plant development and its environment.

- *Influence of environmental conditions and horticultural practices on the branching and axillary flowering structures of fruit tree shoots.* (Yann Guédon, Evelyne Costes [AFEF Team, AGAP], Ted DeJong [UC Davis], Claudia Negron [UC Davis]).

  In the context of a collaboration with Claudia Negron and Ted DeJong, we studied the influence of water availability and pruning practices [21] on the branching and axillary flowering structures of different categories of almond shoots. Stochastic models (hidden semi-Markov chains) were built for the branching and axillary flowering structures of different categories of almond shoots corresponding to different genetic backgrounds, levels of irrigation and pruning practices.

- *Analyzing growth components in trees.* (Yann Guédon, Yves Caraglio [AMAP], Olivier Taugourdeau [AMAP])

We identified robust indicators that summarize the respective importance of ontogeny and environmental constraints (mainly related to light environment) in forest tree development [26]. In this context, tree growth data correspond to the retrospective measurement of annual shoot characteristics (e.g. length, number of branches) along the main stem. We applied segmentation models to identify tree growth phases. These segmentation models, which are hidden semi-Markov chains, were compared with simple hidden Markov chains that correspond to the environment-driven development assumption. This statistical modelling approach was applied to both evergreen (Corsican pine and silver fir) and deciduous (sessile oak and Persian walnut) tree species growing in contrasted conditions ranging from managed forest stands to unmanaged understoreys. Growth phase duration distributions estimated within these segmentation models characterize the respective importance of ontogeny and environmental constraints in tree development at the population scale and have very contrasted characteristics in terms of shape and relative dispersion between ontogeny-driven and environment-driven tree development. These characteristics may change over tree life, reflecting changes in tree competition. Growth phase duration distributions summarize the joint trajectory of tree ontogeny and environment without requiring tree growth follow-up data for their estimation.

- *Analyzing fruit tree phenology in various climatic conditions* Yann Guédon, Jean-Michel Legave [AFEF team, AGAP], Gustavo Malagui [Universidade Tecnológica Federal do Paraná]

  The responses of flowering phenology to temperature increases in temperate fruit trees have rarely been investigated in contrasting climatic regions. This is an appropriate framework for highlighting varying responses to diverse warming contexts, which would potentially combine chill accumulation declines and heat accumulation increases. To examine this issue, a data set was constituted in apple tree from flowering dates collected for two phenological stages of three cultivars in seven climate-contrasting temperate regions of Western Europe and in three mild regions, one in Northern Morocco and two in Southern Brazil. Multiple change-point models were applied to flowering date series, as well as to corresponding series of mean temperature during two successive periods, respectively determining for the fulfillment of chill and heat requirements. A new overview in space and time of flowering date changes was provided in apple tree highlighting not only flowering date advances as in previous studies but also stationary flowering date series [18]. At global scale, differentiated flowering time patterns result from varying interactions between contrasting thermal determinisms of flowering dates and contrasting warming contexts. This may explain flowering date advances in most of European regions and in Morocco vs. stationary flowering date series in the Brazilian regions. A notable exception in Europe was found in the French Mediterranean region where the flowering date series was stationary. While the flowering duration series were stationary whatever the region, the flowering durations were far longer in mild regions compared to temperate regions. Our findings suggest a new warming vulnerability in temperate Mediterranean regions, which could shift towards responding more to chill decline and consequently experience late and extended flowering under future warming scenarios.

- *Investigating how architectural development interfere with epidemics and epidemic control* (Christian Fournier, Corinne Robert [Ecosys, INRA], Guillaume Garin [ITK, Montpellier], Bruno Andrieu [Ecosys, INRA], Christophe Pradal)

  Sustainable agriculture requires the identification of new, environmentally responsible strategies of crop protection. Modelling of pathosystems can allow a better understanding of the major interactions inside these dynamic systems and lead to innovative protection strategies. In particular, functional–structural plant models (FSPMs) have been identified as a means to optimize the use of architecture-related traits. A current limitation lies in the inherent complexity of this type of modelling, and thus the purpose of this work is to provide a framework to both extend and simplify the modelling of pathosystems using FSPMs. Complex models are disassembled into separate *knowledge sources* originating from different specialist areas of expertise and these can be shared and reassembled into multidisciplinary models. This year, we worked on four application studies that used the framework. In the frame of the PhD of Guillaume Garin, we perform a validation of the wheat septoria model, an analysis of the influence of the wheat architecture on the competition between septoria and brown rust, and a sensitivity analysis of the response of the severity of septoria to architectural traits. In the frame of the Echap project, we use the wheat-septoria model to indentify optimal date of pesticide application. All these studies allows to populate the framework with consistent example of application, and lead to the development of operational modules that allows the fitting and validation of pathosystem models with experimental data.

6.2. Meristem functioning and development

In axis 2 work focuses on the creation of a *virtual meristem*, at cell resolution, able to integrate the recent results in developmental biology and to simulate the feedback loops between physiology and growth. The approach is subdivided into several sub-areas of research.

6.2.1. Data acquisition and design of meristem models

- *Improvement of the MARS-ALT pipeline robustness* Meristem, laser microscopy, image reconstruction, cell segmentation, automatic lineaging
  Participants: Léo Guignard, Christophe Godin, Christophe Pradal, Grégoire Malandain [Morpheme, Inria], Gaël Michelin [Morpheme, IPL Morphogenetics, Inria], Guillaume Baty, Sophie Ribes [IBC, UM], Jan Traas [RDP, ENS], Patrick Lemaire [CRBM, CNRS], Yassin Refahi [RDP, ENS-Lyon / Sainsbury Lab, Cambridge, UK].

  This research theme is supported by a PhD FRM grant, Jan Traas's ERC, Inria ADT programme and the Morphogenetics Inria Project Lab.

  The MARS-ALT (Multi-Angles Registration and Segmentation - Automatic Lineage Tracking) software pipeline [6] automatically performs a segmentation at cell resolution from 3D or 2D voxel images where the membranes/walls are marked (by a die for example) and makes it possible to follow the lineage of these cells through time.

  This year, the ALT tracking pipeline has been reformulated by using a generic cell modeling approach (enabling for example more than one cell division), and both stability and robustness were improved. The modeling approach is generic and can be used on other kind of data (nuclei, human cells, ...). These trials will be conducted during the year. Moreover, the architecture of the image processing components has been modified (plugin approach) and integrated with the TissueLab platform. Some vizualisation tools have been improved, and the platform includes a module allowing an interaction with data (Alizon Konig, master internship). This point enables an efficient creation of gold standard to validate segmentation results.

  This year, we also finalize the development of a a new segmentation and tracking pipeline, ASTEC (Adaptive Segmentation and Tracking of Embryonic Cells). ASTEC is a one-pass algorithm (in contrast to MARS-ALT, that perform first the segmentation and then the tracking in two-passes) that is best suited for movies with numerous close time-points acquired at high spatio-temporal resolution. This pipeline takes advantage of information redundancy across the movies and biological knowledge on the segmented organism to constrain and improve the segmentation and the tracking. We used this one-pass algorithm to segment and track all cell shapes of a developing embryo of the marine invertebrate *Phallusia mammillata*. As a result we obtained the full track of the shapes of all the cells from the 64 cell stage up to the early tailbud stage (1030 cells undergoing 640 division events followed across 180 time-points through 6 hours of development imaged every 2 minutes, Figure 2).

  Based on this quantitative digital representation, we systematically identified cell fate specification events up to the late gastrula stage. Computational simulations revealed that remarkably simple rules integrating measured cell-cell contact areas with spatio-temporal expression data for extracellular signalling molecules are sufficient to explain most early cell inductions. This work suggests that in embryos developing with stereotyped cell shapes and positions (like *Phallusia mammillata* embryos), the genomic constraints for precise gene expression levels are relaxed, thereby allowing rapid genome evolution.

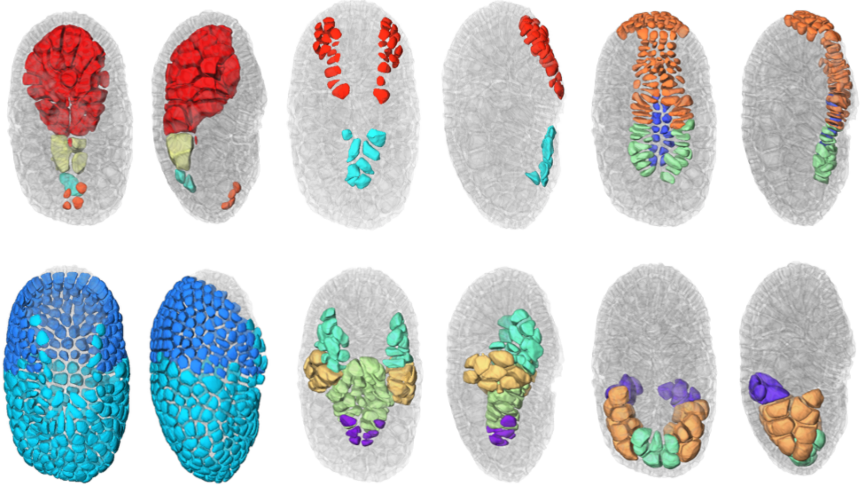


Figure 2. 3D projection of the segmented embryo at the early tailbud stage. The cells are colored by tissue type. The cells are slightly eroded to allow their distinction. The other cells of the embryo are in transparent grey. The doral and lateral sides are shown.

- *Creating mesh representation of cellular structures*
  Participants: Guillaume Cerutti, Sophie Ribes, Christophe Godin, Géraldine Brunoud [RDP, ENS], Carlos Galvan-Ampudia [RDP, ENS], Teva Vernoux [RDP, ENS], Yassin Refahi [RDP, ENS, Sainsbury Lab].

  This research theme is supported the HFSP project Biosensors.

  To produce a more efficient data structure accounting for the geometry of cellular tissues, we studied the problem of reconstructing a mesh representation of cells in a complex, multi-layered tissue structure, based either on membrane/wall images segmented using MARS or on nuclei images of shoot apical meristems. The construction of such mesh structures for plant tissues is currently a missing step in the existing image analysis pipelines.

  We developed tools to reconstruct a 3D cell complex representing the tissue, based on the dual simplicial complex of cell adjacencies. This set of tetrahedra is optimized from a reasonable initial guess to match the adjacencies in the tissue, which proved to produce a very faithful reconstruction[39]. We also developed a set of methods to triangulate such reconstructions, and enhance the quality of triangular mesh representations of plant tissue, simultaneously along several criteria [28].

  These tools can produce light discrete representations of the cell interfaces that enables fast visualization, information projection, and quantitative analysis of the tissue, and have given way to to some of the first biomechanical simulations on real-world data.

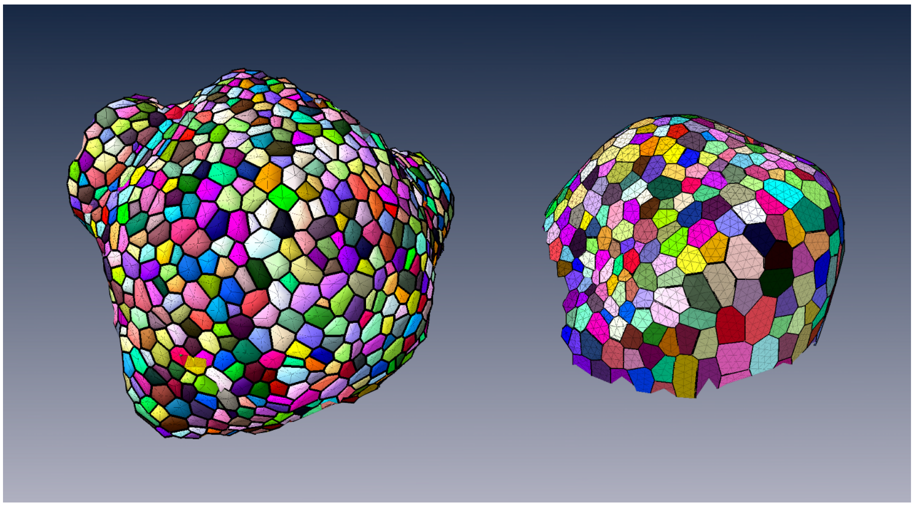


Figure 3. Triangular mesh representations of shoot apical meristem and flower meristem tissues obtained from MARS segmentations

- *Design of 3D digital atlases of tissue development*
  Participants: Sophie Ribes, Yassin Refahi [RDP, ENS, Sainsbury Lab], Guillaume Cerutti, Christophe Godin, Christophe Pradal, Christophe Pradal, Frédéric Boudon, Gregoire Malandain [RDP, ENS], Gaël Michelin [RDP, ENS], Guillaume Baty, Jan Traas [RDP, ENS], Teva Vernoux [RDP, ENS], Patrick Lemaire [CRBM, CNRS], Françoise Monéger [RDP, ENS].

  This research theme is supported the Inria Project Lab Morphogenetics, the ADT Mars-Alt and the HFSP project Biosensors.

  To organize the various genetic, physiological, physical, temporal and positional informations, we build a spatialized and dynamic database [67]. This database makes it possible to store all the collected information on a virtual 3D structure representing a typical organ. Each piece of information has to be located spatially and temporally in the database. Tools to visually retrieve and manipulate the information, quantitatively through space and time are being developed. For this, the 3D structure of a typical organ has been created at the different stages of development of the flower bud. This virtual structure contains spatial and temporal information on mean cell numbers, cell size, cell lineages, possible cell polarization (transporters, microtubules), and gene expression patterns. Such 3D digital atlas is mainly descriptive. However, like for classical databases, specific tools make it possible to explore the digital atlas according to main index keys, in particular spatial and temporal keys. Both a dedicated language and a 3D user interface are being designed to investigate and query the 3D virtual atlas. Current developments of this tool consist in using directly the segmented images produced from laser microscopy to build the atlas. To better represent the development of a biological population, a method to compute an "average" structure is investigated.

6.2.2. Shape analysis of meristems

Participants: Jonathan Legrand, Pierre Fernique, Frédéric Boudon, Yann Guédon, Christophe Godin, Pradeep Das [RDP, ENS], Arezki Boudaoud [RDP, ENS].

At cellular resolution, we studied the organization of cells in the meristems. The MARS-ALT pipeline provides rich spatio-temporal data sets for analyzing the development of meristems. A first step consisted of designing a dedicated graph structure for efficiently representing the spatial (adjacency between cells) and temporal (cell division) relationships between cells. Various variables can be attached either to the vertices (e.g. cell volume, inertia axes) or the edges (e.g. wall surface, distance between cell centroids). This graph may be augmented by new variables resulting from various spatial or temporal filtering (e.g. cell volumetric growth). Looking at homogeneous regions in the variable value space, cellular patterns can be identified.

Considering the highly-structured nature of our data (time and space structuring) and the potential diversity and heterogeneity of possible cell descriptors, we developed two complementary approaches:

- A first one that favours the spatial structuring: In this approach, the cell neighbourhood and the cell descriptors are jointly taken into account in a clustering approach whose objective is to identify a small number of clusters corresponding to well-defined cell identities. Once the cells have been labelled using the clustering algorithm, cell generation distributions are estimated on the basis of the labelled lineage trees.
- A second one that favours the temporal structuring: In this approach, the data of interest are lineage forest and the only spatial structuring taken into account corresponds to siblings with respect to a given parent cell. In a first step, cell identities are inferred on the basis of the cell descriptors taking into account lineage relationships using hidden Markov tree models and the spatial regions that emerge from the cell identity labelling are then characterized. This second approach is supported by the fact that cell topology is only affected by division which makes highly relevant the local spatial information taken into account in this approach.

6.2.3. Mechanical models of plant tissues

Participants: Jean-Philippe Bernard, Olivier Ali, Christophe Godin, Benjamin Gilles, Frédéric Boudon, Ibrahim Cheddadi, Jan Traas [ENS-Lyon], Olivier Hamant [ENS-Lyon], Arezki Boudaoud [ENS-Lyon].

This research theme is supported by the Inria Project Lab Morphogenetics and the Jan Traas's ERC.

The rigid cell walls that surround plant cells are the main load-bearing structures in plant tissues. These walls are submitted to stresses due to cell turgor pressure. Above some threshold, these stresses cause deformation in the cell walls and triggers wall **irreversible expansion** (*synthesis*). Shape changes of plant tissues are therefore tightly related to the turgidity of cells and to the mechanical state and the molecular composition of the underlying cell walls. We developped a conceptual and numerical framework to model the mechanical structure of cell walls and their deformation by turgor pressure in 3-dimensions. This framework was used to study the interplay between post-transcriptional regulation, biochemistry, and mechanics within growing plant tissues. This work has been published this year in Plos Computational Biology [13].

In this first step, all mechanical and structural quantities are defined at the tissular scale. This is made possible by abstracting the connection between the actual molecular composition of the walls and the various signalling cascade at play during growth. To extend this approach, we also started to develop a mechanobiological approach relating the irreversible expansion of the walls to molecular mechanisms happening within them, based on the thermodynamical equilibrium of the pectin-based matrix within the wall. We propose that at the molecular scale expansion of this matrix is based on the adsorption of newly synthetized pectin molecules. This adsorption mechanism is regulated by the mechanical stresses applied on the wall. We show that this mechanism belongs to a class of biochemical / biomechanical processes commonly appearing in the dynamics of supra-molecular load-bearing structures: the force-driven polymerization processes. A preliminary version of these ideas (the 1D case) is currently under review in Trends In Plants Sciences.

We also considered to extend the original modeling approach to situations where entire organ dynamics should be modeled over large time lapse (several days) (PhD work of Jean-Philippe Bernard). In our first approach, the mechanical model relies on a finite element method (FEM) to describe the deformation of the tissue. In FEM, the tissue is represented by a mesh. The positions of the vertices at each time step are estimated from a linear system. If the tissue is big or if the mesh is fine, the linear system can be large and thus leads to computational overheads. An alternative way to classical FEM is to use a meshless method where the deformation of the tissue can be characterized by a linear combination of deformations of a finite and small set of frames. Because shape functions are no longer defined on each element but on the whole tissue, they have to be updated at each growth step by estimating a new rest configuration. With meshless method, the discretization of the system can be dynamically updated parsimoniously according to the precision required to model the emergence of shapes. With an uniform distribution of the frames within the volume, our method still leads to computational overheads. However, since the meristem initiates a branching structure at a macroscopic scale, we combined our mechanical model at tissular resolution with classical method used to generate branching structures at macroscopic scales. For this, we use the information of the plant branching structure to distribute the frames along the plant's axes. This allows us to use curvilinear shape functions while describing the branching structure growth using L-systems. This multi-scale framework allows us to define developmental rules which can initiate new organs at the surface of the meristematic dome by softening locally the meristem dome and thus creatin new growing initia. First very encouraging results were obtained this year that demonstrate the feasibility of the approach.

6.2.4. Gene regulatory networks: Design of a genetic model of inflorescence development.

Participants: Eugenio Azpeitia, Christophe Godin, François Parcy, Etienne Farcot.

This research theme is supported by the Inria Project Lab Morphogenetics.

Modeling gene activities within cells is of primary importance since cell identities correspond to stable combination of gene expression.

We studied the regulatory network that controls the flowering transition during morphogenesis. To overcome the network complexity and integrate this regulation during ontogenesis, we have developed a first model of the control of floral initiation by genes, and in particular the situation of cauliflower mutants, in which the meristem repeatedly fails in making a complete transition to the flower. Three different network models were done and validate. A first Boolean version, a second fuzzy logic and an ODEs models were studied. The models are able to correctly recover the gene steady states observed in the meristems during the flower transitions, the gene transitions and the mutant effects. Importantly, the model is able to explain the cauliflower mutants. This work couples models at different scales, since the gene regulatory network is used as a decision module in an L-system model of the inflorescence architecture. This mixed model has led us to make different hypotheses about gene interactions and hormonal regulation. First predictions about gene actors controlling the passage to flower could be verified. Some links between gene regulation and plant growth have been identified. These links can be experimentally tested which could lead to a first integrated picture of flower development.

Finally, given that the cauliflower have different morphologies (i.e. regular and romanesco cauliflower morphologies) we explored the effect of changes in the L-system parameter values over the cauliflower morphology. Interestingly, we discovered by exploring the model that variations in the regulation of some phyllotactic parameters can produce the different cauliflower morphologies and explain other reported differences among them. Predictions were made using the model and experimental validations of this hypothesis are curren being tested. All our results could provide a comprehensive understanding of how does genes and plant architecture are linked in a dynamical way.

6.2.5. Modelling the influence of dimerisation sequence dissimilarities on the auxin signalling network

Participants: Jonathan Legrand, Yann Guédon, Teva Vernoux [ENS-Lyon].

Auxin is a major phytohormone involved in many developmental processes by controlling gene expression through a network of transcriptional regulators. In *Arabidopsis thaliana*, the auxin signalling network is made of 52 potentially interacting transcriptional regulators, activating or repressing gene expression. All the possible interactions were tested in two-way yeast-2-hybrid experiments. Our objective was to characterise this auxin signalling network and to quantify the influence of the dimerisation sequence dissimilarities on the interaction between transcriptional regulators. We applied model-based graph clustering methods relying on connectivity profiles between transcriptional regulators. Incorporating dimerisation sequence dissimilarities as explanatory variables, we modelled their influence on the auxin network topology using mixture of linear models for random graphs. Our results provide evidence that the network can be simplified into four groups, three of them being closely related to biological groups. We found that these groups behave differently, depending on their dimerisation sequence dissimilarities, and that the two dimerisation sub-domains might play different roles. We proposed the first pipeline of statistical methods combining yeast-2-hybrid data and protein sequence dissimilarities for analyzing protein- protein interactions. We unveil using this pipeline of analysis the transcriptional regulator interaction modes.

6.2.6. Model integration

Participants: Frédéric Boudon, Christophe Godin, Guillaume Baty, Guillaume Cerutti, Jean-Louis Dinh, Jan Traas.

This research theme is supported by the Morphogenetics Inria Project Lab.

Our approach consists of building a programmable tissue which is able to accept different modeling components. This includes a central data structure representing the tissue in either 2-D or 3-D, which is able to grow in time, models of gene activity and regulation, models of signal exchange (physical and chemical) between cells and models of cell cycle (which includes cell division). An introduction to the modeling of some main components of such integrated system was published as a book chapter in the series of Ecole de Physique des Houches [43]. For each modeling component, one or several approaches are investigated in depth, possibly at different temporal and spatial scales, using the data available from the partners (imaging, gene networks, and expression patterns). Approaches are compared and assessed on the same data. The objective of each sub-model component will be to provide plugin components, corresponding to simplified versions of their models if necessary, that can be injected in the programmable tissue platform. This work is developed in collaboration with the RDP group at ENS-Lyon [70] and the CPIB group in Nottingham, UK [53].

One key aspect of our approach is the development of a computer platform dedicated to programming virtual tissue development, TissueLab. This platform, based on *OpenAlea*, will be used to carry out integration of the different models developed in this research axis. In the past year, progress has been made in defining a generic tissue data structure that could be used in this platform. Currently, robust geometric operations such as division are implemented and tested. Moreover, a redesign of the structure based on more elaborated formalisms such as combinatorial maps is being investigated. A 2D version is being developed in the context of Jean-Louis's Dinh PhD thesis, and will be described in a forthcoming book chapter.

6.3. Multi-scale models and analysis: from cells to plant architecture (and back)

6.3.1. Modeling water transport in roots

Participants: Mikaël Lucas [IRD], Christophe Pradal, Christophe Godin, Yann Boursiac Bpmp ,, Christophe Maurel [BPMP].

This research theme is supported by the ANR project HydroRoot.

A model of Arabidopsis thaliana root hydraulics at the cellular level was developed in the OpenAlea modeling platform. The model relies on the integration throughout root architecture of elementary hydraulic components. Each component integrates local radial and axial water flows. Axial hydraulic conductivity is calculated according to Poiseuille's law, based on local size of xylem vessels. Radial hydraulic conductivity is determined in part by aquaporin activity and was set constant throughout root architecture in the first model versions. In its current state, the model is parameterized using architectural, tissular and physiological data that were experimentally determined in the Aquaporin group at UMR BPMP. The architectural reconstruction of the root system is based on a tridimensional multi-scale tree graph (MTG). The current model is capable of predicting the water flow that is transported by a root system in the standard experimental conditions used in the Aquaporin group. This model was used to perform sensitivity analyses and determine the respective contributions to root hydraulic dynamics of various biological parameters (axial and radial hydraulic conductivites, root architecture). One major finding is that the root hydraulic conductivity (Lpr) computed from the model is highly dependent on root architecture. This is due to the limiting role of axial (xylem) conductance, one feature that had been neglected in previous representations of root water transport. The radial hydraulic conductivity may primarily be limiting in conditions of Lpr inhibition, since its increase from values in control roots has marginal effects on Lpr. A new set of experimental data including root diameter repartitions in wild-type plants, and xylem vessel diameters in mutants with altered xylem morphology (irx3, esk1) will be used to challenge the model. Root cell hydraulic conductivities will also be measured in these and aquaporin mutant phenotypes. Our aim is to check whether, based on anatomical and morphological data, the model can properly predict the radial hydraulic conductivity of these genotypes.

As the simulations may be time consuming and results sometimes difficult to interpret on complex branching systems, we started to investigate new methods to compute efficiently hydraulic conductivities and corresponding flows on complex root systems using architecture compression technics developed in the 1st axis of the project. First results show that very efficient computations of complex hydraulic architectures can be derived from the use of these compression techniques on idealized root architectures. These encouraging results provide a new abstraction that will be used in combination with the detailed modeling approach described above to break down the complexity of the analysis these huge branching systems.

6.3.2. Mechanical modeling of fruit growth

Participants: Ibrahim Cheddadi [Inra, Avignon], Mik Cieslak [U. Calgary], Frédéric Boudon, Valentina Baldazzi [Inra, Avignon], Nadia Bertin [Inra, Avignon], Michel Genard [Inra, Avignon], Christophe Godin.

This research theme is supported by the Agropolis project MecaFruit3D.

Fruits and plants in general are large scale hydraulic systems in which growth is closely linked to water fluxes: thanks to osmotic pressure difference, the cells are able to absorb water from their environment and therefore increase their volume; as the cells are bounded by rigid walls, this results in both hydrostatic pressure (the so-called turgor pressure) in the cell and tension in the cell walls; above a threshold, synthesis of new cell wall material occurs and relaxes the tension. This process allows cells to grow, and along with cell division, is responsible for plant growth. In fruits, phloem and xylem vascular networks provide the water fluxes necessary for growth, while the osmotic pressure is mainly regulated by sugar intake from the phloem. The goal of this project is to combine a description of water and sugar fluxes at the fruit scale (see section 4) with a modelling of growth at cell level, as described above.

As a first step in this direction, we have developed a bidimensional multicellular model that couples, on the one hand, water fluxes between cells (symplastic pathway) and between cells and intercellular space (apoplastic pathway), and on the other hand, mechanical properties of the cell walls and mechanical equilibrium of this complex system. Existing multicellular models for plant growth overlook this coupling. From a mathematical point of view, it corresponds to a coupling between (1) the ordinary differential equations that describe fluxes and cell walls properties and (2) the highly non linear system of equations that describes the mechanical equilibrium of the cell walls.

We have developed a numerical method for this coupled system, that allows to simulate in a reasonable amount of time a hundred of connected cells. The non linear system of equations (2) is the bottleneck to reach a higher number of cells; in order to overcome this, we plan to use the framework developed for the mechanical modelling of meristems (see section 6.2.3) and adapt it to this system. This will also allow to address tridimensional tissues.

Numerical simulations exhibit a highly non linear behaviour with respect to the governing parameters. We have identified two clearly distinct growth regimes: one regime that allows large growth heterogeneities by amplifying the effect of differences between cells, and conversely another regime that smoothes differences out and yields a homogeneous growth. On the biological level, the first regime is well adapted to morphogenesis, whereas the second one is well adapted to homothetic growth after the differentiated tissues have been created. A publication of these completely new results is in preparation.

We plan to compare this model to experimental results of the tomato fruit at the tissue level. In the longer term, a continuous version of this multicellular model could be an interesting way to build a model at the fruit scale.

6.3.3. Analyzing root growth and branching

Participants: Beatriz Moreno Ortega, Sixtine Passot, Yann Guédon, Laurent Laplaze [IRD, DIADE], Mikaël Lucas [IRD, DIADE], Bertrand Muller [INRA, LEPSE].

This research theme is supported by two PhD programmes.

New 2D and 3D root phenotyping plateforms are emerging with associated image analysis toolbox (e.g. Smart-Root, RhizoScan) and the high-level analysis these complex phenotyping data requires new computational investigation methods.

Here, we aim at developing a pipeline of methods for analyzing root systems at three scales:

1. tissular scale to identify and characterize the division, elongation and mature zones along a root using piecewise heteroscedastic linear models. To this end, we introduced a new slope heuristic for the selection of the number of zones in cell length series [29] [36].
2. individual root scale to analyze the dynamics of lateral root elongation. We investigated the use of semi-Markov switching linear models for classifying roots on the basis of the identification of phases within growth rate profiles,
3. root system scale to analyze the branching structure.

This pipeline of analysis methods will be applied to different species (maize, millet and *arabidopsis*) and for different biological objectives (study of genetic diversity for millet and of metabolic and hormonal controls of morphogenesis for maize).

6.3.4. Analyzing shoot and leaf elongation

Participants: Maryline Lièvre, Yann Guédon, Leo Guignard, Christine Granier [INRA, LEPSE].

This research theme is supported by one PhD programme and the labex Agro project "Integrated model of plant organ growth".

This study is based on the observation that there is a lack of methods enabling the integrated analysis of the processes controling the vegetative development in *Arabidopsis thaliana*.

The changes in leaf size and shape during ontogeny associated with the heteroblastic development is a composite trait for which extensive spatio-

temporal data can be acquired using phenotyping platforms such as PHENOPSIS. However, only part of the information contained in such data is exploited and developmental phases are usually defined using a selected organ trait. We introduced new methods for identifying developmental phases in *Arabidopsis* rosette using various traits and minimum a priori assumptions. A first pipeline of analysis was developed, combining image analysis and statistical models to integrate morphological, shape, dimensional and expansion dynamics traits for the successive leaves of the *Arabidopsis* rosette. Dedicated segmentation models called semi-Markov switching models were built for selected genotypes in order to identify rosette developmental phases. Four successive developmental phases referred to as seedling, juvenile, transition and adult were identified for the different genotypes. We showed that the degree of covering of the leaf abaxial surface with trichomes is not sufficient to define these developmental phases. Using our pipeline of analysis, we were able to identify the supplementary seedling phase and to uncover the structuring role of various leaf traits. This enabled us to compare on a more objective basis the vegetative development of *Arabidopsis* mutants.

We developed a second pipeline of analysis methods combining a semi-automatic method for segmenting leaf epidermis images based on the ilastik software, and the analysis of the obtained cell areas using a gamma or inverse Gaussian mixture models whose component parameters are tied by a scaling rule. These mixture models allowed us to estimate the distribution of the number of endocycles. We highlighted in this way that the mean number of endocycles changes drastically with leaf rank. We extended the inference approach to take into account not only complete cell areas but also censored cell areas (corresponding to cells that intercept the edges of the images). We also investigated possible temporal interpretations of endoreduplication using stochastic processes.

6.3.5. A stochastic model of phyllotaxis

Participants: Yassin Refahi, Christophe Godin, Etienne Farcot, Teva Vernoux [RDP, ENS].

This research theme has been supported by IBC and the Inria Project Lab Morphogenetics.

The geometric arrangement of lateral organs along plant stems, named phyllotaxis, shows a variety of striking patterns with remarkable regularities and symmetries. This has interested biologists, physicists, mathematicians and computer scientists for decades. These studies have lead to a commonly accepted standard interpretation of phyllotaxis that postulates that organs inhibit the formation of new organs in their vicinity. At a molecular scale, these inhibitory fields have been shown to result from the spatio-temporal distribution of the plant hormone auxin. This model theoretically explains a large part of the diversity of phyllotactic patterns observed in plants.

Recently, our colleagues from ENS-Lyon observed intriguing perturbation in *arabidopsis* mutants. These perturbations were also present, to a lesser extent in the wild type. In a series of works [74], [64], [2], we could show that these perturbations patterns in both wild-type and mutant plants could be explained by permutations in the order of insertion along the stem of 2 or 3 consecutive organs. After closer inspection, we realized that the mutated gene encodes a protein diffusing from the organs and creating a field around the organs that regulates the plastochron. We could demonstrate that in the mutant, the absence of this field leads to co-initiations and subsequently to the observed permutations.

To proceed further and find a mechanistic interpretation of this phenomenon, we developed a stochastic extension of the standard model of phyllotaxis. We first analyzed the properties of the inhibitory fields created by the existing primordia on the initiation of new promordia, and concluded that the angular positions of organs are very robust to perturbations while plastochrons may be dramatically affected. This suggested that there exists a strong decoupling between space and time in the patterning process. To account for this observation, we modeled the perception of the initiation signal by cells using stochastic processes coupled with the intensity of inhibitory fields and showed that the observed permutation patterns emerge spontaneously from this purely local processes. This model recapitulates accurately the classical phyllotactic patterns and, in addition, produces realistic pattern disorders at higher organization levels as a result of stochasticity in signal perception. We show that these subtle disorders surprisingly reveal key information on the functioning of the developmental system and can therefore be regarded as *biological watermarks* of the system. In genetically or environmentally modified plants, these biological watermarks inform us on the molecular mechanisms that have been affected in the experiment. Our theoretical analysis allows us to predict the specific pattern variations that would arise from perturbations of the signaling pathways involved in lateral inhibition signaling at the shoot apex. A paper describing this model has been submitted recently for publication.

6.3.6. The role of auxin and sugar in rose bud outgrowth control

Participants: Jessica Bertheloot [INRA, Angers], Frédéric Boudon, Christophe Godin.

Auxin in the stem is known to be a key regulator of apical dominance. Over the last decades, many studies have been undertaken to understand its action mode, which is indirect because auxin in the main stem does not enter into the bud. Recently, apical dominance over basal buds in pea has been related to low sugar availability caused by high sugar demand of growing apical organs. Auxin and sugar are two signals regulating the entrance of bud into sustained growth in opposite ways. In the last year, it has also been demonstrated that sugar effect on bud outgrowth was preceded by a modification of the hormonal levels involved in bud outgrowth, which suggests that auxin and sugar pathways do interact in a non- trivial way. However, auxin and sugar effects have been studied separately until now. In this work, we investigate what is the combined effect of sugar and auxin on bud outgrowth, and how they integrate to regulate bud entrance into sustained growth. For this, a series of experiments has been carried out on a single-node cuttings of Rosa hybrida grown in vitro in which different combinations of sugar and auxin levels have been tested. A model of the regulatory networks controling stem-bud molecular interaction is currently being developed.

6.4. Generic methodological results

In the context of our research work on biological questions, we develop concepts and tools in mathematics, statistics and computer science. This paragraph is intended to put emphasis on the most important results obtained by the team during the current year in these disciplins, independently of their biological application.

6.4.1. Scientific workflows

Participants: Christophe Pradal, Sarah Cohen-Boulakia, Christian Fournier, Didier Parigot [Inria, Zenith], Patrick Valduriez [Inria, Zenith].

6.4.1.1. OpenAlea scientific workflows

Analyzing biological data may involve very complex and interlinked steps where several tools are combined together. Scientific workflow systems have reached a level of maturity that makes them able to support the design and execution of such in-silico experiments, and thus making them increasingly popular in the bioinformatics community (e.g. to annotate genomes, assemble NGS data, ...) . However, in some emerging application domains such as system biology, developmental biology or ecology, the need for data analysis is combined with the need to model complex multi-scale biological systems, possibly involving multiple simulation steps. This requires the scientific workflow to deal with retro-action to understand and predict the relationships between structure and function of these complex systems. In collaboration with the Zenith EPI, we have proposed a conceptualisation of OpenAlea workflows [34] by introducing the concept of higher-order dataflows as a means to uniformly combine classical data analysis with modeling and simulation. Ongoing work include deploying OpenAlea workflows on a Grid technology using the SciFloware middleware in close collaboration with Zenith within IBC and INRA Phenome projects.

6.4.1.2. Querying Scientific workflows repositories

Several workflow systems have developped scientific workflow repositories (e.g., repositories of Galaxy workflows at IBC, or repositories of OpenAlea workflows). Such repositories have grown to sizes that call for advanced methods to support workflow discovery, in particular for similarity search. Effective similarity search requires both high quality algorithms for the comparison of scientific workflows and efficient strategies for indexing, searching, and ranking of search results. Yet, the graph structure of scientific workflows poses severe challenges at each of these steps. We present a complete system for effective and efficient similarity search in scientific workflow repositories, based on the Layer Decompositon approach to scientific workflow comparison. Layer Decompositon specifically accounts for the directed dataflow underlying scientific workflows and, compared to other state-of-the-art methods, delivers best results for similarity search at comparably low runtimes. Stacking Layer Decomposition with even faster, structure-agnostic approaches allows us to use proven, off-the-shelf tools for workflow indexing to further reduce runtimes and scale similarity search to sizes of current repositories [25]. Very efficient and powerful ranking methods have been used in this work. We based our choice on the large scale study of algorithms for rank aggregation with ties we performed [56].

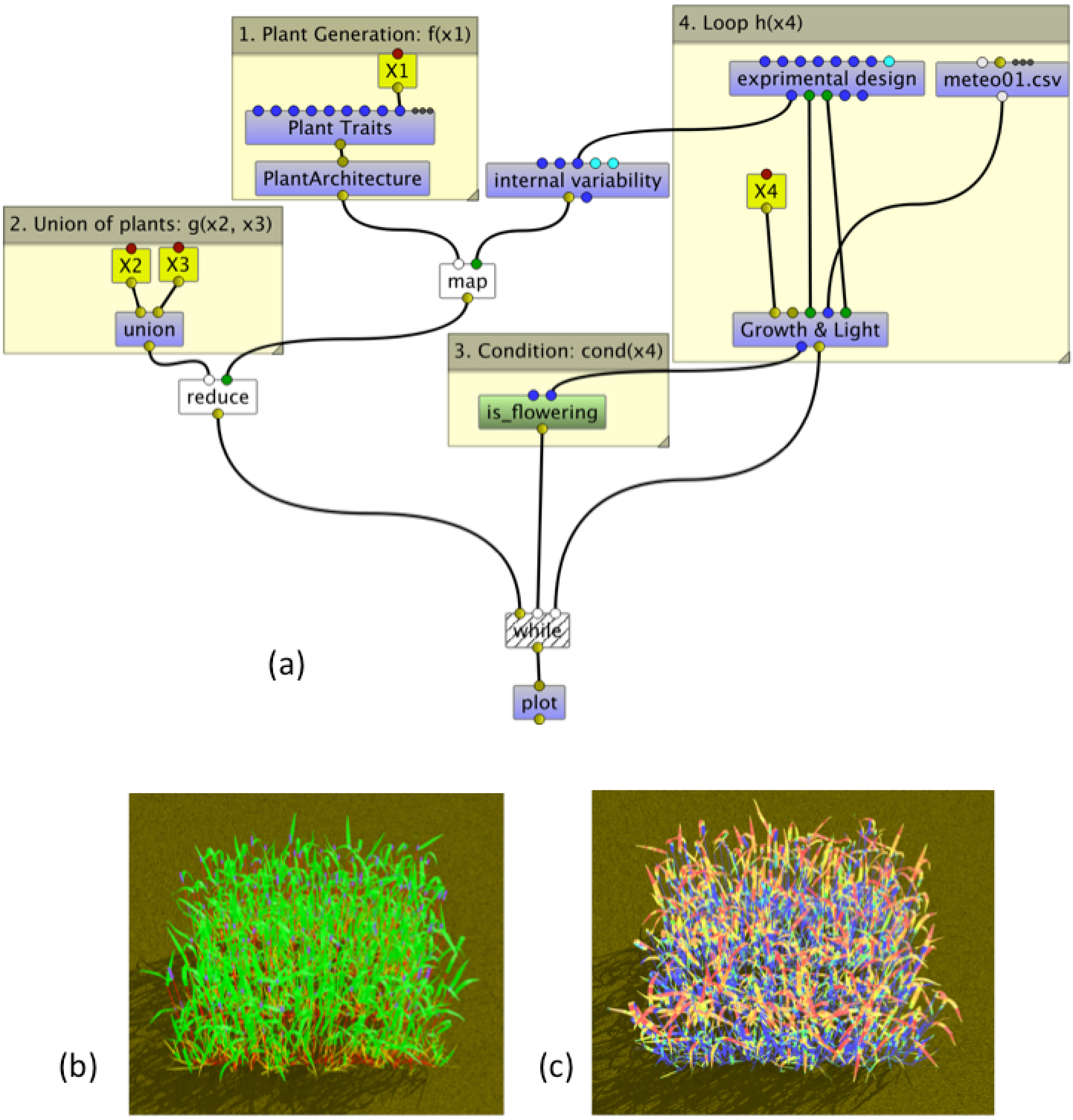


Figure 4. (a) OpenAlea workflow [34] for simulating Maize and Wheat crop performance based on phenotypic and environment data, and two image outputs (b and c). Colors represent the organ's type in (b) and the amount of intercepted light in (c).

6.4.2. Statistical modeling

Participants: Yann Guédon, Jean Peyhardi.

We develop statistical models and methods for identifying and characterizing developmental patterns in plant phenotyping data. Phenotyping data are very diverse ranging from the tissular to the whole plant scale but are often highly structured in space, time and scale. Problems of interest deal with the definition of new family of models specifically adapted to plant phenotyping data and the design of new methods of inference concerning both model structure, model parameters and latent structure. This is illustrated this year by [17] and [22].

6.4.3. Lossy compression of tree structures

Participants: Christophe Godin, Romain Azaïs, Jean-Baptiste Durand, Alain Jean-Marie.

the degree of self-nestedness of a tree as the edit-distance between the considered tree structure and its nearest embedded self-nested version. Indeed, finding the nearest self-nested tree of a structure without more assumptions is conjectured to be an NP-complete or NP-hard problem. We thus introduced a lossy compression method that consists in computing in polynomial time for trees with bounded outdegree the reduction of a self-nested tree that closely approximates the initial tree. This approximation relies on an indel edit distance that allows (recursive) insertion and deletion of leaf vertices only. We showed in a conference paper accepted at DCC'2016 [46] with a simulated dataset that the error rate of this lossy compression method is always better than the loss based on the nearest embedded self-nestedness tree [7] while the compression rates are equivalent. This procedure is also a keystone in our new topological clustering algorithm for trees. In addition, we obtained new theoretical results on the combinatorics of self- nested structures. The redaction of an article is currently in progress.

WIMMICS Project-Team

7. New Results

7.1. Users modeling and designing interaction

7.1.1. Exploratory search

Participants: Emilie Palagi, Alain Giboin.

Contrary to lookup search engines that help users to retrieve specific items (e.g., names, numbers, short statements, or specific documents), Exploratory Search Systems (ESSs) are search engines that help users to explore a topic of interest. Exploratory search (ES) tasks are open-ended, multi-faceted, and iterative like learning or topic investigation [59]. Currently, the evaluation methods of ESSs are not entirely adapted to the special features of ES tasks, and do not effectively assess that ESSs support users in performing those tasks. Our research goal is to elaborate methods that effectively lead to this assessment. Two research actions were undertaken this year to contribute to achieve this goal.

7.1.1.1. Design of an exploratory-search-oriented protocol for testing an image search algorithm based on user's eye movements

Participants: Emilie Palagi, Alain Giboin.

This work was undertaken in the context of the VISIIR ANR project [0], leaded by the MinD team (I3S, UNS), with Stéphanie Lopez and Frédéric Precioso. One of the objectives of VISIIR is to design an interactive image search system based on user's eye movements. Detected by an eye-tracker, these movements allow the system to infer the image that the user is going to select; VISIIR's aim is to replace user's mouse clicks as a selection mode by implicit *eye-clicks*. Since ES behaviors can be observed in image search tasks, we designed an eye-tracking user test protocol on Discovery Hub [0] in order to: 1) increase our understanding of the ES process (at the cognitive and perceptual-motor levels); 2) verify if identified characteristics of gaze trajectories allow to infer the images that the user is going to select in ES tasks (as opposed to lookup search tasks).

7.1.1.2. Design of a user-centered evaluation method of exploratory search systems based on a model of the exploratory search process

Participants: Emilie Palagi, Alain Giboin, Fabien Gandon.

(with Raphaël Troncy, Eurecom)

This work was undertaken in the context of the PhD of Emilie Palagi. In [41] we introduced our approach for designing a user-centered evaluation method for ESSs. Our method takes into account users's ES behavior and is based on a cognitive model of an ES task. We will specially work on Discovery Hub (Wimmics project – Inria) and 3cixty [0] (EURECOM project) ESSs.

7.1.2. Sentiment Analysis

Participant: Andrea Tettamanzi.

Together with Célia da Costa Pereira (I3S, UNS) and Mauro Dragoni of FBK, Trento, who visited our team for three months from April to June 2014, we have further refined our approach to concept-level sentiment analysis based on fuzzy logic [12].

7.1.3. Recommendation of Pedagogical Resources Adapted to User Profile and Context

Participants: Oscar Rodriguez Rocha, Catherine Faron-Zucker.

---

[0] http://www.agence-nationale-recherche.fr/?Projet=ANR-13-CORD-0009

[0] http://discoveryhub.co/

[0] https://www.3cixty.com/

In the framework of the Semantic Educloud project, we developed a Web ontology for the description and representation of serious games. Such ontology describes the functional and design elements of the game, the profile and virtual context of the players and furthermore the datasets from the Web of data that the game can query. The ontology has been evaluated through a prototype, which is basically a serious game quiz based on DBpedia. As future work, it is planned to implement state-of-the-art recommendation algorithms of Linked Data resources that take into account the context and players' profile. Furthermore an integration with the EDUCLOUD platform is considered: EDUCLOUD is an emerging initiative in Sophia Antipolis, for the implementation of a platform of digital educational content accessible through the cloud from a 3D portal of resources, and any interface devices (tablets, smartphones, PCs). [44]

7.2. Communities and social interactions analysis

7.2.1. Community Detection and Interest Labeling

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

7.2.1.1. Topic Modeling Based Overlapping Community Detection

Based on previous work, we conducted more experiments to evaluate the effectiveness and efficiency of the proposed tag tree based method. We used perplexity score to evaluate the performance of topic extraction. We got consistent performance when applying the model on a Flickr dataset. This work has been published in IEEE/WIC/ACM Web Intelligence 2015 [32] and Social Network Analysis and Mining Journal [13].

7.2.1.2. Temporal Analysis of User and Topic

By jointly modeling topic, expertise, time and activity, we were able to retrieve many meaningful latent information from the user generated contents. We proposed a method to track the dynamics of topics and users. It can also track the dynamics with a specific granularity of time level such as, yearly, monthly, daily and hourly. Besides, the model can overcome a comparison problem of LDA based model by modeling the reverse distribution.

7.2.1.3. Topic labeling

The output of topic model is normally a bag of words. Each topic consists of closely related words. An interesting question is to assign one or more topic label to this set in order to indicate the general meaning of a bag of words. By integrating the original dataset with linked open data sources, we are now planning to propose a generic method to automatically label the detected topics.

7.2.2. Semantic Modeling of Social, Spatiotemporal and Dedicated Networks

Participants: Amel Ben Othmane, Nhan Le Thanh, Michel Buffa, Andrea Tettamanzi, Serena Villata.

We have been working on modeling a multi-agent based recommender system. The aim of such system is to recommend a list of activities (plans) according to user preferences in order to achieve a goal. For this purpose, we propose a multi-context framework based on the well-known agent Belief-Desire-Intention (BDI) architecture [58]. First, we extend the BDI model with additional contexts in order to handle sociality. Second, we use a possibilistic approach based on the work of Da Costa Pereira & Tettamanzi [55], to reason about beliefs, desires, goals and intentions. Further, we use ontologies to represent and reason about plans and intentions. The proposed framework is detailed in a long paper that will be presented in the 8th International Conference on Agents and Artificial Intelligence in 2016 [18].

7.2.3. Collaborative Software Development Platforms

Participant: Isabelle Mirbel.

The collaborative nature of software development helped in the emergence of several online collaborative software development platforms (CSDPs). These platforms enable distributed teams of contributors to participate in the development of the various hosted projects. In such a context, the identification of relevant contributors is very important for handling efficiently the abundant requirements. However, this can be really challenging because of the fairly large number of involved contributors, especially in some distinguished projects. Moreover, the contributor profiles on a CSDP are often inadequately informative, which makes them an unqualified resource for learning about the contributors.

In this context, we proposed to identify contributors by their reputation on a CSDP. Our approach calculates reputation scores using a belief calculus, called subjective logic, according to contributors' performed roles. Knowing the reputation of anonymous contributors would enable project members to reduce the uncertainty in their future interactions with them. Moreover, we use concept lattices to classify contributors by their reputation scores, which enable us to have a comparable view on the considered contributors. Consequently, we can produce a roadmap to examine new requirements thus supporting their effective communication and prioritization.

7.2.4. Logical Foundations of Cognitive Agents

Participants: Andrea Tettamanzi, Serena Villata.

Together with Célia da Costa Pereira (I3S, UNS), we have continued an investigation about the issue of trust in multi-agent systems, and we proposed a computational model of trust based on the content of messages and on the characteristics of their sources [27].

7.2.5. Combining Argumentation Theory and Normative Reasoning with Natural Language Processing

Participants: Serena Villata, Elena Cabrio.

Together with Cristian Cardellino and Laura Alonso Alemany from the University of Cordoba (Argentina), we applied different Active Learning strategies to Information Extraction from licenses in English, with highly repetitive text, few annotated or unannotated examples available, and very fine precision needed. We showed that the most popular approach to active learning, i.e., uncertainty sampling for instance selection, does not provide a good performance in this setting. We showed that we can obtain a similar effect to that of density-based methods using uncertainty sampling, by just reversing the ranking criterion, and choosing the most certain instead of the most uncertain instances. The results of this research have been published at the CICLing [24] and JURIX [23] international conferences.

In another work, together with Alessio Palmero Aprosio (FBK Trento, Italy), we have worked on an extension of QAKiS, the system for open domain Question Answering over linked data, that allows to query DBpedia multilingual chapters. Such chapters can contain different information with respect to the english version, e.g. they provide more specificity on certain topics, or fill information gaps. In particular, we have introduced and evaluated the RADAR 2.0 framework for information reconciliation over language-specific DBpedia chapters. The framework is composed of three main modules: a module computing the confidence score of the sources depending either on the length of the related Wikipedia page or on the geographical characterization of the queried entity, a module retrieving the relations holding among the elements of the results set, and finally a module computing the reliability degree of such elements depending on the confidence assigned to the sources and the relations among them. This third module is based on bipolar argumentation theory to return the acceptability degrees. A demo of the RADAR framework is available online [0]. This contribution has been submitted to the Semantic Web Journal and is under review.

Moreover, we have proposed the BEGincs (BEG-Inconsistencies) framework, which translates a bipolar entailment graph into an argumentation graph. It then provides to the annotators sets of arguments that are supposed to be consistent. If it is not the case, the Textual Entailment system wrongly assigned some relations. Moving from single pairs to an overall graph allows for the detection of inconsistencies otherwise undiscovered. BEGincs does not identify the precise relation causing the inconsistency, but provides annotators with the consistent arguments sets, they are supported in narrowing the causes of inconsistency. The results of this research have been published at the CLIC conference [43].

7.2.6. Argumentation theory and its applications

Participants: Elena Cabrio, Serena Villata, Fabien Gandon, Andrea Tettamanzi.

---

[0] http://qakis.org/qakis2

Together with Celia da Costa Pereira (UNS), we have proposed a framework to measure the acceptability of an information in a multiagent system, according to *(i)* the agent's goals and the information source's goals, *(ii)* the credibility, for the agent, of the incoming information and *(iii)* the agent's beliefs (or perceptions) about the context (or situation) in which it operates. The results of this research have been published at the AAMAS international conference [27].

Moreover, together with Sahbi Benlamine, Maher Chaouachi and Claude Frasson (U. of Montreal) we have presented an empirical evaluation of the relationship between the argumentative structures of human debates and the emotions felt by the debate participants. Argumentation is often seen as a mechanism to support different forms of reasoning such that decision-making and persuasion, but all these approaches assume a purely rational behavior of the involved actors. However, humans are proved to behave differently, mixing rational and emotional attitudes to guide their actions, and it has been claimed that there exists a strong connection between the argumentation process and the emotions felt by people involved in such process. We assess this claim by means of an experiment: during several debates people's argumentation in plain English is connected and compared to the emotions automatically detected from the participants. The results of this research have been published at the IJCAI international conference [19], and submitted to the Cognitive Science journal (under review).

7.2.7. Natural Language Argumentation on Twitter

Participants: Tom Bosc, Elena Cabrio, Serena Villata.

A great amount of textual data is published on social media every day. For example, there are about 500 million new tweets per day on Twitter. These data reflect the opinion and thoughts of a large population and are thus potentially useful to decision-makers and marketers, among others. But processing them is challenging because of their large quantity as well as their noisiness : poor quality of writing, redundancy, presence of advertisement, etc.

The goal of this project is to build a pipeline to automatically analyze messages exchanged on Twitter and build informative and synthetic views. We study tweets under the angle of argumentation theory. First of all, the algorithm filters in argumentative tweets. Then, it describes how tweets relate to one another : tweets may support or attack other tweets, or be neutral. Finally, a visualisation of the interactions between tweets is produced. Individual parts of the pipeline are machine learning models that are trained using datasets that are crafted specially for the project. Importantly, datasets span several domains (politics, society topics, product announcements) to ensure that the approach is generic enough and will generalize to unseen topics.

7.3. Vocabularies, Semantic Web and linked data based knowledge representation

7.3.1. SPARQL Template Transformation Language

Participants: Olivier Corby, Catherine Faron-Zucker, Fabien Gandon, Fuqi Song.

We designed and developed a generic software environment to generate Semantic Web Servers and Linked Data Navigators [0] on top of the STTL SPARQL Template Transformation Language. We designed STTL transformations from RDF to HTML that enable to set up hypertext Linked Data Navigators on local or remote (e.g. DBpedia) triple stores. This work was published at ISWC, WebIST, LNBIP and IC [26], [25], [45], [39].

We extended STTL in order to perform rule based constraint checking. Templates return boolean true (resp. false) when constraint cheking succeeds (resp. fails). We applied this extension on OWL profile conformance checking and we tested with success OWL RL, OWL EL and OWL QL profiles.

7.3.2. SPARQL Function Language

Participants: Olivier Corby, Catherine Faron-Zucker.

---

[0] http://corese.inria.fr

We started the design of a Function Language on top of SPARQL filter language. We added the `function` statement that enables users to define extension functions directly in the filter language. We added statements to the filter languages such as `let` local variables, `for` loop and `list` datatype and we integrated `select` and `construct` queries in the language. Extension functions are directly available into SPARQL queries. This solves the problem of extension function interoperability. We were able to design custom datatypes such as roman numbers, custom aggregates such as median and standard deviation, extension functions to compute the week day of a given date, approximate search functions, recursive functions with the service clause, etc. [50].

7.3.3. Graph Pattern Matching

Participants: Olivier Corby, Fuqi Song.

We proposed a heuristics-based query planning approach which allows reducing SPARQL query executing time. This approach has been developed and integrated to Corese plaform. The relevant work and results have been published at conference KES 2015 [35].

We developed a component that can improve the storage capacity of Corese software, generally speaking this approach stores large RDF literals into the file system instead of in memory. The experiments are performed based on the data set of BSBM [54] and the results suggested that with this component, it can save up to 40% RAM space without slowing down the query execution time.

We implemented and integrated similarity measurement algorithms to Corese software in order to enable approximate semantic search. The main objective is to return approximate results when there are no results in the data source corresponding to the query.

7.3.4. Dynamic Application Scheme Composition

Participant: Isabelle Mirbel.

Dynamic service composition has emerged as a promising approach to build complex runtime-adaptable applications. In this context, new approaches for bottom-up opportunistic assembly of services have emerged. However, these approaches may lead to meaningless and useless compositions. Therefore, we advocate an approach in which bottom-up discovery of services is coupled with top-down user's requirements elicitation.

In our approach, application schemes publish available behaviors from basic component assembly. Our user's requirements elicitation framework, based on previous work, offers the capability to capture high-level end-user's requirements in an iterative and incremental way and to turn them into queries to retrieve application scheme's descriptions. We adopt semantic Web languages and models as a unified framework to deal with end-user's requirements and application scheme's descriptions in order to take advantage of their reasoning and traceability capabilities. We extended previous work about requirement's modeling by providing means to represent and reason on AND and OR operators as well as contextual data. Moreover, relying on the STTL language (see Section 7.3.1 , we proposed two transformations for runtime composition: the first transformation aims at detecting the possible compositions with regards to the available applications schemes; the second one aims at building a BPMN modeling to achieve user's requirements.

7.3.5. Semantic Web Languages And Techniques for Digital Humanities

Participants: Serena Villata, Elena Cabrio, Catherine Faron-Zucker, Franck Michel.

In the framework of the Zoomathia project, we conducted three complementary works. Their results have been published at the SW4SH international workshop [22][37][38]. First, together with Cécile Callou, Chloé Martin and Johan Montagnat (UNS), we started working on the construction of a thesaurus to support multidisciplinary studies on the transmission of zoological knowledge throughout historical periods, combining the analysis of ancient literature, iconographic and archaeozoological resources. We constructed a SKOS thesaurus based on the TAXREF taxonomical reference designed to support studies in Conservation Biology.

Second, together with Molka Tounsi (UNS), and Arnaud Zucker (UNS), we have introduced a methodology to (i) extract pertinent knowledge from medieval texts using Natural Language Processing methods, (ii) semantically enrich semi-structured zoological data and publishing it as an RDF dataset and its vocabulary, linked to other relevant Linked Data sources, and (iii) reason on this linked RDF data to help epistemologists, historians and philologists in their analysis of these ancient texts.

Third, together with Arnaud Zucker, we have proposed to adopt argumentation theory together with Semantic Web languages and techniques to provide an overall view of conflicting critiques over ancient texts, and to detect what are the different competing viewpoints and what are the strongest arguments emerging from the debate. An ontology for argumentative documents is used to annotate ancient texts, and an example of such annotation is provided about the topic of the Eternity of the species in Aristotle.

Moreover, together with Ahmed Missaoui (UNS) and Sara Tonelli (FBK Trento, Italy), we have presented the process performed to map the metadata from the Verbo-Visual-Virtual Project to the Linked Open Data cloud and the related data enrichment. Although the work was largely inspired by past efforts by other cultural heritage institutions, they face new challenges, partly related to the small size of the collection, with little-known artists and few information available from other online sources, and partly to the integration of Natural Language Processing techniques to enrich the metadata. The results of this research have been published at the AIUCD international conference.

7.3.6. *Autonomous Learning of the Meaning of Objects*

Participants: Valerio Basile, Elena Cabrio, Fabien Gandon.

The goal of ALOOF (CHIST-ERA) project [0] is to enable robots to tap into the ever-growing amount of knowledge available on the Web, by learning from there about the meaning of previously unseen objects, expressed in a form that makes them applicable when acting in situated environments. By searching the Web, robots will be able to learn about new objects, their specific properties, where they might be stored and so forth. To achieve this, robots need a mechanism for translating between the representations used in their real-world experience and those on the Web.

In this direction, we are building a *machine reading* pipeline to extract formally encoded knowledge from unstructured text. By combining linguistic and semantic analysis of natural language with entity linking and formal reasoning techniques, our system is capable of extracting meaningful knowledge about entities with URIs in the Linked Open Data (e.g., from DBpedia) and their relationships, encoded in standard Semantic Web fashion, i.e., RDF triples. We then employ the machine reading software to harvest the Web, targeting informative natural language resources such as educational Websites, to create a large-scale meaning bank of common sense knowledge.

7.3.7. *Social Media Intelligence and Linked Knowledge*

Participants: Farhad Nooralahzadeh, Elena Cabrio, Fabien Gandon.

Automated Natural Language Processing (NLP), Web Open Data (Linked Open Data) and social networks are the three topics of the SMILK ANR LabCom including their coupling studied in three ways: texts and Linked Data, Linked Data and social resources, texts and social resources. It is a Joint laboratory between the Inria research institute and the VISEO company to develop research and technologies on the one hand, retrieve, analyze, and reason about linking data from textual Web resources and other to use open Web data taking into account the social structures and interactions in order to improve the analysis and understanding of textual resources.

In this context, we have developed the entity discovery tools by adopting the semantic spreading activation, then we integrated it with the SMILK framework. The goal of this work was to semantically enrich the data by linking the mentions of named entities in the text to the corresponding known entities in knowledge bases. In our approach multiple aspects are considered: the prior knowledge of an entity in Wikipedia (i.e. the keyphraseness and commonness features that can be precomputed by crawling the Wikipedia dump), a set of

---

[0] http://www.dis.uniroma1.it/~aloof/

features extracted from the input text and from the knowledge base, along with the correlation/relevancy among the resources in Linked Data. More precisely, this work explores the *collective ranking approach* formalized as a weighted graph model, in which the mentions in the input text and the candidate entities from knowledge bases are linked using the local compatibility and the global relatedness. Experiments on the datasets of the Open Knowledge Extraction (OKE) [0] challenge with different configurations of our approach in each phase of the linking pipeline reveal its optimum mode. We investigate the notion of semantic relatedness between two entities represented as sets of neighbors in Linked Open Data that relies on an associative retrieval algorithm, with consideration of common neighborhood. This measure improves the performance of prior link-based models and outperforms the explicit inter-link relevancy measure among entities (mostly Wikipedia-centric). Thus, our approach is resilient to non-existent or sparse links among related entities.

7.3.8. Ontology-Based Workflow Management Systems

Participants: Tuan Anh Pham, Nhan Le Thanh.

The main objective of this PhD work is to develop a Shared Workflow Management System (SWMS) using ontology engineering. Everybody can share a semi-complete workflow which is called “Workflow template”, and other people can modify and complete it to use it in their system. This customized workflow is called “Personalized workflow”. The challenges of a SWMS are to be simple, easy to use, user friendly and not too heavy. But it must have all functions of a WMS. There are three major challenges in this work: How to allow the users to customize the workflow template to correspond to their requirements, with changes compliant with the predefined rules in the workflow template? How to build an execution model to evaluate step by step a personalized workflow [34][33].

7.3.9. Semantic Mappings with a Control Flow-Based Business Workflow

Participants: Thi Hoa Hue Nguyen, Nhan Le Thanh.

The aim of this PhD work is to improve the Coloured Petri Nets (CPNs) and Ontology engineering to support the development of business process and business workflow definitions of various fields. To realize this objective, we first propose an ontological approach for representing business models in a meta-knowledge base. We introduce four basic types of manipulation operations on process models used to develop and modify business workflow patterns. Second, we propose a formal definition of semantic constraints and an O(n3)-time algorithm for detecting redundant and conflicting constraints. By relying on the CPN Ontology and sets of semantic constraints, workflow processes are semantically created. Finally, we show how to check the semantic correctness of workflow processes with the SPARQL query language [34].

7.4. Analyzing and Reasoning on Heterogeneous Semantic Graphs

7.4.1. RDF Mining

Participants: Andrea Tettamanzi, Catherine Faron-Zucker, Fabien Gandon, Tran Duc Minh, Claudia d’Amato.

We carried on our investigation in an approach to RDF mining based on grammatical evolution and possibility theory, whose aim is to mine large RDF graphs by automatically generating and testing OWL 2 axioms based on the known facts. In particular, we addressed the problem of scaling up the scoring heuristics based on falsification and possibility theory we have recently proposed [36].

7.4.2. Data and Knowledge Integration and Extraction

Participant: Andrea Tettamanzi.

Together with Somsack Inthasone of the National University of Laos, Nicolas Pasquier and Célia da Costa Pereira of I3S, we completed a survey on biodiversity and environment data mining [16].

---

[0] https://github.com/anuzzolese/oke-challenge

7.4.3. Scalable Uncertainty Management

Participant: Andrea Tettamanzi.

Within the framework of the CNR PEPS GéoIncertitude, we proposed and studied the properties of uncertain logical gates in possibilistic network, using a problem of human geography as a motivating example and testbed [28].

7.4.4. Natural Language Question Answering

Participants: Andrea Tettamanzi, Elena Cabrio, Catherine Faron-Zucker, Amine Hallili.

We extended previous work on answering N-relation natural language questions in the commercial domain by combining an approach to learning regular expressions based on genetic programming [21].

7.4.5. Events Detection in Twitter

Participants: Amosse Edouard, Elena Cabrio, Nhan Le Thanh.

We analyze Twitter data in the objective of identifying events reported by Twitter users. Specially we have worked on two main aspects: an approach for classifying tweets as either related or not related to events and secondly we have studied an approach for disambiguating geographic entities in tweets.

We have worked on an approach for separating event-related content from the rest of Twitter messages. We have combined technics from Natural Language Processing (NLP) and Machine Learning (ML) for building a classifier model that aims at classifying tweets into two mutually exclusive classes. First of all, we apply a Named Entity Recognizer to the tweets in order to identify the occurrences of named entities and special Twitter features such as hashtags, shortened URLs or user mentions. In a second step, the named entities are replaced by their generic class in the DBpedia Ontology; we do so by using SPARQL to query the DBpedia Knowledge Base to extract the class related to each entity. Third, we use the modified content as examples to train a binary classifier. Our evaluation using different classifiers such as Naive Bayes and Long Short Term Memory have shown promising results in term of performance compared to the state of the art.

We have also worked on an approach for identifying geographic entities in Twitter. This task is challenging for two main reasons: first, a geographic term can be related to either geographic or non geographic entities (Paris can be a person or a place) and second, many geographic places might have the same name (Paris can be either the capital of France or a city in Texas). We have proposed an approach based on distant-supervision and ontology matching for identifying and disambiguate ambiguous geographic terms.

ZENITH Project-Team

7. New Results

7.1. Big Data Integration

7.1.1. CloudMdsQL, a query language for heterogeneous data stores

Participants: Carlyna Bondiombouy, Boyan Kolev, Oleksandra Levchenko, Patrick Valduriez.

The blooming of different cloud data management infrastructures, specialized for different kinds of data and tasks, has led to a wide diversification of DBMS interfaces and the loss of a common programming paradigm. The CoherentPaaS European project addresses this problem, by providing a common programming language and holistic coherence across different cloud data stores.

In this context, we have started the design of a Cloud Multi-datastore Query Language (CloudMdsQL), and its query engine. CloudMdsQL is a functional SQL-like language, capable of querying multiple heterogeneous data stores, e.g. relational, NoSQL or HDFS) [19], [31]. The major innovation is that a CloudMdsQL query can exploit the full power of the local data stores, by simply allowing some local data store native queries to be called as functions, and at the same time be optimized. Our experimental validation, with three data stores (graph, document and relational) and representative queries, shows that CloudMdsQL satisfies the five important requirements for a cloud multidatastore query language. In [32], we extend CloudMdsQL to allowing the ad-hoc usage of user defined map/filter/reduce operators in combination with traditional SQL statements, to integrate relational data and big data stored in HDFS and accessed by a data processing framework like Spark.

7.1.2. Semantic Data Integration using Bio-Ontologies

Participant: Pierre Larmande.

The AgroPortal project [49] aims at developing and supporting a reference ontology repository for the agronomic domain. The ontology portal features ontology hosting, search, versioning, visualization, comment, with services for semantically annotating data with the ontologies, as well as storing and exploiting ontology alignments and data annotations. All of these within a fully semantic web compliant infrastructure. The main objective of this project is to enable straightforward use of agronomic related ontologies, avoiding data managers and researchers the burden to deal with complex knowledge engineering issues to annotate the research data. Thus, we specifically pay attention to the requirements of the agronomic community and the specificities of the crop domain. AgroPortal will offer a robust and stable platform that we anticipate will be highly valued by the community.

7.1.3. Access and Integration of Molecular Biology Data

Participants: Sarah Cohen-Boulakia, Patrick Valduriez.

The volumes of molecular biology data available on the web are constantly increasing. Accessing and integrating these data is crucial for making progress in biology. In [26], we provide all the necessary pointers to identify the reference databases capable of providing bioinformatic data for molecular biology. We also discuss the problems posed by the exploitation of these very highly heterogeneous and distributed data. Finally, in order to guide a prospective user on the choice of one of these systems, we provide an overview of the systems that provide unified access to these data.

7.2. Distributed Indexing and Searching

7.2.1. Diversified and Distributed Recommendation for Scientific Data

Participants: Esther Pacitti, Maximilien Servajean.

Recommendation is becoming a popular mechanism to help users find relevant information in large-scale data (scientific data, web). To avoid redundancy in the results, recommendation diversification has been proposed, with the objective of identifying items that are dissimilar, but nonetheless relevant to the user's interests.

We propose a new diversified search and recommendation solution suited for scientific data (i.e., plant phenotyping, botanical data) [22] We first define an original profile diversification scoring function that enables to address the problem of returning redundant items, and enhances the quality of diversification. Through experimental evaluation using two benchmarks, we showed that our scoring function gives the best compromise between diversity and relevancy. Next, to implement our new scoring function, we propose a basic Top-k threshold-based algorithm that exploits a candidate list to achieve diversification and several techniques to improve performance. First, we simplify the scoring model to reduce its computational complexity. Second, we propose two techniques to reduce the number of items in the candidate list, and thus the number of diversified scores to compute. Third, we propose different indexing scores that take into account the diversification of items and an adaptive indexing approach to reduce the number of accesses in the index dynamically based on the queries workload. The experimentation results show that our techniques yield a major reduction of response time, up to 12 times compared to a baseline greedy diversification algorithm.

We also address distributed and diversified recommendation in the context of P2P and multisite cloud [23]. We propose a new scoring function (usefulness) to cluster relevant users over a distributed overlay. Our experimental evaluation using different datasets shows major gains in recall (order of 3 times) compared with state-of-the-art solutions.

7.3. Scientific Workflows

7.3.1. Scientific Workflows: combining data analysis and simulation

Participant: Sarah Cohen-Boulakia.

While scientific workflows are increasingly popular in the bioinformatics community in some emerging application domains such as ecology, the need for data analysis is combined with the need to model complex multi-scale biological systems, possibly involving multiple simulation steps. This requires the scientific workflow to deal with retro-action to understand and predict the relationships between structure and function of these complex systems. OpenAlea (openalea.gforge.inria.fr) developped by the EPI Virtual plants is the only scientific workflow system able to uniformly address the problem, which made it successful in the scientific community.

For the first time, we proposed a conceptualisation of OpenAlea in [42]. We introduce the concept of higher-order dataflows as a means to uniformly combine classical data analysis with modeling and simulation. We provide for the first time the description of the OpenAlea system involving an original combination of features. We illustrate the demonstration on a high-throughput workflow in phenotyping, phenomics, and environmental control designed to study the interplay between plant architecture and climatic change. Ongoing work include deploying OpenAlea on a Grid technology using the SciFloware middleware.

7.3.2. Processing Scientific Workflows in Multi-site cloud

Participants: Ji Liu, Esther Pacitti, Patrick Valduriez.

As the scale of the data increases, scientific workflow management systems (SWfMSs) need to support workflow execution in High Performance Computing (HPC) environments. Because of various benefits, cloud emerges as an appropriate infrastructure for workflow execution. However, it is difficult to execute some scientific workflows in one cloud site because of geographical distribution of scientists, data and computing resources. Therefore, a scientific workflow often needs to be partitioned and executed in a multisite environment.

In [21], we define a multisite cloud architecture that is composed of traditional clouds, e.g., a pay-per-use cloud service such as Amazon EC2, private data-centers, e.g. a cloud of a scientific organization like Inria, COPPE or LNCC, and client desktop machines that have authorized access to the data-centers. We can model this architecture as a distributed system on the Internet, each site having its own computer cluster, data and programs. An important requirement is to provide distribution transparency for advanced services (i.e., workflow management, data analysis), to ease their scalability and elasticity. Current solutions for multisite clouds typically rely on application specific overlays that map the output of one task at a site to the input of another in a pipeline fashion. Instead, we define fully distributed services for data storage, intersite data movement and task scheduling.

7.3.3. Data-centric Iteration in Dynamic Workflows

Participant: Patrick Valduriez.

Dynamic workflows are scientific workflows supporting computational science simulations, typically using dynamic processes based on runtime scientific data analyses. They require the ability of adapting the workflow, at runtime, based on user input and dynamic steering. Supporting data-centric iteration is an important step towards dynamic workflows because user interaction with workflows is iterative. However, current support for iteration in scientific workflows is static and does not allow for changing data at runtime.

In [17], we propose a solution based on algebraic operators and a dynamic execution model to enable workflow adaptation based on user input and dynamic steering. We introduce the concept of iteration lineage that makes provenance data management consistent with dynamic iterative workflow changes. Lineage enables scientists to interact with workflow data and configuration at runtime through an API that triggers steering. We evaluate our approach using a novel and real large-scale workflow for uncertainty quantification on a 640-core cluster. The results show impressive execution time savings from 2.5 to 24 days, compared to non-iterative workflow execution. We verify that the maximum overhead introduced by our iterative model is less than 5% of execution time. Also, our proposed steering algorithms are very efficient and run in less than 1 millisecond, in the worst-case scenario.

7.3.4. Analyzing Related Raw Data Files through Dataflows

Participant: Patrick Valduriez.

Computer simulations may ingest and generate high numbers of raw data files. Most of these files follow a de facto standard format established by the application domain, e.g., FITS for astronomy. Although these formats are supported by a variety of programming languages, libraries and programs, analyzing thousands or millions of files requires developing specific programs. DBMS are not suited for this, because they require loading the raw data and structuring it, which gets heavy at large-scale. Systems like NoDB, RAW and FastBit, have been proposed to index and query raw data files without the overhead of using a DBMS. However, they focus on analyzing one single large file instead of several related files. In this case, when related files are produced and required for analysis, the relationship among elements within file contents must be managed manually, with specific programs to access raw data. Thus, this data management may be time-consuming and error-prone. When computer simulations are managed by a SWfMS, they can take advantage of provenance data to relate and analyze raw data files produced during workflow execution. However, SWfMS register provenance at a coarse grain, with limited analysis on elements from raw data files. When the SWfMS is dataflow-aware, it can register provenance data and the relationships among elements of raw data files altogether in a database which is useful to access the contents of a large number of files. In [24], we propose a dataflow approach for analyzing element data from several related raw data files. Our approach is complementary to the existing single raw data file analysis approaches. We validate our approach with the Montage workflow from astronomy and a workflow from Oil and Gas domain as I/O intensive case studies.

7.4. Scalable Query Processing

7.4.1. Scalable Query Processing with Big Data

Participants: Reza Akbarinia, Miguel Liroz, Patrick Valduriez.

The popular MapReduce parallel processing framework is inefficient in case of data skew, which makes the reduce side done by a few worker nodes.

In [28], [20], we propose FP-Hadoop, which makes the reduce side of MapReduce more parallel. We extend the MapReduce programming model to allow the collaboration of reduce workers on processing the values of an intermediate key, without affecting the correctness of the final results. In FP-Hadoop, the reduce function is replaced by two functions: intermediate reduce and final reduce. There are three phases, each phase corresponding to one of the functions: map, intermediate reduce and final reduce phases. In the intermediate reduce phase, the function, which usually includes the main load of reducing in MapReduce jobs, is executed by reduce workers in a collaborative way, even if all values belong to only one intermediate key. This allows performing a big part of the reducing work by using the computing resources of all workers, even in case of highly skewed data. We implemented a prototype of FP-Hadoop by modifying Hadoop's code, and conducted extensive experiments over synthetic and real datasets. The results show that FP-Hadoop makes MapReduce job processing much faster and more parallel, and can efficiently deal with skewed data. We achieve excellent performance gains compared to native Hadoop, e.g. more than 10 times in reduce time and 5 times in total execution time.

7.5. Data Stream Mining

7.5.1. Summarizing Uncertain Data Streams

Participants: Reza Akbarinia, Florent Masseglia.

Probabilistic data management has shown growing interest to deal with uncertain data. In [29], we focus on probabilistic time series with high volumes of data, which requires efficient compression techniques. To date, most of the work on probabilistic data reduction uses synopses that minimize the error of representation wrt. the original data. However, in most cases, the compressed data will be meaningless for usual queries involving aggregation operators such as SUM or AVG. We propose *PHA* (Probabilistic Histogram Aggregation), a compression technique whose objective is to minimize the error of such queries over compressed probabilistic data. We incorporate the aggregation operator given by the end-user directly in the compression technique, and obtain much lower error in the long term. We also adopt a global error aware strategy in order to manage large sets of probabilistic time series, where the available memory is carefully balanced between the series, according to their individual variability.

7.6. Scalable Data Analysis

7.6.1. Parallel Mining of Maximally Informative k-Itemsets in Big Data

Participants: Saber Salah, Reza Akbarinia, Florent Masseglia.

The discovery of informative itemsets is a fundamental building block in data analytics and information retrieval. While the problem has been widely studied, only few solutions scale. This is particularly the case when i) the data set is massive, and/or ii) the length K of the informative itemset to be discovered is high. In [45], we address the problem of parallel mining of maximally informative k-itemsets (miki) based on joint entropy. We propose PHIKS (Parallel Highly Informative K-itemSets) a highly scalable, parallel mining algorithm. PHIKS renders the mining process of large scale databases (up to terabytes of data) succinct and effective. Its mining process is made up of only two compact, yet efficient parallel jobs. PHIKS uses a clever heuristic approach to efficiently estimates the joint entropies of miki having different sizes with very low upper bound error rate, which dramatically reduces the runtime process. PHIKS has been extensively evaluated using massive, real-world data sets. Our experimental results confirm the effectiveness of our approach by the significant scale-up obtained with high featuresets length and hundreds of millions of objects.

7.6.2. Frequent Itemset Mining in Massively Distributed Environments

Participants: Saber Salah, Reza Akbarinia, Florent Masseglia.

While the problem of Frequent itemset mining (FIM) has been thoroughly studied, few solutions scale. This is mainly the case when i) the amount of data tends to be very large and/or ii) the minimum support (MinSup) threshold is very low. In [46], we study the effectiveness and leverage specific data placement strategies for improving parallel FIM (PFIM) performance in MapReduce, a highly distributed computation framework. By offering a clever data placement and an optimal organization of the extraction algorithms, we show that the itemset discovery effectiveness does not only depend on the deployed algorithms. We propose ODPR (Optimal Data-Process Relationship), a solution for fast mining of frequent itemsets in MapReduce. Our method allows discovering itemsets from massive datasets, where standard solutions do not scale.

In [44], we propose a highly scalable PFIM algorithm, namely Parallel Absolute Top Down (PATD). PATD renders the mining process of very large databases (up to Terabytes) simple and compact. Its mining process is made up of only one parallel job, which dramatically reduces the mining runtime, communication cost and energy power consumption overhead, in a distributed computational platform. Based on a clever and efficient data partitioning strategy, namely Item Based Data Partitioning (IBDP), PATD mines each data partition independently, relying on an absolute minimum support (AM inSup) instead of a relative one. Through an extensive experimental evaluation using real-world data sets, we show that PATD is significantly more efficient and scalable than alternative approaches.

7.6.3. Scalable Mining of Closed Frequent Itemsets

Participants: Mehdi Zitouni, Reza Akbarinia, Florent Masseglia.

Mining big datasets poses a number of challenges which are not easily addressed by traditional mining methods, since both memory and computational requirements are hard to satisfy. One solution is to take advantage of parallel frameworks, such as MapReduce, using ordinary machines. In [48], we address the issue of mining closed frequent itemsets (CFI) from big datasets in such environments. We introduce a new parallel algorithm, called CloPN , for CFI mining. One important feature of CloPN is to use a prime number based approach to transform the data into numerical form, and then to mine closed frequent itemsets by using only multiplication and division operations. We carried out exhaustive experiments over big real world datasets to assess the performance of CloPN. The results show that our algorithm is very efficient in CFI mining from large real world datasets with up to 53 million articles.

7.6.4. Chiaroscuro

Participants: Tristan Allard, Florent Masseglia, Esther Pacitti.

The advent of on-body/at-home sensors connected to personal devices leads to the generation of fine grain highly sensitive personal data at an unprecendent rate. However, despite the promises of large scale analytics there are obvious privacy concerns that prevent individuals to share their personnal data. In [30], we propose Chiaroscuro, a complete solution for clustering personal data with strong privacy guarantees. The execution sequence produced by Chiaroscuro is massively distributed on personal devices, coping with arbitrary connections and disconnections. Chiaroscuro builds on our novel data structure, called Diptych, which allows the participating devices to collaborate privately by combining encryption with differential privacy. Our solution yields a high clustering quality while minimizing the impact of the differentially private perturbation. Our study show that Chiaroscuro is both correct and secure.

7.6.5. Large-scale Recognition of Visual and Audio Entities

Participants: Valentin Leveau, Alexis Joly, Patrick Valduriez.

We improved our work on the retrieval of visual identities by introducing a supervised classification layer on top of the large-scale instance-based matching layer. We introduce a new match kernel based on the inverse rank of the Shared Nearest Neighbors (SNN) combined with local geometric constraints [40]. To avoid overfiting and reduce processing costs, the dimensionality of the resulting over-complete representation is further reduced by hierarchically pooling the raw consistent matches according to their spatial position in the training images. The final image representation is obtained by concatenating the resulting feature vectors at several resolutions. Learning from these representations using a logistic regression classifier is shown to provide excellent fine-grained classification performance. In [38], we transpose our new SNN match kernel to the case of audio contents (applied to bird sounds recognition). Thus, the spatial pooling of geometrically consistent visual matches is replaced by a temporal pooling of temporally consistent audio matches. The resulting classification system obtained the second best results at the LifeCLEF bird identification challenge 2015 [36], the largest challenge of this kind ever organized (1000 bird species, 33K audio recordings).

7.6.6. Crowd-sourced Biodiversity Data Production through Pl@ntNet

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

Initiated in the context of a citizen sciences project with botanists of the AMAP laboratory and the Tela Botanica social network, Pl@ntNet [18] is an innovative collaborative plateform focused on image-based plant identification as a mean to enlist new contributors and boost the production of biodiversity data and knowledge. Since 2010, several hundreds of thousands of geo-tagged and dated plant photographs were collected and revised by tens of thousands of novice, amateur and expert botanists. A content-based identification tool, available as both web and mobile applications, is synchronized with the growing data and allows any user to query or enrich the system with new observations. As a concrete new result, the cumulative number of downloads of the iPhone or Android app did reach 1M in October 2015. One of the main novelty in 2015 was the introduction of deep learning technologies in order to improve classification performance as well as the quality and speed of the content-based image retrieval.

A comparative study that we conducted in the context of the LifeCLEF [0] plant identification challenge did actually confirm that deep convolutional neural networks definitely outperforms the best fine-grained classification models on the aggregation of hand-crafted visual features [33]. Thus, we integrated this technology in the Pl@ntNet platform and exploited it in two ways: (i) for extracting more relevant (local and global) visual features to be indexed and searched within our efficient content-based indexing and retrieval framework (SnoopIm software) (ii) for reranking the species returned by the content-based search engine so as to increase the average reciprocal rank of the correct species while keeping a good level of interpretability of the returned results.

7.6.7. Crowd-sourced Biodiversity Data Production through LifeCLE

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

We continued sharing the data produced by the Pl@ntNet platform with the international research community through the animation of the LifeCLEF research platform and the set-up of three new challenges, one related to plant images, one to bird sounds and one to fish videos. More than 200 research groups registered to at least one of the challenges and about 20 of them crossed the finish lines by running their system on the final test data. A synthesis of the results is published in the LifeCLEF 2015 overview paper [37] and more detailed analyses are provided in technical reports for the plant task [35] and the bird task [36]. We also report on an experimental study aimed at evaluating how state-of-art computer vision systems perform in identifying plants compared to human expertise [15],. A subset of the evaluation dataset used within LifeCLEF 2014 plant identification challenge was shared with volunteers of diverse expertise, ranging from leading experts of the targeted flora to inexperienced test subjects. In total, 16 human runs were collected and evaluated comparatively to the 27 machine-based runs of LifeCLEF challenge. The main outcome of the experiment was that machines are still far from outperforming the best expert botanists but they are clearly competing with some experienced botanists specialists of other floras.

---

[0] www.lifeclef.org