INRIA

Project-Team VISTA

Vision Spatio-Temporelle et Apprentissage

Rennes - Bretagne Atlantique

THEME COG

Activity Report

2007

Table of contents

1. Team

Head of project-team

Patrick Bouthemy [DR, head of project team until 30/06/2007, HdR]

Patrick Pérez [DR, head of project team since 01/07/2007, HdR]

Administrative assistant

Huguette Béchu [TR Inria, shared with Temics project-team]

Research scientists

Patrick Bouthemy [DR, head of Inria Rennes Research Center, since 01/07/2007, HdR]

Charles Kervrann [Inra research scientist, on Inria secondment (mise à disposition)]

Ivan Laptev [CR]

Étienne Mémin [Associate professor, Ifsic, University of Rennes 1, on sabbatical leave at Unv. of Bueno Aires (UBA) since 01/04/2007, HdR]

Jean-Pierre Le Cadre [DR CNRS, HdR]

External collaborators

Bruno Cernuschi-Frias [Visiting professor, UBA, Inria associate team FIM, June 2006]

Thomas Corpetti [CR CNRS at Costel (University of Rennes 2)]

Jian-Feng Yao [Full professor, Irmar, UFR Maths, University of Rennes 1, HdR]

Ph-D students

Kamel Aouichat [Cifre grant, Video Metrix, from 01/03/06]

Christophe Avenel [ENS grant, from 01/10/07]

Vijay Badrinarayanan [Cifre grant, Thomson-RD, from 01/03/06]

Jérôme Boulanger [Inria grant, until 31/01/07]

Aurélie Bugeau [Research Ministry grant, until 15/11/07]

Tomas Crivelli [Faculty of Engineering, University of Buenos-Aires, "thèse en co-tutelle"]

Emilie Dexter [DGA grant, from 01/10/06]

Alexandre Hervieu [Inria grant, FP6-NOE Muscle]

Adrien Ickowicz [Inria grant, from 01/11/06 to 01/10/07, then DGA grant, from 01/10/07]

Matthieu Fradet [Cifre grant, Thomson-RD, from 01/03/07]

Nicolas Papadakis [Inria grant, FP6-IST Fluid project until 15/11/07]

Thierry Pécot [Inra grant, from 01/10/06]

Cécile Simonin [DGA grant, from 01/10/05]

Post-doctoral fellows

Imran Junejo [Inria post-doc since 01/11/2007]

Gwénaëlle Piriou [Lecturer (ATER) Ifsic, University of Rennes 1, until 30/07/07]

Inria technical staff

Nicolas Gengembre [FP6-IST Fluid project from 01/06/06 until 31/05/07]

Pham Nam Trung [FP6 Pegase project from 15/11/07]

Patrick Heas [FP6-IST Fluid project until 30/09/07]

2. Overall Objectives

2.1. Overall Objectives

Vista research work is concerned with various types of spatio-temporal images, (mainly video images, but also meteorological satellite images, video-microscopy, X-ray images). We are investigating methods to analyze dynamic scenes, and, more generally, dynamic phenomena, within image sequences. We address the full range of problems raised by the analysis of such dynamic contents with a focus on image motion analysis issues: denoising, motion detection, motion estimation, motion-based segmentation, tracking, motion recognition and

interpretation with learning. We usually rely on statistical approaches, resorting to: Markov models, Bayesian inference, robust estimation, particle filtering, learning. Application-wise, we focus our attention on three main domains: content-aware video applications, meteorological imaging and experimental visualization in fluid mechanics, biological imaging. For that, a number of collaborations, academic and industrial, national and international, are set up.

3. Scientific Foundations

3.1. Motion estimation and motion segmentation with MRF models

Assumptions (i.e., data models) must be formulated to relate the observed image intensities to motion, and other constraints (i.e., motion models) must be added to solve problems like motion segmentation, optical flow computation, or motion recognition. The motion models are supposed to capture known, expected or learned properties of the motion field ; this implies to somehow introduce spatial coherence or more generally contextual information. The latter can be formalized in a probabilistic way with local conditional densities as in Markov models. It can also rely on predefined spatial supports (e.g., blocks or pre-segmented regions). The classic mathematical expressions associated with the visual motion information are of two types. Some are continuous variables to represent velocity vectors or parametric motion models. The others are discrete variables or symbolic labels to code motion detection (binary labels), motion segmentation (numbers of the motion regions or layers) or motion recognition output (motion class labels).

In the past years, we have addressed several important issues related to visual motion analysis, in particular with a focus on the type of motion information to be estimated and the way contextual information is expressed and exploited. Assumptions (i.e., data models) must be formulated to relate the observed image intensities to motion, and other constraints (i.e., motion models) must be added to solve problems like motion segmentation, optical flow computation, or motion recognition. The motion models are supposed to capture known, expected or learned properties of the motion field ; this implies to somehow introduce spatial coherence or more generally contextual information. The latter can be formalized in a probabilistic way with local conditional densities as in Markov models. It can also rely on predefined spatial supports (e.g., blocks or pre-segmented regions). The classic mathematical expressions associated with the visual motion information are of two types. Some are continuous variables to represent velocity vectors or parametric motion models. The others are discrete variables or symbolic labels to code motion detection (binary labels), motion segmentation (numbers of the motion regions or layers) or motion recognition output (motion class labels). We have also recently introduced new models, called mixed-state models and mixed-state auto-models, whose variables belong to a domain formed by the union of discrete and continuous values. We briefly describe here how such models can be specified and exploited in two central motion analysis issues: motion segmentation and motion estimation.

The brightness constancy assumption along the trajectory of a moving point $p(t)$ in the image plane, with $p(t) = (x(t), y(t))$, can be expressed as $dI(x(t), y(t), t)/dt = 0$, with I denoting the image intensity function. By applying the chain rule, we get the well-known motion constraint equation:

$$r(p, t) = \mathbf{w}(p, t).\nabla I(p, t) + I_t(p, t) = 0 \ , \tag{1}$$

where ∇I denotes the spatial gradient of the intensity, with $\nabla I = (I_x, I_y)$, and I_t its partial temporal derivative. The above equation can be straightforwardly extended to the case where a parametric motion model is considered, and we can write:

$$r_\theta(p, t) = \mathbf{w}_\theta(p, t).\nabla I(p, t) + I_t(p, t) = 0 \ , \tag{2}$$

where θ denotes the vector of motion model parameters.

One important step ahead in solving the motion segmentation problem was to formulate the motion segmentation problem as a statistical contextual labeling problem or in other words as a discrete Bayesian inference problem. Segmenting the moving objects is then equivalent to assigning the proper (symbolic) label (i.e., the region number) to each pixel in the image. The advantages are mainly two-fold. Determining the support of each region is then implicit and easy to handle: it merely results from extracting the connected components of pixels with the same label. Introducing spatial coherence can be straightforwardly (and locally) expressed by exploiting MRF models. Here, by motion segmentation, we mean the competitive partitioning of the image into motion-based homogeneous regions. Formally, we have to determine the hidden discrete motion variables (i.e., region numbers) $l(i)$ where i denotes a site (usually, a pixel of the image grid; it could be also an elementary block). Let $l = \{l(i), i \in S\}$. Each label $l(i)$ takes its value in the set $\Lambda = \{1, .., N_{reg}\}$ where N_{reg} is also unknown. Moreover, the motion of each region is represented by a motion model (usually, a 2D affine motion model of parameters θ which have to be conjointly estimated; we have also explored non-parametric motion modeling [79]). Let $\Theta = \{\theta_k, k = 1, .., N_{reg}\}$. The data model of relation (2) is used. The *a priori* on the motion label field (i.e., spatial coherence) is expressed by specifying a MRF model (the simplest choice is to favor the configuration of the same two labels on the two-site cliques so as to yield compact regions with regular boundaries). Adopting the Bayesian MAP criterion is then equivalent to minimizing an energy function E whose expression can be written in the general following form:

$$E(l, \Theta, N_{reg}) = \sum_{i \in S} \rho_1[r_{\theta_{l(i)}}(i)] + \sum_{i \sim j} \rho_2[l(i), l(j)] \ , \tag{3}$$

where $i \sim j$ designates a two-site clique. We first considered [75] the quadratic function $\rho_1(x) = x^2$ for the data-driven term in (3). The minimization of the energy function E was carried out on l and Θ in an iterative alternate way, and the number of regions N_{reg} was determined by introducing an extraneous label and using an appropriate statistical test. We later chose a robust estimator for ρ_1 [88], [89]. It allowed us to avoid the alternate minimization procedure and to determine or update the number of regions through an outlier process in every region.

Specifying (simple) MRF models at a pixel level (i.e., sites are pixels and a 4- or 8-neighbor system is considered) is efficient, but remains limited to express more sophisticated properties on region geometry or to handle extended spatial interaction. Multigrid MRF models [82] is a means to address somewhat the second concern (and also to speed up the minimization process while usually supplying better results). An alternative is to first segment the image into spatial regions (based on gray level, color or texture) and to specify a MRF model on the resulting graph of adjacent regions [80]. The motion region labels are then assigned to the nodes of the graph (which are the sites considered in that case). This allowed us to exploit more elaborated and less local *a priori* information on the geometry of the regions and their motion. However, the spatial segmentation stage is often time consuming, and getting an effective improvement on the final motion segmentation accuracy remains questionable.

By definition, the velocity field formed by continuous vector variables is a complete representation of the motion information. Computing optical flow based on the data model of equation (1) requires to add a motion model enforcing the expected spatial properties of the motion field, that is, to resort to a regularization method. Such properties of spatial coherence (more specifically, piecewise continuity of the motion field) can be expressed on local spatial neighborhoods. First methods to estimate discontinuous optical flow fields were based on MRF models associated with Bayesian inference (i.e., minimization of a discretized energy function). A general formulation of the global (discretized) energy function to be minimized to estimate the velocity field **w** can be given by:

$$E(\mathbf{w}, \zeta) = \sum_{p \in S} \rho_1[r(p)] \ + \sum_{p \sim q} \rho_2[\|\mathbf{w}(p) - \mathbf{w}(q)\|, \zeta(p'_{p \sim q})] + \sum_{A \in \chi} \rho_3(\zeta_A) \ , \tag{4}$$

where S designates the set of pixel sites, $r(p)$ is defined in (1), $S' = \{p'\}$ the set of discontinuity sites located midway between the pixel sites and χ is the set of cliques associated with the neighborhood system chosen on S'. We first used quadratic functions and the motion discontinuities were handled by introducing a binary line process ζ [81]. Then, robust estimators were popularized leading to the introduction of so-called auxiliary variables ζ now taking their values in $[0, 1]$ [87]. Multigrid MRF are moreover involved, and multiresolution incremental schemes are exploited to compute optical flow in case of large displacements. Dense optical flow and parametric motion models can also be jointly considered and estimated, which enables to supply a segmented velocity field [86]. Depending on the followed approach, the third term of the energy $E(\mathbf{w}, \zeta)$ can be optional.

3.2. Fluid motion analysis

Analyzing fluid motion is essential in number of domains and can rarely be handled using generic computer vision techniques. In this particular application context, we study several distinct problems. We first focus on the estimation of dense velocity maps from image sequences. Fluid flows velocities cannot be represented by a single parametric model and must generally be described by accurate dense velocity fields in order to recover the important flow structures at different scales. Nevertheless, in contrast to standard motion estimation approach, adapted data model and higher order regularization are required in order to incorporate suitable physical constraints. In a second step, analyzing such velocity fields is also a source of concern. When one wants to detect particular events, to segment meaningful areas, or to track characteristic structures, dedicated methods must be devised and studied.

Since several years, the analysis of video sequences showing the evolution of fluid phenomena has attracted a great deal of attention from the computer vision community. The applications concern domains such as experimental visualization in fluid mechanics, environmental sciences (oceanography, meteorology, ...), or medical imagery.

In all these application domains, it is of primary interest to measure the instantaneous velocity of fluid particles. In oceanography, one is interested to track sea streams and to observe the drift of some passive entities. In meteorology, both at operational and research levels, the task under consideration is the reconstruction of wind fields from the displacements of clouds as observed in various satellite images. In medical imaging, the issue can be to visualize and analyze blood flow inside the heart, or inside blood vessels. The images involved in each domain have their own characteristics and are provided by very different sensors. The huge amount of data of different kinds available, the range of applicative domains involved, and the technical difficulties in the processing of all these specific image sequences explain the interest of the image analysis community.

Extracting dense velocity fields from fluid images can rarely be done with the standard computer vision tools. The latter were originally designed for quasi-rigid motions with stable salient features, even if these techniques have proved to be more and more efficient and provide accurate results for natural images [87], [86]. These generic approaches are based on the brightness constancy assumption of the points along their trajectory ($\frac{df}{dt} = 0$), along with the spatial smoothness assumption of the motion field. These estimators are defined as the minimizer of the following energy function:

$$\int_{\Omega} \rho[\nabla f \cdot w + \frac{\partial f}{\partial t}]d\mathbf{s} + \alpha \int_{\Omega} \rho(\|\nabla w\|)d\mathbf{s}. \tag{5}$$

The penalty function ρ is usually the L_2 norm, but it may be substituted for a robust function attenuating the effect of data that deviate significantly from the brightness constancy assumption [87], and enabling also to implicitly handle the spatial discontinuities of the motion field.

Contrary to usual video image sequence contents, fluid images exhibit high spatial and temporal distortions of the luminance patterns. The design of alternative approaches dedicated to fluid motion thus constitutes a widely-open research problem. It requires to introduce some physically relevant constraints which must be embedded in a higher-order regularization functional [76]. The method we have devised for fluid motion involves the following global energy function:

$$\int_\Omega \rho \left\{ f\left(\mathbf{s} + \mathbf{w}, t+1\right) \exp\left(div\ \mathbf{w}\right) - f(\mathbf{s}, t) \right\} d\mathbf{s} + \alpha \int_\Omega \left(\|\nabla div\ \mathbf{w}\|^2 + \|\nabla curl\ \mathbf{w}\|^2 \right) d\mathbf{s}.$$

The first term comes from an integration of the continuity equation (assuming the velocity of a point is constant between instants t and $t + \Delta t$). Such a data model is a "fluid counterpart" of the usual "Displaced Frame Difference" expression. Instead of expressing brightness constancy, it explains a loss or gain of luminance due to a diverging motion. The second term is a smoothness term designed to preserve divergence and vorticity blobs. This regularization term is nevertheless very difficult to implement. As a matter of fact, the associated Euler-Lagrange equations consist in two fourth-order coupled PDE's, which are tricky to solve numerically. We proposed to simplify the problem by introducing auxiliary functions, and by defining the following alternate smoothness function:

$$\int_\Omega |div\ \mathbf{w} - \xi|^2 + \lambda \rho(\|\nabla \xi\|) d\mathbf{s} + \alpha \int_\Omega |curl\ \mathbf{w} - \zeta|^2 + \lambda \rho(\|\nabla \zeta\|) d\mathbf{s}. \qquad (6)$$

The new auxiliary scalar functions ξ and ζ can be respectively seen as estimates of the divergence and the curl of the unknown motion field, and λ is a positive parameter. The first part of each integral enforces the displacement to comply with the current divergence and vorticity estimates ξ and ζ, through a quadratic goodness-of-fit enforcement. The second part associates the divergence and the vorticity estimates with a robust first-order regularization enforcing piece-wise smooth configurations. From a computational point of view, such a regularizing function only implies the numerical resolution of first-order PDE's. It may be shown that, at least for the L_2 norm, the regularization we proposed is a smoothed version of the original second-order div-curl regularization.

Once given a reliable description of the fluid motion, another important issue consists in extracting and characterizing structures of interest such as singular points or in deriving potential functions. The knowledge of the singular points is precious to understand and predict the considered flows, but it also provides compact and hierarchical representations of the flow [77]. Such a compact representation enables for instance to tackle difficult tracking problems. As a matter of fact, the problem amounts here to track high dimensional complex objects such as surfaces, level lines, or vector fields. As these objects are only partially observable from images and driven by non linear 3D laws, we have to face a tough tracking problem of large dimension for which no satisfying solution exists at the moment.

3.3. Object tracking with non-linear probabilistic filtering

Tracking problems that arise in target motion analysis (TMA) and video analysis are highly non-linear and multi-modal, which precludes the use of Kalman filter and its classic variants. A powerful way to address this class of difficult filtering problems has become increasingly successful in the last ten years. It relies on sequential Monte Carlo (SMC) approximations and on importance sampling. The resulting sample-based filters, also called particle filters, can, in theory, accommodate any kind of dynamical models and observation models, and permit an efficient tracking even in high dimensional state spaces. In practice, there is however a number of issues to address when it comes to difficult tracking problems such as long-term visual tracking under drastic appearance changes, or multi-object tracking.

The detection and tracking of single or multiple targets is a problem that arises in a wide variety of contexts. Examples include sonar or radar TMA and visual tracking of objects in videos for a number of applications (e.g., visual servoing, tele-surveillance, video editing, annotation and search). The most commonly used framework for tracking is that of Bayesian sequential estimation. This framework is probabilistic in nature, and thus facilitates the modeling of uncertainties due to inaccurate models, sensor errors, environmental noise, etc. The general recursions update the posterior distribution of the target state $p(\mathbf{x}_t|\mathbf{y}_{1:t})$, also known as the filtering distribution, where $\mathbf{y}_{1:t} = (\mathbf{y}_1 \cdots \mathbf{y}_t)$ denotes all the observations up to the current time step, through two stages:

$$\textbf{prediction step:} \qquad p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) = \int p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})d\mathbf{x}_{t-1}$$

$$\textbf{filtering step:} \qquad p(\mathbf{x}_t|\mathbf{y}_{1:t}) = \frac{p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t-1})}{p(\mathbf{y}_t|\mathbf{y}_{1:t-1})},$$

where the prediction step follows from marginalization, and the new filtering distribution is obtained through a direct application of Bayes' rule. The recursion requires the specification of a dynamic model describing the state evolution $p(\mathbf{x}_t|\mathbf{x}_{t-1})$, and a model for the state likelihood in the light of the current measurements $p(\mathbf{y}_t|\mathbf{x}_t)$. The recursion is initialized with some distribution for the initial state $p(\mathbf{x}_0)$. Once the sequence of filtering distributions is known, point estimates of the state can be obtained according to any appropriate loss function, leading to, e.g., Maximum *A Posteriori* (MAP) and Minimum Mean Square Error (MMSE) estimates.

The tracking recursion yields closed-form expressions in only a small number of cases. The most well-known of these is the Kalman Filter (KF) for linear and Gaussian dynamic and likelihood models. For general non-linear and non-Gaussian models the tracking recursion becomes analytically intractable, and approximation techniques are required. Sequential Monte Carlo (SMC) methods [78], [84], [83], otherwise known as particle filters, have gained a lot of popularity in recent years as a numerical approximation strategy to compute the tracking recursion for complex models. This is due to their efficiency, simplicity, flexibility, ease of implementation, and modeling success over a wide range of challenging applications.

The basic idea behind particle filters is very simple. Starting with a weighted set of samples $\{w_{t-1}^{(n)}, \mathbf{x}_{t-1}^{(n)}\}_{n=1}^{N}$ approximately distributed according to $p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})$, new samples are generated from a suitably designed proposal distribution, which may depend on the old state and the new measurements, i.e., $\mathbf{x}_t^{(n)} \sim q(\mathbf{x}_t|\mathbf{x}_{t-1}^{(n)}, \mathbf{y}_t)$, $n = 1 \cdots N$. Importance sampling theory indicates that a consistent sample is maintained by setting the new importance weights to

$$w_t^{(n)} \propto w_{t-1}^{(n)} \frac{p(\mathbf{y}_t|\mathbf{x}_t^{(n)})p(\mathbf{x}_t^{(n)}|\mathbf{x}_{t-1}^{(n)})}{q(\mathbf{x}_t^{(n)}|\mathbf{x}_{t-1}^{(n)}, \mathbf{y}_t)}, \quad \sum_{n=1}^{N} w_t^{(n)} = 1,$$

where the proportionality is up to a normalizing constant. The new particle set $\{w_t^{(n)}, \mathbf{x}_t^{(n)}\}_{n=1}^{N}$ is then approximately distributed according to $p(\mathbf{x}_t|\mathbf{y}_{1:t})$. Approximations to the desired point estimates can then be obtained by Monte Carlo techniques. From time to time it is necessary to resample the particles to avoid degeneracy of the importance weights. The resampling procedure essentially multiplies particles with high importance weights, and discards those with low importance weights.

In many applications, the filtering distribution is highly non-linear and multi-modal due to the way the data relate to the hidden state through the observation model. Indeed, at the heart of these models usually lies a data association component that specifies which part, if any, of the whole current data set is "explained" by the hidden state. This association can be implicit, like in many instances of visual tracking where the state specifies a region of the image plane. The data, e.g., raw color values or more elaborate descriptors, associated to this region only are then explained by the appearance model of the tracked entity. In case measurements are the sparse outputs of some detectors, as with edgels in images or bearings in TMA, associations variables are added to the state space, whose role is to specify which datum relates to which target (or clutter).

In this large context of SMC tracking techniques, two sets of important open problems are of particular interest for Vista:

- selection and on-line estimation of observation models with multiple data modalities: except in cases where detailed prior is available on state dynamics (e.g., in a number of TMA applications), the observation model is the most crucial modeling component. A sophisticated filtering machinery will not be able to compensate for a weak observation model (insufficiently discriminant and/or insufficiently complete). In most adverse situations, a combination of different data modalities is necessary. Such a fusion is naturally allowed by SMC, which can accommodate any kind of data

model. However, there is no general means to select the best combination of features, and, even more importantly, to adapt online the parameters of the observation models associated to these features. The first problem is a difficult instance of discriminative learning with heterogeneous inputs. The second problem is one of online parameter estimation, with the additional difficulty that the estimation should be mobilized only parsimoniously in time, at instants that must be automatically determined (adaptation when the entities are momentarily invisible or simply not detected by the sensors will always cause losses of track). These problems of feature selection, online model estimation, and data fusion, have started to receive a great deal of attention in the visual tracking community, but proposed tools remain ad-hoc and restricted to specific cases.

- multiple-object tracking with data association: when tracking jointly multiple objects, data association rapidly poses combinatorial problem. Indeed, the observation model takes the form of a mixture with a large number components indexed by the set of all admissible associations (whose enumeration can be very expensive). Alternatively, the association variables can be incorporated within the state space, instead of being marginalized out. In this case, the observation model takes a simpler product form, but at the expense of a dramatic dimension increase of the space in which the estimation must be conducted.

  In any case, strategies have thus to be designed to keep low the complexity of the multi-object tracking procedure. This need is especially acute when SMC techniques, already often expensive for a single object, are required. One class of approach consists in devising efficient variants of particle filters in the high-dimensional product state space of joint target hypotheses. Efficiency can be achieved, to some extent, by designing layered proposal distributions in the compound target-association state space, or by marginalizing out approximately the association variables. Another set of approaches lies in a crude, yet very effective approximation of the joint posterior over the product state space into a product of individual posteriors, one per object. This principle, stemming from the popular JPDAF (joint probabilistic data association filter) of the trajectography community, is amenable to SMC approximation. The respective merits of these different approaches are still partly unclear, and are likely to vary dramatically from one context to another. Thorough comparisons and continued investigation of new alternatives are still necessary.

4. Application Domains

4.1. Application Domains

We are dealing with the following application domains (mainly in collaboration with the listed partners) :

- Content-aware video applications (Thomson, FT-RD, INA);
- Experimental fluid mechanics (Cemagref) and meteorological imagery (LMD). We are also leading the FET-IST European project Fluid (with see paragraph 7.6.1) and are in the Inria associate team FIM with the University of Buenos-Aires (see paragraph 8.3.1);
- Biological imagery (Inra, Curie Institute, Biology Dpt of University of Rennes 1)
- Surveillance (Onera, Thales, collaborations are nevertheless considered only from an academic viewpoint). The main addressed issues are search and surveillance, navigation, distributed tracking with a sensor network.

4.1.1. *Content-aware video applications*

The amount of video footage is constantly increasing due to the dissemination of video cameras, the broadcasting of TV programs by multiple means, the seamless acquisition of personal videos,...The exploitation of video material, whatever its usage, requires automatic (or at least semi-automatic) tools to process video contents. A wide range of applications can be envisaged dealing with editing, analyzing, annotating, browsing and authoring video contents. Video indexing and retrieval for audio-visual archives is, for instance, a major

application, which is receiving lot of attention. Other needs include the creation of enriched videos, the design of interactive video systems, the generation of video summaries, and the development of re-purposing frameworks (specifically, for 3G mobile phones and Web applications). For most of all these applications, tools for segmenting videos, detecting events or recognizing actions are usually required.

We are mainly interested in the processing of videos which are shot (and broadcast) in the audiovisual domain, more specifically, sports videos but also TV shows or dance videos. Amateur videos of similar content can also be within our concern. On one hand, sports videos raise difficult issues, since the acquisition process is weakly controlled and content exhibits high complexity, diversity and variability. On the other hand, motion is tightly related to sports semantics. Besides, the exploitation of sports videos forms an obvious business target. We have developed several methods and tools in that context addressing issues such as shot change detection, camera motion estimation and characterization, object tracking, motion modeling and recognition, event detection, video summarization. Beside this main domain of applications, we are also investigating gesture analysis problems. An on-going project in particular aims at monitoring automatically car drivers' attention.

4.1.2. *Experimental fluid mechanics and meteorological imagery*

Concerning the analysis of fluid flows from image sequences, we focus mainly on the domains of experimental fluid mechanics and meteorological imaging. We aim at designing new methods allowing us to extract kinematic or dynamical descriptors of fluid flows from image sequences. We have to face an huge amount of high resolution image sequences. These data reveal in a more and more accurate way the spatio-temporal evolution of flow structures in a non intrusive way. The kinds of data involved in these applicative domains may be various, depending on the experimental imaging set-up and/or the image sensor used. Very specific applications may be tackled for some type of images, but general and common goals can nevertheless be defined in term of motion analysis. Image motion estimation aims at providing instantaneous measurements of the flow velocity and at bringing to physicists kinematic elements allowing them to analyze complex fluid flows. In both domains, the estimation of velocity flow fields from an image sequence is routinely performed with local methods which rely on the computation of average displacements by cross-correlation over small search windows. Despite sophisticated block-matching schemes have been designed in order to cope with intrinsic difficulties of particle-seeded images or atmospheric satellite images, these approaches can hardly cope with low contrast visualization techniques such as Schlieren images or images of the MSG (Meteosat Second Generation satellite) water vapor channel. These methods are not convenient also to get dense velocity fields accurate enough at different scales and for spatially varying motions in order to exhibit for instance the relevant flow features. Besides, the incorporation of fluid flow dynamic laws (almost inescapable in a near future with upcoming high time resolution image sequences) cannot be really handled with local correlation methods. As a matter of fact, no spatial and temporal coherency can be handled with such processing techniques as they operate entirely in a data-driven way allowing no inclusion of physical prior knowledge (related to the basic equations of fluid mechanics). From that point of view, motion analysis techniques developed in computer vision are particularly relevant as they combine model-driven variational smoothness functions with data-driven terms.

On such a basis, as for the meteorological domain, the first objective we are pursuing consists in designing techniques for an accurate estimation of the atmospheric wind fields. Such a goal should require fine sophisticated schemes incorporating physical models of the atmosphere. The second goal is to propose methods for tracking cloud systems of importance, which are useful when one aims at monitoring potentially dangerous events such as convective clouds, hurricane, tornadoes, etc. These two issues have potentially a great impact on weather forecasting, risk prevention, or enhancement of global atmospheric circulation model assimilation.

As for experimental fluid mechanics, we are investigating new methods for the analysis of complex fluid flows from image sequences. A large range of applications is concerned for instance with turbulent flows in aerodynamics, aeronautics, heat transfer, etc. Applications involving flow control are of particular interest (flow separation delay, mixing enhancement, drag reduction,...). These applications need enhanced visualization and sound numerical techniques such as low-order modeling with reduced dynamical models. The processing of

real data and the accuracy enhancement of spatio-temporal measurements may together bring improvements in the modeling of turbulent flows which is traditionally solely based on initial conditions captured through experimental conditions.

4.1.3. Biological imagery

Recent progresses in molecular biology and light microscopy make henceforth possible the acquisition of multi-dimensional data (3D + time) at one or several wavelengths (multispectral imaging) and the observation of intra-cellular molecular dynamics at sub-micron resolutions. Automatic image processing methods to study molecular dynamics from image sequences are therefore of major interest, for instance, for membrane transport involving the movement of small particles from donor to acceptor compartments within the living cell.

The challenge is then to track GFP tags (fluorescent proteins for labeling) with high precision in movies representing several gigabytes of image data. The data are collected and processed automatically to generate information on partial or complete trajectories. In our research work, we are developing methods to perform the computational analysis of these complex 3D image sequences since the capabilities of most commercial image analysis tools for automatically extracting information are rather limited and/or require a large amount of manual interactions with the user.

Quantitative analysis of data obtained by fast 4D wide-field microscopy combined with deconvolution and Green Fluorescence Protein (GFP)-tagging allows one to enlighten the role of specific proteins on HeLa human cell lines. Among these proteins, some are member of the family of Rab-GTPases that bind reversibly to specific membranes within the cells. In our study, we aim at designing computational and statistical models to understand membrane trafficking and, more precisely to better elucidate the role of Rab family proteins inside their multiprotein complexes. We mainly focus on the analysis of transport intermediates (vesicles) that deliver cellular components to appropriate places within cells. Methods have been developed for specific Rab6a and Rab6a' proteins - involved in the regulation of transport from the Golgi apparatus to the endoplasmic reticulum. These small proteins are propelled by molecular motors (kinesin/dynein) to move material along microtubules (polymers). A second application concerns the CLIP 170 protein involved in the kinetochores anchorage (in the segregation of chromosomes to daughter cells, the chromosomes appear to be pulled via a so-called kinetochore attached to chromosome centromeres). This year, we have developed computational methods to estimate the growth (polymerization) of microtubules and catastrophes (depolimerization) organized by centrosomes observed in DIC (Differential Interference Contrast) microscopy.

5. Software

5.1. Motion2d software - parametric motion model estimation

Participants: Fabien Spindler (Lagadic team), Patrick Bouthemy.

Motion2D is a multi-platform object-oriented library to estimate 2D parametric motion models in an image sequence. It can handle several types of motion models, namely, constant (translation), affine, and quadratic models. Moreover, it includes the possibility of accounting for a global variation of illumination. The use of such motion models has been proved adequate and efficient for solving problems such as optic flow computation, motion segmentation, detection of independent moving objects, object tracking, or camera motion estimation, and in numerous application domains, such as dynamic scene analysis, video surveillance, visual servoing for robots, video coding, or video indexing. Motion2D is an extended and optimized implementation of the robust, multi-resolution and incremental estimation method (exploiting only the spatio-temporal derivatives of the image intensity function) we defined several years ago [88]. Real-time processing is achievable for motion models involving up to 6 parameters (for 256x256 images). Motion2D can be applied to the entire image or to any pre-defined window or region in the image. Motion2D is released in two versions :

- Motion2D Free Edition is the version of Motion2D available for development of Free and Open Source software only (no commercial use). It is provided free of charge under the terms of the Q Public License. It includes the source code and makefiles for Linux, Solaris, SunOS, and Irix. The latest version (last release 1.3.11, January 2005) is available for download.
- Motion2D Professional Edition provided for commercial software development. This version also supports Windows 95/98 and NT.

More information on Motion2D can be found at http://www.irisa.fr/vista/Motion2D and the software can be donwloaded at the same Web address.

5.2. d-Change software - motion detection

Participants: Fabien Spindler (Lagadic team), Patrick Bouthemy.

D-change is a multi-platform object-oriented software to detect mobile objects in an image sequence acquired by a static camera. It includes two versions : the first one relies on Markov models and supplies a pixel-based binary labeling, the other one introduces rectangular models enclosing the mobile regions to be detected. It simultaneously exploits temporal differences between two successive images of the sequence and differences between the current image and a reference image of the scene without any mobile objects (this reference image is updated on line). The algorithm provides the masks of the mobile objects (mobile object areas or enclosing rectangles according to the considered version) as well as region labels enabling to follow each region over the sequence.

5.3. DenseMotion software - Estimation of 2D dense motion fields

Participants: Thomas Corpetti, Patrick Heas, Etienne Mémin.

This code allows the computation from two consecutive images of a dense motion field. The estimator is expressed as a global energy function minimization. The code enables the choice of different data model and different regularization functional depending on the targeted application. Generic motion estimator for video sequences or dedicated motion estimator for fluid flows can be specified. This estimator allows in addition the users to specify additional correlation based matching measurements. It enables also the inclusion of a temporal smoothing prior relying on a velocity vorticity formulation of the Navier-Stoke equation for Fluid motion analysis applications. The different variants of this code correspond to research studies that have been published in IEEE transaction on Pattern Analysis and machine Intelligence, Experiments in Fluids, IEEE transaction on Image Processing, IEEE transaction on Geo-Science end Remote Sensing. The binary of this code can be freely downloaded on the FLUID web site http://fluid.irisa.fr.

5.4. 2DLayeredMotion software - Estimation of 2D independent mesoscale layered atmospheric motion fields

Participants: Patrick Heas, Etienne Mémin, Nicolas Papadakis.

This software enables to estimate a stack of 2D horizontal wind fields corresponding to a mesoscale dynamics of atmospheric pressure layers. This estimator is formulated as the minimization of a global energy function. It relies on a vertical decomposition of the atmosphere into pressure layers. This estimator uses pressure data and classification clouds maps and top of clouds pressure maps (or infra-red images). All these images are routinely supplied by the EUMETSAT consortium which handles the Meteosat and MSG satellite data distribution. The energy function relies on a data model built from the integration of the mass conservation on each layer. The estimator also includes a simplified and filtered shallow water dynamical model as temporal smoother and second-order div-curl spatial regularizer. The estimator may also incorporate correlation-based vector fields as additional observations. These correlation vectors are also routinely provided by the Eumetsat consortium. This code corresponds to research studies published in IEEE transaction on Geo-Science and Remote Sensing. It can be freely downloaded on the FLUID web site http://fluid.irisa.fr.

5.5. 3DLayeredMotion software - Estimation of 3D interconnected layered atmospheric motion fields

Participants: Patrick Heas, Etienne Mémin.

This software extends the previous 2D version. It allows (for the first time to our knowledge) the recovery of 3D wind fields from satellite image sequences. As with the previous techniques, the atmosphere is decomposed into a stack of pressure layers. The estimation relies also on pressure data and classification clouds maps and top of clouds pressure maps. In order to recover the 3D missing velocity information, physical knowledge on 3D mass exchanges between layers has been introduced in the data model. The corresponding data model appears to be a generalization of the previous data model constructed from a vertical integration of the continuity equation. This research study has been recently accepted for publication in IEEE trans. on Geo-Science and Remote Sensing. A detailed description of the technique can be found in an Inria research report. The binary of this code can be freely downloaded on the FLUID web site http://fluid.irisa.fr.

5.6. Low-Order-Motion - Estimation of low order representation of fluid motion

Participants: Anne Cuzol, Nicolas Gengembre, Etienne Mémin.

This code enables the estimation of a low order representation of a fluid motion field from two consecutive images. The fluid motion representation is obtained using a discretization of the vorticity and divergence maps through regularized Dirac measure. The irrotational and solenoidal components of the motion fields are expressed as linear combinations of basis functions obtained through the Biot-Savart law. The coefficient values and the basis function parameters are obtained as the minimizer of a functional relying on an intensity variation model obtained from an integrated version of the mass conservation principle of fluid mechanics. Different versions of this estimation are available. The code which includes a Matlab user interface can be downloaded on the FLUID web site http://fluid.irisa.fr. This program corresponds to a research study that has been published in the International Journal on computer Vision.

5.7. VTrack software - generic interactive visual tracking platform

Participants: Nicolas Gengembre, Patrick Pérez.

As part of a past research contract with FT-RD, we have developed an interactive tracking platform (Windows Visual C++ development with Microsoft MFC and Intel OpenCV). It includes both state-of-the-art generic tracking methods (template matching, feature tracking, kernel-based tracking with global color characterization, particle filtering) and original developments, as well as a number of visualization features for enhanced experimental and demonstration experiences. The flexible architecture and the rich HCI allow easy design, implementation and test of novel trackers.

5.8. ObjectDet software - learning and detection of visual object classes

Participant: Ivan Laptev.

ObjectDet is an open source efficient C++ implementation of object detection that extends our previous method [85]. The software achieves object detection at the approximate rate of 10 frames per second on 320×240 images on a modest PC. The accuracy of the method was ranked among the top ones in The PASCAL Visual Object Classes Challenges 2006 and 2007 (VOC2006, VOC2007). The detection is achieved with a "scanning window" classifier applied to different positions and scales of the image. The underlying AdaBoost classifier is trained from histogram features computed on rectangle-annotated object images. Variations in object views can be handled by training separate classifiers for different views of the object. Different types of histogram features including Histograms of Oriented Gradient (HOG), second-order derivative histograms and color histograms are implemented and can be used in a complementary way for increased performance.

Earlier version of the software with pre-trained classifiers is available for download from http://www.irisa.fr/vista/Equipe/People/Laptev/download.html. An updated release including the module for object training is planed to be made available before the end of 2007. Linux and Windows platforms are supported.

5.9. SAFIR-nD - image denoising software

Participants: Charles Kervrann, Jérôme Boulanger.

The SAFIR-nD software written in C++, JAVA and MATLAB, enables to remove additive Gaussian and non-Gaussian noise in a still 2D or 3D image or in a 2D or 3D image sequence (with no motion computation). The method is unsupervised. It is based on a pointwise selection of small image patches of fixed size in (a data-driven adapted) spatial or space-time neighborhood of each pixel (or voxel). The main idea is to associate with each pixel (or voxel) the weighted sum of intensities within an adaptive 2D or 3D (or 2D or 3D + time) neighborhood and to use image patches to take into account complex spatial interactions. The neighborhood size is selected at each spatial or space-time position according to a bias-variance criterion. The algorithm requires no tuning of control parameters (already calibrated with statistical arguments) and no library of image patches. The method has been applied to real noisy images (old photographs, JPEG-coded images, videos, ...) and is exploited in different biomedical application domains (fluorescence microscopy, video-microscopy, MRI imagery, X-ray imagery, ultrasound imagery, ...). This algorithm outperforms most of the best published denoising methods for still images or image sequences.

5.10. FAST-2D-SAFIR software - fast denoising of large 2D images

Participant: Charles Kervrann.

The FAST-2D-SAFIR software written in C++ enables to remove mixed Gaussian-Poisson noise in large 2D images, typically $10^3 \times 10^3$ pixels, in few seconds. The method is unsupervised and is a simplified version of the method related to the SAFIR-nD software. The method is based on a locally piecewise constant modeling of the image with an adaptive choice of a window around each pixel. The restoration technique associates with each pixel the weighted sum of data points within the window. The method has been applied to real microarray images routinely used in medical practices.

6. New Results

6.1. Image sequence processing and modeling

6.1.1. *Biological image sequence denoising*

Participants: Charles Kervrann, Patrick Bouthemy, Jérôme Boulanger.

New video-microscopy technology enables to acquire 4-D data that require the design and the development of specific image denoising methods able to preserve details and discontinuities in both the $(x - y - z)$ space dimensions and the time dimension. Images are noisy due to the weakness of the fluorescence signal in time-lapse recording. Accordingly, we have developed an original and efficient spatio-temporal filtering method for significantly increasing the signal-to-noise ratio (SNR) in noisy fluorescence microscopic image sequences where small particles have to be tracked from frame to frame. The proposed "adaptive window approach" is conceptually simple, being based on the local estimation of a weighted average of the intensities (for the considered regression model) within an adaptively and locally selected space-time window size (neighborhood). We use statistical 4-D data-driven criteria to automatically select the size of the adaptive space-time neighborhood. At each pixel, we estimate the weighted average by iteratively increasing a space-time window to achieve an optimal compromise between bias and variance corresponding to the minimization of the pointwise L_2 risk of the local estimator. The method involves also a patch-based similarity step to fix the weights. The proposed algorithm complexity is actually controlled by simply limiting the size of the largest window and the patch size.

In addition, theoretical properties of the non-parametric estimator have been proved. Recently, we have shown that the proposed estimation procedure can be interpreted as a steepest descent algorithm related to the fixed point solution corresponding to the minimization of a global energy function involving non-local terms and local image contexts described by patches. We have applied this method to noisy synthetic and real 4-D images where a large number of small fluorescently labeled vesicles move in regions close to the Golgi apparatus within the cell. Preliminarily, the assumed Poisson noise is transformed into a Gaussian noise using an original variance stabilization procedure based on a generalized Anscombe transform. The SNR is shown to be drastically improved and enhanced images can then be correctly segmented. The objective is to report evidences about the lifetime kinetics of specific Rabs for membrane transport. This novel approach can be further used for biological studies where dynamics have to be analyzed in molecular and subcellular bio-imaging. The patch-based method combined with Radon transform has been used to denoise images with a very low number of photon counts ($\approx$ 1 photon/pixel). Finally, let us also point out that we have applied our adaptive patch-based denoising method to usual video sequences, and it was demonstrated that it outperforms other recent methods.

6.1.2. *Non-parametric patch-based estimation vs. Bayesian non-local means filter for image representation and denoising*
Participants: Charles Kervrann, Jérôme Boulanger, Patrick Pérez.

Partial Differential equations (PDE), wavelets-based methods and neighborhood filters were proposed as locally adaptive machines for noise removal. Recently, Buades, Coll and Morel proposed the non-local means (NL-means) filter for image denoising. This method replaces a noisy pixel by the weighted average of other image pixels with weights reflecting the similarity between local neighborhoods of the pixel being processed and the other pixels. The NL-means filter was proposed as an intuitive neighborhood filter but theoretical connections to diffusion and non-parametric estimation approaches are also given by the authors. This year, we proposed another bridge, and showed that the NL-means filter also emerges from the Bayesian approach with new arguments. Based on this observation, we show how the performance of this filter can be significantly improved by introducing adaptive local dictionaries and a new statistical distance measure to compare patches. The new Bayesian NL-means filter is better parametrized and the amount of smoothing is directly provided by the noise variance (estimated from image data) given the patch size. We compared this method to the non-parametric patch-based method developed last year, also inspired from the NL-means filter. Experimental results are given for real images with artificial Gaussian noise added, and for images with real image-dependent noise (electronic microscopy, ultrasound imagery, ...). This modeling has been jointly investigated with the VisAGeS Team - U746. More recently, we have started to adapt this new formulation for change detection between image pairs.

6.1.3. *Network tomography for tracking in fluorescence microscopy imaging*
Participants: Thierry Pécot, Charles Kervrann, Patrick Bouthemy, Jérôme Boulanger.

GFP-tagging and time-lapse fluorescence microscopy can be considered as an investigation tool mainly used to observe molecular dynamics and interactions in live cells at both the microscopic and the nanoscopic scales. Hence, it is imperative to develop novel image analysis techniques that are necessary for the measurement of dynamics of biological processes observed in image sequences. This motivates our present research effort which is to develop novel techniques based on recent advances in Network Tomography (NT), a new field which we believe will benefit greatly from the wealth of statistical theory, to extract quantitative measurements from nD data. Indeed, object tracking using conventional techniques can be very hard or impossible in applications where more than one hundred objects interact. NT-based approaches, devoted to statistical traffic analysis, simplifies the tracking process because it only requires detection of an object as it moves from one region to another and avoids the difficult data association problem. In this approach, a dynamic scene formed by moving particles along a dense set of microtubules, is modeled as a network of interconnected regions of interest. In such a network, each node represents a sub-cellular location selected by a biologist or expert. A connection between two nodes is called a path and each path consists of one or more unidirectional or bidirectional links, that is physical links (microtubules) connected by intermediate routers. Broadly speaking,

network inference involves estimating network performance parameters based on traffic measurements at a limited subset of the nodes. In traffic intensity estimation, the measurements consist of counts of objects that pass through nodes in the network. Based on these measurements, the goal is to estimate how particles traffic originated from a source node to a destination node along a path which generally passes through several nodes. In this approach, it is not necessary to track the moving objects, we just need to determine when an object reaches a node, something that is generally easier than estimating a continuous trajectory. The measurements are usually the number of vesicles successfully detected at each destination region receiver or the path time of the vesicle between the source and each destination. The inherent randomness in both link-level and path-level measurements motivates the adoption of a statistical framework. Application of this method has been demonstrated with promising results for the Rab6 protein, a GTPase involved in the regulation of intracellular membrane trafficking, on real and artificial image sequences.

6.1.4. *Motion detection in adverse conditions*
Participants: Aurélie Bugeau, Patrick Pérez.

Detecting individual moving objects in videos that are shot by either still or mobile cameras is an old problem, which is routinely addressed in a number of real applications such as tele-surveillance. There are, however, a number of interesting instances of this motion analysis problem that are not satisfactorily handled by existing techniques. One class of such problems is the extraction of certain types of moving regions. In the context of activity analysis in dynamically cluttered environments for instance, the problem is the one of separating out foreground moving objects of interest from other uninteresting moving objects in the background. This might be addressed by characterizing the spatial and/or temporal content of surrounding clutter. It is an acute problem of this sort that we have faced in the Behaviour ACI project (see paragraph 8.2.2) where the detection and the tracking of driver's moving hands and face is corrupted by the exterior view through the side car window. Various approaches to this challenging problem can be considered. We have introduced the following new paradigm: sparse motion fields defined on a sub-sampled grid of points that don't belong to the dominant motion are first estimated with Lucas-Kanade technique robustified via a statistical p-test. These points are clustered in an unsupervised fashion based on multi-dimensional motion-color features using a novel non-parametric variable-bandwith density estimator that explicitly exploits the heterogeneity of the multi-dimensional input space (see 6.1.5). This method has demonstrated its potential on a number of complex dynamic scenes of various types (including real driver sequences obtained with handheld camera).

6.1.5. *Variable bandwidth kernel estimation and clustering in high dimensional feature spaces*
Participants: Aurélie Bugeau, Patrick Pérez.

Motivated by the problem of partitioning complex feature vectors (related to position, motion, color and texture) in the context of motion analysis, we addressed the generic issue of clustering data points in a high-dimensional heterogeneous input space. Starting from recent developments on kernel density estimation and iterative mode finding techniques for such non-parametric densities, we have investigated the critical issue of determining automatically the multi-dimensional kernels at work. To this end, we have proposed a new variable bandwidth approach based on the so-called "balloon" density estimator, and developed an associated mode finding iterative algorithm, with proved convergence. Experiments on the problem of space-range segmentation of static color images allowed us to validate this methodological development. The application of the automatic clustering method thus obtained to the problem of sparse motion-color clustering was then conducted, leading to results that are qualitatively excellent.

6.1.6. *Background modeling for video surveillance*
Participants: Kamel Aouichat, Patrick Pérez.

Despite the relative simplicity of typical surveillance setups (fixed camera observing a single scene over long periods), and the importance of research and developments conducted in the past years in this field, a number of problems remain largely open regarding the detection, identification and tracking of objects in such setups. In the context of several surveillance applications of interest for VideoMetrix (such as the robust tracking of cars in roundabouts for directional counting purposes), the Ph.D work of K. Aouichat aims at exploring

various problems. One is the dynamic modeling of background and foreground appearances at each pixel, as a strong cue for foreground object detection and tracking. It turns out that well established techniques, such as the celebrated one of Grimson and Stauffer based on joint modeling of both appearances with a single Gaussian mixture model, are rather disappointing in operational conditions. Based on this type of pixel-wise joint modeling of appearance, we are currently exploring various probabilistic extensions (coupling of GMM modeling and kernel density estimation, Bayesian formulation of the pixel-wise foreground detection).

6.1.7. *New theoretical results on mixed-state MRF models*

Participants: Patrick Bouthemy, Bruno Cernuschi-Frias, Jian-Feng Yao, Tomas Crivelli, Gwénaëlle Piriou.

Mixed-state models provide a generalization of existing statistical models applied in motion analysis dealing with random variables that take exclusively discrete or continuous values, to the case where both types of information is present and can be displayed by a motion measurement. In the last years of the research conducted in the context of the FIM project, Markov random fields with mixed states have shown to be a powerful non-linear representation of motion textures, with many applications in dynamic content recognition. Thus, a complete characterization and understanding of the theory of mixed-state models is crucial for the evolution of the research work. The equivalence between general Markov random fields and Gibbs distribution was exploited for obtaining new theoretical results. For general conditional models responding to a mixed-state probability density it was shown that the shape of the global energy for the Gibbs formulation, can be decomposed into one term accounting for the discrete part of the model, and a second term related to the continuous part. This decomposition theorem permits to define conditional mixed states models in a very simple way, and is a generalization to previous formulations and results of mixed-state auto-models, where some conditions and constraints were needed in order to know the shape of the field.

The problem of the partition function calculation in Gibbs distributions was also addressed obtaining some general results for its calculation, with direct application to dynamic content recognition (segmentation, detection, classification etc). These are not restricted to mixed-state models and it should provide an efficient method for dealing with this intricate and fundamental problem in the theory of Markov random fields. One of the premises of the proposed models is the ability of (motion texture) discrimination. Associated to this, the necessity of measuring similarity between mixed-states distributions, led to obtain new results for computing the Kullback-Leibler divergence between parametric statistical models. The possibility of obtaining this pseudo-distance is crucial in classification applications.

6.1.8. *Motion texture modeling and applications*

Participants: Patrick Bouthemy, Bruno Cernuschi-Frias, Tomas Crivelli, Jian-Feng Yao.

We have introduced new mixed-state models for the temporal modeling of motion textures. Now, we propose to describe a sequence of motion maps, defining local conditional interaction between motion random variables given at different instants, instead of the previous studied scheme, that was purely spatial. A mixed state Markov chain framework was defined, assuming causal dependence, as a natural extension to the time axis. The necessity of considering the time evolution and temporal properties of motion measurements is evident when we want to tackle applications like tracking motion textures, sequence reconstruction, prediction, detection and sequence segmentation. We have analyzed and compared the performance of this approach against spatial models in motion texture segmentation problems: temporal models are, usually, easier to handle, due to the property of causality. We have addressed the problem of motion texture classification. Based on real sequences obtained from the DynTex dynamic texture database, we obtained promising results of over 90% of classification rate for several different classes of motion textures and hundreds of samples. The process was based only on the parametric representation of motion textures, and a similarity measure between statistical models as explained before. No additional or complementary processing was done to improve performance. Consequently, these results have shown that the model is able to discriminate different dynamic phenomena, and we should be able to embed it in a more complex classification strategy in order to achieve greater classification rates.

The results obtained for motion texture modeling and mixed-state distribution were conducted in the context of Tomas Crivelli's Ph-D thesis within a "co-tutelle" program between University of Rennes 1 and UBA. Tomas Crivelli has spent a two-month stage in Rennes during May-June 2007.

6.1.9. *Joint background estimation and moving object detection*

Participants: Gwenaëlle Piriou, Patrick Bouthemy, Jian-Feng Yao.

We are investigating a new approach to tackle motion detection based on background subtraction. We aim at jointly estimating the static background and detecting the moving objects. To this end, we have defined conditional mixed-state models. The two targeted outputs are represented by the same mixed-state variable allowing a full expression of the different interactions and an efficient inference stage. More specifically, let us consider an image sequence where objects or people are moving through a static background. The goal is to obtain an accurate detection of these moving entities as well as a regularized view of the background. Usually, one applies different methods to solve these two problems in a sequential manner (one after another). Here, we propose a unified framework to deal with these two issues simultaneously. A mixed-state configuration space is introduced where at each pixel a purely symbolic value represents moving objects (their masks) while intensity values account for background recovering. Actually, this regularization is achieved dynamically over time by continuously updating the estimation of a "reference background". Spatial interactions are taken into account by a conditional MRF for the mixed-state values to be reconstructed. In particular, no *a priori* probabilistic model is assumed for the static background nor for the moving objects. Furthermore, the approach allows the observed data, namely intensity values as well as computed change measurements, to contribute to different energy terms of the conditional MRF in a very flexible way. Preliminary results have been obtained so far.

6.2. Motion estimation and matching

6.2.1. *Schlieren image velocimetry*

Participants: Guillermo Artana, Étienne Mémin.

We have addressed the problem of estimating the motion of fluid flows visualized with the Schlieren technique. Such an experimental visualization system, well known in fluid mechanics, enables the visualization of unseeded flows. It thus allows the capture of phenomena which are impossible to visualize with particle seeding such as natural convection, phonation flow, breath flow, as well as the visualization of large scale structures. Since the resulting images exhibit very low intensity contrasts, classical motion estimation methods based on the brightness constancy assumption (correlation-based approaches, optical flow methods) are inefficient. In order to extract motion fields from these specific images, we have introduced a new energy function composed of i) a specific data model accounting for the fact that the observed luminance is related to the gradient of the fluid density, and ii) a specific constrained div-curl regularization term. The minimization of this energy provides what we believe to be the only existing motion estimator that works properly on Schlieren images.

6.2.2. *Large scales fluid motion estimator*

Participant: Étienne Mémin.

During the PhD thesis of Anne Cuzol we have worked on the definition of a low-dimensional fluid motion estimator. This estimator is based on the Helmholtz decomposition which consists in representing the velocity field as the sum of a divergence-free component and a curl-free one. In order to provide a low-dimensional solution, both components have been approximated using a discretization of the vorticity (curl of the velocity vector) and divergence maps through regularized Dirac measures. The resulting so-called irrotational (resp. solenoidal) field is then represented by a linear combination of basis functions obtained by a convolution product of the Green kernel gradient and the vorticity map (resp. the divergence map). The coefficient values and the basis function parameters are obtained by minimizing a function formed by an integrated version of the mass conservation principle of fluid mechanics. This fluid motion estimation method has also been applied to medical imagery in order to estimate the growing of multiple sclerosis lesions. This last study has been conducted in cooperation with P. Hellier (Visages project-team). This work has been published in the International Journal on Computer Vision.

6.2.3. *Potential function estimation and finite mimetic differences*
Participant: Étienne Mémin.

In collaboration with Mannheim University (group of Professor Schnoerr) we have studied a direct estimation from two consecutive images of the potential functions associated to a fluid flow (respectively the *stream* function and the *velocity* potential). The estimation has been defined on the basis of a high order regularization scheme and has been implemented through mimetic difference methods. These approaches guaranty the discretization to preserve basic relationships of continuous vector analysis. The considered scheme appeared to be numerically much more stable and leads to an improve accuracy compared to previous discretization schemes based on auxiliary div-curl variables. This approach has been published in the Journal on Mathematical Imaging and Vision.

6.2.4. *2D and 3D atmospheric motion layer estimation*
Participants: Patrick Heas, Étienne Mémin, Nicolas Papadakis.

We have explored the problem of estimating mesoscale dynamics of atmospheric layers from satellite image sequences. Due to the intrinsic sparse 3-dimensional nature of clouds and to occluded areas between different cloud layers, the estimation of an accurate dense motion field is an intricate issue. Relying on a physically-sound vertical decomposition of the atmosphere into layers, we have proposed two dense motion estimators for the extraction of multi-layer horizontal (2D) and 3D wind fields. These estimators are expressed as the minimization of a global function including a data-driven term and a spatio-temporal smoothness term. A robust data term relying on shallow-water mass conservation model has been proposed to fit sparse observations related to each layer. In the 3D case, the layers are interconnected through a term modeling mass variations at the layers surfaces frontiers

A novel spatio-temporal regularizer derived from the shallow-water momentum conservation model has been considered to enforce temporal consistency of the solution along time. These constraints are combined with a robust second-order regularizer preserving divergent and vorticity structures of the flow. A two-level motion estimation scheme has been settled to overcome the limitations of the multiresolution incremental estimation scheme when capturing the dynamics of fine mesoscale structures. This alternative approach relies on the combination of correlation and optical-flow observations. An exhaustive evaluation of the novel method has been first performed on a scalar image sequence generated by Direct Numerical Simulation of a turbulent bi-dimensional flow. Based on qualitative experimental comparisons, the method has also been assessed on a Meteosat infrared image sequence. These pieces of work have been recently accepted in two distinct issues of IEEE Transactions on Geo-Science on Remote Sensing.

6.2.5. *Atmospheric motion estimation with pressure image assimilation*
Participants: Étienne Mémin, Nicolas Papadakis, Heas Patrick, Thomas Corpetti.

The complexity of dynamical laws governing 3D atmospheric flows associated to incomplete and noisy observations makes the recovery of atmospheric dynamics from satellite images sequences very difficult. We have faced the challenging problem of estimating physical sound and time-consistent horizontal motion fields at various atmospheric depths for a whole image sequence. Based on a vertical decomposition of the atmosphere, we have proposed two dense motion estimators relying on different multi-layer dynamical models. Both estimators use a framework derived from data assimilation and are applied on noisy and incomplete pressure difference observations derived from satellite images.

In the first model, dense pressure difference maps are reconstructed according to shallow-water model on each cloud layer. While performing this reconstruction, the variational process estimates the average horizontal wind fields of the multi-layer model. The second model relies on a simplified vorticity-divergence form of the previous multi-layer shallow-water model. In this case, average horizontal motion fields are estimated for each layer without reconstructing pressure maps. While the simplified model is not as precise as the exact shallow-water model, the latter estimator exploits image spatio-temporal structures and succeeds to characterize motion at finer spatial scales. The performance of both methods has been assessed on synthetic examples before their validation on real world meteorological satellite image sequences.

6.2.6. *Dynamically consistent motion estimation*
Participants: Étienne Mémin, Nicolas Papadakis, Thomas Corpetti.

We have concentrated on the dynamically consistent estimation of the motion fields over a sequence of images by explicitly imposing a dynamical law. This has been applied for both fluid images and usual videos sequences. Most of the techniques developed for fluid motions are limited to frame to frame estimation and do not use the underlying physical laws. Geophysical flows are quite well described by appropriate physical models. As a consequence in such contexts, physic-based approaches can be very powerful for analyzing incomplete and noisy image data, in comparison to standard statistical methods. The approach that we have developed exploits recipes related to optimal control theory that allows performing the estimation of an unknown state function according to a given dynamical model and to noisy and incomplete measurements.

The observations are measured on images either by a Lucas-Kanade approach (if the confidence is high on the dynamical law) or with a more sophisticated robust optical flow estimation. For fluid applications, our approach was based on the incompressible vorticity velocity formulation of the Navier-Stokes equation with an additive control variable, u that aims at representing deviations from the pure vorticity transport model. It allows us also dealing with compressible flows associated to low divergence value (intrinsically or at the observed scale).

For video sequences, no universal physical law can be stated for general videos showing moving objects of different natures. We have then assumed over a short range of time that the velocity is transported by itself up to a Gaussian discretization error.

This method has been validated on synthetic and real image sequences and we have proved that it allows us to cope with several delicate situations (such as the absence of data) which are not well managed with usual estimators.

6.2.7. *Motion layer extraction for video object removal in post-production*
Participants: Matthieu Fradet, Patrick Pérez.

Object removal in the context of post-production is especially challenging as very high quality is required. The first part of our study, conducted in the context of the Ph.D work of M. Fradet at Thomson R&D addresses rigid objects only. We propose to use dense motion estimator and to investigate a motion layer-based approach. This approach is characterized by a motion segmentation of the surrounding area of the "hole" (the selected region including the object to be removed). Each layer is described by a color model and a parametric motion model. Temporal constraints are used to reinforce temporal consistence. We propose to refine predicted motion boundaries by using a graph-cut based technique. For each extracted layer a set of compensated reference frames (also called mosaics) is generated by warping all the frames together with their motion parameters. Holes in frames in which the unwanted object is removed are filled by superposition of the synthesized reference layers, using the motions cumulated between the current frame and the reference ones. Our motion-based segmentation method is still limited by the fact that the motion layer repairing step requires accurate boundaries. In the next step, we plan to develop novel interactive methods, compliant with the post-production context, allowing the user to easily and quickly introduce useful side information.

6.2.8. *Motion-based intra-scene and inter-scenes video alignment*
Participants: Émilie Dexter, Patrick Pérez, Ivan Laptev.

Independent motion is a strong cue for detecting, segmenting and recognizing objects and activities in image sequences. Motion-based segmentation, however, is known to be a hard problem in scenes with motion parallax and in scenes with multiple moving objects. Non-parametric segmentation of independently moving objects using color-motion features is one of our research strands in this domain (see paragraph 6.1.4). Another strand concerns the exploitation of not only the presence but also the type of motion as an informative segmentation cue. We are investigating such cues in the particular context of retrieving dynamic repetitiveness across videos. One form of this problem is the alignment of video sequences with similar type of activities, using appropriate local motion descriptors and global geometric constraints. Another form, which we have started to study in the

context of the Ph.D. research of Émilie Dexter, is the problem of multi-camera video alignment. The ambition here is to address in its full generality the problem of aligning, both in space and time, multiple unsynchronized views of the *same dynamic scene* as provided by uncalibrated moving cameras. To this end, we have started to investigate the view invariance of various motion descriptors extracted from coarse optical flows (part-based histograms of motion orientations, of acceleration, of motion rotational and divergence). First experiments on Perception project-teams' multi-view data set for human motions have shown the relevance of this approach. A more systematic study, including experiments on more complex data (in particular with moving cameras) and the use of machine learning techniques to choose both final descriptors and associated similarity metrics, will constitute the next step. An alternative approach based on simple geometric relationships between different sets of points tracked independently in two different views (hence without inter-view matching), and factorization techniques, is also under study on motion capture data.

6.3. Tracking

6.3.1. *Stochastic filtering for fluid motion tracking*

Participants: Nicolas Gengembre, Étienne Mémin, Nicolas Papadakis.

We have proposed a recursive Bayesian filter for tracking velocity fields of fluid flows. The filter combines an Îto diffusion process associated to 2d vorticity-velocity formulation of Navier-Stokes equation and discrete image error reconstruction measurements. In contrast to usual filters designed for visual tracking problems, our filter combines a continuous law for the description of the vorticity evolution with discrete image measurements. We resort to a Monte-Carlo approximation based on particle filtering. The designed tracker provides a robust and consistent estimation of instantaneous motion fields along the whole image sequence. In order to handle a state space of reasonable dimension for the stochastic filtering problem, the motion field is represented as a combination of adapted basis functions. The basis functions are derived from a mollification of Bio-Savart integral and a discretization of the vorticity and divergence maps of the fluid vector field. The output of such a tracking is a set of motion fields along the whole time range of the image sequence. As the time discretization is much finer than the frame rate, the method provides consistent motion interpolation between consecutive frames. In order to reduce further the dimensionality of the associated state space when we are facing a large number of motion basis functions, we have explored a new dimensional reduction approach based on dynamical systems theory. The study of the stable and unstable directions of the continuous dynamics enables to construct an adaptive dimension reduction procedure. It consists in sampling only in the unstable directions, while the stable ones are treated deterministically.

When the likelihood of the measurement can be modeled as Gaussian law, we have also investigated the use of so-called ensemble Kalman filtering for fluid tracking problems. This kind of filters introduced for the analysis of geophysical fluids is based on the Kalman filter update equation. Nevertheless, unlike traditional Kalman filtering setting, the covariances of the estimation errors, required to compute the so called Kalman gain, relies on an ensemble of forecasts. Such a process gives rise to a Monte Carlo approximation for a family of non-linear stochastic filters enabling to handle state spaces of large dimension. We have recently proposed an extension of this technique that combines sequential importance sampling and the propagation law of ensemble Kalman filter. This technique leads to an ensemble Kalman filter with an improve efficiency. This strategy appears to be a generalization of the so called optimal importance sampling proposed within the PhD of Elise Arnaud in the context of partial conditional Gaussian trackers. The framework describing partial conditional Gaussian modeling for visual tracking has been published in the International Journal of Computer Vision. The bridge of this setup with ensemble Kalman filters is described in the PhD manuscript of Nicolas Papadakis.

The tracking experiments have been conducted on fluid experimental image sequences provided by the CEMAGREF and ONERA and also on satellite meteorological sequences supplied by the LMD (Laboratoire de Météorologie Dynamique) and the EUMETSAT consortium.

6.3.2. *Variational data assimilation for visual tracking*

Participants: Thomas Corpetti, Étienne Mémin, Nicolas Papadakis.

We studied a variational framework for the tracking of high dimensional features in image sequences. This framework relies on variational data assimilation principles as developed in environmental sciences to analyze geophysical flows. We have first devised a data assimilation technique for the tracking of closed curves and their associated motion fields. The proposed approach enables a continuous tracking along an image sequence of both a deformable curve and its associated velocity field. Such an approach has been formalized through the minimization of a global spatio-temporal continuous cost functional, with respect to a set of variables representing the curve and its related motion field. The resulting minimization sequence consists in a forward integration of an evolution law followed by a backward integration of an adjoint evolution model. The latter PDE includes a term related to the discrepancy between the state variables evolution law and discrete noisy measurements of the system. The closed curves are represented through implicit surface modeling, whereas the motion is described either by a vector field or through vorticity and divergence maps according to the type of targeted application. The efficiency of the approach has been demonstrated on two types of image sequences showing deformable objects and fluid motions.

We have investigated the use of this framework to realize a temporal Bayesian smoothing of fluid flow velocity fields. The velocity measurements are assumed to be supplied by an optical flow estimator or PIV velocity measurements. These noisy measurements are smoothed according to the vorticity-velocity formulation of Navier-Stokes equation. As previously, following optimal control recipes, the associated minimization is conducted through an iterative process involving a forward integration of our dynamical model followed by a backward integration of an adjoint evolution law. Both evolution laws are implemented with a second order non-oscillatory scheme. The approach has been validated on a synthetic sequence of turbulent 2d flow provided by Direct Numerical Simulation (dns) and on a real meteorological satellite image sequence depicting the evolution of a cyclone.

More recently, assimilation techniques for the direct estimation of fluid velocity fields from images have been devised. These techniques rely on a brightness variation model of the intensity function. They do not include anymore motion measurements provided by external motion estimators. The resulting estimator allows us to recover very accurate fluid motion fields and enables to track very accurately the vorticity map along an image sequence.

These works have been described in two papers published at ICCV'07.

6.3.3. *Low order dynamical system assimilation and POD representation*
Participants: Guillermo Artana, Etienne Mémin, Nicolas Papadakis.

We have proposed improvements to the construction of low order dynamical systems (lods) for incompressible turbulent flows. The reduced model is obtained by means of a Proper Orthogonal Decomposition (pod) basis extracted from experimental data. This decomposition is obtained from a truncated singular value decomposition of the fluid velocity fields auto-correlation matrix. The pod modes are then used to formulate an ordinary differential equations (ode) system or dynamical system which contains the main features of the flow. This is achieved by applying a Galerkin projection to the Navier-Stokes Equations. Usually, the obtained lods presents stability problems due to modes truncation and numerical uncertainties, especially when working on experimental data. We performed the model closure with a variational data assimilation technique. This technique allows us to correct the dynamic system coefficients as well as to identify and restore the noisy experimental data used to extract the pod basis. It is based on techniques from optimal control. In this case the initial condition and the temporal mode values constitute the control parameters. The measurements are given through the POD decomposition. This work has been published in the Journal of Turbulence.

6.3.4. *Stochastic filtering technique for the tracking of closed curves*
Participants: Christophe Avenel, Etienne Mémin, Patrick Pérez.

We aim at proposing a filtering methodology for the visual tracking of closed curves. Opposite to works of the literature related to this issue, we consider here a curve dynamical model based on a continuous time evolution law with different noise models. This led us to define three different stochastic differential equations that capture the uncertainty relative to curve motions. This new approach provides a natural understanding of

classical level-set dynamics in terms of such uncertainties. These evolution laws have been combined with various color and motion measurements to define probabilistic state space models whose associated Bayesian filters can be handled with particle filters. This on going work will be continued within extensive curve tracking experiments and extended to the tracking of other very high dimensional entities such as vector fields and surfaces.

6.3.5. *Robust visual tracking without prior*

Participants: Vijay Badrinarayanan, Patrick Pérez.

In this research we are interested in the problem of tracking arbitrary entities along videos of arbitrary type and quality. Such a tracking cannot rely, as classically done, on *a priori* information regarding both the appearance of the entities of interest (shape, texture, key views, etc.) and their visual motion (kinematical constraints, expected dynamics relative to the camera, etc.). The first crucial step is then the definition and the estimation of the reference appearance model on which the tracking, no matter its precise form, will rely on. Roughly, two extreme types of representations are routinely used in the literature: detailed pixel-wise appearance models subject to rapid fluctuations (e.g, intensity template instantaneously refreshed) and rough color models very persistent over time (e.g., color histogram instantiated at initialization time and kept unchanged). They are both interesting and complementary. For these reasons, it is appealing to fuse them within the same probabilistic tracker. In order to address this fusion problem in a principled way, we have investigated, in the context of a Cifre convention with Thomson-RD, a unifying information-theoretic approach to the problem. We have first proposed a unified computation of individual trackers entropies, whether posterior state distribution is approximated by samples (particle filter) or by grid discretization (correlation surface of standard feature trackers). This permits the self-assessment of each individual tracker and, in the case of a bundle of feature trackers, the design of an automatic mechanism for generating births and deaths of such trackers. The most recent step of this work has led to the definition of a probabilistic multi-cue tracker constructed by combining our randomized template tracker with a color-based particle filter. Our approach is based on deriving simple binary confidence measures for each tracker which aid priority based switching between the two fundamental cues for state estimation. Thereby the state of the object is estimated from one of the two distributions associated to the cues at each tracking step. This switching also brings about interaction between the cues at irregular intervals in the form of cross sampling. Within this scheme, we tackle the important aspect of dynamic target model adaptation under randomized template tracking which, by construction, possesses the ability to adapt to changing object appearances. Further, to track the object through occlusions we interrupt sequential resampling and achieve relock using the color cue. Quantitative comparisons on the CMU VIVID online evaluation system have demonstrated that our tracking system out-performs state of art trackers.

6.3.6. *Joint segmentation and tracking using graph cuts*

Participants: Aurelie Bugeau, Patrick Pérez.

The problem of jointly segment and track objects in videos remains largely open, despite a number of recent developments on image segmentation techniques (in particular level-set techniques that rely on continuous view of the image support and the graph-cut techniques that operate directly at the pixel level using combinatorial optimization). As described in 6.3.4, we have just started to explore the combination of level-set models with stochastic dynamic filtering. On the discrete side, we have proposed a novel graph-cut approach that permits to both track and segment multiple objects in videos using min-cut/max-flow optimizations. To this end we have introduced objective functions that combine low-level pixel-wise measures (color, motion), high-level observations obtained via an independent detection module, motion prediction and contrast-sensitive contextual regularization. The minimization of these cost functions simultaneously allows "detection-before-track" tracking (track-to-observation assignment and automatic initialization of new tracks) and segmentation of tracked objects, even in case of partial and total occlusions or of dropped frames. Experiments have been run on sequences from PETS 2006 corpus, where stillness of the camera allows background subtraction to provide the high-level observations (connected components of foreground detection masks). These experiments first demonstrate the ability of the method to detect, track and precisely segment persons and groups as they enter and traverse the field of view, possibly with occlusions. Additional

experiments also demonstrate that a second stage of minimization allows the segmentation of individual persons when spatial proximity makes them merge at the foreground detection level. Final experiments have exploited the output of the original motion detection and clustering technique described in 6.1.4 as external observations. They have provided excellent results on very complex dynamic scenes, including video of drivers.

6.4. Dynamic event modeling, learning and recognition

6.4.1. *Human action detection in video*

Participants: Ivan Laptev, Patrick Pérez.

Actions of people and their interactions with the environment are among the most informative events of feature films, TV programs, documentaries, personal video footage and so forth. The automatic interpretation of human actions is, hence, essential for many emerging applications such as video search in public and commercial video databases (YouTube, BBC Motion Gallery). The problem of human action recognition is complicated due to many factors including individual variations of people in expression, posture, motion and clothing; perspective effects and camera motions to mention just a few. In addition, the available research-oriented human action databases are few and limited in the content. As a consequence, recognition of human actions in the past has mostly been addressed in controlled settings or restricted scenarios.

In our ongoing work we address the challenging and not yet explored issue of recognizing and localizing human actions in *generic* videos such as movies. We currently consider short human actions with reasonably well-defined temporal structure existing for many action classes such as "drinking", "hand shaking" and "kissing" as well as for interactions of people with the environment: "entering a car"; "answering a phone"; "opening a door". To attack recognition of such actions in natural scenes, we particularly focus on two problems: (i) within-class action variability and (ii) action variation due to view changes. Both problems are treated in the framework of recent machine learning techniques in combination with efficient video descriptors.

More specifically, we have considered action classes "drinking" and "smoking" and manually collected a unique dataset with natural human actions in movies. On the training subset of annotated actions we have trained an AdaBoost space-time window classifiers using histogram video descriptors. A similar classifier has previously shown state-of-the-art performance for object detection in still images [85]. Here we have treated actions as space-time objects and extended [85] to action detection in video. We in particular investigated the influence of motion and shape descriptors on action recognition, we compared performance to several existing methods and we improved state-of-the-art in action detection considerably using introduced "Keyframe priming". We have evaluated action detection on the movie "Coffee and Cigarettes" (2003) and, hence, for the first time addressed the full range of problems associated with action recognition in movie scenario. Results of this work have been reported in an article at ICCV'07.

Human actions are usually not performed in isolation but in the context of other people, particular objects and scenes. To address and to benefit from these constraints in the future, we continued work on object detection. Specifically, we extended our previous method [85] by introducing histogram features in terms of color and second order derivatives. We have also trained and applied classifiers for frequent object combinations such as "person+X", X∈{"horse", "bicycle", "motorbike"}. The competitive results of this method have been demonstrated in the recent PASCAL Visual Object Classes Challenge 2007 (VOC2007) where our method received best rank in the person detection task.

Annotated video data with natural human actions is essential for the future progress in action recognition. Manual annotation of natural actions, however, is very time consuming and scales poorly to the large number of action classes. In the ongoing work we investigate methods that will provide automatic but noisy and incomplete action annotation. To make use of this data we plan development of weakly supervised action training methods in the future. Another direction for future research is explicit treatment of view invariance for action detection.

6.4.2. *Video content recognition using trajectories*

Participants: Patrick Bouthemy, Jean-Pierre Le Cadre, Alexandre Hervieu.

Content-based exploitation of video documents is of continuously increasing interest in numerous applications, e.g., for retrieving video sequences in huge TV archives, creating automatic video summarization of sports TV programs, or detecting specific actions or activities in video-surveillance. Considering 2D trajectories computed from image sequences is attractive since they capture elaborate space-time information on the viewed actions. Methods for tracking moving objects in an image sequence are now available to get reliable enough 2D trajectories in various situations. These trajectories are given as a set of consecutive positions in the image plane (x, y) over time. Our approach takes into account both the trajectory shape (geometrical information related to the type of motion) and the speed change of the moving object on its trajectory (dynamics-related information). Due to the trajectory features we have specified (local differential features combining curvature and motion magnitude), the designed method can be invariant to translation (namely, not affected by the location of the trajectory in the image plane), to rotation (motion direction in the image plane), to scaling (distance of the viewed action to the camera). We have tackled three important tasks related to dynamic video content understanding within the same trajectory-based framework. The first one is clustering trajectories extracted from videos. An unsupervised solution was developed. The second considered problem is recognizing (or retrieving) events in videos. Semantic classes of dynamic video contents are first learned from a set of representative training trajectories. The third task is detecting unexpected events by comparing the test trajectory to representative trajectories of known classes of events. Our method relies on the statistical HMM framework where the HMM states are given by quantized values of the considered trajectory features. All the involved parameters are properly estimated or specified. Appropriate similarity measures between HMMs are exploited (both to compare two trajectories or to evaluate the distance of a test trajectory to a given trajectory cluster). We have conducted an important set of comparative experiments both on synthetic examples and real videos (sports TV programs, Formula-1 race videos and ski videos) with classification ground truth. We have shown that our method supplies accurate results and offers better performance and usability than other approaches such as SVM classification, histogram comparison or LCSS distance.

6.4.3. *Extracting information from a video sensor network*
Participants: Jean-Pierre Le Cadre, Adrien Ickowicz.

Recent trends lead to consider *globally* networks of (video) sensors. These networks can be relatively large, so we have to face specific problems. How can we use the data at the sensor level, how to represent the information collected at the sensor level, how to fuse? A first step consists in extracting spatio-temporal informations from video sensors. Of course, these sensors are generally uncalibrated and asynchronous. So we have to consider rather rough informations. Roughness can go up to reduce the information to proximity and to a binary information about object motion (closing or not). A first step has been to consider the use of the estimated *cpa* times (*cpa*: closest point approach) for estimating the parameters of the target trajectories. This study is relatively simple but has the great advantage to put in evidence the basic requirements and the limits of this approach. In a second step, we considered the estimation of the *cpa* times from a sequence of images.
The limits of the above approach are quite evident. If it is able to exploit a temporal contrast, there is a strong need to use a spatio-temporal contrast at the (binary) sensor network level. Actually, it has been shown that the separation problem we have to solve present strong similarities with the optimization problems we have to solve in a SVM context. The benefits of this approach are multiple: it is well adapted to (robust) tracking and the combinatorial problems which plagued multitarget tracking are fundamentally reduced. For the tracking step, particle filtering is the natural way since they can easily include complex priors, non-linear measurements as well as separation properties, within a hierarchical context.

6.4.4. *On the effect of data contamination on track purity*
Participants: Jean-Pierre Le Cadre, Adrien Ickowicz.

A fundamental problem in multi-target tracking is to evaluate the performance of the association algorithms. However, it is quite obvious that tracking and association are completely entangled.

Effects of misassociation are considered in a simple (linear) multiscan framework so as to provide closed-form expressions of the probability of correct association. We focus on the development of explicit approximations of this probability. Via rigorous calculations the effect of dimensioning parameters (number of scans, false measurement positions or densities) is analyzed, for various modelings of the false measurements. Remarkably, it is possible to derive very simple expressions of the probability of correct association which are independent of the scenario kinematic parameters. Multiple extensions and applications render it quite attractive for a wide variety of contexts (close targets, clutter, intentionally generated false measurements, ECM, etc.).

6.4.5. *Planning surveillance*
Participants: Jean-Pierre Le Cadre, Cécile Simonin.

Many real-world applications require the optimization of hierarchical problems. This is especially true when resources are scarce and the space of search is large. In this problem the target is hidden, according to a prior probabilistic density. Various search means, called sensors, are available. They have different capacities of search and limited range. A way of considering this problem is to turn it into a hierarchical problem by splitting it into two interconnected optimization levels:

- a global level: find the best allotment of sensors to size-restricted search zones (a sensor is allotted to a unique zone)
- for every sensor, find the best resource sharing in order to have an optimal surveillance of the allotted zone

The space of search is thus divided in zones, in order to be able to explore efficiently a whole zone by means of a unique sensor. Each zone is partitioned into units; a unit is an area in which every points have the same properties, according to the difficulty of detection (altitude, vegetation, etc.) (*See Fig. 1.*). Each sensor has its own coefficient of visibility above a unit. The two-leveled hierarchical problem described above is easy to solve in the case where the allotment of sensors to zones is injective. Indeed, it can be solved by coupling together linear programming (LP) (for solving the discrete optimization problem) and continuous (convex) for the local level.

But in the general case, where the (1:1) hypothesis is abandoned, the complexity of the problem increases dramatically. In order to overcome this difficulty, the global level is optimized via a rare events simulation method, namely the Cross-Entropy (CE) algorithm. This approach has been extended to the multiperiod multizone and multisensor search for a Markovian moving target. The key is to use the two-level algorithm we describe above within a forward and backward framework. Despite the complexity of the problem, the whole algorithm performs quite satisfactorily and converges rapidly (only a few iterations).

Up to now, a common feature of search optimization methods was that they were devoted to the search for a unique target. Practically, this is far to be a realistic assumption. So, we have put important efforts on the search for multiple targets. The functional we use is the maximum of elementary rewards. So, we have to solve a minimax problem. The objective functional is still convex but is no longer differentiable everywhere. Moreover, the points where the functional is not differentiable are the natural "candidates". Generally, this problem can be solved via sub-gradient methods but it is far more efficient here to have recourse to duality. The algorithm is both feasible and optimal. Moreover, it has been successfully extended to the multitarget and multiperiod search.

6.4.6. *Path planning for navigation in a geographic information system*
Participant: Jean-Pierre Le Cadre.

The problem we considered is the optimization of the navigation of an intelligent mobile in a real world environment, described by a map. The map is composed of features representing natural landmarks in the environment. The vehicle is equipped with sensors which allows it to obtain landmark parameter estimates. These measurements are correlated with the map so as to estimate the mobile position. The optimal trajectory must be designed in order to control a measure of the performance for the filtering algorithm used for mobile navigation. As the mobile state and the measurements are random, a well-suited measure can be a functional of the Posterior Cramer-Rao Bound. In many applications, it is crucial to be able to estimate accurately the state of the mobile during the execution of the plan. So it seems necessary to couple the planning and the execution stages.

A classical tool is the constrained Markov Decision Process (MDP) framework. However, our optimality criterion is based on the Posterior Cramer-Rao bound, and the nature of the objective function for path planning makes it impossible to perform complete optimization within the MDP framework. Indeed, the reward in one stage of our MDP depends on all the history of the trajectory. To overcome this problem, the Cross-Entropy method, originally used for *rare-events* simulation, is a valuable tool. Its principle is to translate a "classical" optimization method into an *associated stochastic problem* and then to solve it adaptively as the simulation of rare events. This approach has been tested on various (simple) geographic environments and performs satisfactorily. This year, a large part of our efforts, conducted in the context of F. Celeste's PhD work, has been devoted to the derivation of closed-form approximations of the information we can gain from an elementary motion. Using them, it is possible to immerse the problem within an optimal control framework and to use efficiently the maximum principle.

7. Contracts and Grants with Industry

7.1. Cifre grant with Thomson-R&D on visual tracking

Participants: Patrick Pérez, Vijay Badrinarayanan.

no. Inria 2029, duration 36 months.

This contract started in March 2006 is associated with the supervision of V. Badrinarayanan's thesis funded by a Cifre grant. It concerns the problem of robust tracking of arbitrary objects in arbitrary videos. The first goal is the design of novel probabilistic ingredients to improve the robustness of existing tracking tools, with a first contribution on information-theoretic uncertainty assessment in probabilistic tracking as a generic tool for multiple cue fusion and intermittent adaptation. The second goal concerns the application, and possibly the specialization, of proposed generic techniques to tasks of interest to Thomson. Two scenarios are especially targeted: blurring of selected objects (typically faces) in TV news for business unit Thomson Grass Valley, and object colorization in film post-production for business unit Technicolor. In both cases, robust tracking tools (tracking a bounding box in the first case and the precise object outline in the second case) allowing the partial automatization of painstaking tasks are sought. Part of the work already conducted has given rise to the filing of the following European patent: "Method for tracking an object in a sequence of images and device implementing said method" by Vijay. B, P. Pérez, F. Le Clerc, L. Oisel.

7.2. Cifre grant with Thomson-R&D on video inpainting

Participants: Patrick Pérez, Matthieu Fradet.

no. Inria 2210, duration 36 months.

This contract started in March 2007 is associated with the supervision of M. Fradet's thesis funded by a Cifre grant. It concerns the problem of semi-automatic object removal for film and television post-production. The first goal is, for a given static mask to be filled-in, to design a local motion analysis approach (estimation and segmentation) that allows the interpolation of motion information within the region and the effective filling of the region based on current and surrounding frames as well as measured and interpolated motion information. The second step will aim at combining previous tool with tracking tools developed in 7.1 in order to allow the removal of a moving object selected by the user in one or several key frames. Two applicative scenarios are especially targeted: removing of logos in TV broadcasts for business unit Thomson Grass Valley, and object removal in film post-production for business unit Technicolor.

7.3. CEA-Cesta contract: Target tracking in a dense environment

Participants: Jean-Pierre Le Cadre, Adrien Ickowicz.

no. Inria 1542, duration 7 months.

The aim of the multitarget tracking is to associate elementary measurements corresponding to feasible trajectories. This association step is made jointly with a tracking step and both are completely entangled. This means that this problem is largely different from classical target tracking. There is fundamentally uncertainty about the origin of the measurements. To solve such problems, a wide variety of methods are available. Roughly, they can be divided in two categories: the probabilistic methods (e.g., JPDAF, PMHT), and combinatorial ones. However, a major problem remains for initializing multitarget algorithms. While integer programming or flow approaches have been developed for solving it rigorously, a basic tool is to limit the arborescence complexity via merging and pruning (the MHT). In contrast to the "elementary" target tracking framework, there is a strong need for defining convenient tool for the performance of data association. The probability of correct association and the track purity index are sensible tools. In this context, we have shown that a linear regression framework allows us to conduct explicit calculations. More precisely, the probability of correct association has been derived as an explicit function of the scenario parameters: scan number, mean track distance, measurement variance, probability of detection, etc. By this way, it is possible to derive a measure of track-to-track interaction. In a dense target environment, the problem becomes still more complicated since we have to investigate the effects of permutations.

7.4. DGA (REI)- contract: dynamic scene analysis with camera networks

Participants: Jean-Pierre Le Cadre, Adrien Ickowicz, Patrick Bouthemy.

no. Inria 2338, duration: 30 months.

This contract deals with surveillance of large zones via a network of video sensors. Of course, sensor outputs can be treated in a centralized architecture. However, centralized architectures suffer from serious drawbacks. Communication constraints (e.g. bandwidth) are frequently evoked, but still more fundamentally we have to face many problems inherent of this architecture, like:

- Sensor calibration, positioning and synchronisation.
- False alarms, multiple objects, occlusions, etc.

Overall, there is strong need for extracting a global picture at the network level. This means that we have to focus on the level of information we can extract at the sensor level and how to fuse them. Work has been done on the first point, using both simulated and real video sequences. The less informative level of information is the binary one. However, there is a fundamental difference between a $\{0, 1\}$ information and $\{-, +\}$ information. A "general" $\{0, 1\}$ information corresponds to a detection / non-detection information. Such architecture has been widely studied in a distributed detection framework, but is not well suited to our context. However $\{0, 1\}$ information is especially interesting if the detection process includes geographic constraints like proximity, field-of-view, etc. The $\{-, +\}$ information corresponds to a motion information: the object gets closer or is going far away. At the network level, this is a very rich information which can present definite advantages (robustness, multi-target tracking). However, its interest depends on the network

density. So, it is also necessary to consider various and complementary decentralized architectures according to the sensing capabilities, the target behaviors and, overall, the combinatorial complexity of the problem. Work has been done for defining processings and architectures adapted to this context.

An internship work (André Silva de Oliveira) has been devoted to the processing of image sequences for extracting the binary information. After considering the divergence of a local motion model, its estimation and use on real data, we turned toward a temporal analysis of the bearing information. More precisely, bearing rates and bearing rate changes give us an estimate of the (local) target behavior. It is thus possible to derive a local estimate of the ratio $\frac{\dot{r}}{r}$, through purely passive measurements, at the sensor level. Although this analysis is purely local, it performs satisfactorily on simulated sequences. Its performance has been investigated and the method appears relatively robust to an imperfect knowledge of the focal length.

7.5. Grant from Dupont De Nemours

Participant: Etienne Mémin.

University of Rennes contract, duration 24 months

In 2006, Dupont De Nemours company has provided an excellency unrestricted grant to E. Mémin to support his activities in the field of "Developing computational and visualization capabilities to extract object motion fields and fluid flow fields from high-speed imaging".

7.6. European initiatives

7.6.1. FP6-IST FLUID

Participants: Nicolas Gengembre, Patrick Heas, Étienne Mémin, Nicolas Papadakis.

no. Inria 737, duration 36 months

The FLUID project is a FP6 STREPS project labeled in the Future and Emerging Technologies Open scheme program. The goal of this selective Information Society Technologies program of basic researches is to enable a range of ideas for future and emerging technologies to be explored and realised.

The fluid project has started in November 2004. E. Mémin is the scientific coordinator of the project. This 3-year project aims at studying and developing new methods for the estimation, the analysis and the description of complex fluid flows from image sequences. The consortium is composed of five academic partners (Inria, Cemagref, University of Mannheim, University of Las Palmas de Gran Canaria and the LMD, "Laboratoire de Météorologie Dynamique") and one industrial partner (La Vision company) specialized in PIV (Particle Image Velocimetry) system. The project gathers computer vision scientists, fluid mechanicians and meteorologists. The first objective of the project consists in studying novel and efficient methods to estimate and analyze fluid motions from image sequences. The second objective is to guarantee the applicability of the developed techniques to a large range of experimental fluid visualization applications. To that end, two specific areas are considered: meteorological applications and experimental fluid mechanics for industrial evaluation and control. From the application point of view, the project particularly focus on 2D and 3D wind field estimation, and on 2D and 3D particle image velocimetry. A reliable structured description of the computed fluid flow velocity field will further allow us to address the tracking of turbulent structures in the flows.

During the third year of this project we have pursue the effort on the design of methodologies that aim at coupling fluid dynamical model and image data. These couplings have been settled either within a stochastic filtering framework or either within a variational assimilation framework.

Concerning the stochastic filtering, we have defined a filter that allows us to built a bridge between two different sequential Monte-Carlo techniques Ensemble Kalman Filter routinely used in environmental sciences and the particle filter used in signal and image processing. A very promising technique leading to a gain of robustness for a cheaper computational cost has emerged. This technique has been assessed on 2D turbulent flows.

As for the variational assimilation framework, we have proposed several schemes that enables the estimation of coherent motion fields with respect to a given dynamical law. This framework allows the tracking from image sequences of features leaving in space of infinite dimension such as curves and motion.

We have also applied this framework to the estimation of low order dynamical systems. Such a system is obtained through a Galerkin projection of the Navier-Stokes equation on a modal representation of the flow under concern using a truncated singular value decomposition of the autocorrelation matrix of experimental velocity measurements (Proper Orthogonal Decomposition). The proposed estimation relies on an optimal control of the initial condition and the coefficient of the unknown dynamical system. This new method appears to be very stable and allows the estimation of a larger number of modes to represent the flow.

In a meteorological context we have investigated the particular issue of wind fields estimation for a stratified atmosphere into layers. We have also proposed for the very first time solutions to estimate the vertical component of wind fields from 2D satellite pressure images organized in successive layers.

Cooperations with the Cemagref Rennes and the LMD (Laboratoire de Météorologie Dynamique) have enabled us to assess the relevance of the proposed methods either in the context of experimental fluid mechanics or for meteorological applications.

7.6.2. *FP6 PEGASE: helicoPter and aEronef naviGation Airborne System Experimentation*

Participants: Patrick Bouthemy, Jean-Pierre Le Cadre.

no. Inria 1832, duration 36 months

PEGASE is a multisciplinary European project involving 15 partners (industrial and academic) gathering the major actors of the domain. It is headed by Dassault Aviation. The kick-off-meeting was held in October 2006. For civil aviation, it is widely recognized that approaches, landings and take-offs, or more generally, maneuvers or navigation in the terminal zone, are among the most critical tasks in aircraft operation. PEGASE is a feasibility study of a new navigation system which should allow a three-dimensional truly autonomous approach and guidance for airports and helipads and improves the integrity and accuracy of GNSS differential navigation systems. The purpose of the PEGASE project is to prepare the development of an autonomous, all weather conditions, localization and guidance system based upon correlation between vision sensors output and a ground reference database. The work package WP6 will be led by Inria (Lagadic project-team) and is gathering academic labs (Inria - project-teams Icare, Lagadic and Vista -, CNRS, EPFL, ETHZ, ITJSI) and industrial partners (Dassav, EADS, Euroimage,...). Its aim is to develop new methods in image processing, visual tracking and visual servoing to implement the functionalities required in the PEGASE navid system, in connection with other WPs. Concurrent image processing methods will be implemented and tested on real image sequences, and synthesized image sequences. Its kick-off-meeting has been held in Turin (March 2007) and an important work has been done for defining flight models, real image sequences, etc. The Vista team is more specifically involved in two tasks: the tracking of points of interest on the first hand and the aircraft positioning on the second one. For the second task, inputs are the tracked points of interest and the flight model is used to have an accurate estimate of the aircraft trajectory.

7.6.3. *FP6-IST NEO Muscle*

Participants: Ivan Laptev, Patrick Bouthemy, Alexandre Hervieu.

no. Inria 104A04950, duration 48 months

The Vista team is involved in the FP6 Network of Excellence MUSCLE ("Multimedia Understanding through Semantics, Computation and Learning") started in April 2004. It gathers over forty research groups all over Europe from public institutes, universities or research labs of companies. Due to the convergence of several strands of scientific and technological progress, one is witnessing the emergence of unprecedented opportunities for the creation of a knowledge driven society. Indeed, databases are accruing large amounts of complex multimedia documents, networks allow fast and almost ubiquitous access to an abundance of resources and processors have the computational power to perform sophisticated and demanding algorithms. However, progress is hampered by the sheer amount and diversity of the available data. As a consequence,

access can only be efficient if based directly on content and semantics, the extraction and indexing of which is only feasible if achieved automatically. MUSCLE aims at creating and supporting a pan-European Network of Excellence to foster close collaboration between research groups in multimedia datamining on one hand and machine learning on the other hand, in order to make breakthrough progress toward different objectives.

Vista was a part of the MUSCLE showcase project "Content-Based Copy Detection for Videos and Still Images" and provided implementation of space-time interest point detection for the video copy detection demonstrator. I. Laptev collaborated with INRIA-Imedia on the evaluation of methods for video copy detection and published a joint paper. I. Laptev took part in the MUSCLE "Visual Saliency" e-team collaboration and visited KTH for one week in Jan. 2007 where he worked on object recognition. P. Bouthemy and I. Laptev organized a special session devoted to MUSCLE research at the International Workshop on Content-based Multimedia Indexing (CBMI'2006) held in Bordeaux, France, June 2007. Vista contributed to several WP reports.

7.6.4. FP6-IST TN Visiontrain
Participants: Patrick Bouthemy, Alexandre Hervieu.

no. Inria 850, duration 48 months

Visiontrain is a Marie Curie Research Training Network (belonging to the Computational and Cognitive Vision Systems chapter) which started in May 2005. Visiontrain addresses the problem of understanding vision from both computational and cognitive points of view. The research approach will be based on formal mathematical models and on the thorough experimental validation of these models. In order to achieve these ambitious goals, 11 academic partners plan to work cooperatively on a number of targeted research objectives: *(i)* computational theories and methods for low-level vision, *(ii)* motion understanding from image sequences, *(iii)* learning and recognition of shapes, objects, and categories, *(iv)* cognitive modeling of the action of seeing, and *(v)* functional imaging for observing and modeling brain activity. A. Hervieu and P. Bouthemy participated to the Visiontrain meeting hold in Utrecht in May 2007 where A. Hervieu gave a talk on "Trajectory-based video event recognition". A. Hervieu and T. Pecot attended the second Visiontrain one-week thematic (winter) school held at Les Houches Physics School in March 2007 and devoted to "Computational and Neurophysiological Models for Visual Perception".

8. Other Grants and Activities

8.1. Regional initiatives

8.1.1. PRIR Britanny council grant : Imafluid
Participants: Etienne Mémin, Patrick Heas, Nicolas Papadakis.

no. Inria 103C18930, duration 36 months.

This project granted by the Brittany council aims within a collaboration with the CEMAGREF Rennes at developing new methods for the estimation of dense motion fields of fluid flows. The purpose of this project is also to assess the accuracy of several estimation schemes on several known typical experimental flows observed through different image modalities. In this context we have worked, in collaboration with A. Cuzol, on methods allowing an effective collaboration of correlation techniques (PIV methods) and variational dense motion estimators. We investigate also the use of dynamical fluid priors to enforce a temporal consistency along time of velocity estimates. As a last issue we are studying how to incorporate within the estimation scheme the effect of small scales of the flow corresponding to unobservable sub-grid spatial resolution.

8.2. National initiatives

8.2.1. ANR STANDS-MSG
Participants: Thomas Corpetti, Etienne Mémin.

duration 36 months.

This ANR project entitled "Spatio-temporal Analysis of deformable structures in Meteosat Second Generation images" aims at developing methods for the analysis of deformable structures in meteorological images. More precisely, within this project we will focus on two meteorological phenomenon: the convective cells and sea breeze circulation. The first type of cloud system is responsible of dangerous meteorological events such as strong showers. Their monitoring is thus very important. See breezes influence deeply the climate of coastal regions. The comprehension of the daily and seasonal evolution of see breeze fronts is of great importance for local weather forecasting. The goal of this project will be to propose tools based on appropriate physical evolution laws for the tracking and analysis of these events. This project involves computer vision scientists from different groups, climatologists and meteorologists.

8.2.2. ACI Ministry grant: Behaviour project

Participants: Patrick Pérez, Aurélie Bugeau.

no. Inria 104C08130, duration 36 months.

The Behaviour project was granted in October 2004 by the collaborative ACI program on Security and Computer Science. It involves Compiègne University of Technology (Heudiasyc lab) as the prime, along with PSA-Peugeot-Citroën (Innovation and Quality group) and Vista. The main applicative goal is visual monitoring of car drivers, based of videos shot inside the car, such that hypo-vigilant behaviors (mainly drowsiness and distraction) can be detected. To this end, the project aims at providing new tools to perform automatically the recognition of a wide range of elementary behavioral items such as blinks and eye direction, yawn, nape of the neck, posture, head pose, interaction between face and hands, facial actions and expressions, control of the car radio, or mobile phone handling. Before trying to achieve such fine grain activity recognition, one has to select and extract relevant spatio-temporal features to apply subsequent learning on. While UTC is focusing on robust extraction and tracking of facial features in frontal views (shot through the wheel), we are attacking the complementary problem of detection and tracking of mobile items (especially head and hands) in arbitrary driver views. Although the problem seems classic, the specificity of videos under concern makes it very difficult (drastic changes of appearance and prolonged occlusions; low contrast of sequences shot at night; presence of very complex dynamic visual content through window in daylight). In this context, new motion detection, tracking and matching techniques have been studied last year. Further investigation of the first item (detection) has been conducted this year, with novel non-parametric tools for extracting interesting motion regions within highly complex dynamical contents (see paragraph 6.1.4). The last contribution consists in the coupling of this first grid-based extraction step with state-of-art graph-cut techniques for more complete pixel-based extraction of moving regions of interest, while avoiding corruption by outside elements seen through the window (see 6.3.6).

8.2.3. ACI Ministry grant: project related to the development of new fluorescent probes
Participant: Charles Kervrann.

CNRS *contract, duration 36 months.*

This project, labeled within the DRAB ACI program, was contracted in October 2004. It involves two other teams: UMR-CNRS 6026 ("Interactions Cellulaires et Moléculaires" Laboratory - "Structure et Dynamique des Macromolécules" team, University of Rennes 1) and UMR-CNRS 6510 ("Synthèse et Électrosynthèse Organiques" Laboratory - "Photonique Moléculaire" team, University of Rennes 1). The project aims at characterizing the +TIPs (plus-en tracking proteins) at the extremities "+" of microtubules and their dynamics using new fluorescent probes (Quantum Dots). New image analysis methods are developed for tracking fluorescent molecules linked to microtubules. We have focused on particles detection in images corrupted by Poisson noise.

8.2.4. ARC INRIA: DynaMIT project
Participant: Charles Kervrann.

no. Inria 2075, duration 24 months.

This project, labeled within the ARC Inria program, was contracted in January 2006. The Vista team is the prime contractor of the project DYNAMIT which associates the following other groups: MIA (Mathématiques et Informatique Appliquées) Unit from Inra Jouy-en-Josas, Curie Institute ("Compartimentation et Dynamique Cellulaires" Laboratory, UMR CNRS-144 located in Paris) and EPFL (Ecole Polytechnique Fédérale de Lausanne, "Biomedical Imaging Group") . In this project, we develop new methods dedicated to the analysis of nD microscopy data and to the modeling of molecular and macromolecular mechanisms at the cell level. Our main objective is then to provide computational methods and mathematical models to automatically extract, organize and model dynamic information observed in temporal series of images in multi-dimensional (nD) microscopy. The central problem addressed by this project concerns the roles played by different molecular motors in Rab6 dynamics and a rich set of data (mostly image sequences in video-microscopy) will support the analysis. The data to be considered are twofold: a/ data related to a wild type golden standard (no motor inactivated); b/ data related to perturbed situations (at least one motor inactivated). Taking into account the possible dependences between motors, an experimental design will result in many sets of dynamic acquisitions of Rab6A/A' traffic. Moreover, data in relation with the cytoskeleton will also be considered.

8.2.5. STIC-AMSUD: Modeling voice production

Participant: Etienne Mémin.

duration 36 months. This project aims at studying both, theoretically and experimentally, the phenomena involved in vocal fold oscillations, towards the elaboration of performing vocal production models, for application in voice and speech technology. It gathers partners from Brazil (PUCRS, UFF), Argentina (UBA) and France (LIMSI,ICP, IRISA).

8.2.6. Other involvements

- The Vista team is involved in the French network GDR ISIS, "Information, Signal and Images".

- C. Kervrann participates in the network GDR 2588, "Microscopie Fonctionnelle du Vivant".

- **Collaboration with Cesta, Bordeaux**
  Participant: Jean-Pierre Le Cadre.

  Target acquisition is a common problem for narrow-beam tracking radars. During the target acquisition stage, the radar must operate in a search mode over a limited volume of space. This limited volume corresponds to the prior uncertainty on the target location. Typically, a cued electronic beam scanning radar must seek the target in a 3-dimensional growing error basket. Therefore, the radar needs to fix a sequence of pulses or looks in successive appropriate directions. This sequence, determined over a fixed temporal horizon, should optimize the chances to detect the moving target, once or more times. There are classic acquisition search patterns for agile beam radars, such as rectangular raster scans, fence or ellipsoidal search patterns, which can be dedicated to various operational configurations. However, these semi-empirical patterns do not necessarily provide the best search. Other patterns could offer a higher probability of detection of the target or could require less resource or energy. The only way to determine a search pattern is to study a case where one must allocate integer search efforts into the cells. Then, the search consists of successive cell search moves which depend on the target probability of presence. Generally, B&B (Branch and Bound) methods are well-known exact optimization methods that consist in enumerating cleverly the solution space. Also called implicit enumeration methods, they aim at dividing the solution space in smaller and smaller subsets, most of them being eliminated by bound calculus before being constructed explicitly. In Hohzaki and Iida work, the B&B approach was developed above all in the conditionally deterministic target dynamic case, i.e., when the target dynamic is deterministic given the initial target dynamic state value, i.e., its position, speed, etc. We applied the Hohzaki B&B framework to the target acquisition search pattern issue. It was illustrated by the acquisition of a ballistic target by a narrow-beam sensor. The main assumption is effectively checked: the target dynamic is conditionally (to its ballistic coefficient) deterministic. In this way, optimized search patterns have been obtained for ballistic target acquisition.

8.3. Bilateral international co-operation

8.3.1. Collaboration with University of Buenos-Aires, Inria associate team FIM

Participants: Guillermo Artana, Bruno Cernuschi-Frias, Tomas Crivelli, Patrick Bouthemy, Étienne Mémin, Jian-Feng Yao, Patrick Heas, Nicolas Papadakis.

The Inria associate team FIM ("Fluidos e Imágenes de Moviemento") is concerned with the analysis of fluid flow from image sequences. It was created in December 2004. This long-term and intensive cooperation involves two groups from the Engineering Faculty of the University of Buenos-Aires: the Signal processing group headed by Professor Bruno Cernuschi-Friàs and the Fluid Mechanics group headed by Professor Guillermo Artana. Two main themes are investigated. The first one deals with experimental visualization and embeds modeling, motion measurement and analysis of fluid flows. The second one is concerned with the modeling, segmentation and recognition of dynamic textures in videos of natural fluid scenes (sea-waves, rivers, smoke, moving foliage, etc...).

Concerning the first topic, we have continued our work on motion estimation in image sequences supplied by a Schlieren device. This device allows visualizing the density variation of fluid flows through the changes of refraction index of a light beam. This technique enables to image unseeded flows and is therefore suitable to analyze large-scale experiments or flows that are difficult to visualize with particles such as breath flows or natural convection. We have extended the imagery device. We are now able to visualize flows experiments in a square region of one meter per one meter. This has to be compared to typical PIV analysis windows that are at most of size 15 cm by 15 cm. During the visit of Patrick Heas we have improved the Schlieren image velocimetry software developed the two previous years. It includes now a collaborative scheme allowing incorporating correlation based measurement and a spatio-temporal smoothing function based on the vorticity transport equation. These two ingredients allow us to cope with large displacements and also to enforce a temporal consistency of the solution. Patrick Heas has also started to work with Guillermo Artana and Etienne Mémin on the design of physically grounded smoothness functional for the estimation of fluid flows motion fields. This approach that is currently under development defines the smoothing function from the invariant of the velocity gradient tensor. Such approach should enable in the one hand estimating fluid motion fields with improve accuracy (especially in the small scales of the motion) and in the other hand to characterize directly from images of the flow turbulent regions such as vortex tubes, areas of pure straining, or vortex sheets.

Guillermo Artana, Juan D'Adamo, Etienne Mémin and Nicolas Papadakis have improved the variational assimilation techniques for the estimation of Low Order Dynamical System from image sequence. Such a system is obtained through a Galerkin projection of the Navier-Stokes equation on a modal representation of the flow under concern using a proper orthogonal decomposition. The assimilation method is not anymore based on two successive steps. It consists now of a unique process where both the initial condition and the coefficient of the unknown dynamical system are estimated. This new method is much more stable and allows the estimation of a larger number of modes.

Patrick Bouthemy, Bruno Cernuschi-Frias, Tomas Crivelli, and Jian-Feng Yao have continued the study of the so called mixed-state models in the context of the FIM project, for the modeling of dynamic textures in videos of natural scenes (such as views of rivers, sea-waves, moving foliage, fire, steam, smoke). Such models were introduced in the context of image motion analysis and are useful to represent information that can take both discrete values accounting for symbolic states, and real values corresponding to continuous measurements. Several new results where obtained regarding the theoretical and practical applications of the model.

Theoretical results: Mixed-state models provide a generalization of existing statistical models applied in motion analysis dealing with random variables that take exclusively discrete or continuous values, to the case where both types of information are present and can be displayed by a motion measurement. In the last years of the research conducted in the context of the FIM project, Markov random fields with mixed states have shown to be a powerful non-linear representation of motion textures, with many applications in dynamic content recognition. Thus, a complete characterization and understanding of the theory of mixed-state models is crucial for the evolution of the research work. The equivalence between general Markov random fields and Gibbs distribution was exploited for obtaining new theoretical results. For general conditional models

responding to a mixed-state probability density it was shown that the shape of the global energy for the Gibbs formulation, can be decomposed into one term accounting for the discrete part of the model, and a second term related to the continuous part. This decomposition theorem permits to define conditional mixed states models in a very simple way, and is a generalization to previous formulations and results of mixed-state auto-models, where some conditions and constraints were needed in order to know the shape of the field. The problem of the partition function calculation in Gibbs distributions was also addressed obtaining some general results for its calculation, with direct application to dynamic content recognition (segmentation, detection, classification etc). These are not restricted to mixed-state models and it should provide an efficient method for dealing with this intricate and fundamental problem in the theory of Markov random fields. One of the premises of the proposed models is the ability of (motion texture) discrimination. Associated to this, the necessity of measuring similarity between mixed-states distributions, led to obtain new results for computing the Kullback-Leibler divergence between parametric statistical models. The possibility of obtaining this pseudo-distance is crucial in classification applications.

Modeling: we have introduced new mixed-state models for the temporal modeling of motion textures. Now, we propose to describe a sequence of motion maps, defining local conditional interaction between motion random variables given at different instants, instead of the previous studied scheme, that was purely spatial. A mixed state Markov chain framework was defined, assuming causal dependence, as a natural extension to the time axis. The necessity of considering the time evolution and temporal properties of motion measurements is evident when we want to tackle applications like tracking motion textures, sequence reconstruction, prediction, detection and sequence segmentation. We have analyzed and compared the performance of this approach against spatial models in motion texture segmentation problems: temporal models are, usually, easier to handle, due to the property of causality.

Applications: We have addressed the problem of motion texture classification. Based on real sequences obtained from the DynTex dynamic texture database, we obtained promising results of over 90% of classification rate for several different classes of motion textures and hundreds of samples. The process was based only on the parametric representation of motion textures, and a similarity measure between statistical models as explained before. No additional or complementary processing was done to improve performance. Consequently, these results have shown that the model is able to discriminate different dynamic phenomena, and we should be able to embed it in a more complex classification strategy in order to achieve better classification rates.

The results obtained for motion texture modeling and mixed-state distribution were conducted in the context of Tomas Crivelli's Ph-D thesis within a "co-tutelle" program between University of Rennes 1 and UBA. Tomas Crivelli has spent a two-month stage in Rennes during May-June 2007 and he will spent another two-month stay at the end of the year.

8.3.2. *Advanced Research Program (PRA) France-China*

Participants: Jian-Feng Yao, Patrick Bouthemy.

This collaboration with the research group headed by Dr. Véronique Prinet at LIAMA (Sino-French Laboratory for Computer Sciences, Automation and Applied Mathematics, Beijing) is founded jointly by the French Ministry of Foreign Affairs and the Chinese Ministry of Science and Technology. It started in June 2006. It also involves the Ariana project-team (Inria Sophia-Antipolis, X. Descombes). The main objective of the collaborative research program is to build efficient Markov models and algorithms for modeling transformations of geometric structures in satellite image sequences. Potential applications include inspections of urban area or forest modifications from satellite images. A master thesis, entitled "Structural change detection on urban areas from high-resolution satellite images" was conducted during the period March-August 2007. C. Cassia, a Ph. D student of V. Prinet, spent one month in Rennes in November 2007.

8.3.3. *Visiting scientists*

- Bruno Cernuschi-Frias (Prof. University of Buenos-Aires) spent two months in our team in the context of the Inria Associate team FIM.

9. Dissemination

9.1. Leadership within scientific community

- *Editorial boards of journals*
  - J.-P. Le Cadre is Area Editor of Journal of Advances in Information Fusion (ISIF);
  - P. Pérez is Associate Editor for the IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI).

- *Conference organization*
  - P. Bouthemy and I. Laptev were invited to organize a special session with contributions from MUSCLE NoE at the International Workshop on Content-based Multimedia Indexing (CBMI'2006) held in Bordeaux, France, June 2007.
  - E. Mémin has organized with J.P. Bonnet (LEA Poitier), C. Schnoerr (U. Mannheim) and C. Troppea (T.U. Darmstadt) a prospective seminar in Dagstuhl, Germany entitled "Experimental fluid mechanics, computer vision and pattern recognition Pattern recognition and Fluid mechanics: a vision of the future". This seminar has gathered numerous European groups working on the analysis, visualization and simulation of fluid flows. This seminar has been a great success. It has been the opportunity, for three different communities with common objectives, to present state-of-art techniques and to share different perspective points of view.
  - Ch. Kervrann has organized with A. Trubuil (NIA Inra) a workshop entitled *"Mathematics Applied to Biology"* in conjunction with the Congress Fédération Réaumur des Sciences du Vivant, October 2007.

- *Technical program committees of conferences*
  - P. Bouthemy: general co-chairman of ICCV'2007, TPC member of ACVIS'2007, CBMI'2007, CIVR'2007, IBPRIA'2007, ICIAP'2007, ICME'2007, OTCBVS, WIAMIS'2007.
  - C. Kervrann: TPC member of TAIMA'2007, ORASIS'2007, MIAAB'2007, RFIA'2008, ICPR''2008.
  - I. Laptev: TPC member of ACCV'2007, ICCV'2007.
  - J.-P. Le Cadre: TPC member and award comittee member Fusion'2007 (Quebec), TPC member of ICIF'2007 (China), technical chairman of COGIS'2007 (Stanford).
  - E. Mémin: TPC member of SSVM'2007.
  - P. Pérez: TPC member of SIGGRAPH'2007, CVPR'2007, EUROGAPHICS'2007, ICASSP'2007, ICME'2007, IROS'2007, RFIA'2008.

- *Ph.D. reviewing*
  - P. Bouthemy: C. Dorea (UPC Barcelona), A. Benoit (LIS, Grenoble)
  - J.-P. Le Cadre: S. Boutoille (ULCO, Calais), A. Ziadi (LAAS Toulouse).
  - E. Mémin: T. Isambert (Univ. Paris 5), W. Rekkik (Univ. Paris 6), L. Igual (Univ. Pompeu Fabra, Barcelona)
  - P. Pérez: K. Smith (EPFL), A. Ganoun (Univ. Orléans), N. Thome (Univ. Lyon 2), A. Herbulot (Univ. Nice Sophia-Antipolis),

- *Project reviewing, consultancy, administrative responsibilities*

– P. Bouthemy is director of the Inria centre in Rennes and of Irisa since July 2007. He is member of the Board of the scientific association AFRIF (Association Française pour la Reconnaissance et l'Interprétation des Formes). P. Bouthemy was heading the committee of the AFRIF'2007 prize of the best (French) thesis in pattern recognition and image processing. He is member of the Board of the scientific association GRETSI. He was member of the committee of the EEA prize of the best (French) thesis in signal and image processing. He serves as a regular expert for the MRIS("Mission pour la Recherche et l'Innovation Scientifique") of the French Defense Agency (DGA). He also served as a reviewer for the ANR program call. Until September 2007, he was the Inria "main contact" in the preparation of the French-German Quaero program on multimedia indexing and retrieval.

– P. Bouthemy and J.-P. Le Cadre are deputy members of the committee ("Commission de spécialistes") of the 61th section (Signal Processing and Automation) at Universiy of Rennes 1.

– J.-P. Le Cadre is a member of the evaluating instance for "space, observation, intelligence and UAV" (DGA).

– C. Kervrann is member of the Scientific Council of the Biometry and Artificial Intelligence Department of Inra since 2006. He served as a reviewer for the creation of a research group at the Pasteur Institute (Paris) and was a part of the evaluation committee of Cemagref IRM-food department in Rennes in 2007. He is member of the AERES committee for Cemagref and member of the animation committee of the PIXEL microscopy platform at university of Rennes 1.

– P. Pérez is vice president of the Inria-Rennes project-team committee ("Comité des projets") and deputy member of Inria evaluation board ("Commission d'évaluation"). He is member of the direction team of Irisa/Inria-Rennes ("Équipe de direction") and member of the scientific and technological orientation council (COST, workgroup on large scale inititiaves) of Inria. Also, in 2007, he was president of the recruitment committee for Inria-Rennes researcher positions. P. Pérez conducted one week of consultancy for Bertin Technologies, in Feb 2007, on computer vision techniques to assist video surveillance.

– J.-F. Yao is member of the executive committee of MAS, a section of the SMAI. He is also member of the committee (Commission de spécialistes) of the 26th section (Applied mathematics) at University of Rennes 1, University of Rennes 2 and University of South-Brittany.

9.2. Teaching

- Master STI "Signal, Telecommunications, Images", University of Rennes 1, (E. Mémin : statistical image analysis, P. Bouthemy: image sequence analysis, J.-P. Le Cadre : distributed tracking, data association, estimation via MCMC methods, C. Kervrann : geometric modeling for shapes and images).

- Master of Computer Science, Ifsic, University of Rennes 1 (P. Pérez: motion analysis; P. Bouthemy: video indexing).

- DIIC INC, Ifsic, University of Rennes 1 (P. Heas: Markov models for image analysis; Th. Corpetti: PDEs for image processing; P. Bouthemy: motion analysis)

- Master PIC and ENSPS Strasbourg, (P. Bouthemy : image sequence analysis).

- ENSAI Rennes, 3rd year (C. Kervrann, P. Pérez : statistical models and image analysis : particle filtering and target tracking).

- ENS Cachan, Brittany, 1st year (P. Pérez: introduction to image processing and analysis)

- Graduate student trainees and interns :

- B. Belmudez (ENSPS Strasbourg, co-supervised by J.-F. Yao, G. Piriou and P. Bouthemy, work on detection of structural changes in urban areas from high-resolution satellite images using mixed-state Markov models).
- J.-A. Silva de Oliveira (INSA Rennes, co-supervised by J.-P. Le Cadre and P. Bouthemy, work on object detection and tracking using a camera network).
- S. Loya (Ifsic and Master STI, Rennes 1, supervised by C. Kervrann, work on microtubule extremities tracking in differential interference contrast video-microscopy).
- Ch. Avenel (Engineering Master in Computer Science, ENS Cachan Bretagne, co-supervised by E. Mémin and P. Pérez, work on free curve tracking with particle filtering).
- A. Dame (INSA RENNES, co-supervised by P. Pérez and F. Lamarche [Bunraku team]), work on detection of people in crowed indoor scene).

- External thesis supervision :
  - F. Celeste (DGA-CEP) supervised by J.-P. Le Cadre;
  - A. Lehuger (FT-RD, Rennes) supervised by P. Pérez;

9.3. Participation in seminars, invitations, awards

- C. Kervrann was an invited speaker at the Scientific Meeting of the Curie Institute on "Analysis and modeling for fluorescence video-microscopy " (Paris, February 2007). C. Kervrann gave the following invited talks: "Estimation adaptative et approche bayésienne pour le débruitage d'images nD à partir de motifs locaux", GREYC seminar (Caen university, March 2007), "On N-dimensional image analysis and time-lapse fluorescence microscopy: descriptors and modeling for intra-cellular dynamics and trafficking", GDR 2588 meeting (Action Thématique *"noyau"*) (Toulouse, January 2007), "Descriptors and modeling for intra-cellular dynamics and trafficking", joint GDR 2588 / SDV Cnrs meeting (Paris, January 2007), ¨Modélisation dynamique pour l'étude d'une machine moléculaire", Inra MICALIS project seminar (Jouy-en-Josas, October 2007), "Non-parametric patch-based estimation vs. Bayesian non-local means filter for image representation and denoising", MAS seminar of the Ecole Centrale de Paris (Chatenay-Malabry, October 2007). He presented two posters at the National Meeting of Inria ARC projects (Rennes, October 2007) and at the ACI IMPBIO Meeting (Paris, October 2007).
- I. Laptev was a keynote speaker at "Journée Détection et Reconnaissance d'Objets dans des images", GDR ISIS (Paris, July 2007). I. Laptev also gave the following invited talks: "Object Detection with Boosted Histogram Features" at "Symposium on Machine Learning in Image and Document", 10th Anniversary of LIAMA (Beijing, China, Jan. 2007); "From objects to actions: detection by boosted histogram classifiers" at The Rank Prize Funds "Mini-Symposium on Interacting with Still and Moving Images - From Signals to Semantics" (Windermere, UK, July 2007) as well as at LEAR vision seminar, INRIA Rhône-Alpes (Grenoble, May 2007). I. Laptev participated in The PASCAL Visual Object Classes Challenge 2007 (VOC2007) and presented competitive results for the detection of visual object classes.
- E. Mémin has been invited to give a talk on Fluid flows analysis from image sequence at the Argentinian Academy of Sciences.
- V. Auvray has received the 2007 best Ph-D thesis prize delivered by Fondation Métivier for Ph-D thesis work conducted in collaboration with an industrial partner.

10. Bibliography

Major publications by the team in recent years

[1] E. ARNAUD, E. MÉMIN, B. CERNUSCHI-FRIAS. *Conditional filters for image sequence based tracking - Application to point tracking*, in "IEEE Trans. on Image Processing", vol. 14, n⁰ 1, 2005, p. 63–79.

[2] F. BAVENCOFF, J.-M. VANPEPERSTRAETE, J.-P. LE CADRE. *Constrained bearings-only target motion analysis via Monte Carlo Markov chain methods*, in "IEEE Trans. on Aerospace and Electronic Systems", vol. 42, n$^{\text{o}}$ 4, 2006, p. 1240–1263.

[3] P. BOUTHEMY, C. HARDOUIN, G. PIRIOU, J.-F. YAO. *Mixed-state auto-models and motion texture modeling*, in "Journal of Mathematical Imaging and Vision", vol. 25, n$^{\text{o}}$ 3, 2006, p. 387–402.

[4] T. CORPETTI, D. HEITZ, G. ARROYO, E. MÉMIN, A. SANTA-CRUZ. *Fluid experimental flow estimation based on an optical-flow scheme*, in "Int. J. Experiments in Fluid", vol. 40, n$^{\text{o}}$ 1, 2006, p. 80–97.

[5] T. CORPETTI, E. MÉMIN, P. PÉREZ. *Dense estimation of fluid flows.*, in "IEEE Trans. on Pattern Analysis and Machine Intelligence", vol. 24, n$^{\text{o}}$ 3, March 2002, p. 365-380.

[6] C. HUE, J.-P. LE CADRE, P. PÉREZ. *Posterior Cramer-Rao bounds for multi-target tracking*, in "IEEE Trans. on Aerospace and Electronic Systems", vol. 42, n$^{\text{o}}$ 1, 2006, p. 37–50.

[7] C. KERVRANN, J. BOULANGER. *Optimal spatial adaptation for patch-based image denoising*, in "IEEE Trans. on Image Processing", vol. 15, n$^{\text{o}}$ 10, 2006, p. 2866–2878.

[8] C. KERVRANN, A. TRUBUIL. *Optimal level curves and minimizers of cost functionals in image segmentation*, in "Journal of Mathematical Imaging and Vision", vol. 17, n$^{\text{o}}$ 2, 2002, p. 153–174.

[9] I. LAPTEV. *On space-time interest points*, in "Int. J. Computer Vision", vol. 64, n$^{\text{o}}$ 2, 2005, p. 107–123.

[10] P. PÉREZ, J. VERMAAK, A. BLAKE. *Data fusion for visual tracking with particles*, in "Proc. IEEE", vol. 92, n$^{\text{o}}$ 3, 2004, p. 495–513.

[11] T. VEIT, F. CAO, P. BOUTHEMY. *An a contrario decision framework for region-based motion detection*, in "International Journal on Computer Vision", vol. 68, n$^{\text{o}}$ 2, 2006, p. 163–178.

Year Publications

Doctoral dissertations and Habilitation theses

[12] J. BOULANGER. *Estimation non-paramétrique et contributions à l'analyse de séquences d'images. Modélisation, simulation et estimation du trafic intra-cellulaire dans des séquences de vidéo-microscopie*, Ph. D. Thesis, université de Rennes 1, Mention Traitement du Signal et des Télécommunications, janvier 2007, http://www.irisa.fr/vista/Papers/2007_these_boulanger.pdf.

[13] A. BUGEAU. *Détection et suivi d'objets en mouvement dans des scènes complexes, application à la surveillance des conducteurs*, Ph. D. Thesis, université de Rennes 1, Mention Traitement du Signal et des Télécommunications, décembre 2007.

[14] N. PAPADAKIS. *Assimilation de données images: application au suivi de courbes et de champs de vecteurs*, Ph. D. Thesis, université de Rennes 1, Mention Mathématiques et applications, novembre 2007, http://www.irisa.fr/vista/Papers/2007_these_papadakis.pdf.

Articles in refereed journals and book chapters

[15] E. ARNAUD, E. MÉMIN. *Partial linear Gaussian models for tracking in image sequences using sequential Monte-Carlo methods*, in "International Journal of Computer Vision", vol. 74, n° 1, 2007, p. 75-102, http://www.irisa.fr/vista/Papers/2007_ijcv_arnaud.pdf.

[16] J. BOULANGER, C. KERVRANN, P. BOUTHEMY. *Space-time adaptation for patch based image sequence restoration*, in "IEEE Trans. on Pattern Analysis and Machine Intelligence", vol. 8, n° 6, 2007, p. 1096-1102, http://www.irisa.fr/vista/Papers/2007_pami_boulanger.pdf.

[17] J. BOULANGER, T. PÉCOT, P. BOUTHEMY, J. SALAMERO, J.-B. SIBARITA, C. KERVRANN. *Simulation and estimation of intra-cellular dynamics and trafficking*, S. WONG, R. MACHIRAJU, J. RITTSCHER (editors), Artech Publishing House, 2007.

[18] T. BRÉHARD, J.-P. LE CADRE. *Hierarchical particle filter for bearings only tracking*, in "IEEE Trans. on Aerospace and Electronic Systems", vol. 43, n° 4, 2007, http://www.irisa.fr/vista/Papers/2007_AES_brehard.pdf.

[19] F. CAO, J. DELON, A. DESOLNEUX, P. MUSÉ, F. SUR. *A unified framework for detecting groups and application to shape recognition*, in "Journal of Mathematical Imaging and Vision", vol. 27, n° 2, 2007, p. 91-119.

[20] P. COUPÉ, P. YGER, S. PRIMA, P. HELLIER, C. KERVRANN, C. BARILLOT. *An optimized blockwise Non-Local means denoising filter for 3D Magnetic Resonance Images*, in "IEEE Trans. on Medical Imaging", 2007.

[21] T. CRIVELLI, B. CERNUSCHI-FRIAS, P. BOUTHEMY. *Texturas de movimiento: campos markovianos mixtos y segmentacion*, in "RIAI Revista Iberoamericana de Automática e Informática Industrial", vol. 4, n° 4, 2007, p. 1-7.

[22] A. CUZOL, P. HELLIER, E. MÉMIN. *A low dimensional fluid motion estimator*, in "International Journal of Computer Vision", vol. 75, n° 3, 2007, p. 329-349, http://www.irisa.fr/vista/Papers/2007_ijcv_cuzol.pdf.

[23] J. D'ADAMO, N. PAPADAKIS, E. MÉMIN, G. ARTANA. *Variational assimilation of POD low-order dynamical systems*, in "Journal of Turbulence", vol. 8, n° 9, 2007, p. 1-22.

[24] F. DAMBREVILLE, F. CELESTE, C. SIMONIN. *6*, in "Modélisation probabiliste de politiques et leurs optimisations pour la planification de capteurs", P. SIARRY (editor), Hermes, 2007, p. 163-193.

[25] F. DAMBREVILLE, J.-P. LE CADRE. *Constrained minimax optimization of continuous search efforts for the detection of a stationary target*, in "Naval Research Logistics", vol. 54, n° 6, 2007, p. 589-601, http://www.irisa.fr/vista/Papers/2007_NRL_dambreville.pdf.

[26] P. DODIN, F. MINVIELLE, J.-P. LE CADRE. *A branch-and-bound algorithm applied to optimal radar search pattern*, in "Aerospace Science and Technology", vol. 11, n° 4, 2007, p. 279-288, http://www.irisa.fr/vista/Papers/2007_AST_dodin.pdf.

[27] P. DODIN, F. MINVIELLE, J.-P. LE CADRE. *Estimating the ballistic coefficient of a re-entry vehicle*, in "IET Proc. Radar, Sonar and Navigation", vol. 1, n° 3, 2007, p. 173-183, http://www.irisa.fr/vista/Papers/2007_rsn_dodin.pdf.

[28] P. HÉAS, E. MÉMIN. *3D motion estimation of atmospheric layers from image sequences*, in "IEEE Trans. Geoscience and Remote Sensing", 2007, http://www.irisa.fr/vista/Papers/2007_IEEEGRS_HeasMemin.pdf.

[29] P. HÉAS, E. MÉMIN, N. PAPADAKIS, A. SZANTAI. *Layered estimation of atmospheric mesoscale dynamics from satellite imagery*, in "IEEE Trans. Geoscience and Remote Sensing", 2007, http://www.irisa.fr/vista/Papers/2007_IEEEGRS_Heas.pdf.

[30] C. KERVRANN, J. BOULANGER. *Local adaptivity to variable smoothness for exemplar-based image denoising and representation*, in "International Journal of Computer Vision", 2007, http://www.irisa.fr/vista/Papers/2007_ijcv_kervrann.pdf.

[31] I. LAPTEV, B. CAPUO, C. SCHULTZ, T. LINDEBERG. *Local velocity-adapted motion events for spatio-temporal recognition*, in "Computer Vision and Image Understanding", vol. 108, n⁰ 3, 2007, p. 207-229, http://www.irisa.fr/vista/Papers/2007_cviu_laptev.pdf.

[32] J.-P. LE CADRE. *7*, in "Optimisation des émissions pour la trajectographie et la poursuite de cibles mobiles", P. SIARRY (editor), Hermes, 2007, p. 195-200, http://www.irisa.fr/vista/Papers/2007_IC2_lecadre.pdf.

[33] R. VENKATESH BABU, P. PÉREZ, P. BOUTHEMY. *Robust tracking with motion estimation and local Kernel-based color modeling*, in "Image and Vision Computing", vol. 25, n⁰ 8, 2007, p. 1205-1216.

[34] J. YUAN, C. SCHNÖRR, E. MÉMIN. *Discrete orthogonal decomposition and variational fluid flow estimation*, in "Journal of Mathematical Imaging and Vision", vol. 28, n⁰ 1, 2007, p. 67-80, http://www.irisa.fr/vista/Papers/2007_jmiv_yuan.pdf.

Publications in Conferences and Workshops

[35] V. BADRINARAYANAN, P. PÉREZ, F. LE CLERC, L. OISEL. *On uncertainties, random features and object tracking*, in "Proc. Int. Conf. Image Proc. (ICIP'07), San Antonio, TX", septembre 2007.

[36] V. BADRINARAYANAN, P. PÉREZ, F. LE CLERC, L. OISEL. *Probabilistic color and adaptive multi-feature tracking with dynamically switched priority between cues*, in "Proc. Int. Conf. Comp. Vis.(ICCV'07), Rio de Janeiro, Brazil", October 2007.

[37] A. BUGEAU, P. PÉREZ. *Detection and segmentation of moving objects in highly dynamic scenes*, in "Proc. Int. Conf. Computer Vision and Pattern Recog. (CVPR' 07), Minneapolis, MI", June 2007, http://www.irisa.fr/vista/Papers/2007_cvpr_bugeau.pdf.

[38] A. BUGEAU, P. PÉREZ. *Joint Tracking and Segmentation of Objects*, in "Proc. Conf. Advanced Concepts for Intelligent Vision Systems (ACIVS' 07), Delft, the Netherlands", August 2007, p. 628-639, http://www.irisa.fr/vista/Papers/2007_acivs_bugeau.pdf.

[39] A. BUGEAU, P. PÉREZ. *Sélection de la taille du noyau pour l'estimation à noyau dans des espaces multidimensionnels hétérogènes*, in "Proc. Colloque GRETSI, Troyes, France", septembre 2007, http://www.irisa.fr/vista/Papers/2007_gretsi_bugeau.pdf.

[40] F. CELESTE, F. DAMBREVILLE, J.-P. LE CADRE. *Evaluation of a robot learning and planning via extreme value theory*, in "Proc. Conf. on Information Fusion (Fusion'07), Quebec, Canada", July 2007, http://www.irisa.fr/vista/Papers/2007_fusion_celeste.pdf.

[41] T. CORPETTI, N. PAPADAKIS, E. MÉMIN. *Dense estimation of motion fields on Meteosat Second Generation images using a dynamical consistency*, in "IEEE Int. Geoscience and Remote Sensing Symp (IGARSS'07), Barcelona, Spain", July 2007.

[42] T. CRIVELLI, B. CERNUSCHI-FRIAS, P. BOUTHEMY. *Modelos temporales de texturas de movimiento con estados mixtos*, in "Symposium on Technology (AST 2007), 36 JAIIO, Mar del Plata, Argentina", August 2007.

[43] T. CRIVELLI, B. CERNUSCHI-FRIAS, P. BOUTHEMY. *Una teoria sobre campos aleatorios mixtos con aplicacion al analisis de movimiento en imagenes*, in "XII Reunion de Trabajo en Procesamiento de la Informacion y Control (RPIC 2007), Rio Gallegos, Argentina", October 2007.

[44] D. HEITZ, P. HÉAS, V. NAVAZA, J. CARLIER, E. MÉMIN. *Collaboration corrélation-variationnelle pour une estimation robuste du flot-optique*, in "18ème Congrès Français de Mécanique (CFM)", June 2007, http://www.irisa.fr/vista/Papers/2007_CFM_Heitz.pdf.

[45] D. HEITZ, P. HÉAS, V. NAVAZA, J. CARLIER, E. MÉMIN. *Spatio-temporal correlation-variational approach for robust optical flow estimation*, in "Symposium on Particle Image Velocimetry (PIV'07)", June 2007, http://www.irisa.fr/vista/Papers/2007_piv_heitz.pdf.

[46] A. HERVIEU, P. BOUTHEMY, J.-P. LE CADRE. *A HMM-based method for recognizing dynamic video contents from trajectories*, in "Proc. Int. Conf. Image Proc. (ICIP'07), San Antonio, TX", septembre 2007, http://www.irisa.fr/vista/Papers/2007_icip_hervieu.pdf.

[47] A. HERVIEU, P. BOUTHEMY, J.-P. LE CADRE. *Reconnaissance d'évènements dans des vidéos par l'analyse de trajectoires à l'aide de modèles de Markov*, in "Proc. Colloque GRETSI, Troyes, France", septembre 2007, http://www.irisa.fr/vista/Papers/2007_gretsi_hervieu.pdf.

[48] P. HÉAS, K. KRISSIAN, E. MÉMIN, A. SZANTAI. *Reconstruction and visualization of 3D wind fields from satellite image sequences*, in "15th Satellite Meteorology and Oceanography Conf. of the American Meteorological Soc., Amsterdam, The Netherlands", September 2007.

[49] P. HÉAS, E. MÉMIN, N. PAPADAKIS. *A consistent spatio-temporal motion estimator for atmospheric layers*, in "Proc. Conf. Scale-Space and Variational Meth. (SSVM'07), Ischia, Italy", June 2007, p. 251-263, http://www.irisa.fr/vista/Papers/2007_ssvm_heas.pdf.

[50] P. HÉAS, N. PAPADAKIS, E. MÉMIN, A. SZANTAI. *Motion estimation of 2D atmospheric layers from satellite image sequences*, in "15th Satellite Meteorology and Oceanography Conf. of the American Meteorological Soc., Amsterdam, The Netherlands", September 2007.

[51] A. ICKOWICZ, J.-P. LE CADRE, P. MINVIELLE. *On the effect of data contamination for multitarget tracking, Part I*, in "Proc. Conf. on Information Fusion (Fusion'07), Quebec, Canada", July 2007, http://www.irisa.fr/vista/Papers/2007_fusion_ickowicz_a.pdf.

[52] A. ICKOWICZ, J.-P. LE CADRE, P. MINVIELLE. *On the effect of data contamination for multitarget tracking, Part II*, in "Proc. Conf. on Information Fusion (Fusion'07), Quebec, Canada", July 2007, http://www.irisa.fr/vista/Papers/2007_fusion_ickowicz_b.pdf.

[53] C. Kervrann, J. Boulanger, P. Coupé. *Bayesian non-local means filter, image redundancy and adaptive dictionaries for noise removal*, in "Proc. Conf. Scale-Space and Variational Meth. (SSVM' 07), Ischia, Italy", June 2007, p. 520-532, http://www.irisa.fr/vista/Papers/2007_ssvm_kervrann.pdf.

[54] I. Laptev, P. Pérez. *Retrieving actions in movies*, in "Proc. Int. Conf. Comp. Vis.(ICCV'07), Rio de Janeiro, Brazil", October 2007, http://www.irisa.fr/vista/Papers/2007_iccv_laptev.pdf.

[55] J. Law-To, L. Chen, A. Joly, I. Laptev, O. Buisson, V. Gouet-Brunet, N. Boujemaa, F. Stentiford. *Video Copy Detection: a Comparative Study*, in "Proc. CIVR'07, Amsterdam, The Netherlands", October 2007, p. 371–378, http://www.irisa.fr/vista/Papers/2007_civr_law-to.pdf.

[56] N. Papadakis, T. Corpetti, E. Mémin. *Dynamically consistent optical flow estimation*, in "Proc. Int. Conf. Comp. Vis.(ICCV'07), Rio de Janeiro, Brazil", October 2007, http://www.irisa.fr/vista/Papers/2007_iccv_papadakis_b.pdf.

[57] N. Papadakis, P. Héas, E. Mémin. *Image assimilation for motion estimation of atmospheric layers with shallow-water model*, in "Proc. Asian. Conf. Comp. Vis.(ACCV'07), Tokyo, Japan", November 2007, http://www.irisa.fr/vista/Papers/2007_accv_papadakis.pdf.

[58] N. Papadakis, P. Héas, E. Mémin. *Motion estimation of 2D atmospheric layers with variational assimilation techniques*, in "15th Satellite Meteorology and Oceanography Conf. of the American Meteorological Soc., Amsterdam, The Netherlands", September 2007.

[59] N. Papadakis, E. Mémin, T. Corpetti. *Variational estimation of 2D time consistent dense motion from image sequence*, in "Proc. Eur. Geoscience Union, Nonlinear processes in geophysics, Data assimilation in the presence of nonlinearities, Vienna, Austria", April 2007.

[60] N. Papadakis, E. Mémin, J. D'Adamo, G. Artana. *Assimilation variationelle de systèmes dynamiques d'ordre faible*, in "3e Congrès national de mathématiques appliquées et industrielles (SMAI 2007), Grenoble, France", June 2007.

[61] N. Papadakis, E. Mémin. *A variational framework for spatio-temporal smoothing of fluid motions*, in "Proc. Conf. Scale-Space and Variational Meth. (SSVM'07), Ischia, Italy", June 2007, p. 603-615, http://www.irisa.fr/vista/Papers/2007_ssvm_papadakis.pdf.

[62] N. Papadakis, E. Mémin. *Variational optimal control technique for the tracking of deformable objects*, in "Proc. Int. Conf. Comp. Vis.(ICCV'07), Rio de Janeiro, Brazil", October 2007, http://www.irisa.fr/vista/Papers/2007_iccv_papadakis_a.pdf.

[63] T. Pécot, J. Boulanger, C. Kervrann, P. Bouthemy. *Network tomography for trafficking simulation and analysis in fluorescence microscopy imaging*, in "Proc. IEEE Int. Symp. on Biomedical Imaging: from nano to macro (ISBI'07), Washington, D.C.", April 2007, p. 268-271, http://www.irisa.fr/vista/Papers/2007_isbi_pecot.pdf.

[64] T. Pécot, J. Boulanger, C. Kervrann, P. Bouthemy. *Tomographie de réseau appliquée à la simulation et à l'analyse de trafic dans des séquences d'images de microscopie à fluorescence*, in "Proc. Colloque GRETSI, Troyes, France", septembre 2007.

[65] C. SIMONIN, J.-P. LE CADRE, F. DAMBREVILLE. *The cross-entropy method for solving a variety of hierachical search problems*, in "Proc. Conf. on Information Fusion (Fusion'07), Quebec, Canada", July 2007, http://www.irisa.fr/vista/Papers/2007_fusion_simonin.pdf.

[66] T. VEIT, F. CAO, P. BOUTHEMY. *Space-time a contrario clustering for detecting coherent motion*, in "IEEE Int. Conf. on Robotics and Automation, ICRA'07, Rome, Italy", April 2007, p. 33-39, http://www.irisa.fr/vista/Papers/2007_icra_veit.pdf.

Internal Reports

[67] A. BUGEAU, P. PÉREZ. *Bandwidth selection for kernel estimation in mixed multi-dimensional spaces*, Research Report, n° 6286, INRIA, September 2007, https://hal.inria.fr/inria-00171686.

[68] A. BUGEAU, P. PÉREZ. *Detection and segmentation of moving objects in complex scenes*, Research Report, n° 6282, INRIA, September 2007, https://hal.inria.fr/inria-00170360.

[69] A. BUGEAU, P. PÉREZ. *Track and Cut: simultaneous tracking and segmentation of multiple objects with graph cuts*, Research Report, n° 6337, INRIA, 10 2007, https://hal.inria.fr/inria-00181865.

[70] B. CERNUSCHI-FRIAS. *Mixed states markov random fields with symbolic labels and multidimensional real values*, Research Report, n° 6255, INRIA, July 2007, https://hal.inria.fr/inria-00165642.

[71] T. CRIVELLI, B. CERNUSCHI-FRIAS, P. BOUTHEMY, J.-F. YAO. *Mixed-state models in image motion analysis: theory and applications*, Research Report, n° 6335, INRIA, September 2007, https://hal.inria.fr/inria-00182150.

[72] P. HÉAS, M. DATCU. *Supervised learning on graphs of spatio-temporal similarity in satellite image sequences*, Research Report, n° 6299, INRIA, September 2007, http://hal.inria.fr/inria-00173280.

[73] P. HÉAS, N. PAPADAKIS, E. MÉMIN. *Time-consistent estimators of 2D/3D motion of atmospheric layers from pressure images*, Research Report, n° 6292, INRIA, September 2007, http://hal.inria.fr/inria-00172330.

[74] N. PAPADAKIS, E. MÉMIN. *A variational method for joint tracking of curve and motion*, Research Report, n° 6283, INRIA, September 2007, http://hal.inria.fr/inria-00171087.

References in notes

[75] P. BOUTHEMY, E. FRANÇOIS. *Motion segmentation and qualitative dynamic scene analysis from an image sequence*, in "Int. Journal of Computer Vision", vol. 10, n° 2, April 1993, p. 157-182.

[76] T. CORPETTI, E. MÉMIN, P. PÉREZ. *Dense estimation of fluid flows.*, in "IEEE Trans. on Pattern Analysis and Machine Intelligence", vol. 24, n° 3, March 2002, p. 365-380.

[77] T. CORPETTI, E. MÉMIN, P. PÉREZ. *Extraction of singular points from dense motion fields: an analytic approach*, in "Journal of Mathematical Imaging and Vision", vol. 19, n° 3, 2003, p. 175-198.

[78] A. DOUCET, S. GODSILL, C. ANDRIEU. *On sequential Monte Carlo sampling methods for Bayesian filtering*, in "Statistics and Computing", vol. 10, n° 3, 2000, p. 197-208.

[79] R. FABLET, P. BOUTHEMY. *Non-parametric scene activity analysis for statistical retrieval with partial query*, in "Journal of Mathematical Imaging and Vision", vol. 14, n° 3, May 2001, p. 257-270.

[80] M. GELGON, P. BOUTHEMY. *A region-level motion-based graph representation and labeling for tracking a spatial image partition*, in "Pattern Recognition", vol. 33, n° 4, April 2000, p. 725-745.

[81] F. HEITZ, P. BOUTHEMY. *Multimodal estimation of discontinuous optical flow using Markov random fields*, in "IEEE Trans. on Pattern Analysis and Machine Intelligence", vol. 15, n° 12, December 1993, p. 1217-1232.

[82] F. HEITZ, P. PÉREZ, P. BOUTHEMY. *Multiscale minimization of global energy functions in some visual recovery problems*, in "CVGIP : Image Understanding", vol. 59, n° 1, 1994, p. 125-134.

[83] M. ISARD, A. BLAKE. *CONDENSATION–Conditional Density Propagation for Visual Tracking*, in "Int. Journal of Computer Vision", vol. 29, n° 1, 1998, p. 5-28.

[84] G. KITAGAWA. *Monte Carlo filter and smoother for non-Gaussian nonlinear state space models*, in "J. Computational and Graphical Stat.", vol. 5, n° 1, 1996, p. 1-25.

[85] I. LAPTEV. *Improvements of object detection using boosted histograms*, in "Proc. British Machine Vision Conf. (BMVC'06), Edinburgh, U.-K.", September 2006.

[86] E. MÉMIN, P. PÉREZ. *Hierarchical estimation and segmentation of dense motion fields*, in "Int. Journal of Computer Vision", vol. 46, n° 2, February 2002, p. 129-155.

[87] E. MÉMIN, P. PÉREZ. *Dense estimation and object-based segmentation of the optical flow with robust techniques*, in "IEEE Trans. on Image Processing", vol. 7, n° 5, May 1998, p. 703-719.

[88] J.-M. ODOBEZ, P. BOUTHEMY. *Robust multiresolution estimation of parametric motion models*, in "Journal of Visual Communication and Image Representation", vol. 6, n° 4, December 1995, p. 348-365.

[89] J.-M. ODOBEZ, P. BOUTHEMY. *Direct incremental model-based image motion segmentation for video analysis*, in "Signal Processing", vol. 6, n° 2, 1998, p. 143-155.