INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Project-Team VISTA

Vision Spatio-Temporelle et Apprentissage

Rennes - Bretagne-Atlantique

THEME COG

Activity Report

2008

Table of contents

1. Team

Research Scientist

Patrick Bouthemy [DR, head of Inria Rennes Research Center, HdR]
Thomas Corpetti [CR CNRS at Costel (University of Rennes 2), external collaborator]
Patrick Heas [CR, from 30/09/08]
Charles Kervrann [Inra research scientist, on Inria secondment (mise à disposition)]
Ivan Laptev [CR]
Jean-Pierre Le Cadre [DR CNRS, HdR]
Étienne Mémin [DR, on sabbatical leave at Unv. of Bueno Aires (UBA) until 01/09/2008, HdR]
Patrick Pérez [DR, head of project-team, HdR]

Technical Staff

Florent Dutrech [2ACI DGA PEA project from 01/03/08]
Guillaume Neveu [FTRD CRE project from 01/09/08]
Pham Nam Trung [FP6 Pegase project from 15/11/07]

PhD Student

Christophe Avenel [ENS grant, from 01/10/07]
Vijay Badrinarayanan [Cifre grant, Thomson-RD, from 01/03/06]
Emilie Dexter [DGA grant, from 01/10/06]
Sophie Blestel [MENRT grant, since 01/10/08]
Pierre Cariou [Cemagef grant, since 01/09/08]
Tomas Crivelli [Faculty of Engineering, University of Buenos-Aires, “thèse en co-tutelle”]
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]
Thierry Pécot [Inra from 01/10/06 to 31/03/08, Inria-Curie Institute grant from 01/04/08]
Cécile Simonin [DGA grant, from 01/10/05]

Post-Doctoral Fellow

Imran Junejo [Inria post-doc, from 01/11/07]
Anatole Chessel [Inria post-doc, from 01/01/08]

Administrative Assistant

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

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 [77]). 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 [73] 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 [86], [87]. 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 [80] 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 [78]. 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 $\mathbf{w}$ 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 [79]. Then, robust estimators were popularized leading to the introduction of so-called auxiliary variables ζ now taking their values in $[0, 1]$ [85]. 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 [84]. 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 [85], [84]. 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:

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

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 [85], 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 [74]. The method we have devised for fluid motion involves the following global energy function:

$$\int_{\Omega} \rho \left\{ f\left(\mathbf{s}+\mathbf{w}, t+1\right) \exp\left(div\ \mathbf{w}\right) - f(\mathbf{s}, t) \right\} d\mathbf{s} + \alpha \int_{\Omega} (\|\nabla div\ \mathbf{w}\|^2 + \|\nabla curl\ \mathbf{w}\|^2) d\mathbf{s}. \tag{6}$$

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:

$$\int_{\Omega} |div\ \mathbf{w} - \xi|^2 + \lambda \rho(\|\nabla \xi\|) d\mathbf{s} + \alpha \int_{\Omega} |curl\ \mathbf{w} - \zeta|^2 + \lambda \rho(\|\nabla \zeta\|) d\mathbf{s}. \tag{7}$$

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 [75]. 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:

$$\begin{aligned} \textbf{prediction step:} \quad & p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) = \int p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})d\mathbf{x}_{t-1} \\ \textbf{filtering step:} \quad & p(\mathbf{x}_t|\mathbf{y}_{1:t}) = \frac{p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t-1})}{p(\mathbf{y}_t|\mathbf{y}_{1:t-1})}, \end{aligned} \tag{8}$$

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 [76], [82], [81], 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

$$w_t^{(n)} \propto w_{t-1}^{(n)} \frac{p(\mathbf{y}_t|\mathbf{x}_t^{(n)})p(\mathbf{x}_t^{(n)}|\mathbf{x}_{t-1}^{(n)})}{q(\mathbf{x}_t^{(n)}|\mathbf{x}_{t-1}^{(n)}, \mathbf{y}_t)}, \qquad \sum_{n=1}^{N} w_t^{(n)} = 1, \tag{9}$$

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 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 with deconvolution, confocal spinning-disk microscopy, Total Internal Reflectance Microscopy (TIRF), Fluorescence Recovery After Photobleaching (FRAP) combined with 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 interaction estimation between Rab11 and Langerin proteins, and dynamic estimation of Rab6a and Rab6a' proteins - involved in the regulation of transport from the Golgi apparatus to the endoplasmic reticulum.

Moreover, microscopic imaging at both the light and electron microscopic level provides multiscale unique information on protein localization and interactions, and extends and enriches that obtained from molecular and biochemical techniques. The 3D reconstruction of macromolecular structures form 2D EM (Electronic Microscopy) images of vitrified biological samples (Cryo-EM) has some advantages over other imaging techniques since it has proved to be an effective technique to investigate the structure of native cells with macromolecular resolution and preserve the whole integrity of the cell. Nevertheless, the high magnification available with EM comes with a limited field of view. Also, the very low contrast of unstained and vitrified biological specimens and the need to minimize the exposure to electron radiation make the identification of specific structures a difficult and time consuming task. Therefore one needs a gentle and time efficient way to locate structures of interest, improve image contrast and remove noise for a better interpretation of the image contents. We currently investigated image segmentation methods (3-year PhD grant) to analyze microtubule dynamics observed in Cryo-EM in collaboration with University of Rennes 1 - UMR 6026 (D. Chretien). These macromolecules with about 13 protofilaments formed with the $\alpha\beta$-tubulin dimers, are known to involved intracellular transporting, cell motility, meiosis and mitosis. In the coming 3 years, we will study interactions between neighboring protofilaments at the extremity of microtubules and the biological function of open sheet.

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 [86]. 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, Guillaume Neveu, 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 [83]. 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 is planned to be applied to real microarray images routinely used for disease diagnosis.

6. New Results

6.1. Image sequence processing and modeling

6.1.1. Non-parametric regression for fluorescence microscopy image sequence denoising

Participants: Charles Kervrann, Patrick Bouthemy.

[In collaboration with J. Boulanger (Curie Institute)]

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, in collaboration with UMR 144 CNRS / Curie Institute, 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 method exploits 3D+time information to improve the signal-to-noise ratio of images corrupted by mixed Poisson-Gaussian noise. A variance stabilization transform is first applied to the image-data to introduce independence between the mean and variance. This pre-processing requires the knowledge of parameters related to the acquisition system, also estimated in our approach. In a second step, we propose an original statistical patch-based framework for noise reduction and preservation of space-time discontinuities. In the continuous setting, 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. The size of each neighborhood is optimized to improve the performance of the pointwise estimator. In our experiments, the SNR is shown to be drastically improved and enhanced images can then be correctly segmented. In fluorescence video-microscopy, recent experiments demonstrated also that this method can be used as a pre-processing stage for image deconvolution, and allows us to increase the frame-rate of a factor 10 with the same SNR values. Finally, this novel approach can be used for biological studies where dynamics have to be analyzed in molecular and subcellular bio-imaging.

6.1.2. Patch-based redundancy analysis for change detection in an image pair

Participants: Charles Kervrann, Patrick Pérez.

To develop better change detection algorithms, new models able to capture all the spatio-temporal regularities and geometries seen in an image pair are needed. In contrast to the usual pixel-wise methods, a recent line of work consists in modeling semi-local interactions from image patches. Therefore, we proposed also a patch-based formulation for detecting occlusions and other local or regional changes in an image pair. The redundancy property observed in similar images is exploited to detect unusual spatio-temporal patterns in the scene. By introducing scores to compare patches and false alarm rates, a detection algorithm can be derived for dynamic scene analysis with no optical flow computation. From binary local decisions, we propose a collaborative decision rule that uses the total number of detections made by individual neighboring pixels. Our patch-based approach is robust to many types of variations, such as local appearance change, motion and scale variation. Experimental results on several applications including background subtraction, defect detection in video inspection of manufactured objects or detection of changes in satellite images, demonstrate that the method performs well at detecting occlusions, meaningful regional changes and space-time corners, and is especially robust in the case of low signal-to-noise ratios.

6.1.3. Network tomography for tracking in fluorescence microscopy imaging

Participants: Thierry Pécot, Charles Kervrann, Patrick Bouthemy.

[In collaboration with J. Boulanger (Curie Institute)]
The study of protein dynamics is essential for understanding the multi-molecular complexes at subcellular levels. Green Fluorescent Protein (GFP)-tagging and time-lapse fluorescence microscopy enable to observe molecular dynamics and interactions in live cells and consequently to make progress in knowledge about protein dynamics. Original image analysis methods are then required to process challenging 2D or 3D image sequences. Tracking methods that estimate the whole trajectories of moving objects have been successfully experimented but can be applied for tracking a limited number of objects (a few dozens). To address the tracking problem of several hundreds of objects, we propose instead an original framework that provides general information about molecule transport, that is about traffic flows between origin and destination regions detected in the image sequence. Traffic estimation can be accomplished by adapting the recent advances in Network Tomography commonly used in network communications. Our estimation method is inspired from the Network Tomography (NT) concept, introduced by Vardi in network communications and further applied

to video surveillance. 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. This statistical method allows us to provide a global description of traffic flows. We just need to count the number of "objects/vesicles" in different image regions at each time step. In collaboration with UMR 144 CNRS / Curie Institute, we extended the usual NT concept to non-binary routing from geodesic paths given the image sequence. Unlike previous approaches, the new formulation can be considered as a probabilistic minimal paths modeling for object tracking. We showed that the origin-destination (OD)paths are not the minimal paths between the two extremities but formed as as a set of minimal paths joining intermediate points. We also we proposed an estimation/optimization framework to derive counting measurements from image intensity (fluorescence). The traffic flow problem is also solved with additional parsimonious constraints. This approach has been developed for real image sequences and Rab proteins, known to be involved in the regulation of intracellular membrane trafficking.

6.1.4. Patch-based Markov models for event detection in fluorescence microscopy imaging

Participants: Thierry Pécot, Charles Kervrann.

Motivated by the problem of spatial detection of fluorescence irregularities and meaningful events in temporal image sequences obtained in time-lapse fluorescence (wide-field) microscopy, we proposed an original patch-based Markov modeling to detect spatial irregularities in fluorescence images with low false alarm rates. In contrast to the usual pixel-wise MRF methods, the redundancy property and patch-based representation can be exploited to detect unusual spatial patterns seen in the scene. We detect further the locations where redundancy is low, that is molecule concentrations against a nearly uniform background ideally. In the presence of noise, patch-wise Markov models produce potential maps which are more regular than those obtained with pixel-wise Markov models. In collaboration with UMR 144 CNRS / Curie Institute, our goal was to robustly determine the "origin" and "destination" regions (OD regions) involved in vesicle trafficking. These stable features should be detected in each frame whereas the vesicles are only detected in a few frames in the image sequence. A temporally cumulated detection has been then derived to detect reliably the OD regions/"stocking" areas.

6.1.5. Repetitive and transient event detection in fluorescnce video-microscopy

Participant: Charles Kervrann.

Endocytosis-recycling is an essential cellular trafficking process regulating the proper distribution and function of a large set of molecules, such as lipids, receptors, or adhesion transmembrane proteins. This dynamic process also participates to the homeostasis of intracellular membrane compartments. Progresses in imaging dynamics behaviors of molecules including fast video microscopy and the application of evanescent wave microscopy have allowed to image intracellular vesicular movements, exocytosis and endocytosis of fluorescent-tagged proteins. In parallel, statistical image analysis has emerged as a basic methodology in the study of many biological phenomena. However, spatio-temporal analysis of transient events occurring at different sites of the cell has not been systematically performed. In addition, more formal tests are required in testing biological hypotheses, rather than visual inspection combined with more or less manual statistical analysis. For an unbiased quantification of repetitive and transient events, such as those observed during the trafficking of molecules traveling through the endosomal-recycling network of cells, their automatic detection become necessary. While requiring particular adjustments, our proposed approach is versatile enough, to be applicable to diverse although complementary modes of microscopy. This is illustrated for fast both video and TIRF imaging techniques in collaboration with UMR 144 CNRS / Curie Institute and RICAM (Austria). The proposed detection method can be decomposed into three main steps: i) a first pre-processing step is dedicated to the normalization of the image sequence; ii) the second step is the patch-based detection procedure to detect unusual patterns; iii) a third post-processing step allows us to cluster and count detected events in space and time. While focusing here on one particular Lectin receptor that constitutively recycles from internal compartments to the plasma membrane, it could be translated to many other studies of membrane trafficking in health and diseases such as diabetes, neurological, pigmentation or lysosomal defects.

6.1.6. Dynamic background subtraction in fluorescence video-microscopy

Participants: Charles Kervrann, Anatole Chessel, Thierry Pécot, Patrick Bouthemy.

In collaboration with UMR 144 CNRS / Curie Institute, the main idea was to produce new descriptors able to capture spatio-temporal features of moving particles or group of particles with high variable motions. Nevertheless, the membrane (vesicles) and cytosolic components must be separated since the useful component for traffic estimation correspond to vesicles. Previous parametric and non-parametric estimation methods have been tested for background subtraction (Boulanger et al., 2006). This year, we proposed a promising and simple and original method based on partial differential equations (PDE) and convex hull modeling. The membrane and cytosolic components are then analyzed respectively by the Network Tomography approach for both the vesicle traffic estimation and the free-diffusion estimation in the cytosol. Recently, we have also investigated the mixed-state MRF modeling framework to improve the background subtraction with more flexibility. Finally, this approach will be served to co-localize several molecules belonging either to the background or traffic components and will be compared to other simulation studies in FRAP (Fluorescence Recovery After Photobleaching) imaging we are currently developing.

6.1.7. Change Detection in Satellite Images

Participants: Ivan Laptev, Patrick Pérez.

The Earth surface is constantly changing. Automatic detection of such changes is required for the map updating, environmental studies and in other domains. In this work we study the structural changes caused by human activities in urban areas using temporal sequences of satellite images. We propose to detect structural changes as opposed to other change types by learning appropriate distance metrics for image descriptors. This work was done in collaboration with LiangLiang He and Veronique Prinet from INRIA/Liama lab in Beijing, China during the internship of LiangLiang He in VISTA, April-July 2008.

6.1.8. Detection and segmentation of moving objects in complex scenes

Participants: Guillaume Neveu, Florent Dutrech, Patrick Pérez.

[In collaboration with A. Bugeau (UPF, Barcelona)]
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 applicative contexts where this motion analysis problem is not satisfactorily handled by existing techniques. In the context of activity analysis in dynamically cluttered environments (dynamic background, crowded scenes, etc.) for instance, the problem is the one of separating out foreground moving objects of interest from other uninteresting moving objects in the background. We have proposed a completely automatic system to address this difficult task. It involves three main steps. First, a set of moving points is selected within a sub-grid of image pixels. A multi-cue descriptor is associated to each of these points. Clusters of points are then formed using a variable bandwidth mean shift technique with automatic bandwidth selection. Finally, segmentation of the object associated to a given cluster is performed using graph cuts. Experiments and comparisons to other motion detection methods on challenging sequences demonstrate the performance of the proposed method for video analysis in complex scenes. We are continuing the fine-tuning of this system to further improve its robustness in different applicative contexts: the detection of people in crowed street scenes, with an application to pedestrian monitoring near a train station (ANR-CIPEBUS), and in sport tv broadcast, with an application to video annotation (7.7); the detection of small moving objects in noisy infra-red images, with an application to detection and identification of target in military contexts (7.6). We are in particular: evaluating alternative ways to tune automatically or semi-automatically the selection of the various kernels involved in the first multi-feature clustering step; comparing different ways to compute the dominant motion due to the mouvement of the camera (either based on motion-2d software (5.1 or on a robust regression on sparse KLT vectors)); adapting the form and parameters of the segmentation objective function in case of small size objects in noisy images.

6.1.9. Simultaneous motion detection and background reconstruction with mixed-states models

Participants: Tomas Crivelli, Patrick Bouthemy.

[In collaboration with B. Cernuschi-Frias (Univ. Bueno Aires), G. Piriou and J.-F. Yao (Univ. Rennes)]

The natural extension of the mixed-state framework to a more general case, consists in considering state spaces where the discrete elements are now symbolic values, corresponding to abstract labels. In this context, the proposed framework constitutes a new way of investigating simultaneous decision-estimation problems, allowing to exploit the interaction between values (states) of different nature (discrete-continuous), but that are closely related.

A particular instance of these types of problems is simultaneous motion detection and background reconstruction. In fact, one of the most used methods for motion detection is background subtraction: given a background image, motion detection is solved by computing the difference between the current image and such a reference image. The main problem here is that this reference is not known, and in a general case it is not given or it is not possible to learn it from a training stage, so it has to be estimated. Thus, together with the (motion) detection problem we deal with an estimation problem.

A mixed-state Conditional random field (CRF) was defined where the symbolic part corresponds to a label indicating a positive motion detection and the continuous part corresponds to the background intensity to be estimated. The key is that estimating this field is equivalent to simultaneously solving the moving points and the intensity for the rest. The CRF approach allows introducing arbitrary observations in an energy function to be minimized, with respect to mixed-state values. Here, two sources of information were exploited which contribute to make the overall scheme complete, accurate and powerful. First, an a-contrario approach was used for estimating at each image location the belief of the presence of a moving point, through a detection term. Then the intensity difference with respect to background estimates is included through a reconstruction term, accounting for background subtraction. Finally, smoothness terms allow obtaining compact regions of moving points and at the same time, homogeneous intensity regions for the background, filtering image noise.

Experimental comparisons with existing motion detection methods were performed, showing improved results in the sense of reduction of false positives and negatives, and compactness of segmented regions. Moreover, the simultaneous reconstruction of the background allows solving the video inpainting problem, by a temporal interpolation of the reference image.

6.2. Motion estimation and matching

6.2.1. Schlieren image velocimetry

Participant: Étienne Mémin.

[In collaboration with G. Artana (Univ. Bueno Aires)]
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. 3D atmospheric motion layer estimation

Participants: Patrick Heas, Étienne Mémin.

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 a dense motion estimators for the extraction of multi-layer 3D wind fields. This estimator is 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. The layers are interconnected through a term modeling mass variations at the layers surfaces frontiers

The method has been assessed on Meteosat infrared image sequence. This work has been accepted in IEEE Transactions on Geo-Science on Remote Sensing.

6.2.3. Atmospheric motion estimation with pressure image assimilation

Participants: Étienne Mémin, Heas Patrick, Thomas Corpetti.

[In collaboration with N. Papadakis (UPF, Barcelona)]
The complexity of the laws of dynamics governing 3D atmospheric flows associated with incomplete and noisy observations make the recovery of atmospheric dynamics from satellite image sequences very difficult. We addressed 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 proposed a dynamically consistent atmospheric motion estimator relying on a multi-layer dynamic model. This estimator is based on an optimal control scheme with uncertainty terms (weak constraint variational data assimilation) and is applied on noisy and incomplete pressure difference observations derived from satellite images. The dynamic model is a simplified vorticity-divergence form of a multi-layer shallow-water model. Average horizontal motion fields are estimated for each layer. The performance of the proposed technique has been assessed using synthetic examples and using real world meteorological satellite image sequences. In particular, it is shown that the estimator enables exploiting the finest spatio-temporal image structures and succeeds in characterizing motion at small spatial scales of the image grid. This work has been recently accepted for publication in the journal Tellus Series A: Dynamic Meteorology and Oceanography.

6.2.4. Motion estimation techniques for turbulent fluid flows

Participants: Patrick Heas, Étienne Mémin.

[In collaboration with G. Artana and P. Minini (Univ. Bueno Aires), D. Heitz (Cemagref, Rennes)]
In order to estimate the more accurately as possible motion fields of turbulent flows, we have proposed a technique relying on a self-similarity constraint based on fractal models of turbulence proposed by Kolmogorov (1941). This constraint provides us implicitly a multi-scale smoother enabling a closure to the optic-flow estimation problem. Regularization is achieved by constraining motion increments to behave as a fractal process. The associate constrained minimization problem results in a collection of first-order optic-flow smoothing functions acting at different scales. The problem has been optimally solved by taking advantage of Lagrangian duality. Such a technique has besides the advantage to allow the estimation of the so-called smoothness parameter that balances the fidelity to the photometric data term and the a priori constraint on the solution (here the respect of Kolmogorov energy cascade). Since, the fractal model parameters observed in real cases can deviate from the Kolmogorov predictions, a Bayesian learning stage has been added in the process. We are currently evaluating this technique on particle image sequences and on a meteorological image sequence. We plan also to assess it on an experimental setup allowing generating real 2D turbulent flows.

6.2.5. Dynamic consistent correlation-variational motion estimation

Participants: Heas Patrick, Étienne Mémin.

[In collaboration with J. Carlier and D. Heitz (Cemagref, Rennes)]

We have proposed a novel collaborative motion estimation scheme dedicated to the measurement of velocity in fluid experimental flows through image sequences. The proposed technique satisfies the Navier-Stokes equations and combines the robustness of correlation techniques with the high density of global variational methods. It can be considered either as a reenforcement of fluid dedicated optical flow methods towards robustness, or as an enhancement of correlation approaches towards dense information. This results in a physics-based technique that is robust with respect to noise and outliers, while providing a dense motion field. The method has been applied on synthetic images and on real experiments in turbulent flows carried out to allow a thorough comparison with a state of the art variational and correlation methods.This work has been recently accepted for publication in the journal Experiments in Fluids.

6.2.6. 3D flows reconstruction from image data

Participants: Pierre Cariou, Étienne Mémin.

[In collaboration with D. Heitz (Cemagref, Rennes)]
In the continuity of the PhD of Nicolas Papadakis, we want to study techniques for the complete reconstruction of 3D flows from image data sequences. In a first approach we will study the extension to 3D of our different 2D fluid motion estimators from 3D volumetric data obtained through tomographic reconstruction. In a second time, we wish to explore the reconstruction of flows observed simultaneously on several 2D plans. The inclusion of a flow dynamical model will allow us to filter erroneous data and also to reconstruct the missing unobserved components. For computational reasons only large scale models or reduced model are conceivable. We will focus both on experimental flows and on atmospheric flows. In this later case, we will explore techniques for the reconstruction of atmospheric layers observed from satellite images. This will be a continuation of the works conducted on the estimation of layered 2D/3D atmospheric winds.

6.2.7. Dynamically consistent motion estimation

Participants: Étienne Mémin, Thomas Corpetti.

[In collaboration with N. Papadakis (UPF, Barcelona)]
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.

6.2.8. Stochastic uncertainty models for motion estimation

Participants: Thomas Corpetti, Étienne Mémin.

We have worked on a stochastic interpretation of the motion estimation problem. The usual optical flow constraint equation (that assumes that the points keeps their brightness under time), embedded for instance within a Lucas-Kanade estimator, can indeed be seen as the minimization of a stochastic process under some strong constraints (the luminance as a function of a stochastic process is a martingal with isotropic diffusion). This constraint can be relaxed by imposing weaker assumption on the luminance function and also in introducing anisotropic intensity-based uncertainty assumptions. The amplitude of these uncertainties are jointly computed with the unknown velocity at each point of the image grid. Different estimators have been designed depending on the various hypothesis assumed for the luminance function. The substitution of our new observation terms on a simple Lucas-Kanade estimator improves significantly the quality of the results (on the basis of optical flow datasets with ground truth available on the web) and leads to a local motion estimator without any parameters. We believe that the definition of a dense estimator based on the very same ideas will be very competitive.

6.2.9. Semi-automatic video segmentation and manipulation with motion layer mosaics

Participants: Matthieu Fradet, Patrick Pérez.

We have developed a new method for motion segmentation based on reference motion layer mosaics. We assume that the scene is composed of a set of layers whose motion is well described by parametric models. This usual assumption is compatible with the notion of motion layer mosaic, which allows a compact representation of the sequence with a small number of mosaics only. We segment the sequence using a reduced number of distant image-to-mosaic comparisons instead of a larger number of close image-to-image comparisons. Apart from computational advantage, another interest lies in the fact that motions estimated between distant images are more likely to be different from one region to another than when estimated between consecutive images. This helps the segmentation process. The segmentation is obtained by graph cut minimization of a cost function which includes an original image-to-mosaic data term. At the end of the segmentation process, it may happen that the obtained boundaries are not precisely the expected ones. Often the user has no other possibility than modifying manually every segmentation one after another or than starting over all again the process with different parameters. We propose an original easy way for the user to manually correct the possible errors on the mosaics themselves. These corrections are then propagated to all the images of the corresponding video interval thanks to a second segmentation pass.

6.2.10. Motion-based intra-scene and inter-scenes video alignment

Participants: Émilie Dexter, Ivan Laptev, Patrick Pérez.

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. 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 "alignement" of multi-camera videos. 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 exploit the stability across views of temporal self-similarity (related to autocorrelation) that can be extracted from each video, based on several types of low-level features (spatial gradients, optical flows, KLT pieces of point tracks, see 6.4.2). Promising synchronization results have been obtained on a variety of data, ranging from mocap data and simple video data with static cameras to complex sport shots from tv broadcasts with moving cameras. Using dynamic time warping, the time mapping from one sequence to the other is only constrained to be monotonic, but can be arbitrary otherwise. This flexibility, not permitted by state-of-art synchronization methods, has been further exploited to synchronize video of the same class of human action seen from drastically different view points and performed by different people (hence involving different speeds and accelerations).

6.3. Tracking

6.3.1. Stochastic filtering for fluid motion tracking

Participant: Étienne Mémin.

[In collaboration with A. Cuzol (UBS, Vannes) and N. Papadakis (UPF, Barcelona)]
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. This work has been recently accepted for publication in IEEE Journal on Pattern Analysis and Machine Intelligence.

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 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, Claire Thomas.

[In collaboration with N. Papadakis (UPF, Barcelona)]
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 approach has been applied for the tracking of deformable object and for the tracking of meteorological systems such as cyclone or convective cells.

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 in Journal of Mathematical Imaging and Vision and SIAM Journal on Imaging Science.

6.3.3. Reduced order models for wake flows

Participant: Étienne Mémin.

[In collaboration with G. Artana, J. D'adamo (Univ. Bueno Aires)], J. Carlier, D. Heitz (Cemagref, Rennes)] In this work we have studied techniques allowing a robust estimation of reduced order model (ROM) for wake flows from flow velocity field experimental data. The proposed method relies on a proper orthogonal decomposition method (POD) and a Galerkin projection. The POD technique allows the extraction of a reduced number of functions or modes that describes at best the flow's velocity field (in the sense of a mean kinetic energy reconstruction error). Galerkin projection of Navier Stokes equation onto these modes supplies a system of ordinary non linear equations. The sole dimension reduction from Galerkin procedure does not assure the convergence and stability of solutions. Recently an efficient POD data assimilation scheme we previously proposed allowed us to handle this instability problem in the time range of the data sequence. In order however to provide an unconditionally stable system and to be able to provide a stable flow prediction beyond the data sequence time range, recent studies on hydrodynamic stability in wake flows suggest model modifications of the POD-Galerkin ROM in agreements with center manifold theory. In this context we have proposed to couple a data assimilation strategy with center manifold theory to achieve improvements for the POD-based ROM from velocity fields experimental data. For the moment the method has been validated only on noise free synthetic data provided by direct numerical simulations (DNS).

6.3.4. Stochastic filtering technique for the tracking of closed curves

Participants: Christophe Avenel, Etienne Mémin, Patrick Pérez.

We have proposed 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.

This work puts forth a probabilistic graphical framework to track un-occluded objects undergoing large out of image plane rotations and/or presenting large scale variations in video sequences. The proposed scheme incorporates measurements from an ensemble of local patch trackers and inter-patch geometric layout to arrive at a sample based approximation of the state posterior. Following this, the geometric layout is updated on-line using the Iterative Conditional Estimation technique. These steps are iterated until convergence to arrive at the final state posterior. In contrast to off-line training based schemes the proposed framework imposes no prior on the geometric layout and instead relies on on-line update of the geometric layout, thus broadening the scope of usage. Amongst other advantages, the scheme implicitly estimates the scale of the target and also adapts to varying target appearances to enable tracking under a fair degree of out of the image plane rotations. The tracking abilities of this scheme is put to test on several challenging videos with scale changes, out of the image plane rotations, illumination changes and motion jerks.

6.3.6. Simultaneous tracking and segmentation of multiple objects with graph cuts

Participant: Patrick Pérez.

[In collaboration with A. Bugeau (UPF, Barcelona)]
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 are currently investigating a combination of level-set models with stochastic dynamic filtering. On the discrete side, we have developed a new method to both track and segment multiple objects in videos using min-cut/max-flow optimizations.We introduce 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. One novelty is that external observations are used without adding any association step. The observations are image regions (pixel sets) that can be provided by any kind of detector. The minimization of appropriate cost functions simultaneously allows detection-before-track tracking (track-to-observation assignment and automatic initialization of new tracks) and segmentation of tracked objects. When several tracked objects get mixed up by the detection module (e.g., a single foreground detection mask is obtained for several objects close to each other), a second stage of minimization allows the proper tracking and segmentation of these individual entities despite the confusion of the external detection module.

6.3.7. Person detection and tracking in archive vidos

Participants: Ivan Laptev, Patrick Pérez.

The rapidly growing video archives with movies and documentaries to their large extent are focused on people, their behavior, relations and actions. Automatic reasoning about this data would benefit from reliable detection and tracking of people in video. Towards this goal we initiated the work on object and person tracking in archive video. To detect temporally consistent tracks of objects, we combined per-frame object detections with KLT point tracks. The approach relies on Agglomerative Clustering and takes advantage of the pre-recorded nature of archive video. The system is able to reliably detect tracks of objects and people in video across minor occlusions and gradual pose variations. In the future we will improve robustness of the system by integrating alternative image features and the on-line learning of visual appearance. Associating identity of people and objects across video shots, scenes and different videos is a long-term extension of this work. This work was conducted within the internship of A. Mittal (ITT Roorkee, India)

6.4. Dynamic event modeling, learning and recognition

6.4.1. Learning realistic human actions from movies

Participant: Ivan Laptev.

[In collaboration with M. Marszałek and C. Schmid (LEAR, Inria Grenoble), B. Rozenfeld (Bar-Ilan University, Israel)]
We addressed recognition of natural human actions in diverse and realistic video settings. This challenging but important subject has mostly been ignored in the past due to several problems one of which is the lack of realistic and annotated video datasets. Our first contribution was to address this limitation and to investigate the use of movie scripts for automatic annotation of human actions in videos. We evaluated alternative methods for action retrieval from scripts and showed benefits of a text-based classifier. Using the retrieved action samples for visual learning, we next turned to the problem of action classification in video. We presented a new method for video classification that builds upon and extends several recent ideas including local space-time features, space-time pyramids and multichannel non-linear SVMs. The method has been shown to improve state-of-the-art results on the standard KTH action dataset by achieving 91.8% accuracy. Given the inherent problem of noisy labels in automatic annotation, we particularly investigated and showed the high tolerance of our method to annotation errors in the training set. We finally apply the method to learning and classifying challenging action classes in movies and show promising results for eight classes of human actions: "Answer Phone", "Get Out Car", "Hand Shake", "Hug Person", "Kiss", "Sit Down", "Sit Up", "Stand Up". The results of this work were presented at Int. Conf. Computer on Vision and Pattern Recognition (CVPR), Anchorage, Alaska in June 2008.

6.4.2. *Cross-View Action Recognition from Temporal Self-Similarities*

Participants: Imran Junejo, Émilie Dexter, Ivan Laptev, Patrick Pérez.

Appearance of dynamic events and human actions in video depends on the position and motion of the camera. The aim of the work here is to construct methods and representations to enable view-independent event reasoning in image sequences. We conducted the study of motion recognition and proposed new view-invariant motion representations. The main objective was to address appearance variations in image sequences due to view-point changes without resorting to classic multi-view geometry methods. Such methods relying on explicit structure recovery or stereo-matching often become unreliable for videos with non-rigid (human) motion and/or low compression quality. To overcome this problem we have developed a new motion representation based on self-similarity of image sequences over time. We have made initial tests of this representation on the tasks of multi-view action recognition and multi-view temporal sequence synchronization. We have obtained good results that favorably compare to results of other methods on the multi-view IXMAS action dataset. The results of this work were presented at European Conference of Computer Vision (ECCV), Marseille, France in October 2008. In future work we plan to further investigate the proposed representations and to applying them to more realistic datasets and new problems such as action class discovery in video.

6.4.3. *Video content recognition using trajectories*

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

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 summaries 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. 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, to rotation and to scaling. We have tackled three important tasks related to dynamic video content understanding within the same trajectory-based framework: unsupervised clustering of trajectories; supervised recognition of events; detection of 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.

Another way of research is to consider a content-based approach for activity-based temporal segmentation of videos. To this end, hierarchical Semi-Markov Chains (HSMCs) are introduced and used to process trajectories of several interacting moving objets. In these models, the temporal interactions between trajectories are taken into account and used to segment observed activities into relevant phases. Our method has been evaluated on trajectories extracted from squash and handball videos. It compares favorably with other methods, especially with Hierarchical Hidden Markov Models (HHMMs). Application of such interaction-based models has also been extended to other tasks such that 3D gesture and action recognition, and temporal segmentation of actions. The results show that taking into account the interactions between body parts movements may be of great interest for such applications.

6.4.4. 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.

6.4.5. Planning surveillance

Participants: Jean-Pierre Le Cadre, Cécile Simonin-Champolion.

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. 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.

A second way of research addresses the problem of cross-cueing. In this setup, the goal is to maximizes chances to detect a target and to confirm the detection, either by considering multiple means of search or by a unique mean of search at multiple time periods. Extensions to multiple periods and Markovian target is done within a common Forward and Backward (FAB) framework.

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 is 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. grant with Curie Insitute on membrane transport/traffic estimation

Participants: Charles Kervrann, Thierry Pécot.

no. xxx, duration 11 months.

This contract started in July 2008 is associated with the supervision of T. Pécot's thesis funded by Curie Institute for 11 months (previously INRA & INRIA). It concerns the problem of traffic estimation and membrane transport modeling in cell biology. In this study, we have investigated an alternative approach to conventional tracking methods and based on the concept of Network Tomography (NT) (Vardi, 1996). In this approach a dynamic scene made up moving particles along a dense microtubule set, is modeled as a network of interconnected regions of interest. In such a network, a connection between two nodes is called a path and each path consists of one or more unidirectional or bidirectional links. Each link can represent a chain of 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 an object through a dynamic scene, just to determine when an object reaches a node, something that is generally easier than estimating a continuous trajectory. This approach simplifies the tracking process because it only requires detection of an object as it moves from one region to another. In exchange, it gives up the ability to estimate an object's state as the motion occurs. Instead, it determines mean traffic intensities based on statistics accumulated over a period of time. It only provides the total number of trips made for which statistics were collected. The measurements are usually the number of vesicles successfully detected at each destination region receiver or the vesicle time between the source and each destination. The inherent randomness in both link-level and path-level measurements motivates the adoption of statistical methodologies.

7.2. 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.

7.3. 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 7.2 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.

7.4. 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.

7.5. DGA (REI): Dynamic scene analysis with camera networks

Participants: Jean-Pierre Le Cadre, Adrien Ickowicz.

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 synchronization.
- False alarms, multiple objects, occlusions, etc.

Overall, there is a 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.

Work 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.

Concerning the information processing at the sensor level, work has been done in three directions. First, the $\{-,+\}$ informations can be summarized via the times of CPA (closest point approach) on the various sensors. A complete analysis of the target motion analysis has been done in this framework, with various models of target motion. Then, we turned toward the analysis of the spatio-temporal $\{-,+\}$ informations. An original method based on a separation principle allows us to obtain an estimate of the target motion parameters via multiple SVM. Second, we considered random target trajectories and target tracking. To that aim, a specific method based on multiple corrections has been developed.

7.6. DGA (2ACI): Detection, recognition and identification of targets in infra-red sequences

Participants: Florent Dutrecht, Patrick Pérez.

no. Inria 2920, duration: 20 months.

The general context of this contract is the detection, the recognition and the identifications of various types of targets in infra-red videos, for different scenarios that interest DGA: camera fixed on a pole (possibly subject to small vibrations) for the surveillance of critical environments; camera mounted a ground vehicle, or held by soldiers; camera on drones. Other partners are Univ. of Caen (F. Jurie), Willow Inria project-team (J. Ponce) and Bertin technologies. We are in charge of the development and the quantitative assessment of various techniques for the detection of moving targets and their tracking through time.

7.7. CRE Orange Labs: Tools for motion-based detection and tracking in videos

Participants: Guillaume Neveu, Patrick Pérez.

no. Inria 2666, duration: 22 months.

In this contract with Orange Labs (Lannion, then Rennes), we aim at extending the tracking platform developed in a preceding contract to other low-level motion analysis tasks, namely: robust dominant motion estimation, automatic extraction of points with consistent color and motion, detailed segmentation of image region associated to these point clusters and tracking of these regions through time. The applications concern the analysis of tv programs, sport broadcasts in particular.

7.8. MS-Inria Video data mining for environmental and social sciences

Participants: Ivan Laptev, Etienne Mémin, Patrick Pérez.

duration 36 months.

This contract takes place in the context of the joint Microsoft-Inria research laboratory. It involves two other Inria project-teams, Willow and Lear. The projects is composed of three fairly disjoint sub-projects. Vista contributes to two of them: (i) Mining dynamical remote data with applications in computational ecology and environmental science; (ii) Mining TV broadcasts with applications to sociology. In the first sub-project, we aim at combining various low and mid-level video analysis tools (shot detection, camera motion characterization, visual tracking, object recognition, human action recognition) for the analysis and annotation of human actions and interactions in video segments to assist and provide data for studies of consumer trends in commercials, political event coverage in newscasts, and class- and gender-related behavior patterns in situation comedies, for example. In the second one, we aim at designing new tools for the detection of salient changes in multi-temporal satellite images (with application to assessment of natural damages, consequences of climate modifications, and changes caused by human action in urban or natural environments), and, in the longer term, for the detection, identification and tracking of dynamics meteorological events, with application to risk assessment and weather forecast.

7.9. European initiatives

7.9.1. FP6 PEGASE: helicoPter and aEronef naviGation Airborne System Experimentation

Participants: Jean-Pierre Le Cadre, Nam Trung Pham.

no. Inria 1832, duration 36 months

PEGASE is a multidisciplinary European project involving 15 partners (industrial and academic). It is headed by Dassault Aviation (kick-off-meeting October 2006). It is widely recognized that 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 navigation approach. The purpose of the PEGASE project is to prepare the development of an autonomous localization and guidance system based upon correlation between vision sensors output and a ground reference database. The work package WP6 is 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 navaid system, in connection with other WPs. Concurrent image processing methods are 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. An important work has been done for solving these two tasks. Difficulties due to the flight model have been solved and various observation models have been considered. So, we have now a robust and reliable tracking method which can still be improved by fusing with the results of pose estimation.

8. Other Grants and Activities

8.1. Regional initiatives

8.1.1. PRIR Britanny council grant : Imafluid

Participants: Etienne Mémin, Patrick Heas.

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, Claire Thomas.

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. STIC-AMSUD Modeling voice production

Participant: Etienne Mémin.

duration 36 months.

This project aim 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).

8.2.3. CNRS-INRIA MINCyT: Reduced dynamical modeling

Participants: Patrick Heas, Etienne Mémin.

duration 36 months.

This bilateral collaboration project between Argentina and France is a three years project that gathers the ESPCI (E. Weifreid) the ENS-LMD (M. Farge) and the University of Buenos Aires (G. Artana). It concerns the study of estimation techniques of low order dynamical from image sequences with the aim of providing recursive and adaptable techniques to go toward the possibility of flow control through visual servoing.

8.2.4. ARC INRIA (supported also by INRA): 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 (confocal spinning-disk microscopy, Fluorescence Recovery After Photobleaching microscopy, Fluorescence Lifetime Microscopy Imaging). 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 Rab dynamics and a rich set of data (mostly image sequences in video-microscopy) will support the analysis. The biologist partners use very powerful and novel molecular (RNA interference), mechanical (micro-patterning) and optical (FRAP, photoablation) tools to normalize and perturb the cell activity.

8.2.5. ANR-Predit Cipebus: Intelligent monitoring of street traffic in dense environments with buses

Participant: Patrick Pérez.

duration 36 months.

The partners of this project are Inrets (leader), Citilog, Versailles city council, Thales, Fareco, SVUT. The overall goal is to study new traffic monitoring tools to improve the quality and safety of ground public transportation in dense environments. The experimentations will be conducted on a real site, namely the vicinity of Versailles-Chantiers train station, which is currently being equipped for that purpose (five successive crossings will receive, among other sensors, around 30 video cameras on poles). In this context, we work together with Citilog on the vision system, with a special emphasis on monitoring pedestrians. Our goal is the conception and in-situ validation of monocular algorithms for the detection of (groups of) individuals, their tracking, the analysis of their dynamics as well as the one of nearby vehicles, on and around zebra crossings.

8.2.6. ANR Icos-HD: Scalable Indexing and Compression for High Definition TV

Participant: Patrick Pérez.

duration 36 months.

The partners of this projects are Temics and Texmex Inria project-teams; Labri (Bordeaux), L2S (Nice), OrangeLabs (Rennes). The objective of the project is to develop new solutions of scalable description for High Definition video content to facilitate their editing, their access via heterogeneous infrastructures (terminals, networks). The introduction of HDTV requires adaptations at different levels of the production and delivery chain. The access to the content for editing or delivery requires associating local or global spatio-temporal descriptors to the content. These descriptors must allow the collection of information related to actions, events or activities taking place in the video document and which can happen at different spatial and temporal resolutions. In this project, we aim at devising scalable descriptors that take motion explicitly into account, either on a time-localized basis (spatio-temporal interest points in particular), or on a longer-term basis (partial tracks on points or objects).

8.2.7. ARC INRIA BioSerre: Sensor network for early detection of bio-aggressors in greenhouse cultures

Participant: Patrick Pérez.

duration 36 months.

The other partners are Pulsar Inria projet-team (leader) and Inra. The context of this project is the design of technologies that helps rationalizing the use of pesticides in cultures. The overall aim here is to rely on a network of ip cameras spread over the greenhouse to monitor the flux of bugs. Their early and localized detection would than allow for a reduced and more effective use of pesticides, as well as the use of biological alternatives (other bugs). In this context, we aim at contributing to the design of simple and robust methods fot the detection of small bugs in leaves that are undergoing small hanging motions as well as appearance changes (illumination changes over the day, and actual changes of the leaf's appearance).

8.2.8. QUAERO

Participants: Ivan Laptev, Patrick Pérez.

no. Inria 3184, duration 60 months

Quaero is a European collaborative research and development program with the goal of developing multimedia and multi-lingual indexing and management tools for professional and public applications. The program has been approved by European Commission on 11 March 2008. The program is planed for five years and is supported by the French government through OSEO with a total budget of 200 millions Euro. Quaero consortium involves 24 academic and industrial partners leaded by Thomson. Vista participates in the Work Package 9 on Video Processing (WP9) of QUAERO Core Technology Cluster Project (CTC). Within WP9 Vista leads three tasks: Motion Recognition (Task 9.2), Object Tracking (Task 9.3) and Event Recognition (Task 9.4). Vista is currently hiring non-permanent stuff including one PhD student and one engineer who will be working on motion and event recognition in videos.

8.2.9. 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".
- P. Pérez is the administrative head of the ICOS-Sud project on Movie Restoration, whose participants are Telecom Paris Tech, Univ. Pierre and Marie Curie, ENS Cachan, Univ. of Monte Video.

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.

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...).

9. Dissemination

9.1. Leadership within scientific community

- *Editorial boards of journals*
  - J.-P. Le Cadre is Area Editor (target Tracking) of Journal of Advances in Information Fusion (ISIF);
  - P. Pérez is Associate Editor for the IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI).
- *Technical program committees of conferences*
  - P. Bouthemy: TPC member of ACIVS'2008, CIVR'2008, MMM'2009.
  - C. Kervrann: TPC member of ECCV'2008, MIAAB'2008, RFIA'2008, ACIVS'2008, ICPR'2008.
  - I. Laptev:
  - J.-P. Le Cadre: TPC member and award committee member Fusion'2008 (Cologne), TPC member of RADAR'2008 (Roma).
  - E. Mémin: TPC member of VISAPP '2008.
  - P. Pérez: Area chair of ECCV'2008 and RFIA'2008, chair of a session of ECCV'2008), TPC member of SIGGRAPH'2008, EUROGAPHICS'2008
- *Ph.D. reviewing*
  - J.-P. Le Cadre: A. Boubezoul (LSIS, Marseille), A. Kazem (LAAS Toulouse).
  - E. Mémin: T. Isambert (Univ. Paris 5), C. Baehr (Univ. Toulouse 3), V. Gay-Bellile (Univ. Blaise Pascal)
  - P. Pérez: S. Boltz (I3S/Creative, Univ. Nice), A. el Abed (LiP6, Univ. Paris 6), M. Fontmary (LAAS Toulouse), D. Larlus (Inria Grenoble, INPG).
- *HdR reviewing*
  - E. Mémin: G. Le Besnerais (ONERA, HDR Univ. Paris 13)
- *Project reviewing, consultancy, administrative responsibilities*
  - P. Bouthemy is director of the Inria centre Rennes-Bretagne Atlantique and of Irisa. He is member of the Board of the scientific association AFRIF (Association Française pour la Reconnaissance et l'Interprétation des Formes). He is member of the Board of the scientific association GRETSI. He serves as a regular expert for the MRIS("Mission pour la Recherche et l'Innovation Scientifique") of the French Defense Agency (DGA). He was also member of the steering committee of the CPER Poitou-Charentes program on "Images and Interactivity."

  - 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 a scientific project for the Ville-de-Paris. 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.
  - J.-P. Le Cadre is a member of the evaluating instance for "space, observation, intelligence and UAV" (DGA).
  - 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”).

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).
- 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)
- Collège de Polytechnique, seminar on information fusion: J.-P. Le Cadre made a one day course about data association and sensor management.
- Graduate student trainees and interns :
  - C. Avenel (ENS Cachan, Brittany, supervised by E. Mémin and P. Pérez), “Bayesian sequential tracking of closed curves”.
  - S. Blestel (Master STI, Rennes 1, supervised by C. Kervrann), "3D microtubule detection cryo-EM (Electronic Microscopy)".
  - L.-L. He (Beijing University), supervised by I. Laptev and P. Pérez, "Change detection and classification in satellite image pairs with supervised learning".
  - A. Mittal (ITT Bangalore), supervised by I. Laptev and P. Pérez.
- External thesis supervision :
  - R. Cariou (DGA-CELAR) supervised by J.-P. Le Cadre;
  - C. Guilmart (secondment from "Corps de l'armement"', DGA-ONERA) supervised by P. Pérez;

9.3. Participation in seminars, invitations, awards

- C. Kervrann was an invited speaker at the SIAM Conference on Imaging Science and gave the following talk entitled “Bayesian Non-local Means, Image Redundancy and Adaptive Estimation for Image Representation and Applications” (San-Diego, July 2008). C. Kervrann gave the following invited talk: “Dynamic modeling for molecular machines and membrane transport” (INRA, Paris March 2008 and Jouy-en-Josas, November 2008)”. He participated to the summer school Mifobio'08 (Carqueiranne, September 2008) organized by the GDR 2588 - CNRS. He presented the DYNAMIT project in the oral session at the National Meeting of Inria ARC projects (INRIA Sophia-Antipolis, October 2008).

- I. Laptev gave invited talks at International Workshop On Object Recognition, Lake Como, Italy, May 2008 and in GIPSA-Lab, Grenoble, November 2008. I. Laptev gave invited lectures at the Summer School on Machine Learning, Statistics and Computer Vision, Ezhou, China, July 2008. I. Laptev recieved outstanding review award at ECCV 2008.
- J.-P. Le Cadre presented a poster ("Traitement de l'information pour le suivi de cibles dans un réseau de capteurs") at the REI/DGA meeting in March 2008 and a talk ("Closed-form PCRB) at the GdR Isis day devoted to "Bounds", Sept. 2008, Paris.
- T. Bréhard and J.-P. Le Cadre received the IEEE Barry Carlton Award from the IEEE AESS.
- P. Dodin, P. Minvielle and J.-P. Le Cadre have been selected by the "Chocs" review of the CEA/Dam.
- A. Ickowicz and J.-P. Le Cadre received the Best Student Paper award at ISSNIP'08.
- E. Mémin has been invited to give talks on Fluid flows analysis from image sequence at the National Academy of Sciences of Buenos-Aires, at the LMD (Ecole Normale Supérieure); he gave an invited talk at the SMAI MAS workshop on visual tracking of entities with non-linear dynamics, and an invited talk at the SMART/NISA Workshop at the Univ. of Copenhagen.
- N. Papadakis received the best thesis award of the AFRIF association and the best thesis award of the EADS fundation in the category "Earth and Univers Sciences".
- P. Pérez was invited at the Microsoft Research Workshop on Markov Random Fields, where he was chair of a session, and gave invited lectures at the SMART/NISA Workshop on Probabilistic Tracking (Copenhagen University), and at the ECOS Workshop on Film Restoration (Monte Video University).

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, n° 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° 4, 2006, p. 1240–1263.

[3] P. BOUTHEMY, C. HARDOUIN, G. PIRIOU, J. YAO. *Mixed-state auto-models and motion texture modeling*, in "Journal of Mathematical Imaging and Vision", vol. 25, n° 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° 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° 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° 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° 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^o 2, 2002, p. 153–174.

[9] I. LAPTEV. *On space-time interest points*, in "Int. J. Computer Vision", vol. 64, n^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^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^o 2, 2006, p. 163–178.

Year Publications

Articles in International Peer-Reviewed Journal

[12] J. BOULANGER, C. KERVRANN, P. BOUTHEMY. *A simulation and estimation framework for intracellular dynamics and trafficking in video-microscopy and fluorescence imagery*, in "Medical Image Analysis", 2008.

[13] A. BUGEAU, P. PÉREZ. *Detection and segmentation of moving objects in complex scenes*, in "Computer Vision and Image Understanding", 2008.

[14] A. BUGEAU, P. PÉREZ. *Track and Cut: simultaneous tracking and segmentation of multiple objects with graph cuts*, in "EURASIP Journal on Image and Video Processing (doi:10.1155/2008/317278)", vol. 2008, n^o 317278, 2008, p. 1-14, http://www.irisa.fr/vista/Papers/2008_eurasip_bugeau.pdf.

[15] T. CORPETTI, P. HEAS, E. MÉMIN, N. PAPADAKIS. *Pressure image assimilation for atmospheric motion estimation*, in "Tellus Series A: Dynamic Meteorology and Oceanography", vol. In Press, 2008.

[16] P. COUPÉ, P. HELLIER, S. PRIMA, C. KERVRANN, C. BARILLOT. *3D wavelet sub-bands mixing for image denoising*, in "International Journal of Biomedical Imaging", vol. 8, n^o 3, 2008, p. 1-11.

[17] 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", vol. 27, n^o 4, 2008, p. 325-441.

[18] A. CUZOL, E. MÉMIN. *A stochastic filtering technique for fluid flows velocity fields tracking*, in "IEEE Trans. on Pattern Analysis and Machine Intelligence (accepted for publication)", 2008.

[19] D. HEITZ, P. HEAS, E. MÉMIN, J. CARLIER. *Dynamic consistent correlation-variational approach for robust optical flow estimation*, in "Experiments in Fluids", vol. In Press, 2008.

[20] A. HERVIEU, P. BOUTHEMY, J.-P. LE CADRE. *A statistical video content recognition method using invariant features on object trajectories*, in "IEEE Trans. on Circuits and Systems for Video Technology (Special Issue on "Event Analysis in Videos"", vol. 18, n^o 11, 2008, p. 1533-1543, http://www.irisa.fr/vista/Papers/2008_csvt_hervieu.pdf.

[21] P. HÉAS, E. MÉMIN. *3D motion estimation of atmospheric layers from image sequences*, in "IEEE Trans. on Geoscience and Remote Sensing", vol. 46, n^o 8, 2008, p. 2385-2396, http://www.irisa.fr/vista/Papers/2007_IEEEGRS_HeasMemin.pdf.

[22] C. KERVRANN, J. BOULANGER. *Local adaptivity to variable smoothness for exemplar-based image denoising and representation*, in "International Journal of Computer Vision", vol. 79, n^{o} 1, August 2008, p. 45-69, http://www.irisa.fr/vista/Papers/2007_ijcv_kervrann.pdf.

[23] I. LAPTEV. *Improving object detection with boosted histograms*, in "Image and Vision Computing", 2008.

[24] N. PAPADAKIS, E. MÉMIN. *A variational technique for time consistent tracking of curves and motion*, in "Journal of Mathematical Imaging and Vision", vol. 31, n^{o} 1, 2008, p. 81-103, http://www.irisa.fr/vista/Papers/2008_jmiv_papadakis.pdf.

[25] N. PAPADAKIS, E. MÉMIN. *Variational assimilation of fluid motion from image sequences*, in "SIAM Journal on Imaging Science", 2008, http://www.irisa.fr/vista/Papers/2008_siims_papadakis.pdf.

[26] C. SIMONIN, J.-P. LE CADRE, F. DAMBREVILLE. *A hierarchical approach for planning a multisensor multizone search for a moving target*, in "Computers and Operations Research", 2008.

[27] R. SOSA, E. ARNAUD, E. MÉMIN, G. ARTANA. *Study of the Flow Induced by a Sliding Discharge*, in "EEE Transactions on Dielectrics and Electrical Insulation (accepted for publication)", 2008.

Invited Conferences

[28] C. KERVRANN, P. PÉREZ, J. BOULANGER. *Bayesian Non-local means, image redundancy and adaptive sstimation for image representation and applications*, in "SIAM Conf. on Imaging Science, San Diego, CA", July 2008.

International Peer-Reviewed Conference/Proceedings

[29] V. BADRINARAYANAN, P. PÉREZ, F. LE CLERC, L. OISEL. *Geometric layout based graphical model for multi-Part object tracking*, in "Proc. ECCV'08, 8th Workshop on Visual Surveillance, Marseille, France", October 2008.

[30] J. BOULANGER, J.-B. SIBARITA, C. KERVRANN, P. BOUTHEMY. *Non-parametric regression for patch-based fluorescence microscopy image sequence denoising*, in "Proc. IEEE Int. Symp. on Biomedical Imaging: from nano to macro (ISBI'08), Paris, France", May 2008, p. 748-751.

[31] A. BUGEAU, P. PÉREZ. *Track and Cut: simultaneous tracking and segmentation of multiple objects with graph cuts*, in "Proc. Int. Conf. on Computer Vision Theory and Applications (VISAPP'08), Funchal, Portugal", vol. 2, January 2008, p. 447-454, http://www.irisa.fr/vista/Papers/2008_visapp_bugeau.pdf.

[32] F. CELESTE, F. DAMBREVILLE, J.-P. LE CADRE. *Path planning for multiple features based localization*, in "Proc. Int. Conf. on Inf. in Control, Aut. and Robotics (ICINCO), Funchal, Portugal", May 2008, http://www.irisa.fr/vista/Papers/2008-icinco-celeste.pdf.

[33] A. CHESSEL, R. FABLET, C. KERVRANN, F. CAO. *Otolith image analysis by computer vision: extracting growth rings and recovering shape evolution of accretionnary structures*, in "Proc. Int. Conf. on Bio-inspired Systems and Signal Processing (BIOSIGNALS'08), Funchal, Portugal", vol. 2, January 2008, p. 490-497.

[34] T. CORPETTI, P. HÉAS, E. MÉMIN, N. PAPADAKIS. *Variational pressure image assimilation for atmospheric motion estimation*, in "Proc. Int. Geoscience and Remote Sensing Symp. (IGARSS'08), Boston, MA", July 2008.

[35] P. COUPÉ, P. HELLIER, C. KERVRANN, C. BARILLOT. *Bayesian Non-Local Means-based speckle filtering*, in "Proc. IEEE Int. Symp. on Biomedical Imaging: from nano to macro (ISBI'08), Paris, France", May 2008, p. 291-1294.

[36] T. CRIVELLI, B. CERNUSCHI-FRIAS, P. BOUTHEMY, J. YAO. *Recognition of dynamic video contents based on motion texture statistical models*, in "Proc. Int. Conf. on Computer Vision Theory and Applications (VISAPP'08), Funchal, Portugal", vol. 1, January 2008, p. 283-289.

[37] T. CRIVELLI, B. CERNUSCHI-FRIAS, P. BOUTHEMY, J. YAO. *Temporal modeling of motion textures with mixed-states Markov chains*, in "Proc. Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP'08), Las Vegas, USA", April 2008, p. 881-884.

[38] T. CRIVELLI, G. PIRIOU, P. BOUTHEMY, B. CERNUSCHI-FRIAS, J. YAO. *Simultaneous motion detection and background reconstruction with a mixed-state conditional Markov random field*, in "Proc. Eur. Conf. Computer Vision (ECCV'08), Marseille, France", vol. 1, October 2008, p. 113-126.

[39] M. FRADET, P. PÉREZ, P. ROBERT. *Semi-automatic motion segmentation with motion layer mosaics*, in "Proc. Eur. Conf. Computer Vision (ECCV'08), Marseille, France", vol. 3, October 2008, p. 210-223.

[40] M. FRADET, P. PÉREZ, P. ROBERT. *Time-sequential extraction of motion layers*, in "Proc. Int. Conf. on Image Processing (ICIP'08), San Diego, CA", October 2008, http://www.irisa.fr/vista/Papers/2008-icip-fradet.pdf.

[41] A. HERVIEU, P. BOUTHEMY, J.-P. LE CADRE. *Activity-based temporal segmentation for videos of interacting objects using invariant trajectory features*, in "Proc. Int. Conf. on Image Processing (ICIP'08), San Diego, CA", October 2008, http://www.irisa.fr/vista/Papers/2008-icip-hervieu.pdf.

[42] A. HERVIEU, P. BOUTHEMY, J.-P. LE CADRE. *Video event classification and detection using 2D trajectories*, in "Proc. Int. Conf. on Computer Vision Theory and Applications (VISAPP'08), Funchal, Portugal", vol. 2, January 2008, p. 158-166, http://www.irisa.fr/vista/Papers/2008-visapp-hervieu.pdf.

[43] P. HÉAS, E. MÉMIN, D. HEITZ. *Self-similar regularization of optic-flow for turbulent motion estimation*, in "Proc. ECCV'08, Int Workshop on Machine Learning for Vision-based Motion Analysis, Marseille, France", October 2008, http://www.irisa.fr/vista/Papers/2008_mlvma08_heas.pdf.

[44] P. HÉAS, E. MÉMIN. *Optical-flow for 3D atmospheric motion estimation*, in "Proc. Int. Conf. on Computer Vision Theory and Applications (VISAPP'08), Funchal, Portugal", vol. 2, January 2008, p. 399-406.

[45] A. ICKOWICZ, J.-P. LE CADRE. *A new method for target trajectory estimation within a binary sensor network*, in "Proc. ECCV'08, Multi Camera and Multi modal sensor fusion algorithms and application Workshop), Marseille, France", October 2008.

[46] A. ICKOWICZ, J.-P. LE CADRE. *Target tracking within a binary sensor network*, in "Proc. Int. Conf. Intelligent Sensors, Sensor Networks and Information (ISSNIP'08), Sydney, Australia", December 2008.

[47] A. ICKOWICZ, J.-P. LE CADRE. *Target trajectory estimation within a sensor network*, in "Proc. Int. Conf. on Information Fusion (Fusion 2008), Cologne, Germany", July 2008, p. 1-8.

[48] I. JUNEJO, E. DEXTER, I. LAPTEV, P. PÉREZ. *Cross-view action recognition from temporal self-similarities*, in "Proc. Eur. Conf. Computer Vision (ECCV'08), Marseille, France", vol. 2, October 2008, p. 293-306.

[49] I. JUNEJO, H. FOROOSH. *Estimating geo-temporal location of stationary cameras using shadow trajectories*, in "Proc. Eur. Conf. Computer Vision (ECCV'08), Marseille, France", vol. 1, October 2008, p. 318-331.

[50] I. JUNEJO, H. FOROOSH. *GPS coordinate estimation from calibrated cameras*, in "Proc. Int. Conf. on on Pattern Recognition (ICPR'08), Tampa Bay, Florida", December 2008.

[51] I. JUNEJO, H. FOROOSH. *Practical PTZ camera calibration using givens rotations*, in "Proc. Int. Conf. on Image Processing (ICIP'08), San Diego, CA", October 2008.

[52] I. JUNEJO, H. FOROOSH. *Practical pure pan and pure tilt camera calibration*, in "Proc. Int. Conf. on on Pattern Recognition (ICPR'08), Tampa Bay, Florida", December 2008.

[53] I. JUNEJO, H. FOROOSH. *Refining PTZ camera calibration*, in "Proc. Int. Conf. on on Pattern Recognition (ICPR'08), Tampa Bay, Florida", December 2008.

[54] I. JUNEJO, H. FOROOSH. *Using Solar Shadow Trajectories for Camera Calibration*, in "Proc. Int. Conf. on Image Processing (ICIP'08), San Diego, CA", October 2008.

[55] I. JUNEJO. *Calibration and 3D geometry estimation of a pan-tilt-zoom camera network*, in "Proc. Eur. Signal Processing Conference (EUSIPCO'08), Lausanne, Switzerland", August 2008.

[56] I. JUNEJO. *Camera calibration for uneven terrains by observing pedestrians*, in "Proc. Int. Conf. on on Pattern Recognition (ICPR'08), Tampa Bay, Florida", December 2008.

[57] I. JUNEJO. *Simple shadow based method for camera calibration*, in "Proc. Int. Conf. on Image Analysis and Recognition (ICIAR'08), Póvoa de Varzim, Portugal", June 2008.

[58] T. PÉCOT, C. KERVRANN, S. BARDIN, B. GOUD, J. SALAMERO. *Patch-based Markov models for event detection in fluorescence bioimaging*, in "Int. Conf. on Medical Image Computing and Computer Assisted Intervention (MICCAI'08), New York City, USA", vol. 2, September 2008, p. 95-103.

[59] T. PÉCOT, C. KERVRANN, P. BOUTHEMY. *Minimal paths and probabilistic models for origin-destination traffic estimation in live cell imaging*, in "Proc. IEEE Int. Symp. on Biomedical Imaging: from nano to macro (ISBI'08), Paris, France", May 2008, p. 843-846.

[60] T. PÉCOT, C. KERVRANN, P. BOUTHEMY. *Network tomography-based tracking for intracellular traffic analysis in fluorescence microscopy imaging*, in "Proc. Int. Conf. on Bio-inspired Systems and Signal Processing (BIOSIGNALS'08), Funchal, Portugal", vol. 1, January 2008, p. 154-161.

[61] T. PÉCOT, C. KERVRANN. *Patch-based Markov models for change detection in image sequence analysis*, in "Proc. Int. Workshop on Local and Non-Local Approximation in Image Processing (LNLA'08), Lausanne, Switzerland", August 2008.

[62] C. SIMONIN, J.-P. LE CADRE, F. DAMBREVILLE. *Common framework for optimizing multitarget search and cross-cueing*, in "Proc. Int. Conf. on Information Fusion (Fusion 2008), Cologne, Germany", July 2008, p. 1-8.

National Peer-Reviewed Conference/Proceedings

[63] A. CHESSEL, R. FABLET, C. KERVRANN, F. CAO. *Détection a contrario multi-modale de structures curvilinéaires*, in "Congrès Francophone de Reconnaissance des Formes et Intelligence Artificielle (RFIA'08), Amiens", January 2008, p. 1-8.

[64] N. PAPADAKIS, E. MÉMIN. *Estimation variationnelle et cohérente en temps de mouvements fluides*, in "Congrès Francophone de Reconnaissance des Formes et Intelligence Artificielle (RFIA'08), Amiens", January 2008, p. 1-8.

[65] C. SIMONIN, J.-P. LE CADRE, F. DAMBREVILLE. *Hierarchical approach for planning a multisensor multizone search for a moving target*, in "Proc. Congrès de la Société Francaise de Recherche Opérationnelle et d'Aide à la Décision (ROADEF 2008), Clermont-Ferrand", February 2008.

Workshops without Proceedings

[66] J. BOULANGER, A. GIDON, J. SALAMERO, J.-B. SIBARITA, C. KERVRANN. *Repetitive and transient event detection in fluorescence video-microscopy*, in "Focus on Microscopy (FOM'08), Osaka-Awaji, Japan", April 2008.

[67] A. CHESSEL, J. BOULANGER, S. GASFOU-UZAN, J. SALAMERO, J.-B. SIBARITA, C. KERVRANN. *Processing fluorescence video microscopy by 1D temporal signal analysis*, in "European Light Microscopy Initiative meeting (ELMI'08), Davos, Switzerland", May 2008.

[68] I. LAPTEV, M. MARSZALEK, C. SCHMID, B. ROZENFELD. *Learning realistic human actions from movies*, in "Proc. Int. Conf. Computer Vision and Pattern Recog. (CVPR'08), Anchorage, Alaska", June 2008, p. 1-8.

[69] T. PÉCOT, C. KERVRANN, P. BOUTHEMY, A. CHESSEL, J.-B. SIBARITA, J. SALAMERO, J. BOULANGER. *Network tomography and minimal paths for traffic flow estimation in time-lapse fluorescence microscopy*, in "European Light Microscopy Initiative meeting (ELMI'08), Davos, Switzerland", May 2008.

Scientific Books (or Scientific Book chapters)

[70] J. BOULANGER, T. PÉCOT, P. BOUTHEMY, J. SALAMERO, J.-B. SIBARITA, C. KERVRANN. *7*, in "Simulation and estimation of intra-cellular dynamics and trafficking", S. WONG, R. MACHIRAJU, J. RITTSCHER (editors), Artech Publishing House, 2008, p. 169-190.

[71] E. MÉMIN, P. PÉREZ. *Estimation de mouvement apparent et suivi visuel*, in "Problèmes inverses en imagerie et en vision", A. MOHAMMAD-DJAFARI (editor), Traité IC2, chap. 6, Hermès, 2008, p. 201–252.

Research Reports

[72] J. BOULANGER, C. KERVRANN, J. SALAMERO, J.-B. SIBARITA, P. BOUTHEMY. *Non-parametric regression for patch-based fluorescence microscopy image sequence denoisng*, Technical report, n^{o} 6651, INRIA, September 2008, https://hal.inria.fr/inria-00322321.

References in notes

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

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

[75] 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^{o} 3, 2003, p. 175-198.

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

[77] 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^{o} 3, May 2001, p. 257-270.

[78] 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^{o} 4, April 2000, p. 725-745.

[79] 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^{o} 12, December 1993, p. 1217-1232.

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

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

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

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

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

[85] 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^{o} 5, May 1998, p. 703-719.

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

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