INRIA

Project-Team VISTA

Vision Spatio-Temporelle et Apprentissage

Rennes

THEME COG

Activity Report

2004

Table of contents

1. Team

Head of project-team

Patrick Bouthemy [DR Inria]

Administrative assistant

Stéphanie Lemaile [TR Inria, shared with Lagadic, Temics and Visages teams]

Inria research scientists

Frédéric Cao [CR]

Patrick Pérez [DR, from 01/02/04]

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

CNRS research scientists

Jean-Pierre Le Cadre [DR]

Faculty members

Étienne Mémin [Associate professor, Ifsic, University of Rennes 1]

Jian-Feng Yao [Full professor, Irmar, UFR Maths, University of Rennes 1, on Inria secondment (délégation), until 31/08/04]

Inria technical staff

Fabien Spindler [IR, 20% partial time, Lagadic team]

François Coldefy [Domus Videum project, until 30/06/04]

Brigitte Fauvet [Feria project]

Nicolas Gengembre [FT-RD contract, from 01/12/04]

Post-doctoral fellows

Pierre Dodin [Inria, Cesta contract, from 01/03/2004 until 15/09/2004]

Ivan Laptev [Inria, from 01/12/2004]

Venkatesh Babu Radhakrishnan [Ercim, from 18/06/2004]

Vincent Samson [Inria, IST Lava project]

Ph-D students

Elise Arnaud [Research Ministry grant until 30/09/2004, Inria grant]

Vincent Auvray [Cifre grant, General Electric Healthcare]

Jérôme Boulanger [Inra grant]

Thomas Bréhard [DGA grant]

Aurélie Bugeau [Research Ministry grant, from 01/10/2004]

Anne Cuzol [Research Ministry grant]

Nicolas Papadakis [Inria grant, FP6 Fluid project, from 20/09/04]

Gwénaëlle Piriou [Inria grant, Brittany Region contract]

Thomas Veit [Inria grant, Brittany Region contract (50%)]

2. Overall Objectives

Keywords: *Bayesian estimation*, *Markov models*, *a contrario decision*, *biological imagery*, *dynamic scene analysis*, *experimental fluid mechanics*, *fluid motion analysis*, *image sequence*, *meteorological imagery*, *motion detection*, *motion recognition*, *motion segmentation*, *optic flow*, *particle filtering*, *registration*, *robust estimation*, *statistical learning*, *statistical modeling*, *tracking*, *trajectography*, *video indexing*, *video processing*.

Vista research work is concerned with various types of spatio-temporal images, mainly optical (video images, sometimes IR images), but also acoustic in certain cases (sonar, ultrasound). We design techniques 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: detection, estimation, segmentation, tracking, recognition, interpretation with learning. We

usually rely on a statistical approach, resorting to: Markov models, Bayesian inference, robust estimation, *a contrario* decision, particle filtering, learning. Application-wise, we focus our attention to four main domains: video processing and indexing, meteorological imaging and experimental visualization in fluid mechanics, biological imaging, surveillance and navigation. 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

Keywords: *Markov models*, *energy function*, *motion estimation*, *motion segmentation*, *optic flow*, *parametric motion models*, *robust estimation*.

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 [55]). 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 favour 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 [50] 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 [63][64]. 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-neighbour 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 [58] 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 grey level, colour or texture) and to specify a MRF model on the resulting graph of adjacent regions [56]. 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 [57]. Then, robust estimators were popularized leading to the introduction of so-called auxiliary variables ζ now taking their values in $[0, 1]$ [62]. 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 [61]. Depending on the followed approach, the third term of the energy $E(\mathbf{w}, \zeta)$ can be optional.

3.2. Fluid motion analysis

Keywords: *continuity equation*, *div-curl regularization*, *experimental fluid mechanics*, *fluid motion analysis*, *meteorological image sequences*, *singular points*.

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, analysing 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 [62][61]. 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:

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

The penalty function ρ is usually the L_2 norm, but it may be substituted for a robust function attenuating the effect of data that deviate significantly from the brightness constancy assumption [62], 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 [51]. The method we have devised for fluid motion involves the following global energy function:

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

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:

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

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

Once given a reliable description of the fluid motion, another important issue consists in extracting and characterizing structures of interest such as singular points or in deriving potential functions. The knowledge of the singular points is precious to understand and predict the considered flows, but it also provides compact and hierarchical representations of the flow [52]. 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

Keywords: *data association*, *data fusion*, *importance sampling*, *multi-object tracking*, *particle filter*, *sequential Monte Carlo*.

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:

$$\text{prediction step:} \quad p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) = \int p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})d\mathbf{x}_{t-1}$$
$$\text{filtering step:} \quad p(\mathbf{x}_t|\mathbf{y}_{1:t}) = \frac{p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t-1})}{p(\mathbf{y}_t|\mathbf{y}_{1:t-1})},$$

where the prediction step follows from marginalisation, 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 initialised 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 [54][60][59], otherwise known as particle filters, have gained a lot of popularity in recent years as a numerical approximation strategy to compute the tracking recursion for complex models. This is due to their efficiency, simplicity, flexibility, ease of implementation, and modeling success over a wide range of challenging applications.

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

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

where the proportionality is up to a normalising 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 marginalised 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 marginalising 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.

3.4. Perceptual grouping for image and motion analysis

Keywords: *Helmoltz principle*, *a contrario decision*, *motion detection*, *number of false alarms*, *parameter-free method*, *perceptual grouping*, *shape matching*.

We have recently been interested in automatic detection problems in image and video sequence processing. A fundamental question is to know whether it is possible to automatically direct the attention to some object of interest (in the broad sense). We have been using a general grouping principle, asserting that conspicuous events are those that have a very small probability of occurence in a random situation. We have applied this principle formalized within an *a contrario* decision framework to the detection of moving objects in an image sequence, and to the matching of shapes in images.

For the last few years, we have been interested in developing methods of image and video analysis with no complex *a priori* model. Of course, in this case the purpose is not to analyse and finely describe complex situations. On the contrary, we try to achieve very low-level vision tasks, with the condition that the methods must be very stable and provide a measure of validity of the detected structures. A qualitative principle, called the Helmholtz principle, was developed a few years ago in École Normale Supérieure de Cachan [53], and we used it in low-level motion analysis. This principle basically states that local observations have to be grouped with respect to some qualitative property, if the probability that they share this quality is very small, assuming that the quality is independently and identically distributed on these observations. In some sense, this can be related to a more classical hypothesis testing setting. Let us express it in the context of detection of motion in a given region R of the image. We would like to test the hypothesis H_0 "there is motion in R" against H_1 "there is no motion". The problem is that, usually, we do not have any precise model for H_0. On the opposite,

we model the background model H_1 (the absence of motion), by the fact that the pointwise observations are independent. This is sound since this hypothesis amounts to say that the observations are only due to noise. We then decide that H_0 is true whenever the probability of occurence of the observed values in R are much improbable under the independence hypothesis H_1. This amounts to an *a contrario* decision framework.

More generally, assume given a set of local measures of some quantity Q. We also make local (pointwise or close to pointwise) observations on the images $(O_k)_{k \in K}$, where K is a set of spatial indices. Assume also that, for one reason or another, we can design some group candidates G_1, ..., G_n, of the local observations, that is to say subsets of $\{O_k, k \in K\}$. We also consider an adequacy measure of $X_{G_i}(O_k)$ which we assume small when the quality Q is satisfied by O_k, relatively to G_i. As a simple example, we can consider as G_i the digital segments of the image, O_k a direction field defined at each position and $X_{G_i}(O_k)$ as the difference between the field at position k and the direction of the segment G_i. Finally, let u be an image. We ask the following question: "in u, is Q a good reason to consider G_i as a group?"

Helmholtz Principle. *Assume that u is a realization of a random image U where it is assumed that, anything else being equal, the random variables $O_k(U)$ are independent and identically distributed.*
The group G_i is all the more conspicuous that the probability

$$\mathcal{P}\left(\forall k, X_{G_i}(O_k(U)) \leq X_{G_i}(O_k(u))\right)$$

is small.
>From this qualitative perceptual principle, we can define the number of false alarms of a configuration, which is the expectation of its number of occurrence in the background model of independence. It can be proven, that this number is a very good and robust measure of the meaningfulness of a configuration. We have applied this principle to the detection of good continuations and corners, of straight lines trajectories of subpixel target, and more recently to the detection of moving objects in images and to shape matching.

4. Application Domains

Keywords: *biological imagery*, *defense*, *environment*, *experimental fluid mechanics*, *meteorological imagery*, *multimedia*, *sonar*, *vehicle navigation*, *video indexing*, *video processing*.

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

- Video processing and indexing (INA, Thomson, FT-RD, Xerox);
- Experimental fluid mechanics (Cemagref) and meterological imagery (LMD);
- Biological imagery (Inra);
- Surveillance (Onera, Thales) and vehicle navigation.

5. Software

5.1. Motion2d software - parametric motion model estimation

Participants: Fabien Spindler, 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 proven 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 [63]. 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 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, 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. Dense-Motion software - optical flow computation

Participant: Etienne Mémin.

The Dense-Motion software written in C enables to compute a dense velocity field between two consecutive frames of a sequence. It is based on an incremental robust method encapsulated within an energy modeling framework. The associated minimization is based on a multi-resolution and multigrid scheme. The energy is composed of a data term and a regularization term. The user can choose among two different data models : a robust optical flow constraint or a data model based on an integration of the continuity equation. Two models of regularization can be selected as well : a robust first-order regularization or a second-order Div-Curl regularization. The association of the latter with the data model based on the continuity equation constitutes a dense motion estimator dedicated to image sequences involving fluid flows. It was proven to

supply very accurate motion fields on various kinds of sequences in the meteorological domain or in the field of experimental fluid mechanics.

6. New Results

6.1. Image sequence processing and modeling

6.1.1. *Auto-models with mixed states for motion analysis*

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

When dealing with motion analysis in a sequence of images, the histograms of local motion measures typically present a composite picture. An important peak appears at the origin accounting for regions where no motion is present (which is a significant symbolic information as it is), while a remaining continuous component reveals the actual motion magnitudes in the images. Then, the question arises to find accurate models for this type of data including both discrete and continuous values -we call them *observations with mixed states*-, collected from the image lattice. Other examples could be given of motion variables taking both discrete (or even symbolic) values and continuous ones; let us mention the velocity field including velocity vectors and motion discontinuities. >From a mathematical point of view, we are searching for models for a random field $\{X_s\}$ with the constraint that the marginal distributions of the X_s's are composed with a discrete component and a continuous one. In its most general form and for the discrete component, we may take any distribution with support on a countable set $\{e_1, ..., e_k, ...\}$ of real numbers (or even symbols), as in the case of the Poisson distribution, while for the continuous component any standard distribution could be considered. However, in the present work, we restrict ourselves to distributions from the exponential family with one atomic value $\{0\}$ and a continuous component supported on the interval $(0, \infty)$. The state space, called a *mixed state space E* is then $E = \{0\} + (0, \infty)$ where the origin 0 will play a special role.

Markov random fields models are now a standard tool in image analysis. However, to our knowledge, the existing models deal either with continuous observations, or with discrete observations, but never with observations that can take values of both types. Indeed, such discrete information is usually examined by the introduction of a label process $\{L_s\}$ where, in our case, $L_s = 1$ if a null observation is available at pixel s, i.e., $X_s = 0$, and $L_s = 2$ when no null observation is recorded, i.e., $X_s > 0$. However, the label process is a latent process and the resulting statistical inference methods need in general a restoration of the latent process (i.e., segmentation). This classical approach is possible only upon the cost of a generally huge computation effort. Here, we propose a different approach. The aim is to give a model which automatically deals with the two types of observations, without the introduction (and then the inference) of any latent process. The basic idea is then to introduce mixed-state variables (or distributions) in a random field set-up. More precisely, we follow the construction of auto-models by J. Besag [49]. We have introduced necessary adaptations for mixed-states variables. These new models for random fields are called *auto-models with mixed states*. We have developed an estimation procedure of the model parameters based on the pseudo-likelihood function. Preliminary results are already obtained on the modeling of motion textures corresponding to videos depicting natural scenes such as rivers, sea-waves, foliage, fire, smokes. The considered motion observations are (locally averaged) normal flow magnitudes.

6.1.2. *Motion detection and a contrario decision*

Participants: Frédéric Cao, Patrick Bouthemy, Thomas Veit.

A basic problem in the perception of visual motion is to detect moving objects in images. This may seem elementary but this is not so: this process is neither completely local in space nor in time, there is no general model for image motion. In this work, we want to answer the following question: given a region of an image in a sequence, is it changing significantly between two consecutive instants? To answer this question, consider a local measure of motion, and a set of given regions, which have been independently determined. Let $C(x)$ be this measure, which we assume non negative and increasing with change. Specifically, it is given by

the minimum of the backward and forward displaced frame differences (evaluated with the corresponding estimated dominant motion models). Let n be the number of points of the region. Assume that for some μ, we observe that k points among the n satisfy $C(x) \geq \mu$. The designed solution consists in showing that we can find relations between n, k and μ such that we can be sure that this observation cannot be simply casual. The theory is quite general, and the method may be applied to any type of region given by an image segmentation. More precisely, we can estimate the distribution of $C(x)$ on the image itself, which is a good model for the marginal laws. Of course, this criterion is not independent from one point to another, but this is precisely what we aim at detecting in an *a contrario* framework. When we make this independence assumption, the probability that $C \geq \mu$ for at least k points among the n in a region R is simply the tail of the binomial law

$$B(n, k, p) = \sum_{j=k}^{n} \binom{n}{j} p^j (1-p)^{n-j},$$

where $p = P(C \geq \mu)$ is empirically estimated. Let be R_1, ..., R_N be a set of N given regions (determined independently of C). Let us call the Number of False Alarm (NFA):

$$NFA(R) = N \cdot B(n, k, p).$$

We will say that R is ε-meaningful if $NFA(R) \leq \varepsilon$. We have considered two kinds of spatial segmentation to supply the set of regions R: a) hierarchical partition into blocks, b) computed regions corresponding to maximal meaningful (intensity) level lines. The method performance has been assessed on various real image sequences.

6.1.3. Biological image filtering

Participant: Charles Kervrann.

We address the adaptive image restoration problem and we have developed a non-parametric estimation method that smoothes homogeneous regions and inhibits smoothing in the neighborhood of discontinuities. The proposed *adaptive window approach* is conceptually very simple being based on the key idea of estimating a local regression function with an adaptive choice of the window size (neighborhood) for which the applied model fits the data well. At each pixel, we estimate the regression function by iteratively growing a window and adaptively weighting input data to achieve an optimal compromise between the bias and variance. The proposed algorithm complexity is actually controlled by simply restricting the size of the larger window and setting the window growing factor. The proposed smoothing scheme provides an alternative method to the anisotropic diffusion and bilateral filtering or energy minimization methods. An advantage of the method is that internal parameters can be easily calibrated using statistical arguments. Experimental results demonstrated its potential for image decomposition into white Gaussian noise, texture and piecewise smooth components. In domains like confocal microscopy, it is established that the additive Gaussian noise model is a poor description of the actual photon-limited image recording, compared with that of a Poisson process. This motivates the use of restoration methods optimized for Poisson noise distorted images. We have extended the *adaptive window approach* to Poisson noise reduction in 2D and 3D imaging. Since we do not address the image formation of the confocal fluorescence microscope, ideally modeled as a convolution of the object function with the point spread function, the proposed method can be seen also as a sophisticated pre-filtering method before starting the more complex *deconvolution* process using the Lucy-Richardson algorithm.

6.1.4. Spatio-temporal detection in 4D video-microscopy

Participants: Charles Kervrann, Patrick Bouthemy, Jérôme Boulanger.

We have designed a spatio-temporal filtering method to significantly increase the signal-to-noise ratio (S/R) in noisy fluorescence microscopic image sequences where small particles have to be tracked from frame to frame. New video-microscopy technologies allow one to acquire 4-D data that require the development and implementation of specific image processing methods to preserve details and discontinuities in both the three

$x - y - z$ spatial dimensions and the time t dimension. Particles motion in such noisy image sequences cannot be reliably calculated since objects are small and untextured with variable velocities ; the S/R ratio is also quite low due to the relatively limited amount of light. However, partial trajectories of objects are line-like structures in the spatio-temporal $x - y - z - t$ domain. Image restoration can be then achieved by an adaptive window approach which has been already used to efficiently remove noise in still images and to preserve spatial discontinuities. The proposed spatio-temporal technique associates with each voxel the weighted sum of data points within a space-time window. We use statistical 4-D data-driven criteria for automatically choosing the size of the adaptive growing neighborhood. We have applied this method to noisy synthetic and real 4-D images where a large number of small fluorescently labeled vesicles move in regions close to the golgi apparatus. The S/R ratio is shown to be drastically improved, yielding enhanced objects which even can be segmented. This novel approach will be further used for biological studies where dynamics have to be analyzed in molecular and subcellular bio-imagery.

6.2. Motion estimation and matching

6.2.1. *Fluid motion analysis*

Participants: Étienne Mémin, Anne Cuzol, Nicolas Papadakis.

We have proposed a low-dimensional motion estimator for image sequence depicting fluid flows. The proposed estimator is based on the Helmholtz decomposition of vector fields. This decomposition consists in representing the velocity field as the sum of divergence-free component and a curl-free component. In order to provide a low-dimensional solution, both components are approximated using a discretization of the vorticity and divergence maps through regularized Dirac measure. The resulting so-called irrotational and solenoidal fields are then given by linear combinations of basis functions obtained through a convolution product of the Green kernel gradient and the vorticity map or the divergence map respectively. The coefficient values and the basis function parameters are obtained as the minimizer of a functional relying on an integrated version of mass conservation principle of fluid mechanics. The resulting method is very efficient from a computational point-of-view and supplies in selected areas accurate flow fields. This work will be extended to estimate wind field layers at specific heights in meteorological applications (starting Ph-D by N. Papadakis).

6.2.2. *Transparency motion estimation*

Participants: Patrick Bouthemy, Vincent Auvray.

This work is carried out in collaboration with J. Liénard (General Electric Healthcare, see paragraph 7.4). When considering images governed by the principle of additive transparency, such as those generated by medical X-ray exams, the modeling associated with motion estimation has to be different from the one related to the classical video case. It is in particular impossible to estimate the motion of any object point by assuming that its grayscale will be constant along its trajectory since its representation depends on the respective background and foreground values of its successive positions. Considering a sequence with two layers translating by transparency, it is nevertheless possible to exhibit a fundamental equation that links three successive images of the sequence. Since this equation is function of the unknown velocities of the two layers, it can be used as a criterion to estimate the velocities from the sequence. We have developed a multiresolution estimation framework for two layers translating by transparency from this equation. At each resolution, we linearize the transparency motion fundamental equation to build a criterion which is a fourth-order polynomial with respect to the unknown motion parameters. A conjugate gradient descent is adopted to minimize it. To be able to capture motion more complicated than simple global translations in the image, we apply this algorithm within predetermined image blocks. To keep it very efficient, an initialization step based on a simplex method is introduced, and a postprocessing stage has been added. We are now working on extending this framework in order to be able to decide how many layers (one, two or three) are involved in the considered block. The final goal of this work being the denoising of fluoroscopic image sequences which exhibit a low signal-to-noise ratio, we are also designing a spatio-temporal filtering scheme exploiting the transparency motion estimation stage.

6.2.3. *Spatio-temporal shape matching*

Participant: Frédéric Cao.

The *a contrario* framework has been applied to the problem of shape matching. The problem we address is to decide whether two images have shapes in common. Such a program involves different subtopics: 1) Detection of shapes, or shapes elements; 2) Local and invariant encoding (since recognition must be robust to occlusion and geometrical transformations); 3) Robust matching of shape elements. In collaboration with P. Musé, F. Sur, and J.-M. Morel (CMLA, ENS Cachan) and Y.Gousseau (ENST Paris), we gave insights on the three points with a focus on the last point. To this purpose, assume that we have a database of N_B normalized pieces of curves, that we will call codes. Given a particular code C, which is independently computed (for instance in a query image), can we find all the occurences C in the database, and only them? If we think in terms of distance, these codes will be such that their distance to C will be less than a threshold $\delta(C)$, which depends on C. The problem is that there is no good completely general shape model, and that we cannot model the distance to C so easily. We then use an *a contrario* setting. Assume that codes are the concatenation of n independent chunks X_i. Then, we can compute the marginal distribution of the distances $d(X_i(C), X_i(C'))$ when C' describes the database. Under the independence assumption, the probability that all the chunks of C' are equal to those of C, up to an error δ is: $p(C, \delta) = \prod_{i=1}^{n} P(d(X_i(C), X_i(C')) \leq \delta)$.

Now, it is very simple to prove, that, under the independence assumption, the expected number of database codes which are (accidentally) similar to C, with error δ, is less than $N_B \cdot p(C, \delta)$. Thus, it is possible to choose δ (as a function of C and the database) such that the number of casual occurences of a code resembling C is less than any predefined fixed value. When this value is equal to 1, this means that we accept one such casual code in average in the database. This method was tested in many images and gave very satisfactoy results. However, matching shape elements is not an end in itself and not sufficient to retrieve real shapes. A natural further step is to group these local matches. If we assume that the matching phase is affine invariant (or similarity invariant), we can seek all codes in a query image, matched with codes in a database image, as transformed by the same mapping. This is a clustering problem in the transformation space, and again we have proposed an *a contrario* approach: we assume that points in the transformation space are independent. It is then possible to compute the probability to observe k points among n in a given domain of the transformation space. Not only can we compute detection thresholds by this method, but we have also defined a merging criterion allowing us to give an optimal cut in a hierarchical clustering procedure.

6.3. Tracking

6.3.1. *Point and structure tracker*

Participants: Étienne Mémin, Elise Arnaud, Anne Cuzol.

Point tracking in an image sequence is a basic but essential issue in computer vision. This problem is paradoxically inherently difficult as one can only rely on local photometric or geometric time-invariant characteristics of the point. In addition, any prior dynamic model of a feature point is almost impossible to establish without any *a priori* knowledge on the dynamics of the surrounding object. These difficulties led us to consider systems composed by a state equation and measure equation which both depend on the image sequence data. To handle such systems, we have proposed a new conditional formulation of classical filtering methods. These conditional filters allow us to solve systems whose measures and state equation are estimated from the image data. The model we have considered for point tracking combines a state equation relying on the optical flow constraint and measurements provided by a matching technique. Two different point trackers have been derived for this model. The first one is a linear tracker well-suited to image sequences exhibiting a global dominant motion. The second one is a nonlinear tracker, implemented through particle filtering. The latter allows us to track points whose motion may only be locally described. For this non-linear tracker, we have considered measurement likelihoods that enable to infer an analytic expression of the optimal importance function. This function has a crucial role in the diffusion step of the particle filter algorithm. The likelihood model which we rely on is expressed as a Gaussian mixture of several potential observations of the unknown state. Such a choice enables to robustify the tracker with respect to occlusion or clutter noise.

The developed point trackers have been extended for the tracking of a cloud of interest points distributed over a 3D plane. The resulting plane tracker combines a motion model with a geometric constraint. The system we have devised for this task is conditionally Gaussian and may be solved through Rao-Blackwelised particle filter. This technique allows us to have a set of components for which, conditionally to the others, the stochastic filtering problem can be optimally solved with a Kalman filter. The overall filter therefore amounts to handle a nest of competing Kalman filters. The competition takes place within particle filtering precept. The plane tracker we have proposed in this context is robust to occlusion and allows us to track roughly planar objects. This work on tracking a set of points will be extended to deformable or fluid motions. We are investigating on this basis, the tracking of vector fields described as a linear combination of basic shape functions in the context of the analysis of fluid flows. Nevertheless, contrary to the previous case, where no prior dynamics have been considered, we will rely here on the vorticity-velocity expression of Navier-Stokes equation to predict the vorticity part of the motion field over time.

6.3.2. *Robust visual tracking without prior*
Participant: Patrick Pérez.

In this starting research we are interested in the generic problem of tracking arbitrary entities along videos of arbitrary type and quality. Such a tracking can't rely, as classically done, on *a priori* information regarding both the appearance of the entities of interest (shape, texture, key views, etc.) and their visual motion (kinematic constraints, expected dynamics relative to the camera, etc.). The first crucial step is then the definition and the estimation of the reference model on which the tracking, no matter its precise form, will rely on. We aim in particular at combining complementary representations of the appearance (from detailed pixel-wise appearance model subject to rapid fluctuations to rough color model very persistent over time) and making them evolve on-line for improved robustness. We also aim at addressing explicitly the critical problem of occlusions and at extending obtained tools to multi-object tracking. These different bricks (selection of best visual features, on-line estimation of the models associated to these features, and effective tracking of one or several entities) are defined in a probabilistic filtering framework. Regarding the difficult problem of handling in a principled way a varying (and unknown) number of entities to be tracked, algorithmic studies have been conducted in collaboration with Jaco Vermaak from the University of Cambridge. Different approaches to multi-target data association, all based on the sequential Monte Carlo framework, have been designed and compared, showing different merits in different contexts. This algorithmic research on multi-target tracking will be continued within a British-French program which started in October 2004 (see §8.3.1). On the application front, part of the experimental validations of this research on tracking without prior will be conducted on team sport sequences, in relation with an FTRD contract (see §7.1). In this specific context, we will try to assess how far generic tracking can get without the introduction of application-specific prior (off-line learning of player appearance and movements), whose extraction might be difficult.

6.3.3. *Target motion analysis*
Participants: Jean-Pierre Le Cadre, Thomas Bréhard.

Particle filtering is definitely worthy for target tracking. It allows us to handle non-linearities as well as process noise. However, some problems limit the applicability of this method, especially when the system state is partially observed. This is generally the case for passive measurements. Consequently, important efforts have been put on the following problems: - Particle filter initialization; Estimation of the σ/r ratio; Performance analysis; Particle filtering for GMTI. Particle filter initialization is the major difficulty for a partially observed process. A first way to remedy this problem is to use observable components for track filtering. Unobservable ones are just updated along the diffusion. When they become "sufficiently" observable, then particles automatically converge toward actual states. Generally, observability is related to the observer maneuvers. However, even in the absence of observer maneuver, the entries of the system are not zeroed as long as the target trajectory can be represented by a diffusion. It is in this way, that we have studied an original extension of the BOT (Bearings-Only Tracking) problem named "σ-BOT", where the state covariance σ which represents the maneuverability of the target is unknown. This is an important issue in practice. In this case,

the problem belongs to the class of non linear filtering problems with unknown variance state. Otherwise, the modified polar (MP) coordinate system introduced by Aidala and Hammel is fundamentally relevant in the classical BOT context in particular for the initialization of the particle filter and for deriving a closed-form solution in the deterministic case. We deduced from this framework that $\frac{\sigma}{r(t)}$ named "variance-to-range ratio" is observable even if the range itself is not observable (i.e. the observer is not maneuvering). More generally, it appears that one more time, the MP coordinate system is perfectly adapted to the "σ-BOT" context.

Performance analysis for random state filtering is another important issue. The Posterior Cramér-Rao Bound (PCRB) now is widely used. A recursive form of the PCRB is especially relevant for Markovian states. However, this formulation is no longer valid in the BOT context due to finite support measurements. So, it has been necessary to develop an original framework for calculating the PCRB in this context. Moreover, closed-form expressions of the PCRB terms have been obtained. Tracking a target moving across a network is another important issue; especially for the GMTI context. Again, particle filtering is especially relevant since the propagation of the particles can take into account the network topology and specific target modelling.

6.3.4. *Person tracking and clustering*

Participants: Patrick Pérez, Patrick Bouthemy, Venkatesh Babu Radhakrishnan.

In this work we focus on tracking people in various TV programs such as talk shows or filmed operas. The main goal is to provide tools for the annotation and the query of large collections of such programs. The wide diversity of scenes, in terms of viewpoints and illumination conditions, number and activities of the persons in the scene, and nature of the background, requires to mix the best possible generic tracking tool with an efficient use of the prior available on tracked entities. The latter is based on a real-time face detector: We initialize the tracking by locating human faces in the video sequence, employing a recently reported algorithm. The former is based on mixing rough global color models, which have shown recently their power for robust tracking under severe appearance changes, with frame-to-frame image matching whose ability to estimate instantaneous displacements is unrivaled. More precisely, the proposed system makes use of the SSD tracker combined with Mean-Shift tracker. The three sources of information (intermittent face detection along with continuous color-based detection and matching-based displacement estimation) are seamlessly merged with smoothness prior on trajectories thanks to a Kalman filter. First experiments show the merit of the approach to extract unbroken face trajectories within the diversity of shots under concern. In the video shots, humans are characterized using the color histogram (both marginal and joint) features that were accumulated over tracking, with the tentative assumption that the body occupies a certain pre-determined area right under the location of the face. It will allow a high level annotation of extracted tracks and a character-based query of the annotated video. Let us note that the tracking part of this work will be able to benefit from progresses in generic visual tracking obtained in the research described in §6.3.2.

6.3.5. *Target acquisition*

Participants: Jean-Pierre Le Cadre, Pierre Dodin.

The aim of this study is the optimization of the target acquisition in a sequence of images. Basically, an image is obtained by means of a classical processing (e.g. energy detector) applied to the sensor outputs. The sensor itself can be passive (infrared sensor) or active (radar). First, the problem was to estimate the target trajectory. The main problem is related to the reentry phase. During this phase, the target trajectory is affected both by target factors (ballistic coefficient) and environmental ones (the atmosphere density). Both are unknown and time-varying since they depend on the kinematic (attitude) components. As it stands, the state evolution is thus highly non-linear and time-varying. The problem is then to estimate the kinematic components as well as the ballistic coefficient. Not surprisingly, linearization is hopeless while particle filtering is quite relevant. Actually, a complex Allen oscillatory ballistic profile may be considered to model the variation of the ballistic coefficient. Simpler models use random walks. To investigate the filtering performance, the Posterior Cramér-Rao Bound (PCRB) is the workhorse. The robustness of particle filtering has been shown, while its performance becomes quite close to the PCRB when a sophisticated model of the ballistic coefficient is used. The second part of this study deals with the optimization of the sequence of sensor

looks. This is a sensor management problem and our aim is to optimize the spatio-temporal sequence of looks so as the probability of detecting a target evolving in a bundle of trajectories be (globally) maximized. In this setup, the target motion is conditionally deterministic (e.g. known up to a ballistic coefficient and an initial position) while the search efforts have integer values. As it stands, the problem suffers from combinatorial explosion. However, it has been shown that complexity can be maintained to a quite reasonable level by using a Branch-and-Bound policy for fathoming the arborescence of decisions. The corresponding algorithm performs quite satisfactorily.

6.4. Motion recognition and learning

6.4.1. *Probabilistic motion modeling and event detection*

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

We adopt a statistical approach involving modeling, (supervised) learning and classification issues, to infer "semantic concepts" related to dynamic events in videos, from numerical motion measurements. We have defined original and efficient probabilistic motion models, both for the dominant image motion (assumed to be due to the camera motion) and the residual image motion (related to scene motion). Handling these two sources of motion is indeed important for video event detection. Motion measurements include, on one hand, affine motion models to capture the camera motion, and on the other hand, low-level local (residual) motion features to account for scene motion. The 2D histogram of the velocity vectors provided by the estimated affine model of the dominant motion is represented by a mixture γ^{cam} of 2D Gaussian distributions. On the other hand, we model the distribution of the local residual motion measurements (considered as mixed-state variables involving continuous motion values and the discrete value 0 equivalent to a particular symbolic state "no motion") by a specific mixture model formed by a distribution of the exponential family (we have explored both the Gaussian distribution with support $(0, \infty)$ and the exponential distribution) and a Dirac distribution at 0. The event detection scheme proceeds in two steps to progressively extract the video segments containing the events of interest. The first step consists of a sorting step into two groups ("interesting" versus "not interesting", e.g., "play" versus "no play" for sports video). The second step is stated as a classification problem (ML criterion) and amounts to recognizing the specified dynamic events among the group of segments selected after the first step. We have successfully applied this framework to sports videos that exhibit complex dynamic contents and events naturally related to motion information.

We have also extended the residual motion model by explicitly accounting for the (pixelwise) temporal evolution of the local motion measurements in order to attain a finer motion categorization. We have thus introduced new mixed-state probabilistic causal motion models which can be easily computed from the video data. The originality here is to take into account the mixed-state nature in the definition of the continuous Markov chain modeling the temporal evolution of each residual motion variable. As demonstrated in two applications, retrieval of similar video segments and unsupervised clustering of motion in videos, these models can handle a large variety of dynamic video contents. Comparison between residual motion models is achieved by evaluating the (symmetrized) Kullback-Leibler distance which can be analytically derived with the defined densities.

6.4.2. *Fine-grain analysis of human activities*

Participants: Patrick Pérez, Aurélie Bugeau, Ivan Laptev.

We are interested in the detailed analysis of human behavior in unconstrained set-ups (that is, not in a set up where a person is trying consciously to comply with a predefined small set of gestures, e.g., using a sign language). We assume however that video footages with partial annotations are available for a coherent set of situations, such as sport or office scenes. Hence, this work falls within the scope of semi-supervised learning. The first problem is the one of selecting the more appropriate set of spatial and/or temporal features, which are at the same time sufficiently discriminant and reliably accessible. Both the definition and the extraction of such features might require the use of various elementary object detectors (face, skin color, human silhouettes, etc.) along with tracking tools. The second problem consists in building activity classifiers on top of the selected

features. Finally, efficient ways of computing the classifiers on new input videos must be derived. In connection with the Behaviour ACI project (see §8.2.2), this program will be especially conducted and validated in the context of car driver monitoring from in-habitacle videos.

6.4.3. *Robust clustering and video interpretation*

Participants: Patrick Bouthemy, Vincent Samson.

In the context of video interpretation, a major concern in the learning stage is the high video appearance variability that a given event may exhibit. Therefore, we have to deal with heterogeneous classes while classes may be not so distant from each other. For a given class, observations and consequently computed video features, may vary according to the way the scene is filmed (camera motion, distance to the scene, illumination conditions) and the considered instance of the event of interest. This is the case for example in sports video analysis where a class of a given "play" event is actually reflected by several clusters in the feature space of low-level motion descriptors (depending on the camera parameters and on the athlets being filmed). SVM do not take into account this intra-class variability. Moreover, they are often limited to two-class classification, since multi-class extension is not so straightforward and still an active subject of research. We suggest instead to consider a robust partitional clustering approach applied in parallel to each predefined class in order to capture their internal data structure. Such an approach is intrinsically multi-class since clusters are trained separately on the instances of each class. Furthermore, it is flexible (classes can be handled in any order) and extensible (new classes can be straightforwardly added). It is also incremental with respect to the training data (new sets of training data can be processed just by iterating from the current learning state). Our clustering technique exploits a non-Euclidean distance in order to be robust to the presence of outliers in the training data. We have chosen a hard-redescending robust function, the Tukey's biweight, which depends on a single scale parameter. The algorithm yields an estimated number of clusters for each class and associated prototypes (actually the centers of the subclasses), from which different strategies of classification can be considered.

The core part of any video content analysis is made up of the extraction of appropriate descriptors for the given interpretation task. Considering motion information only, our aim is to get a fine enough characterization of motion activity by extracting localized features from the low-level motion measurements. We have investigated two approaches: 1) Spatio-temporal projections of binary motion detection masks; 2) Selection of spatio-temporal blocks and combination of local motion descriptors. The first one looks for modeling the spatio-temporal distribution of moving pixels. At each time, a binary mask is obtained by thresholding motion measurements: frame differences for a static camera, displaced frame differences for a moving camera. Then, by projecting the obtained bitmaps on the spatial and temporal axes, we obtain three 1D profiles characterizing the location of motion activity during the video sequence. Feature vectors correspond to the k first Fourier coefficient magnitudes of these spatial and temporal profiles, so that they are invariant to absolute location of moving objects. The second approach consists in characterizing a video segment by the combination of local models evaluated on spatio-temporal blocks. The location of the blocks in a video could be either random (uniformly sampled along space and time) or defined by maximizing a simple criterion of motion activity that enables to select the regions of interest in the video sequence. On each selected block, local statistics can be computed and combined to construct a global motion model or feature vector on the whole segment.

6.4.4. *Video abstraction based on audio and visual cues*

Participants: Patrick Bouthemy, François Coldefy.

We have designed two methods for sports video abstraction exploiting both image and audio features. The first one exploits a supervised learning stage. It is well adapted for sports with a regular structure and repetitive steps such as tennis which is moreover filmed according to fixed production rules. This first solution has been developed in collaboration with G. Gravier (Metiss project-team). The second method is concerned with sports defined by time constraints, such as soccer game, exhibiting no regular or periodic patterns in the game process and following no strict rules in the video shooting. We first need to segment the video into homogeneous temporal units. This is achieved by combining shot change detection and camera motion change detection (based on the dominant image motion analysis). In the first method, we extract low-level

visual features evaluating the spatial distribution of the residual motion (related to the scene motion) in each video segment. Audio event detection is carried out in a two-step process. First, the soundtrack is segmented into segments with homogeneous content. Then, the detection of ball hits and applause is performed based on statistical hypothesis testing. In an off-line learning stage, a K-means clustering of these features is previously performed over a training set formed with the kind of segments to be involved in the video abstract. The video abstract results from the selection of specific clusters which allow us to determine the relevant video segments. The second method is an unsupervised one, developed for soccer video, but it could be applied to other similar sports as rugby or american football. An excited commentary is supposed to correspond to an interesting moment of the game. It is simultaneously characterized by an increase of the pitch (or fundamental frequency) within the voiced segments and an increase of the energy supported by the harmonics of the pitch. The pitch is estimated from the autocorrelation function and its local increases are detected from a multiresolution technique. A statistical analysis of (appropriately defined) energy measures is performed to detect the most excited parts of the speech. A deterministic combination of excited speech detection, dominant color identification (here, green value for the playfield), and camera motion analysis is then performed to select the relevant video segments in the processed TV programs.

7. Contracts and Grants with Industry

7.1. ft-rd contract : Probabilistic visual tracking for annotation of team sport broadcasts

Participants: Patrick Pérez, Etienne Mémin, Nicolas Gengembre.

no. Inria, duration : 18 months.

The aim of this contract is to design probabilistic tracking tools to help operators annotate television broadcasts of teams sport (with a special emphasis on rugby games). The type of semi-automatic tracking that will be targeted is especially challenging because of: the drastic changes of appearance of players within a given shot, and from one shot to another; the joint presence of similar players (teammates), with, in worse cases, large groups with major mutual occlusions; the diversity of shot types (viewpoint and camera motion). None of state-of-the-art robust tracking tools are expected to be satisfactory in this difficult operational context. Hence, we will have to design and combine new tools of feature selection, on-line learning, data fusion, and single and multiple-object probabilistic filtering, to achieve our goal. This work has started in December 2004.

7.2. Cesta contract: Target acquisition

Participants: Jean-Pierre Le Cadre, Pierre Dodin.

no. Inria 104C04270, duration 6 months.

The work carried out in this contract supported by CESTA-DGA (Centre d'études scientifique et technique d'Acquitaine) is described in paragraph 6.3.5.

7.3. Ifremer contract : Automatic analysis of otolith images

Participant: Frédéric Cao.

no. Inria 104C03460, duration : 8 months.

Otoliths are small calcareous concretions that can be found in fishes inner ear. They present rings (as tree trunks) which are representative of the growth of individuals. In Anatole Chessel's internship (Master MVA), we try to study a way to interpolate the geometrical information that is partly observed on otoliths.

7.4. Cifre grant with General Electric Healthcare

Participants: Patrick Bouthemy, Vincent Auvray.

no. Inria 103C19960, duration 36 months.

The X-ray medical exams present two main modalities, as far as image quality is concerned: high quality record or limited radiation fluoroscopy. When the clinician decides to work with fluoroscopy (mainly during interventions), image denoising processing is needed to maintain an acceptable signal-to-noise ratio in the images. The currently used noise reduction filters are inefficient where the anatomy or the devices are moving, since temporal filters are considered leading to motion blurring. We therefore believe that the next generation of filters have to include a motion-compensation stage before filtering in the temporal dimension. The goal of the Ph-D thesis work, carried out in partnership with General Electric Healthcare, is to specify the kind of motion estimation required for that type of image sequences, and how this information should be efficiently used to denoise. The main difficulty is coming from the particular nature of the X-ray images governed by the principle of superposition, which amounts to motion transparency issues (see paragraph 6.2.2).

7.5. Cifre grant with Thales Systèmes Aéroportés

Participant: Jean-Pierre Le Cadre.

CNRS *contract 511008, duration 36 months.*

This contract corresponds to the supervision of Frédéric Bavencoff's Ph-D thesis (3rd year). The problem of performing target motion analysis using noisy bearings-only measurements (BOT TMA) derived from a single moving observer is addressed. It is perhaps in the passive sonar environment that BOT TMA is most familiar, though it is a challenging problem for other contexts like surface and airborne ASW (IR sensors) or passive surveillance via Electronic Support Measurement (ESM), where it is instrumental for evaluating threats, for performing data association, or for correlation processing (e.g. track-to-track association). We have restricted ourselves to poorly observable scenarios. In particular, amplitude of the observer maneuver will be very limited. To avoid some specific problems (linearization, convergence) of the EKF, simulation methods (particle filters) seem definitely relevant. It has been recognized that for a non-maneuvering target the problem is fundamentally relevant of non-linear regression for which batch algorithms are the convenient framework. However, inclusion of constraints in the estimation process for BOT TMA is rarely considered. Moreover, practically, results seem to indicate a neat tendency to "push" the solution toward the constraint bounds. A definitely different approach is to consider reduction of dimensionality. In this context, it seems that restricting estimation to "estimable" parameters could be the best we can do. Furthermore, we want to deal with general constraints or multi-modality. The aim of this work was to develop a way to access to the marginal posterior distribution of the r(range) component and to exploit this distribution to build a confidence interval for constrained estimation of r. Then, we have proposed a method for providing an interval for $\hat{r}$ based on Highest Probability Density (HPD) interval. For a given probability content, say $1 - \alpha$, the HPD method enables to obtain a confidence interval for the range r with a probability content of $1 - \alpha$ by using a sample from the marginal posterior density $\pi\left(r \mid (\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_N)\right)$. This sample is obtained by generating a Markov Chain Monte Carlo (MCMC) sample from the marginal posterior distribution, sample that is itself obtained by generating a MCMC sample from the posterior distribution $\pi\left((r, \theta, v, \alpha) \mid (\hat{\theta}_1, \hat{\theta}_2, \cdots, \hat{\theta}_N)\right)$.

7.6. Eumetsat contract : Alternative tracking methods for derivation of atmospheric motion vectors from Meteosat-8

Participant: Étienne Mémin.

no. Inria 104C05630, duration : 6 months.

The main objective of this project is to compare results of operational computation of atmospheric motion vectors currently used at Eutmesat to optic-flow estimators. This comparison is performed on Meteosat-8 data. More precisely, this evaluation work is carried out on image sequences acquired in different channels. A special attention is given to the detection of low-level clouds. Comparison with EMWF forecast model is provided.

7.7. European initiatives

7.7.1. fp5-ist Lava project

Participants: Patrick Bouthemy, Frédéric Cao, Vincent Samson.

no. Inria 102G0424, duration 36 months

The IST project LAVA ("Learning for Adaptable Visual Assistants") started in May 2002 and involves the following partners : XRCE (prime), IDIAP (Switzerland), Lear team from Inria Rhône-Alpes and Vista team, RHUL (UK), University of Graz (Austria), University of Lund (Sweden), et ANU (Australia). It gathers groups from computer vision, machine learning and cognitive sciences and focuses on two key problems : categorizing objects in static images and interpreting events in video. Two formal meetings have been held this year (in Grenoble in January and in Southampton in June) where we have presented our contributions related to two work-packages "Image models, descriptors and kernels" and "Recognition of objects and events". We are particularly interested in the issue of video content analysis which arises in numerous applications such as video summarization, video retrieval or surveillance. Solutions to this problem include first the design and extraction of discriminating features followed by a supervised classification stage. In this context, our contribution is twofold: 1) Recognizing semantic events in video requires to preliminary learn the different classes of video events. We have proposed an original approach based on a robust partitional clustering algorithm applied in parallel to each predefined class in order to capture their internal data structure (see paragraph 6.4.3; 2) The second point concerns the extraction of motion-based numerical features that could characterize the dynamic content of a video. We are currently considering the use of more localized descriptors. These descriptors are extracted either from spatio-temporal profiles of motion-detection binary masks, or from local statistics computed around space-time interest points. A complementary task that we have addressed is the collection of new video databases. In particular, we have collected a set of basket-ball videos. Each video is a short image sequence of several seconds (between 50 and 200 frames) recorded with a static camera involving one or two amateur players. A total of 300 video segments has been stored. Three types of actions have been predefined: "middle shot", "lay-up" and "one-on-one". For each event, different players and variable shooting conditions (pose, illumination) are considered so as to reflect the heterogeneousness of the observations of a given semantic event. The other databases concern amateur tennis events (serves, strokes, steps) as well as categorization of simple human gestures (e.g. hand waving, clapping, nodding/shaking head). Satisfactory results concerning both the learning stage and the classification one have been obtained on these sets of video data.

7.7.2. fp6-ist noe Muscle

Participants: Patrick Bouthemy, Frédéric Cao.

no. Inria 104A04950, duration 48 months

The Vista team is involved in the FP6 Network of Excellence MUSCLE ("Multimedia Understanding through Semantics, Computation and Learning") started in April 2004. It gathers 41 research teams all over Europe from public institutes, universities or research labs of companies. Due to the convergence of several strands of scientific and technological progress, we are witnessing the emergence of unprecedented opportunities for the creation of a knowledge driven society. Indeed, databases are accruing large amounts of complex multimedia documents, networks allow fast and almost ubiquitous access to an abundance of resources and processors have the computational power to perform sophisticated and demanding algorithms. However, progress is hampered by the sheer amount and diversity of the available data. As a consequence, access can only be efficient if based directly on content and semantics, the extraction and indexing of which is only feasible if achieved automatically. MUSCLE aims at creating and supporting a pan-European Network of Excellence to foster close collaboration between research groups in multimedia datamining on one hand and machine learning on the other hand, in order to make breakthrough progress towards different objectives.

We have contributed to State-of-the-Art reports in three workpackages of MUSCLE, by writing chapters respectively on video processing (WP5, Single modality), on the use of image and audio features for video summarization (WP6, Improving Performance via Cross-Modal Integration), on robust classification in video

analysis (WP8, Machine Learning for Multimedia Content). We participated to the Kick-off meeting in Heraklion in May 2004. F. Cao has given a talk in the November meeting in Malaga, entitled "An a contrario approach to hierarchical clustering validity assessment".

7.7.3. *fp6 Fluid project*

Participants: Étienne Mémin, Patrick Bouthemy, Nicolas Papadakis.

no. Inria, duration 36 months

The FLUID project is a FP6 STREPS project labeled in the FET-open program. It has started in November 2004. We are the prime contractor and E. Mémin is the scientific coordinator of the project. This 3-year project aims at studying and developing new methods for the estimation, the analysis and the description of complex fluid flows from image sequences. The consortium is composed of five academic partners (Inria, Cemagref, University of Mannheim, University of Las Palmas de Gran Canaria and the LMD, "Laboratoire de Météorologie Dynamique") and one industrial partner (La Vision company) specialized in PIV (Particle Image Velocimetry) system. The project gathers computer vision scientists, fluid mechanicians and meteorologists. The first objective of the project consists in studying novel and efficient methods to estimate and analyse fluid motions from image sequences. The second objective is to guarantee the applicability of the developed techniques to a large range of experimental fluid visualization applications. To that end, two specific areas are considered: meteorological applications and experimental fluid mechanics for industrial evaluation and control. From the application point of view, the project will particularly focus on 2D and 3D wind field estimation, and on 2D and 3D particle image velocimetry. A reliable structured description of the computed fluid flow velocity field will further allow us to address the tracking of turbulent structures in the flows.

7.8. National initiatives

7.8.1. *rntl Domus Videum project*

Participants: Patrick Bouthemy, François Coldefy.

no. Inria 202C01000, duration 30 months.

Domus Videum is a RNTL project involving Thomson-Multimedia as prime contractor, academic partners, Irccyn, Irin and Irisa (Metiss, Temics and Vista projects-teams), along with public organisms, INA and SFRS. This project was completed in July 2004. Its main objective was to design a video archiving and browsing software prototype dedicated to digital Personal Video Recorders (PVR). The developed prototype enables to store videos on the PVR according to the user profile evaluated dynamically. Then, it provides three ways of browsing the stored TV programs: a short insight of the video for previewing, which contains the important sequences of the video (highlights), a long summary which presents the video more extensively, and a thematic video browsing, the shots being merged in homogeneous audio and visual clusters. These functionalities would help the user to refine his choice before watching a video. The application focuses on sports videos for which short and long summary are particularly adapted and on documentary films. We were concerned with the short video summary module. We have first developed a temporal video segmentation based on the detection of both shot changes and camera motion changes. Video abstraction is obtained by considering audio and visual descriptors (see §6.4.4 for details on the two designed methods, a supervised one and a non-supervised one). Convincing results on tennis and soccer videos have been obtained. Experiments on more than twenty hours of tennis TV programs corresponding to different tournaments have been successfully carried out to supply abstracts containing the winning serves and the best rallies. The unsupervised method has been tested on seven soccer TV programs for a total duration of almost 20 hours, providing abstracts including the goals and the major actions (only one false alarm was obtained).

7.8.2. *riam Feria project*

Participants: Patrick Bouthemy, Brigitte Fauvet, Fabien Spindler.

no. Inria 203C1460, duration : 24 months.

The Feria project belongs to the RIAM programme and is financially supported by the Ministry of Industry. The partners are INA, C&S, Vecsys and NDS companies, Arte, IRIT, Inria (Texmex and Vista teams). The goal of the project is to build a general and open framework allowing the easy development of applications for editing interactive video documents. We are concerned with the design of video processing and representation tools. So far, we have completed several modules related to video analysis: a shot change detection module particularly focused on progressive transitions (dissolve, wipe,...), a camera motion characterization tool, a novel technique for selecting an appropriate set of key-frames to represent the visualized scene while accounting for the camera motion. Besides, we have also developed a face tracking algorithm involving the combination of region (face area) and point (eyes) tracking (based on the method described in [14]) along with some specific constraints, in order to ensure robustness to significant face orientation changes (from front-view to side-view of the face) and to fast motion of the tracked face.

7.8.3. Predit MobiVip project

Participant: Patrick Bouthemy.

no. Inria 203A20050, duration 36 months.

This large project started in December 2003 and is headed by Inria Sophia-Antipolis. It is concerned with the navigation of mobile vehicles in urban environments. Within this project, the work carried out in Irisa (mainly by the project-team Lagadic) consists in designing autonomous vision-based navigation techniques using an image database of the environment.

8. Other Grants and Activities

8.1. Regional initiatives

8.1.1. Brittany Region contract : video summarization

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

no. Inria 401C0576, duration 36 months.

This contract supplies the financial support of the Inria thesis grant allocated to Gwénaëlle Piriou. The thesis started in November 2001 and is described in paragraphs 6.1.1 and 6.4.1.

8.1.2. cper initiative : Aerobio project

Participant: Etienne Mémin.

no. Inria 402C06340 and 403C12560, duration 48 months.

The AEROBIO project associates the Vista team and the Cemagref-Rennes for the study of methods to prevent contamination in food transformation industry. The purpose of this project was to initiate a research program on the analysis of fluid flow from image sequences. Within this framework, we have in particular worked on the physical evaluation of a dense motion estimator dedicated to fluid flows. This statistical evaluation was carried out on a free turbulent shear layer flow.

8.2. National initiatives

8.2.1. aci Ministry grant : Assimage project

Participant: Etienne Mémin.

no. Inria 103C18930, duration 36 months.

The ASSIMAGE project, labeled within the "Masse de données" ACI program, involves three Inria teams (Clime in Rocquencourt, Idopt in Grenoble, and Vista), three Cemagref groups (located in Rennes, Montpellier and Grenoble), the LEGI laboratory and the LGGE laboratory both located in Grenoble. It has started in September 2003. The aim of the ASSIMAGE project is to develop methods for the assimilation of images in mathematical models governed by partial differential equations. The targeted applications concern predictions

of geophysical flows. Our contribution is concerned with the tracking of vortex structures in collaboration with Cemagref Rennes.

8.2.2. aci Ministry grant : Behaviour project

Participants: Patrick Pérez, Patrick Bouthemy, Aurélie Bugeau.

no. Inria 104C08130, duration 36 months.

This project has been contracted in October 2004. The Behaviour project is granted by the ACI program on Security and Computer Science. It involves the Compiègne University of Technology (Heudiasyc lab) as the prime, along with PSA company (Innovation and Quality group) and Vista. The context is the visual monitoring of car drivers, based of videos shot inside the car, such that hypo-vigilant behaviors (mainly drowsiness and distraction) can be detected. The aim of the project is to provide new tools to perform automatically the recognition of a wide range of elementary behavioural items such as blinks and eye direction, yawn, nape of the neck, posture, head pose, interaction between face and hands, facial actions and expressions, control of the car radio, or mobile phone handling. To achieve such a fine grain activity recognition, we intend to combine advanced multi-object tracking tools with techniques to learn probabilistic models in high dimensional spaces with limited training data (video shots within both simulators and real-world environments and manually labeled).

8.2.3. aci Ministry grant : Modyncell5d project

Participants: Charles Kervrann, Patrick Bouthemy, Jérôme Boulanger.

no. Inria 104C08140, duration 36 months.

This project, labeled within the IMPBio ACI program, has been contracted in October 2004. The Vista team is the prime contractor of the project MODYNCELL5D 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 UMR CNRS-6026 ("Interactions Cellulaires et Moléculaires" Laboratory - "Structure et Dynamique des Macromolécules" team, University of Rennes 1). This project aims at extracting motion components by statistical learning and spatial statistics from the computed partial or complete trajectories and compare trajectories among normal-type and wild-type cells. The challenge is to track GFP tags with high accuracy in movies representing several gigabytes of image data and collected and processed automatically to generate information on complex trajectories. Methods should be developed for two proteins: CLIP 170 involved in the kinetochores anchorage (in the segregation of chromosomes to daughter cells, the chromosomes appear to be pulled via a so-called kinetochore attached to chromosome centromeres) ; and Rab6a' involved in the regulation of transport from the Golgi apparatus to the endoplasmic reticulum.

8.2.4. aci Ministry grant : project related to the development of new fluorescent probes

Participant: Charles Kervrann.

CNRS *contract, duration 36 months.*

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

8.2.5. Initiatives supported by cnrs

- DSTIC *specific action 67 : particle filtering methods*
  Participants: Jean-Pierre Le Cadre, Thomas Bréhard.
  Our contribution to that action led by F. Le Gland (Aspi team) is concerned with the tracking of moving entities in sonar/radar or in video.

8.3. Bilateral international co-operation

8.3.1. *Royal Society-cnrs program, France-England*

Participant: Patrick Pérez.

It is a two-year collaboration with Jaco Vermaak from the Eng. Dpt of the University of Cambridge The aim of this project is to pursue and extend the research conducted by J. Vermaak and P. Pérez in the past four years on probabilistic tools for visual tracking and trajectography. The emphasis will now be on the principled handling of a varying (and unknown) number of entities to be tracked. This new collaboration funded jointly by the Royal Society and CNRS has started in October 2004 with a first one-week stay of J. Vermaak with Vista. At this occasion, the first bases for a new data-association based filter, with proper handling of an "existence" binary label in the state space, has been set up. The single-object version of this filter is under development in the context of colour-based visual tracking. The extension to the multiple-object case is planned to rely on an extension of the JPDAF (see paragraph 3.3).

8.3.2. *Collaboration with University of Buenos-Aires*

Participants: Patrick Bouthemy, Étienne Mémin.

We have an (unformal) collaboration with the Faculty of Engineering of the University of Buenos-Aires (teams of Prof. Bruno Cernuschi-Frias, Elec. Dpt, and Guillermo Artana, Fluid Mechanics Dpt). Reciprocal visits have been organized this year (see next item). P. Bouthemy and E. Mémin did a two-weeks visit in October 2004 in Buenos-Aires (with the financial support of the Inria DREI). In particular, they gave two talks, resp., "Contextual image motion analysis: issues and new models for motion detection and motion recognition" and "Motion analysis in fluid imagery". Topics addressed in this collaboration are visual tracking, motion estimation in experimental fluid mechanics, motion recognition.

8.3.3. *Visiting scientists*

- Bruno Cernuschi-Frias spent one month and half in our team as Inria visiting scientist in June and July, and another two weeks in November.

- Short visits by G. Artana (University of Buenos-Aires), A. Doucet (University of Cambridge), J.-O. Eklundh (KTH Stockholm), P. Fua (EPFL Lausanne), S. Marchand-Maillet (University of Geneva), A. Tamtaoui (INPT Rabat), J. Vermak (University of Cambridge).

8.3.4. *Project reviewing and consultancy*

- P. Bouthemy expertised one company project in the domain of digital photography for ANVAR.

- P. Bouthemy and J.-P. Le Cadre are deputy members of the committee of the 61th section at Universiy of Rennes 1.

- F. Cao served as member of the jury of the admission to the École Normale Supérieure de Paris (written test) and École Normale Supérieure de Cachan (written and oral test).

- C. Kervrann is member of the Scientific Council of the Biometry and Artificial Intelligence Department of Inra.

- P. Pérez conducted one week of consultancy for Microsoft Research, Cambridge, UK, in July 2004.

9. Dissemination

9.1. Leadership within scientific community

- The Vista team is involved in the French network GDR ISIS, "Information, Signal and Images" (http://www-isis.enst.fr/).

- C. Kervrann participates in the network GDR 2588, "Microscopie Fonctionnelle du Vivant" (http://www.gdr2588.cnrs.fr.

- P. Bouthemy participates in the Board Committee of Technovision programme launched by the Ministry of Research and by DGA and aiming at supporting evaluation projects of image processing and computer vision techniques (call for proposals published in April 2004 and selection of projects done in September 2004).

- J.-P. Le Cadre participates in the working group "Ultimate performances in trajectography" settled by Thales, and in an evaluation group of the DGA.

- P. Pérez is head of the committee ("Commission Personnel") which oversees all scientific non-permanent hirings in Irisa / Inria-Rennes.

- J.-F. Yao is member of the executive comittee of MAS, a section of the SMAI.

- *Editorial boards of journals*

  - J.-P. Le Cadre is Associate Editor of *Advances in Information Fusion* (ISIF).

- *Conference organization*

  - P. Bouthemy was member of the MICCAI'2004 Local Organizing Committee.
  - F. Cao was member of the Organizing Committee of the workshop MIG'2004 ("Mathématique, Image et Gestaltisme") held in Ile de Berder.
  - C. Kervrann was member of the Organizing Committee of the 5th French-Danish workshop SSIAB'2004 (Spatial Statistics and Image Analysis in Biology) held in Grenoble in May 2004.

- *Technical program committees of conferences*

  - P. Bouthemy : TPC member of ACCV'2004, CVPR'2004, ICPR'2004, ICME'2004, CIVR'2004, Medianet'2004, RFIA'2004, SCVMA'2004, SMVP'2004, CIVR'2005, IBPRIA'2005, CAIP'2005, ICARP'2005, OTCBVS'2005, TAIMA'2005.
  - F. Cao : TPC member of SSIAI'2004, ICIP'2004.
  - C. Kervrann : TPC member of MICCAI'2004.
  - J.-P. Le Cadre : TPC member of Fusion'2004.
  - E. Mémin : TPC member of CVPR'2004, ECCV'2004, CVPR'2005.
  - P. Pérez : TPC member of SIGGRAPH'2004, ICPR'2004, NIPS'2004, CVPR'2005.

9.2. Teaching

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

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

- DIIC INC, Ifsic, University of Rennes 1 (E. Mémin, E. Arnaud : motion analysis; E. Mémin : Markov models for image analysis; E. Mémin is in charge of the INC (Digital image analysis and communication) channel.

- Insa Rennes, Electrical Engineering Dpt (G. Piriou, T. Veit : computer vision).

- Master TIS, University of Cergy and ENSEA (P. Bouthemy : video indexing).

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

- Master MVA, ENS Cachan (F. Cao : image filtering and PDE).

- ENSAI Rennes, 3rd year (C. Kervrann : statistical models and image analysis).

- Graduate student trainees and interns : G. Jeannic (DIIC INC and Master STIR), A. Chessel (Master MVA, ENS Cachan), A. Lehmann (ETH Zürich).

- External thesis supervision :

  - P. Muse et F. Sur (ENS Cachan) co-supervised by F. Cao, A. Chessel (Ifremer Brest) supervised by F. Cao;
  - F. Bavencoff (Thales Airborne Systems, Cifre grant, see §7.5)) and F. Celeste (CTA) supervised by J.-P. Le Cadre;
  - A. Lehuger (FT-RD, Rennes) supervised by P. Pérez;
  - I. Bechar (Research Ministry grant - ACI IMPBio, MIA unit of Inra Jouy-en-Josas) co-supervised by C. Kervrann.

9.3. Participation in seminars, invitations, awards

- P. Bouthemy was invited as key-note speaker at the International Workshop on Spatial Coherence for Visual Motion Analysis held in Prague in May 2004 in conjunction with ECCV'2004 and gave a talk entitled "2D motion description and contextual motion analysis: issues and new models". He gave an invited talk ("Modélisation statistique de contenus spatio-temporels et détection d'événements dans des vidéos") at the Computer Vision for Telecom Days of the CNRT-TIM (National Center of Technological Research - Telecom, Images and Multimedia), Rennes, France, May 2004.

- F. Cao was invited to give a talk in the Mathematics and Image Analysis Conference (MIA'2004), entitled "Extracting meaningful curves from images". He gave an invited talk in the PDE and Numerical Analysis Seminar of J. Dieudonné Laboratory in University of Nice on "Automatic thresholds in shape recognition".

- C. Kervrann was an invited speaker at the Interdisciplinary School "Microscopie Fonctionnelle en Biologie" organized by the GDR-2588 and held in Oléron in September 2004.

- P. Pérez gave an invited talk ("Probabilistic tracking in videos") at the Computer Vision for Telecom Days of the CNRT-TIM, Rennes, France, May 2004. He gave an invited talk ("Pixels and particles: Sequential Monte Carlo for visual tracking") at the Statistical Days of the SMAI (Society for Applied and Industrial Mathematics), session on particle filters, Nancy, France, Sept. 2004. He also gave an invited talk ("Some statistical tools for the analysis of visual motion in image sequences") at the STAR Days (Statistics in Rennes), Rennes, France, Oct. 2004.

- P. Pérez was invited to present his past work on interactive image editing at the Microsoft Digital House, Paris, France, Oct. 2004.

- J.F. Yao gave an invited talk in MIA'2004 Conference (Mathematics and Image Analysis) held in Paris, September 2004. The talk title was "Models for mixed-states data with application to analysis of video sequences". He gave an invited talk in MIG'2004 workshop held in Ile de Berder, June 2004. The talk was entitled "Sur la modélisation des mesures de mouvement dans une séquence d'images". He gave an invited talk in the French-Danish workshop SSIAB'2004 (Spatial Statistics and Image Analysis in Biology) entitled "Motion-based models for analysis and interpretation of video sequences".

10. Bibliography

Major publications by the team in recent years

[1] P. BOUTHEMY, M. GELGON, F. GANANSIA. *A unified approach to shot change detection and camera motion characterization*, in "IEEE Trans. on Circuits and Systems for Video Technology", vol. 9, nº 7, October 1999, p. 1030-1044.

[2] T. CORPETTI, E. MÉMIN, P. PÉREZ. *Dense estimation of fluid flows.*, in "IEEE Trans. on Pattern Analysis and Machine Intelligence", vol. 24, nº 3, March 2002, p. 365-380.

[3] F. DAMBREVILLE, J.-P. LE CADRE. *Detection of a Markovian target with optimization of the search efforts under generalized linear constraints*, in "Naval Research Logistics,", vol. 49, nº 2, February 2002, p. 117-142.

[4] R. FABLET, P. BOUTHEMY. *Motion recognition using non parametric image motion models estimated from temporal and multiscale cooccurrence statistics.*, in "IEEE Trans. on Pattern Analysis and Machine Intelligence", vol. 25, nº 12, December 2003, p. 1619-1624.

[5] C. HUE, J.-P. LE CADRE, P. PÉREZ. *Tracking multiple objects with particle filtering*, in "IEEE Trans. on Aerospace & Electronic Systems", vol. 38, nᵒ 3, July 2002, p. 791-812.

[6] C. KERVRANN. *Learning probabilistic deformation models from image sequences*, in "Signal Processing", vol. 71, 1998, p. 155-171.

[7] P. MUSE, F. SUR, F. CAO, Y. GOUSSEAU. *Unsupervised thresholds for shape matching*, in "IEEE Int. Conf. on Image Processing, Barcelona", vol. Barcelona, September 2003.

[8] 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, nᵒ 5, May 1998, p. 703-719.

[9] J.-M. ODOBEZ, P. BOUTHEMY. *Robust multiresolution estimation of parametric motion models*, in "Journal of Visual Communication and Image Representation", vol. 6, nᵒ 4, December 1995, p. 348-365.

[10] P. PÉREZ, A. CHARDIN, J.-M. LAFERTÉ. *Non-iterative manipulation of discrete energy-based models for image analysis*, in "Pattern Recognition", vol. 33, nᵒ 4, April 2000, p. 573-586.

Books and Monographs

[11] T. CORPETTI. *Analyse d'écoulements fluides à partir de séquences d'images*, Collection Traitement du Signal et de l'Image, Hermes Science Publications, Lavoisier, 2004.

Doctoral dissertations and Habilitation theses

[12] E. ARNAUD. *Méthodes de filtrage pour du suivi dans des séquences d'images - Application au suivi de points caractéristiques*, Ph. D. Thesis, Université de Rennes 1, mention Traitement du Signal et Télécommunications, November 2004.

[13] F. CAO. *Sur quelques problèmes mathématiques de l'analyse des formes*, Habilitation à diriger des recherches, Université de Paris Dauphine, December 2004.

Articles in referred journals and book chapters

[14] E. ARNAUD, E. MÉMIN, B. CERNUSCHI-FRIAS. *Conditional filters for image sequence based tracking - Application to point tracking*, in "IEEE Trans. on Image Processing", to appear, 2004.

[15] F. CAO. *Application of the Gestalt principles to the detection of good continuations and corners in image level lines*, in "Computing and Visualisation in Science", vol. 7, 2004, p. 3-13.

[16] F. CAO, P. MUSÉ, F. SUR. *Extracting meaningful curves from images*, in "J. Mathematical Imaging and Vision", to appear, 2004.

[17] A. CRIMINISI, P. PÉREZ, K. TOYAMA. *Region filling and object removal by exemplar-based inpainting*, in "IEEE Trans. on Image Processing", vol. 13, nᵒ 9, 2004, p. 1200-1212.

[18] F. DAMBREVILLE, J.-P. LE CADRE. *Spatio-temporal multi-mode management for moving target detection*, in "Journal of Information Fusion", vol. 5, 2004, p. 169-178.

[19] M. GELGON, P. BOUTHEMY, J.-P. LE CADRE. *Recovery of the trajectories of multiple moving objects in an image sequence with a PMHT approach*, in "Image and Vision Computing Journal", to appear, 2004.

[20] C. KERVRANN, D. LELAND, L. PARDINI. *Robust incremental compensation of the light attenuation with depth in 3D fluorescence microscopy*, in "J. Microscopy", vol. 214, nº 3, 2004, p. 297–314.

[21] N. PEYRARD, P. BOUTHEMY. *Motion-based selection of relevant video segments for video summarization*, in "Multimedia Tools and Applications Journal", to appear, 2004.

[22] J.-N. PROVOST, C. COLLET, P. ROSTAING, P. PÉREZ, P. BOUTHEMY. *Hierarchical Markovian segmentation of multispectral images for the reconstruction of water depth maps*, in "Computer Vision and Image Understanding", vol. 93, nº 2, 2004, p. 155-174.

[23] P. PÉREZ, J. VERMAAK, A. BLAKE. *Data fusion for visual tracking with particles*, in "Proceedings of IEEE", vol. 92, nº 3, 2004, p. 495-513.

[24] J. VERMAAK, S. GODSILL, P. PÉREZ. *Monte Carlo filtering for multi-target tracking and data association*, in "IEEE Trans. on Aerospace and Electronic Systems", to appear, 2004.

Publications in Conferences and Workshops

[25] E. ARNAUD, E. MÉMIN. *Optimal importance sampling for tracking in image sequences: application to point tracking*, in "European Conference on Computer Vision, ECCV'04, Prague, Czech Republic", May 2004.

[26] E. ARNAUD, E. MÉMIN, B. CERNUSHI FRIAS. *A robust stochastic filter for point tracking in image sequences*, in "Asian Conference on Computer Vision, ACCV'04, Jeju Island, Korea", January 2004.

[27] E. ARNAUD, E. MÉMIN, B. CERNUSHI FRIAS. *Filtrage conditionnel pour la trajectographie dans des séquences d'images - Application au suivi de points*, in "14ème Congrès Francophone AFRIF-AFIA de Reconnaissance des Formes et Intelligence Artificielle, RFIA'04, Toulouse, France", January 2004.

[28] P. BOUTHEMY. *2D motion description and contextual motion analysis: issues and new models*, in "Proc. Int. Workshop on Spatial Coherence for Visual Motion Analysis, SCVMA'2004, Prague, Czech Republic", May 2004.

[29] T. BRÉHARD, J.-P. LE CADRE. *A new approach for the bearings-only problem: estimation of the variance-to-range ratio*, in "7th International Conference on Information Fusion, Stockholm, Sweden", 2004.

[30] T. BRÉHARD, J.-P. LE CADRE. *Un problème de filtrage non linéaire avec variance d'état inconnue: le pistage mono-cible par mesure d'angle seul*, in "XXXVIèmes Journées de Statistique de la SFDS, Montpellier, France", 2004.

[31] F. CAO, R. FABLET. *Automatic morphological detection of otolith nucleus*, in "Proc. Int. Conf. Pattern Recognition, ICPR'04, Cambridge, UK", August 2004.

[32] F. COLDEFY, P. BOUTHEMY, M. BETSER, G. GRAVIER. *Tennis video abstraction from audio and visual cues*, in "Proc. IEEE Int. Workshop on Multimedia Signal Processing, MMSP'2004, Siena, Italy", Sept. 2004.

[33] F. COLDEFY, P. BOUTHEMY. *Unsupervised soccer video abstraction based on pitch, dominant color and camera motion analysis*, in "Proc. ACM Multimedia 2004, New-York", October 2004.

[34] B. FAUVET, P. BOUTHEMY, P. GROS, F. SPINDLER. *A geometrical key-frame selection method exploiting dominant motion estimation in video*, in "International Conference on Image and Video Retrieval, CIVR'04, Dublin, Ireland", July 2004.

[35] C. KERVRANN, S. BA. *Approche non-paramétrique en restauration d'image avec préservation des discontinuités*, in "14ème Congrès Francophone AFRIF-AFIA de Reconnaissance des Formes et Intelligence Artificielle (RFIA'04), Toulouse, France", January 2004.

[36] C. KERVRANN. *An adaptive window approach for image smoothing and structures preserving*, in "European Conference on Computer Vision, ECCV'04, Prague, Czech Republic", May 2004.

[37] C. KERVRANN, A. TRUBUIL. *An adaptive window approach for Poisson noise reduction and structure preserving in confocal microscopy*, in "International Symposium on Biomedical Imaging, ISBI'04, Arlington, VA", April 2004.

[38] F. LAUZE, P. KORNPROBST, E. MÉMIN. *A multiscale approach for linear least-square based optical flow estimation*, in "Proc. British Machine Vision Conference, London", September 2004.

[39] G. PIRIOU, P. BOUTHEMY, J.-F. YAO. *Extraction of semantic dynamic content from videos with probabilistic motion models*, in "European Conf. on Computer Vision, ECCV'04, Prague, Czech Republic", May 2004.

[40] G. PIRIOU, P. BOUTHEMY, J.-F. YAO. *Learned probabilistic image motion models for event detection in videos*, in "Proc. Int. Conf. Pattern Recognition, ICPR'04, Cambridge, UK", August 2004.

[41] G. PIRIOU, F. COLDEFY, P. BOUTHEMY, J.-F. YAO. *Détection supervisée d'événements à l'aide d'une modélisation probabiliste du mouvement perçu*, in "14ème Congrès Francophone AFRIF-AFIA de Reconnaissance des Formes et Intelligence Artificielle, RFIA 2004, Toulouse, France", January 2004.

[42] V. SAMSON, P. BOUTHEMY. *Learning classes for video interpretation with a robust parallel clustering method*, in "Proc. Int. Conf. Pattern Recognition, ICPR'04, Cambridge, UK", August 2004.

[43] T. VEIT, F. CAO, P. BOUTHEMY. *Probabilistic parameter-free motion detection*, in "Conf. Computer Vision and Pattern Recognition, CVPR'04, Washington, DC", June 2004.

[44] L. XIE, P. PÉREZ. *Slightly supervised learning of part-based appearance models*, in "Proc. IEEE Workshop on Learning in Computer Vision and Pattern Recognition, LCVPR'04, Washington, DC", June 2004.

Internal Reports

[45] T. BREHARD, J.-P. LE CADRE. *Initialization of particle filter and posterior Cramer-Rao bound for bearings-only tracking in modified polar coordinate system*, Technical report, n° 1588, IRISA, 2004.

[46] F. CAO, P. MUSÉ, F. SUR. *Extracting meaningful curves from images*, Technical report, n° 5067, INRIA, 2004, http://www.inria.fr/rrrt/rr-5067.html.

[47] P. MUSÉ, F. SUR, F. CAO, Y. GOUSSEAU, J.-M. MOREL. *Accurate estimates of false alarm number in shape recognition*, Technical report, n° 5086, INRIA, 2004, http://www.inria.fr/rrrt/rr-5086.html.

[48] T. VEIT, F. CAO, P. BOUTHEMY. *An a contrario framework for motion detection*, Technical report, n° 5313, INRIA, 2004, http://www.inria.fr/rrrt/rr-5313.html.

Bibliography in notes

[49] J. BESAG. *Spatial interactions and the statistical analysis of lattice systems*, in "J. Royal Statistical Society B", vol. 148, 1974, p. 1-36.

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

[51] 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.

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

[53] A. DESOLNEUX, L. MOISAN, J.-M. MOREL. *A grouping principle and four applications*, in "IEEE Trans. on Pattern Analysis and Machine Intelligence", vol. 25, n° 4, 2003, p. 508-513.

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

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

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

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

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

[59] M. ISARD, A. BLAKE. *CONDENSATION–Conditional Density Propagation for Visual Tracking*, in "Int. Journal of Computer Vision", vol. 29, nº 1, 1998, p. 5-28.

[60] G. KITAGAWA. *Monte Carlo filter and smoother for non-Gaussian nonlinear state space models*, in "J. Computational and Graphical Stat.", vol. 5, nº 1, 1996, p. 1-25.

[61] E. MÉMIN, P. PÉREZ. *Hierarchical estimation and segmentation of dense motion fields*, in "Int. Journal of Computer Vision", vol. 46, nº 2, February 2002, p. 129-155.

[62] 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, nº 5, May 1998, p. 703-719.

[63] J.-M. ODOBEZ, P. BOUTHEMY. *Robust multiresolution estimation of parametric motion models*, in "Journal of Visual Communication and Image Representation", vol. 6, nº 4, December 1995, p. 348-365.

[64] J.-M. ODOBEZ, P. BOUTHEMY. *Direct incremental model-based image motion segmentation for video analysis*, in "Signal Processing", vol. 6, nº 2, 1998, p. 143-155.