INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Project-Team VISTA

Vision Spatio-Temporelle et Apprentissage

Rennes

THEME COG

Activity Report

2005

Table of contents

1. Team

Head of project-team
Patrick Bouthemy [DR Inria]

Administrative assistant
Huguette Béchu [TR Inria, shared with Temics team]

Inria research scientists
Frédéric Cao [CR]
Ivan Laptev [CR, from 01/09/05]
Patrick Pérez [DR]
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, (collaborateur extérieur)]

Inria technical staff
Brigitte Fauvet [Feria project, until 30/06/05]
Nicolas Gengembre [FT-RD contract]
Patrick Heas [Fluid IST project, from 01/06/05]

Post-doctoral fellows
Ivan Laptev [Inria, until 31/08/2005]
Venkatesh Babu Radhakrishnan [Ercim, until 15/03/2005]
Gwénaëlle Piriou [Lecturer (ATER) Ifsic, University of Rennes 1, from 01/10/05]

Ph-D students
Vincent Auvray [Cifre grant, General Electric Healthcare]
Jérôme Boulanger [Inra grant until 31/05/05, Inria grant from 01/06/05]
Thomas Bréhard [DGA grant]
Aurélie Bugeau [Research Ministry grant]
Tomas Crivelli [Faculty of Engineering, University of Buenos-Aires, from 01/09/05]
Anne Cuzol [Research Ministry grant]
Alexandre Hervieu [Inria grant, from 01/11/05]
Nicolas Papadakis [Inria grant, FP6-IST Fluid project]
Gwénaëlle Piriou [Inria grant, Brittany Region contract, until 30/09/05]
Thomas Veit [Inria grant, Brittany Region contract (50%)]

2. Overall Objectives

2.1. 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 video images, but also meteorological satellite images, video-microscopy, x-ray images). We are investigating methods to analyze dynamic scenes, and, more generally, dynamic phenomena, within image sequences. We address the full range of problems raised by the analysis of such dynamic contents with a focus on image motion analysis issues: denoising, motion detection, motion estimation, motion-based segmentation, tracking, motion recognition and interpretation with learning. We usually rely on statistical approaches, resorting to: Markov models, Bayesian inference, robust estimation, *a contrario* decision, particle filtering, learning. Application-wise, we focus our attention on three main domains: content-aware video applications, meteorological imaging and experimental visualization in fluid mechanics, biological imaging. For that, a number of collaborations, academic and industrial, national and international, are set up.

3. Scientific Foundations

3.1. Motion estimation and motion segmentation with mrf models

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 [78]). 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 [71] the quadratic function $\rho_1(x) = x^2$ for the data-driven term in (3). The minimization of the energy function E was carried out on l and Θ in an iterative alternate way, and the number of regions N_{reg} was determined by introducing an extraneous label and using an appropriate statistical test. We later chose a robust estimator for ρ_1 [86], [87]. It allowed us to avoid the alternate minimization procedure and to determine or update the number of regions through an outlier process in every region.

Specifying (simple) MRF models at a pixel level (i.e., sites are pixels and a 4- or 8-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 [81] 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 [79]. 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 [80]. Then, robust estimators were popularized leading to the introduction of so-called auxiliary variables ζ now taking their values in $[0, 1]$ [85]. Multigrid MRF are moreover involved, and multiresolution incremental schemes are exploited to compute optical flow in case of large displacements. Dense optical flow and parametric motion models can also be jointly considered and estimated, which enables to supply a segmented velocity field [84]. Depending on the followed approach, the third term of the energy $E(\mathbf{w}, \zeta)$ can be optional.

3.2. Fluid motion analysis

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 [85], [84]. These generic approaches are based on the brightness constancy assumption of the points along their trajectory ($\frac{df}{dt} = 0$), along with the spatial smoothness assumption of the motion field. These estimators are defined as the minimizer of the following energy function:

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

The penalty function ρ is usually the L_2 norm, but it may be substituted for a robust function attenuating the effect of data that deviate significantly from the brightness constancy assumption [85], and enabling also to implicitly handle the spatial discontinuities of the motion field.

Contrary to usual video image sequence contents, fluid images exhibit high spatial and temporal distortions of the luminance patterns. The design of alternative approaches dedicated to fluid motion thus constitutes a widely-open research problem. It requires to introduce some physically relevant constraints which must be embedded in a higher-order regularization functional [74]. The method we have devised for fluid motion involves the following global energy function:

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

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

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

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 [75]. Such a compact representation enables for instance to tackle difficult tracking problems. As a matter of fact, the problem amounts here to track high dimensional complex objects such as surfaces, level lines, or vector fields. As these objects are only partially observable from images and driven by non linear 3D laws, we have to face a tough tracking problem of large dimension for which no satisfying solution exists at the moment.

3.3. Object tracking with non-linear probabilistic filtering

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:

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

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 [77], [83], [82], 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, to image comparison, 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 [76], 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. We call this methodology 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, to image comparison and to shape matching.

4. Application Domains

4.1. 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) :

- Content-aware video applications (INA, Thomson, FT-RD, PSA);
- Experimental fluid mechanics (Cemagref) and meterological imagery (LMD). We are also leading the FET-IST European project Fluid (with see paragraph 7.5.4) and are in the Inria associate team FIM with the University of Buenos-Aires (see paragraph 8.2.2);
- Biological imagery (Inra, Curie Institute, Biology Dpt of University of Rennes 1)
- Surveillance (Onera, Thales, collaborations are nevertheless considered only from an academic viewpoint). The main addressed issues are search and surveillance, navigation, distributed tracking with a sensor network.

4.1.1. Content-aware video applications

The amount of video footage is constantly increasing due to the dissemination of video cameras, the broadcasting of TV programs by multiple means, the seamless acquisition of personal videos,...The exploitation of video material, whatever its usage, requires automatic (or at least semi-automatic) tools to process video contents. A wide range of applications can be envisaged dealing with editing, analyzing, annotating, browsing and authoring video contents. Video indexing and retrieval for audio-visual archives is, for instance, a major application, which is receiving lot of attention. Other needs include the creation of enriched videos, the design of interactive video systems, the generation of video summaries, and the development of re-purposing frameworks (specifically, for 3G mobile phones and Web applications). For most of all these applications, tools for segmenting videos, detecting events or recognizing actions are usually required.

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

4.1.2. Experimental fluid mechanics and meterological imagery

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

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

As for experimental fluid mechanics, we are investigating new methods for the analysis of complex fluid flows from image sequences. A large range of applications is concerned for instance with turbulent flows in aerodynamics, aeronautics, heat transfer, etc. Applications involving flow control are of particular interest (flow separation delay, mixing enhancement, drag reduction,...). These applications need enhanced visualization and sound numerical techniques such as low-order modeling with reduced dynamical models. The processing of real data and the accuracy enhancement of spatio-temporal measurements may together bring improvements in the modeling of turbulent flows which is traditionally solely based on initial conditions captured through experimental conditions.

4.1.3. Biological imagery

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

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

Applications of the proposed image processing methods to biological problems should provide a new and quantitative way for interpreting the movement of fluorescently labeled membrane transport vesicles. Quantitative analysis of data obtained by fast 4D deconvolution microscopy allows one to enlighten the role of specific Rab proteins on HeLa human cell lines. The role of Rab proteins is viewed as to organize membrane platforms serving for protein complexes to act at the required site withing the cell. Methods have been developed for specific Rab6a and Rab6a' proteins - involved in the regulation of transport from the Golgi apparatus to the endoplasmic reticulum. These small proteins act as molecular motors to move transport intermediates along polymers called microtubules. A second application concerns the CLIP 170 protein involved in the kinetochores anchorage (in the segregation of chromosomes to daughter cells, the chromosomes appear to be pulled via a so-called kinetochore attached to chromosome centromeres) using new fluorescent probes (Quantum Dots).

5. Software

5.1. Motion2d software - parametric motion model estimation

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

Motion2D is a multi-platform object-oriented library to estimate 2D parametric motion models in an image sequence. It can handle several types of motion models, namely, constant (translation), affine, and quadratic models. Moreover, it includes the possibility of accounting for a global variation of illumination. The use of such motion models has been 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 [86]. Real-time processing is achievable for motion models involving up to 6 parameters (for 256x256 images). Motion2D can be applied to the entire image or to any pre-defined window or region in the image. Motion2D is released in two versions :

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

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

5.2. d-Change software - motion detection

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

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

5.3. 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.

5.4. VTrack software - generic interactive visual tracking platform

Participants: Nicolas Gengembre, Patrick Pérez.

As part of the research contract with FT-RD (see paragraph 7.1)), we have started to develop an interactive tracking platform (Windows Visual C++ development with Microsoft MFC and Intel OpenCV). It already includes a number of state-of-the-art generic tracking methods (template matching, kernel-based tracking with global color characterization, particle filtering) as well as a number of visualization features for enhanced experimental and demonstration experiences. The flexible architecture and the rich HCI allow easy design, implementation and test of novel trackers.

6. New Results

6.1. Image sequence processing and modeling

6.1.1. Biological image sequence denoising

Participants: Charles Kervrann, Jérôme Boulanger, Patrick Bouthemy.

New video-microscopy technologies enable to acquire 4-D data that require the development and implementation of specific image denoising methods able to preserve details and discontinuities in both the three $x - y - z$ space dimensions and the time dimension. Actually, they are noisy due to the weakness of the fluorescence signal in time-lapse recording. Accordingly, we have developed an original and efficient spatio-temporal filtering method for significantly increasing the signal-to-noise ratio (SNR) in noisy fluorescence microscopic image sequences where small particles have to be tracked from frame to frame. The motion of the particles cannot be reliably computed since these objects are small and untextured with variable velocities. The image contrast is also quite low due to the relatively limited amount of light. However, partial trajectories of objects are line-like structures in the space-time $x - y - z - t$ domain. Image restoration can then be achieved by extending an adaptive estimation method we previously developed to efficiently remove noise in still images while preserving spatial discontinuities. The proposed "adaptive window approach" is conceptually simple, being based on the local estimation of a regression function with an adaptive choice of the space-time window size (neighborhood) for which the considered regression model fits the data well. We use statistical 4-D data-driven criteria to automatically select the size of the adaptive neighborhood. At each pixel, we estimate the regression function by iteratively increasing a space-time window and adaptively weighting input data to achieve an optimal compromise between bias and variance. It also involves an exemplar-based similarity step to fix the weights. The proposed algorithm complexity is actually controlled by simply restricting the size of the largest window and setting the window growing factor. In addtion, theoretical properties of the estimator have been proved. 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 SNR is shown to be drastically improved and enhanced images can be segmented. The objective is to report evidences about the lifetime kinetics of specific Rabs in different membranes, which may be similar or not for different cell types. This novel approach can be further used for biological studies where dynamics have to be analyzed in molecular and subcellular bio-imaging. For instance, global statistics on the computed velocities and motion directions could be extracted in order to measure the general behaviour of vesicles in the 4D volume. Let us also point out that we have applied our adaptive exemplar-based denoising method to usual video sequences, and it was demonstrated that it outperforms other recent methods.

6.1.2. Coherent motion detection with an a contrario approach

Participants: Frédéric Cao, Thomas Veit, Patrick Bouthemy.

Last year, we investigated how to answer the following question: given a region of an image in a sequence, should it be considered as moving or not? The answer was given by the Helmholtz principle: if a certain well-defined motion quantity were randomly and uniformly spread over the data, then it would be too improbable to observe high concentration of motion as we actually do. This is measured by the *Number of False Alarms* of a region (NFA) which is related to the expected number of regions with the same concentration of motion quantity in a noisy situation. For a region R, the lower $NFA(R)$, the more meaningful the region as a moving region. Now, a question arises when regions are nested: a large region may be detected because it contains one or several moving regions. In this case, what is the best representation of the data in terms of motion? The answer we propose is that the best solution is exactly the one with the smallest NFA. It implies two things: *i)* Given $R_1 \subset R_2$, how to choose between R_1 and R_2. This is simply achieved by comparing $NFA(R_1)$ and $NFA(R_2)$; *ii)* Given R_1, ..., R_m, pairwise disjoint regions contained in R, is it better to keep R or R_1, ..., R_M. We thus need to define the NFA of a m-tuple of regions.

The second part of the work is concerned with the detection of small pieces of trajectories. Given, at each instant, the set of regions detected as moving, how to detect the spatio-temporal coherence between the instantaneous detections? It would allow us to *i)* eliminate detections with no coherence, *ii)* reinforce the detectability of moving regions by integration information through time, *iii)* initialize trajectories. Again, decision is made by observing how far the data is from randomness. Assume that local features are computed in the vicinity of moving regions. This locality is necessary in order to be robust to local perturbations of the images (shadows, occlusion, etc...). Pairs of features from consecutive images are considered. In a short time interval, the trajectory of a moving object is assumed to be close to rectilinear with a constant velocity. Hence, from a pair of features at times $(t, t+1)$, one can compute a candidate velocity and a candidate initial position at time 0. If we now assume that these observations are independent for all instants and any position, we can compute how unlikely it is to observe a concentration of points around a given initial position and velocity. Hence, the problem of finding spatio-temporal coherence, can be turned into a clustering problem where the space of parameters is the space of initial position and velocity. The methodology is inspired by our works on grouping for shape analysis and recognition and also by some work on the analysis of "visual motion" [89].

6.1.3. Motion detection in adverse conditions

Participants: Patrick Pérez, Aurélie Bugeau.

Detecting moving objects in videos shot by either still or mobile cameras is an old problem, which is routinely solved in a number of real applications such as tele-surveillance. There are, however, interesting motion detection problems that are not satisfactorily handled by existing techniques. One class of such problems is the detection of certain types of motion. The case of detecting periodic motions, which is of particular practical interest, is discussed in paragraph 6.2.3. In the context of activity analysis in dynamically cluttered environments, the problem is the one of separating out foreground moving objects from other moving objects. This might be addressed by characterizing the spatial and/or temporal content of surrounding clutter. This is an acute problem of this sort that we are facing in the Behaviour ACI project (see paragraph 8.1.2) where the detection and the tracking of driver's moving hands is corrupted by the exterior view through the side window. Various approaches to this challenging problem are currently investigated. One line of research relies on the spatio-temporal analysis of maps of dominant motion compliance (these maps are produced by Motion2D software, see paragraph 5.1). At a global level, major changes of dynamical activity are searched by statistical tests. At a local level, main 2D+t blobs of independent motion are extracted by simple morphological processing, histogramming and EM-based clustering. Another line aims at estimating and analyzing sparse motion fields on interest points not belonging to the dominant motion (the one of the car interior). Finally, the complete extraction of moving objects with low-contrast and/or uniform intensity, which we also face in this application, remains difficult. For this problem we have started to investigate the novel use of min-cut/max-flow techniques in this context (see internship in paragraph 9.2).

6.1.4. Auto-models with mixed states for motion texture analysis

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

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, any discrete information is usually examined by the introduction of a label process. However, the label process is a latent process and the resulting statistical inference methods need in general to restore it. We have begun to investigate a different approach. The aim is to design a model which automatically deals with the two types of observations, without the introduction 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 have followed the construction of auto-models by J. Besag [70] and extended it to the multi-parameter case. We have introduced necessary adaptations for mixed-state 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. We are exploiting these mixed-state auto-models for the modeling of dynamic textures in videos of natural scenes (such as views of rivers, sea-waves, moving foliage, fire, steam, smokes,...). The considered observations are (locally averaged) normal flow magnitudes which take their values in $\{0\} \cup (0, \infty)$ where 0 is here a discrete (symbolic) value since it accounts for no motion. We call them “motion textures”. We have implemented positive Gaussian mixed-state auto-models defined on a spatial 4-nearest-neighbour system. They involve four parameters only. Simulations have demonstrated that this (simple) model is able to reproduce important texture characteristics such as density, contrast, granularity and spatial orientation, and the influence of each parameter was identified. Experiments on real image sequences have shown that basic properties like isotropy/anisotropy, spatial stationarity/non stationarity are well captured by the model. However, this first implemented model needs to be improved. Current extensions are dealing with larger neighborhoods (8-NN, 2 more parameters), multi-scale models, other continuous distributions (the non zero-mean Gaussian distribution has also been introduced, adding four more parameters, and simulations have proven its interest), in order to model long-distance spatial correlations and correlations between continuous values. The addition of the time dimension in the auto-models is also envisaged. Besides, we are investigating the use of these mixed-state auto-models for the segmentation of motion textures in image sequences.

6.2. Motion estimation and matching

6.2.1. Fluid motion analysis

Participants: Étienne Mémin, Anne Cuzol, Nicolas Papadakis.

Fluid motion estimation is a difficult and important issue in experimental fluid mechanics or in environmental sciences for the study of geophysical flows. In this context, we have investigated different approaches. The dense motion estimator we designed and improved these last years, is formalized as the minimizer of a global energy function. The latter is composed of a data model built from the integration of the continuity equation of fluid mechanics and a smoothness term enforcing smooth curl and divergence blobs. We are considering two important applications: velocity measurement for fluid experimental fluid mechanics and atmospheric wind field estimation in meteorology from satellite image sequences.

For the first application domain, we have addressed this year the problem of estimating the motion of fluid flows visualized with the Schlieren technique. Such an experimental visualization system is well known in fluid mechanics and it enables the visualization of unseeded flows. Since the resulting images exhibit very low intensity contrasts, classical motion estimation methods based on the brightness consistency assumption (correlation-based approaches, optical flow methods) are inefficient. The global energy function we have defined is composed of *(a)* a specific data model accounting for the fact that the observed luminance is related to the gradient of the fluid density, and *(b)* a specific constrained div-curl regularization term. In the second application domain, the goal is to provide computed wind fields for cloud layers at different altitudes. The approach we have proposed consists in exploiting classification data provided by Eumetsat. These classification maps are used as masks in the motion estimation process. Contrary to the classification techniques used routinely in meteorology, our estimation scheme enables to reconstruct a complete dense wind field. Obviously, the output must be taken with caution in large areas where no data are available. Nevertheless, such an approach allows us, without any additional post-processing, to provide small regions with reliable measurements in the considered cloud layer. This is an interesting by-product of our method. Let us point out that it also allows us to easily compute derivative descriptors of the motion field.

We have also proposed a low dimensional motion estimator. It is based on the Helmholtz decomposition which consists in representing the velocity field as a sum of divergence-free component and a curl-free one. In order to provide a low dimensional solution, both components have been approximated using a discretization of the vorticity and divergence maps through regularized Dirac measure. The resulting so-called irrotational (resp. solenoidal) field is then represented by a linear combination of basis functions obtained by a convolution product of the Green kernel gradient and the vorticity map (resp. the divergence map). The coefficient values and the basis function parameters are obtained as the minimizers of a function formed by an integrated version of mass conservation principle of fluid mechanics. This estimator has also been applied to medical imagery for the estimation of multiple sclerosis lesion growing, in cooperation with P. Hellier (Visages team-project). Such a low-dimension motion estimator is very well suited for motion measurement in local targetted regions and provides also a good basis for motion tracking application.

6.2.2. Transparency motion estimation

Participants: Patrick Bouthemy, Vincent Auvray.

This work is carried out in collaboration with J. Liénard (General Electric Healthcare). Motion estimation on transparent image sequences requires the development of specific methods, since the brightness constancy assumption used for classical video sequences does not hold any more. Such a motion estimation scheme applied to medical X-ray images should allow important applications such as quantitative motion measurement, motion stabilization or motion-compensated denoising, the latter being our main goal (see paragraph 7.2). Considering a sequence with two layers moving with transparency, a fundamental motion equation that involves three successive images of the sequence can be inferred, that is function of the unknown velocities of the two layers. We have developed a multiresolution motion estimation framework for two layers moving by transparency, based on this equation. It is particularly robust to the high level of noise present in the medical X-ray sequences and it is computationally tractable for future applications. The layers velocities are first locally estimated using a block-oriented method. Two affine motion fields are then fit to the estimated local velocities (meaning that we assume the motions of the organs can be adequately represented by such parametric models). Finally, the transparency motion equation is directly applied to the affine motion models. Once linearized, it is iteratively minimized to improve the affine motion estimates. We have generated realistic X-ray image sequences to assess the performance of our method with an available ground truth. The results on medium-level radiation images are quite satisfactory, and the performance smoothly decreases with higher noise and scatter (motion magnitude error remains lower than one pixel) unless the MTF effect is incorporated (error of about 1.5 pixels with the highest noise level and scattering effect). Experiments on real clinical image sequences were also reported with convincing results.

We are now extending this framework to process what we have called the “bidistributed transparency” case (each image contains no more than two layers at every pixel, while several layers can be involved in the whole image). A clustering step is first performed to determine the global number of layers and to provide with an initial estimation of the affine motion models for the different layers. A MRF-based framework has been designed to segment the image into the two-layer regions and to update the motion model parameters of each layer. The main part of the remaining research will focus on the use of these velocity estimates to denoise fluoroscopic image sequences which exhibit a low signal-to-noise ratio.

6.2.3. Periodic motion detection and video alignment

Participants: Ivan Laptev, Patrick Pérez.

Independent motion is a strong cue for detecting, segmenting and recognizing objects and activities in image sequences. Motion-based segmentation, however, is known to be a hard problem in scenes with motion parallax and scenes with multiple moving objects. In our work we have addressed this problem by exploiting not only the presence but also the type of motion as an informative cue, and thus, by addressing motion detection and recognition jointly (the two problems being usually treated separately). As one specific type of motions we considered periodic ones, noting their detection can be roughly seen as sequence alignment (a sequence is matched to itself over one or more periods). Hence, we first considered alignment of two video sequences obtained by independently moving cameras. Under assumption of constant translation, fundamental matrices and the homographies were shown to be time-linear matrix functions. These dynamic quantities can be estimated by matching corresponding space-time points with similar local motion and shape. For periodic motion, we matched corresponding points across periods and developed a RANSAC procedure to simultaneously estimate the period and the dynamic geometric transformations between periodic views. Using this method, we demonstrated detection and segmentation of human periodic motion in complex scenes with non-rigid backgrounds, moving camera and motion parallax. In the on-going work we extend this framework to align video sequences with similar type of activities, using appropriate local motion descriptors and global geometric constraints.

6.2.4. Image comparison with an a contrario criterion

Participants: Frédéric Cao, Patrick Bouthemy.

6.3. Tracking

6.3.1. Structure tracker

Participants: Étienne Mémin, Anne Cuzol, Nicolas Papadakis.

A reliable tracking of high dimensional structures is a very difficult task in computer vision. This is particularly true for tracking methods formalized with non-linear Bayesian filtering, since sampling in high-dimension spaces is usually completely inefficient. In that context, we have investigated two different aproaches. The first one is a deterministic approach stemming from the variational data assimilation techniques used for instance in meteorology for global atmospheric circulation model assimilation. We aim at tracking along an image sequence elements involving a high-dimension representation such as dense vector fields and closed curves described by implicit surfaces. The tracking modeling involves an evolution equation with state variables accounting for a motion field or a closed curve representation, and discrete and noisy measurements associated to components of the state variables. The method amounts to minimizing a global spatio-temporal continuous cost function. The minimization is carried out by introducing an adjoint variable and consists of forward integration of an evolution law followed by a backward integration of an adjoint evolution model. The latter includes a term related to the discrepancy between the state prediction and the discrete noisy measurements of the system. The closed curves are represented by implicit surfaces, whereas the motion is described either by a vector field or by vorticity and divergence maps depending on the kind of targeted application. The second approach is the Bayesian recursive filtering one. The problem formulation usually comes to associate a complex multi-modal likelihood of the measurements to a roughly linear dynamic imperfectly describing the feature motion in the image space. In contrast, we propose to rely on an accurate dynamic law either issued from fluid mechanics laws for the tracking of fluid descriptors or from dynamical models estimated on-line from the image sequence for natural scenes.

As for the second approach, two trackers have been developed for two different contexts. The first one is a recursive Bayesian filter for tracking velocity fields of fluid flows. The filter combines an Îto diffusion process, given by a stochastic formulation of the vorticity-velocity form of Navier-Stokes equation, and discrete measurements extracted from the image sequence. The resulting tracker provides robust and consistent estimations of instantaneous motion fields along the whole image sequence. In order to handle a state space of reasonable dimension for the stochastic filtering problem, the motion field is represented as a combination of adapted basis functions. The basis functions are derived from a mollification of Bio-Savart integral and a discretization of the vorticity and divergence maps of the fluid vector field. The output of such a tracking is a set of motion fields within the whole time range of the image sequence. As the time discretization is much finer than the frame rate, the method allows consistent motion interpolation between consecutive frames. The second tracker deals with objects described as cloud of feature points. We have investigated the use of particular particle filters called optimal Rao-Blackwellized particle filters. The designed filter combines an accurate dynamics estimated from the image sequence with a simple likelihood defined as a Gaussian mixture. Such a combination enables to design a tracker relying on the optimal importance function. It allows us in the same time to be reactive to changes of speed and direction of the target and to be robust to occlusions. The tracker also relies on Rao-Blackwell variance reduction principle. Following this principle, we assume a linear dependency between two subgroups of points in the cloud. We make use of this dependency to incorporate a geometric constraint on the points.

6.3.2. Robust visual tracking without prior

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

In this research we are interested in the problem of tracking arbitrary entities along videos of arbitrary type and quality. Such a tracking cannot rely, as classically done, on a priori information regarding both the appearance of the entities of interest (shape, texture, key views, etc.) and their visual motion (kinematical constraints, expected dynamics relative to the camera, etc.).

The first crucial step is then the definition and the estimation of the reference appearance model on which the tracking, no matter its precise form, will rely on. Roughly, two extreme types of representations are routinely used in the literature: detailed pixel-wise appearance models subject to rapid fluctuations (e.g, intensity template instantaneously refreshed) and rough color models very persistent over time (e.g., color histogram instantiated at initialization time and kept unchanged). They are both interesting and complementary. For these reasons, we have combined them in a simple probabilistic tracker, where intensity template matching, refined by gradient ascent on color similarity to a fixed global model, feeds a Kalman filter. Such a simple combination has already permitted to improve robustness compared to trackers based on each of the standard representations alone. We have also shown that the discriminative power of the color representation could be noticeably improved by selecting carefully the selected color space and by performing background color subtraction (which we have done in various ways). These two types of appearance models have however limitations that are only partly circumvented by their combination: template matching assumes simple deformations (e.g, translation) of the whole object; it is also subject to drift, especially in presence of partial or total occlusions; global color characterization with no spatial layout yields inaccurate positioning and behaves badly in case of partial occlusion. To address these problems, we have introduced a random partitioning of the target object into a constellation of parts. These parts are individually tracked and then merged after occluded parts are detected and ignored. Complete occlusions, however, remain a critical problem which must be addressed. Although probabilistic generic trackers have proved better at handling them, as compared to their deterministic counterparts, we are still lacking ways to fully treat the problem. One interesting output of our methodological work on Bayesian tracking with discrete auxiliary variables (see paragraph 6.3.3) is that is can be used to seamlessly incorporate a binary “visibility” variable in existing state space models for tracking. Not only this improves slightly the robustness of tracking under severe occlusions, but it also allows sequential evaluation of the posterior visibility probability within the tracking procedure.

Robust generic tracking tools are of major interest for a wide range of applications dealing with editing, analyzing, annotating, browsing and authoring video contents (see paragraph 4.1). Even in application where a strong prior is “available” (e.g., precise type of videos and/or type of objects of interest), such tools are crucial. On the one hand, they are useful complement to application-specific detection/tracking methods that are often more complex (both in terms of preliminary off-line learning and of on-line application) and sometimes less robust (e.g., false negatives of object-specific detectors). On the other hand, they can facilitate the semi-automatic extraction and labeling of data that application-specific learning modules will require. For these reasons, any progress on this methodological front will be of interest for tracking and recognition problems addressed in the FT-RD contract (semi-automatic tracking of players for annotating team sport videos; see paragraph 7.1), and the Behaviour ACI project (fine-grain analysis of car drivers’ activity; see paragraph 8.1.2).

6.3.3. Multiple-object tracking and Bayesian tracking with auxiliary variables

Participant: Patrick Pérez.

Handling in a principled way a varying (and unknown) number of entities to be tracked is a difficult and important problem, which keeps receiving a great deal of attention from both trajectography and computer vision communities. Our recent contributions to these efforts, in collaboration with the University of Cambridge, included various sequential Monte Carlo techniques, focusing on the data association problem in case the number of objects is fixed. This year, we have pursued this collaboration by addressing the problem of starting and terminating automatically individual tracks as objects enter and exit the volume of observation. To this end we have introduced a binary “existence” variable within a generic multi-object “detect-before-track” state model. We have proposed a general treatment of this augmented model that impacts as little as possible on existing tracking algorithms, so that software can be reused, and that allows implementation with Kalman filters, extended Kalman filters, particle filters, etc. We have applied the proposed framework to color-based tracking of multiple objects, where we adopted a mixture Kalman filter implementation. This work was partly supported by a British-French collaborative program which started in October 2004 (see paragraph 8.2.1).

Based on above-mentioned work on tracking with “existence”, we considered the more general context of sequential estimation problems with an auxiliary discrete process as part of the state space. These discrete variables usually follow a Markovian process and interact with the hidden state either via its evolution model or via the observation process, or both. Examples of such auxiliary variables include depth ordering for occlusion handling, switches between different state dynamics, exemplar indices, etc. We have considered a general model that encompasses all these situations, and show how Bayesian filtering can be rigorously conducted in this general setup. The resulting approach facilitates easy re-use of existing tracking algorithms designed in the absence of the auxiliary process. In particular, we have shown how particle filters can be obtained based on sampling only in the original state space instead of sampling in the augmented space, as it is usually done. We have demonstrated how this framework facilitates solutions to the critical problem of appearance and disappearance of targets, either upon scene entering and exiting, or due to temporary occlusions. We have illustrated this in the context of "track-before-detect" color-based tracking with particle filters.

6.3.4. Target tracking

Participants: Jean-Pierre Le Cadre, Thomas Bréhard.

In this domain, our main contributions are three-fold. The common denominator is non-linear filtering for partially observable systems. First, work has been pursued for particle filter initialization in this context. This is actually a crucial point. Previous work has been extended in a general setup of hierarchical particle filtering, coupled with simulation (MCMC) methods. The resulting algorithm performs satisfactorily and is reliable, even for demanding scenarios (highly maneuvering targets). Performance analysis for random state filtering is another issue. The Posterior Cramér-Rao Bound (PCRB) is now the basic tool. Our main result is the derivation of closed-forms for the PCRB. This is an extension of Riccati equation to this non-linear and partially observed system. Such a result is important. It is mainly due to the use of a particular coordinate system (the logarithmic modified polar coordinate system). It is also possible to handle the incorporation of active measurements in a unique setup. Practically, this result is especially useful for sensor management, in particular for optimizing active measurements scheduling. Theoretical investigations have been validated by simulation results. It is worth stressing that such a result was long-time awaited by the sensor management community. Distributed target tracking is another important area. This means that tracking is made on each receiver and then tracks are transmitted to a centralizer, itself performing tracking from these elementary trackers. While this subject has been long-time worked in a linear context, this is no longer true in our non-linear, partially observed context. The general aim is to perform particle filtering within these decentralized architectures. Two architectures are considered, *with* and *without* feedback, as well as applications to bearings-only tracking with complete or partial state estimates. The corresponding algorithm performs quite satisfactorily and can considerably reduce communication requirement. Moreover, robustness and feasibility can be neatly improved by this way.

6.4. Motion recognition and learning

6.4.1. Video segment comparison and motion recognition with local motion descriptors

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

We have explored the use of local motion descriptors for the recognition of dynamic contents in video. This work was mainly carried out in the internship done by A. Lehmann. First, we have extended the approach we defined last year [88] relying on global probabilistic motion models to classify video segments according to different predefined classes. It exploited simple local motion features (locally-averaged normal flow magnitudes) and occurrence statistics were computed over the whole video segment. The considered probabilistic motion models involve mixed-state variables (i.e., the same variable may take both discrete values, here the 0 value accounting for no motion, and continuous values) and are defined as the mixture of a Dirac distribution at 0 and a continuous distribution from the exponential family. The continuous part was modeled in practice by a zero-mean Gaussian distribution. We have adapted this method to a local scenario. The probabilistic motion models are now inferred from spatio-temporally localized blocks inside the image volume spanned by a video segment. These blocks are selected using a simple criterion based on the displaced frame differences. Two other continuous distributions have been studied: the Lognormal one and the Gamma one, the latter yielding the best experimental classification results. Finally, a method has been specified to classify a video sequence using this local probabilistic motion models; this classification issue has been reformulated as a retrieval problem. A matching of the local models has been defined using the Kullback-Leibler distance. A global distance measure can then be obtained by combining the comparison results of each pair of matched local models. Adding a clustering of the blocks increases the classification performance significantly. The evaluation was carried out using a leave-one-out validation on two video databases (amateur videos of basket-ball and videos of human gestures).

A quite different approach has also been investigated which has led to the design of a novel and efficient dissimilarity measure between video segments. We consider a set of local spatio-temporal descriptors of different types. They are considered to be realizations of an unknown, but class-specific distribution. The similarity of two video segments is calculated by evaluating an appropriate statistic issued from a rank test (Wilcoxon Rank-Sum test). This measure can be computed very efficiently, and it does not require any pairing of the features of the compared video segments and can straightforwardly handle a different number of feature values (or interest points) per segment. Another property of this measure is that in case of similarity, the distribution of the measure converges to $\mathcal{N}(0,1)$ which can be seen as a self-normalization. This simplifies the integration of several descriptors. Furthermore, we have defined a way for learning the discriminative power of the different descriptors and for deducing a self-adaptive combination of the descriptors which performs significantly better than an equal combination. The proposed framework has been tested on two motion classification problems where quite satisfactory results have been obtained using simple local motion features related to object trajectory and scene motion intensity observed in the image sequence. The proposed video-segment similarity criterion can in fact be applied to any kind of features for video comparison, video classification or video retrieval.

6.4.2. Video content recognition using motion trajectories

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

We will only give a brief description of the envisaged work since it is starting in November 2005 with the Ph-D thesis of A. Hervieu. We plan to analyse the dynamic content of a video from local trajectories. We will use feature tracking methods previously developed in the Vista team (see paragraphs 6.3.2 and 5.4) to supply a set of trajectories in a given video segment. The first step will be to structure these trajectories in coherent space-time clusters and to design a graph-based representation of them. After a supervised learning stage, a statistical classification method should allow us to detect events of interest. Trajectories are interesting motion features since they convey elaborated information on the temporal behaviour of the moving entities, which should enable a fine categorization of the dynamic content.

7. Contracts and Grants with Industry

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

Participants: Patrick Pérez, Nicolas Gengembre, Etienne Mémin.

no. Inria 104C1114003, duration : 18 months.

The aim of this contract, started in December 2004, 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 is targeted is especially challenging due to frequent occlusions, drastic changes of players' appearance within and across shots, joint presence of similar players (teammates) in the image, and diversity of shot types (viewpoint and camera motion). In the first part of the project we have focused on the problem of robust visual tracking of a single arbitrary object in adverse conditions. An experimental platform (see paragraph 5.4) has been developed to compare a number of existing techniques for generic visual tracking. As expected, they all exhibit severe limitations on rugby player tracking. Based on the analysis of their failure modes (detailed in a first report, April 2005), we are now investigating within the same experimental platform novel tracking techniques to improve robustness. We are currently focusing on the modeling/on-line learning of the surrounding background as well as on the use of instantaneous motion estimation techniques (both local, to help tracking objects' positions, and global, to help estimating scale changes due to camera zooming). The extension to joint multi-player tracking and to application-specific prior incorporation (e.g., learning of player and field detectors) will be considered in the second part of the contract.

7.2. Cifre grant with General Electric Healthcare

Participants: Patrick Bouthemy, Vincent Auvray.

no. Inria 103C19960, duration 36 months.

This contract started in January 2004. 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).

7.3. Cifre grant with Thales Systèmes Aéroportés

Participant: Jean-Pierre Le Cadre.

CNRS *contract 511008, duration 36 months.*

This contract was related to the supervision of Frédéric Bavencoff's Ph-D thesis, defended in July 2005. This work was centered around information processing for maritime surveillance and three main issues were addressed. The first one is about target motion analysis. Major achievement in this way is constraint inclusion in the estimation part. This is done via simulation methods (MCMC). The benefits of such an approach are of a fundamental importance since: *i)* Providing good estimate of the target range becomes possible, even for unobservable scenarios; *ii)* Estimated confidence intervals are deduced as a by product; *iii)* Multimodality (ghosting) is automatically taken into account. The second one is about the analytic evaluation of the probability of false association. While bounds are widely used for assessing estimation performance for target tracking, a major problem is the performance of data association in a multi-track context. This is not so easy since probability of false association and track parameter estimation are completely entangled. What has been shown is that a linear regression framework allows us to conduct explicit calculations. Probability of false association is then derived as an explicit function of the scenario parameters: e.g. mean track distance, measurement variance, probability of detection, etc. By this way, it is possible to derive a measure of track-to-track interaction. Finally, the third one is related to multi-target tracking in a dense environment and more precisely to investigating the performance of a plot-to-track association. The problem begins with a two-scan two-object one, giving a simple and explicit approximation of the probability of correct association. It is then extended to the N-object one and then to random values of N (Poisson law). Finally, it is shown that it is possible to use the same tools for considering an arbitrary number of scans.

7.4. Grant from Dupont De Nemours

Participant: Etienne Mémin.

Univ. of Rennes contract

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

7.5. European initiatives

7.5.1. fp5-ist Lava project

Participants: Patrick Bouthemy, Frédéric Cao.

no. Inria 102G0424, duration 36 months

The IST project LAVA ("Learning for Adaptable Visual Assistants") started in May 2002 and ended in April 2005. It involved the following partners : XRCE (prime), IDIAP (Switzerland), Inria (Lear team from Inria Rhône-Alpes and Vista team), RHUL (UK), University of Graz (Austria), University of Lund (Sweden), ANU (Australia). It gathered groups from computer vision, machine learning and cognitive sciences and focused on two key problems : categorizing objects in static images and interpreting events in video. Our contribution was two-fold: 1) Recognizing semantic events in video requires to preliminary learn the different classes of video events. We 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; 2) We were also concerned with the extraction of motion-based numerical features that could characterize the dynamic content of a video and with the design of an efficient similarity method between video segments. We have proposed several localized motion descriptors, extracted either from spatio-temporal profiles of motion-detection binary masks, or from local statistics computed around space-time interest points or in localized space-time blocks. We have also designed a novel method to compare video segments which is based on rank tests and does not require the pairing between local descriptors describing the two video segments (see paragraphe 6.4.1). This work was carried out by Alain Lehmann during his internship. Experimental comparison work between different global and local motion descriptors, along with different classification methods, was also conducted. A complementary task that we addressed was the collection of new video databases. More specifically, we collected a set of basket-ball amateur videos (a total of 300 video segments). Three types of actions were 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. A second database was acquired related to the 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.

7.5.2. fp6-ist noe Muscle

Participants: Patrick Bouthemy, Frédéric Cao, Ivan Laptev, Gwénaëlle Piriou, Thomas Veit.

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, one is witnessing the emergence of unprecedented opportunities for the creation of a knowledge driven society. Indeed, databases are accruing large amounts of complex multimedia documents, networks allow fast and almost ubiquitous access to an abundance of resources and processors have the computational power to perform sophisticated and demanding algorithms. However, progress is hampered by the sheer amount and diversity of the available data. As a consequence, access can only be efficient if based directly on content and semantics, the extraction and indexing of which is only feasible if achieved automatically. MUSCLE aims at creating and supporting a pan-European Network of Excellence to foster close collaboration between research groups in multimedia datamining on one hand and machine learning on the other hand, in order to make breakthrough progress towards different objectives.

G. Piriou contributed to workpackage WP8 (machine learning), via her Ph-D thesis work on the probabilistic modeling and the recognition of dynamic contents in videos using mixed state models. T. Veit contributed to WP5 (single modality) by providing new methodological elements for video analysis. F. Cao and P. Bouthemy participated in the Muscle meeting held in Paris, 27-29 April, where Y. Gousseau (ENST Paris) gave a talk on shape recognition (joint work with P. Musé (ENS Cachan), F. Sur (Loria) and F. Cao. P. Bouthemy organized a special Muscle session for the International Workshop on Content-Based Multimedia Indexing, CBMI'2005, held in Riga, June 2005, and gave a talk in that session. The session included 6 talks, from Trinity College, Dublin, UPC Barcelona, University of Thessaloniki, University of Surrey, Bilkent University, and Inria Vista team. I. Laptev participated in the e-team meeting of WP8 (Dublin, 28-29 September) and of WP5 (Rocquencourt, 1-2 December). We contributed to several WP reports.

7.5.3. fp6-ist tn Visiontrain

Participant: Patrick Bouthemy.

no. Inria 850, duration 48 months

Visiontrain is a Marie Curie Research Training Network (belonging to the Computational and Cognitive Vision Systems chapter) which started in May 2005. Visiontrain addresses the problem of understanding vision from both computational and cognitive points of view. The research approach will be based on formal mathematical models and on the thorough experimental validation of these models. In order to achieve these ambitious goals, 11 academic partners plan to work cooperatively on a number of targeted research objectives: *(i)* computational theories and methods for low-level vision, *(ii)* motion understanding from image sequences, *(iii)* learning and recognition of shapes, objects, and categories, *(iv)* cognitive modelling of the action of seeing, and *(v)* functional imaging for observing and modelling brain activity. P. Bouthemy participated in the first meeting held in Grenobla, on September 12-13; he presented the research activities of the Vista team and the work on image comparison based on an *a contrario* decision approach (see paragraph 6.2.4).

7.5.4. fp6 Fluid project

Participants: Étienne Mémin, Patrick Heas, Nicolas Papadakis, Anne Cuzol.

no. Inria 737, duration 36 months

The FLUID project is a FP6 STREPS project labeled in the FET-open program. It 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.

During the first year of the project, we have developed methods for accurate 2D motion estimation from image sequences. We have improved a previously developed dense motion estimator dedicated to fluid flow [2] by considering appropriate temporal initialization and new conditions at the frontiers of the domain. This estimator has been also extended to compute wind fields corresponding to cloud layers at different heights. A low-dimension motion estimator based on adequate basis function stemming from a discretization of the curl and divergence maps has been also designed. This estimator allows us to estimate motion fields described with a reduced set of basis functions. It is particularly well suited to motion measurements over a local domain. It is also very well adapted to tracking purpose. As described in section 6.3.1 this representation of a vector field on the basis of vortex and divergence blob functions has been used to define a Bayesian filter devoted to the tracking of fluid motion fields. A method relying on variational assimilation principles have been investigated to enable the tracking along the time sequence of fluid motion field. This approach couples observed motion fields obtained with correlation techniques or with dense motion estimators with a vorticity-velocity formulation of the Navier-Stokes equation. Besides, we have started to study the estimation of the vertical component of atmospheric velocity fields from satellite images.

A cooperation with the Cemagref Rennes have allowed us to assess the performance of a dedicated optical flow technique to estimate the velocities of a fluid flow from particle images. This evaluation has been conducted on two typical flows: a plane turbulent mixing layer and a near wake of a circular cylinder. It has been shown that dense motion techniques provide reliable dense measurements, which in particular enables to accurately compute second order moments of the flow velocity.

7.6. National initiatives

7.6.1. riam Feria project

Participants: Patrick Bouthemy, Brigitte Fauvet.

no. Inria 203C1460, duration : 24 months.

The Feria project belonged to the RIAM programme and was financially supported by the Ministry of Industry. The partners were INA, C&S, Vecsys and NDS companies, Arte, IRIT, Inria (Texmex and Vista teams). The goal of the project was to build a general and open framework allowing the easy development of applications for editing interactive video documents. It ended in September 2005. We were concerned with the design of video processing and representation tools. We have delivered several modules related to video analysis: a shot change detection module including the handling of progressive transitions, 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, 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. The last contribution in collaboration with other partners was dedicated to scene segmentation and representation based on cross-modal analysis (video images, soundtrack, speech-to-text translations, texts). Extensive experiments were carried out on two video database: opera video and the series of “Grand Echiquier” TV programs, with satisfactory results.

8. Other Grants and Activities

8.1. National initiatives

8.1.1. 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 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. Within this project, two methods have been developed for vortex tracking. The first one relies on stochastic filtering whereas the second one is based on variational data assimilation. Both techniques have been applied to meteorological images and experimental fluid mechanics images. We refer to section 6.3.1 for further description of these techniques.

8.1.2. aci Ministry grant : Behaviour project

Participants: Patrick Pérez, Aurélie Bugeau.

no. Inria 104C08130, duration 36 months.

The Behaviour project was granted in October 2004 by the collaborative ACI program on Security and Computer Science. It involves Compiègne University of Technology (Heudiasyc lab) as the prime, along with

PSA-Peugeot-Citroën (Innovation and Quality group) and Vista. The main applicative goal is visual monitoring of car drivers, based of videos shot inside the car, such that hypo-vigilant behaviors (mainly drowsiness and distraction) can be detected. To this end, the project aims at providing new tools to perform automatically the recognition of a wide range of elementary behavioral items such as blinks and eye direction, yawn, nape of the neck, posture, head pose, interaction between face and hands, facial actions and expressions, control of the car radio, or mobile phone handling. Before trying to achieve such fine grain activity recognition, one has to select and extract relevant spatio-temporal features to apply subsequent learning on. While UTC is focusing on robust extraction and tracking of facial features in frontal views (shot through the wheel), we are attacking the complementary problem of detection and tracking of mobile items (especially head and hands) in arbitrary driver views. Although the problem seems classic, the specificity of videos under concern makes it very difficult (drastic changes of appearance and prolonged occlusions; low contrast of sequences shot at night; presence of very complex dynamic visual content through window in daylight). In this context, new motion detection, tracking and matching techniques are studied (see paragraphs 6.1.3 and 6.3.2). The current focus is on the analysis of outside elements seen through the window, with the aim of preventing them from corrupting the inside activity analysis.

8.1.3. aci Ministry grant : Modyncell5d project

Participants: Charles Kervrann, Jérôme Boulanger, Patrick Bouthemy.

no. Inria 104C08140, duration 36 months.

This project, labeled within the IMPBio ACI program, was 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 information on proteins dynamics from video-microscopy image sequences, using statistical methods. Methods are developed for two target 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); Rab6a' involved in the regulation of transport from the Golgi apparatus to the endoplasmic reticulum. Specific algorithms have been developed for processing 3D image sequences for various tasks such as spatio-temporal image denoising and spot detection.

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

8.1.5. Other involvements

- The Vista team is involved in the French network GDR ISIS, "Information, Signal and Images".
- C. Kervrann participates in the network GDR 2588, "Microscopie Fonctionnelle du Vivant".

- **Collaboration with Ifremer, Brest**
  Participant: Frédéric Cao.
  F. Cao supervises the thesis of A. Chessel (Ifremer) in collaboration with R Fablet (Ifremer). The topic is the analysis of otolith images. Otoliths are small calcareous concretions that can be found in fishes inner ear. They presents rings (as tree trunks) which are representative of the growth of individuals. These images usually have a very low contrast and geometrical information is only partial. A problem is to reconstruct the orientation of the rings. Thus, we tried to find out what kind of interpolation method is sound for extending direction field from partial information. We used an axiomatic approach, similar to the work by Caselles *et al.* [73], [72]. The main novelty is that the data belong to the unit circle. The interpolation must be independent of the parameterization of the circle. We proved that the only second order operator satisfying this requirement is related to the scalar curvature of the level lines of the direction of the reconstructed field. Another interesting operator, which is only invariant with respect to affine parameterization is the so-called ∞-Laplacian operator defined by

  $$\Delta_\infty u = D^2u(Du, Du),$$

  for a real valued function u. It is by now a classical result that if u is solution of $\Delta_\infty u = 0$, then u is a local minimizer of $\|Du\|_\infty$, where Du is the gradient of u. Now if u assumes values in $\mathbb{R}^2$, we remark that trying to minimize $\|Du\|_\infty$ with the constraint $|u| = 1$ everywhere, amounts to locally solve $\Delta_\infty \theta = 0$ where $\theta = \arg u$ is the direction of u.
- **Collaboration with Cesta, Bordeaux**
  Participant: Jean-Pierre Le Cadre.
  We are concerned with the acquisition of the target trajectory of a mobile during its reentry phase. During this phase, the target trajectory is affected both by target factors (drag coefficients) and environmental ones (atmosphere density). Both are unknown and time-varying. The state evolution is thus highly non-linear and stochastic. Not surprisingly, linearization has limited performance while particle filtering is well suited. Actually, a complex Allen oscillatory ballistic profile may be used to model the variations of the ballistic coefficient β. Therefore, a large part of our efforts has been devoted to the estimability analysis of this coefficient. This is done via classical tools like the Cramér-Rao bound and multilinear algebra. Simple expressions have been obtained and validated by simulations. They show that estimability of the β parameter is tightly related to the product of the dynamic pressure with the relative range derivative.

8.2. Bilateral international co-operation

8.2.1. Royal Society-cnrs program, France-England

Participant: Patrick Pérez.

This two-year collaboration with Jaco Vermaak from the Eng. Dpt of the University of Cambridge started in October 2004 with a joint funding (for mutual visits) by the Royal Society and CNRS. The output of this collaboration on probabilistic tracking tools for visual tracking and trajectography is two-fold. A new data-association based filter, with proper handling of an “existence” binary variable has first been proposed in a multi-object “detect-before-track” context, with validation on a color-based visual tracking problem. We have then extended this framework to any sequential estimation problem with an auxiliary discrete process, with illustration on handling both existence and visibility in visual tracking within a “track-before-detect” approach. In the case where particle filtering is mobilized, our approach allows in addition to avoid sampling the auxiliary variables (see paragraph 6.3.3).

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

Participants: Patrick Bouthemy, Étienne Mémin, Anne Cuzol, Nicolas Papadakis, Tomas Crivelli.

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

This collaboration has been instantiated by several mutual visits and longer stays of the different members of the project. Etienne Mémin spent ten days in Buenos-Aires in March 2005. Anne Cuzol did a two-month stay at the University of Buenos-Aires in March-April 2005. On the other side, Guillermo Artana and Bruno Cernuschi-Friàs spent respectively three weeks (mid Sept. - early Oct. 2005) and two-months (end Oct. - end Dec. 2005) in Rennes. Alejandro Gronskis and Tomas Crivelli, Ph-D students, did a three-months and a four-month stay respectively in Rennes from Sept. to Nov., resp. Dec. 2005. B. Cernuschi-Friàs and P. Bouthemy have settled a thesis co-advising agreement between University of Rennes 1 and the University of Buenos-Aires for the Ph-D thesis of Thomas Crivelli on motion texture modeling and recognition. In relation with this collaboration, E. Mémin and G. Artana have submitted a proposal to the Ecos-Sud bilateral cooperation program along with the Aerodynamics laboratory of the University of Poitiers.

G. Artana, E. Mémin, R. Sosa (UBA) and E. Arnaud (formerly Ph-D student in Vista team, now Post-doc at the University of Genova) have developed a motion estimation method devoted to Schlieren images. The Schlieren technique provides a powerful visualization of fluid flows which does not require to seed the flow with particles. It allows large scale experiments and is therefore very relevant for numerous applications. The first experiments we have conducted are very encouraging. G. Artana, A. Cuzol, Juan D'Adamo (UBA) and E. Mémin have investigated the use of model reduction based on proper orthogonal decomposition and a Galerkin projection of the Navier-Stokes equation within the stochastic visual tracking method based on vortex particle. T. Crivelli, P. Bouthemy, J.-F. Yao and B. Cernuschi-Friàs have defined extensions to the modeling of motion texture with mixed-state auto-models, involving the introduction of larger neighborhoods and non zero-mean Gaussian distribution (see paragraph 6.1.4). A motion texture segmentation method is also investigated.

8.2.3. *Visiting scientists*

- Guillermo Artana and Bruno Cernuschi-Frias (Profs University of Buenos-Aires) spent respectively three weeks and two months in our team, Tomas Crivelli and Alejandro Gronskis (Ph-D students of UBA) spent respectively four and three months in our team, all in the context of the Inria Associate team FIM.
- Short visits by Y. Amit (University of Chicago), D. Cremers (University of Bonn), N. Gordon (DSTO Australia), S. Marchand-Maillet (University of Geneva), Erik Sudderth (MIT), J. Vermaak (University of Cambridge).

9. Dissemination

9.1. Leadership within scientific community

- *Editorial boards of journals*
  - J.-P. Le Cadre is Area Editor of *Jal of Advances in Information Fusion* (ISIF).
- *Technical program committees of conferences*

- P. Bouthemy : co-chairman of the TPC of RFIA'2006, TPC member of ICCV'2005, CIVR'2005, ICME'2005, IBPRIA'2005, CAIP'2005, ICARP'2005, OTCBVS'2005, TAIMA'2005, ECCV'2006.
- I. Laptev : TPC member of VISAPP'2006.
- J.-P. Le Cadre : TPC member of Fusion'2005, COGIS'2006.
- E. Mémin : TPC member of CVPR'2005, ECCV'2006, RFIA'2006.
- P. Pérez : TPC member of ICCV'2005, CVPR'2005, Siggraph'2005, Eurographics'2005, Coresa'2005.

• *Ph.D. reviewing*

- P. Bouthemy: B. Besserer, L3I, La Rochelle (HDR thesis); N. Rea, Trinity College, Dublin; M. Barnard, IDIAP, Switzerland; C. Dorea, UPC, Barcelona; M. Gastaud, I3S, Sophia-Antipolis; A. Genovesio (Institut Pasteur, Paris); G. Rochefort, Onera et LSS, Orsay.
- J.-P. Le Cadre: C. Blanc, Lasmea, Clermont-Ferrand; A. Giremus, Irit, Toulouse; A. Ziadi, Laas, Toulouse.
- P. Pérez: F. Pitié, Trinity College, Dublin; F. Flitti, Lsiit, Strabourg; F. Lerasle, Laas, Toulouse.

• *Project reviewing, consultancy, administrative responsibilities*

- P. Bouthemy participates in the Board Committee of Technovision programme launched by the Ministry of Research and by DGA and aiming at supporting benchmarking projects of image processing and computer vision techniques. He is member of the committee of the EEA prize of the best (French) thesis in signal and image processing.
- P. Bouthemy and J.-P. Le Cadre are deputy members of the committee of the 61th section at Universiy of Rennes 1.
- 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 June 2005. P. Pérez is head of the personnel committee ("Commission Personnel") of Irisa, which oversees scientific non-permanent staff hiring at Irisa/Inria-Rennes, member of the direction team of Irisa/Inria-Rennes ("Équipe de direction"), member of the scientific and technological orientation council (COST, workgroup on large scale inititiaves) of Inria. Also, in 2005, P. Pérez was part of the juries for the young researcher competition and the communication assistant competition at Inria-Rennes, and in charge of the post-doc campaign for Inria-Rennes.
- J.-F. Yao is member of the executive comittee of MAS, a section of the SMAI.

9.2. Teaching

- Master STI “Signal, Telecommunications, Images”, University of Rennes 1, (E. Mémin : statistical image analysis, P. Bouthemy : image sequence analysis, J.-P. Le Cadre : distributed tracking, data association, estimation via MCMC methods, C. Kervrann : geometric modeling for shapes and images).
- Master of Computer Science, Ifsic, University of Rennes 1 (E. Mémin : motion analysis; P. Bouthemy : video indexing).
- DIIC INC, Ifsic, University of Rennes 1 (E. Mémin, : 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 (J. Boulanger : computer vision).
- 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, T. Bréhard : particle filtering and target tracking).
- Graduate student trainees and interns :
  - P. Aguillon (supervised by N. Gengembre and P. Pérez, DIIC INC): HCI enrichment for the tracking platform;
  - A. Bouchikhi (supervised by P. Pérez and E. Mémin, Master STI): application of graph-cut techniques to motion detection;
  - V. Kermel (supervised by E. Mémin, Master in Computer Sciences): computation of wind fields in meteorological satellite image sequences.
  - A. Lehmann (supervised by P. Bouthemy and J.-F. Yao, ETH Zürich): comparison of video segments using local motion descriptors.
  - 8-month final internship of V. Badrinarayanan (Thomson, Rennes) supervised by P. Pérez.
- External thesis supervision :
  - A. Chessel (Ifremer Brest) supervised by F. Cao;
  - F. Bavencoff (Thales Airborne Systems, Cifre grant) and F. Celeste (CTA) supervised by J.-P. Le Cadre. F. Bavencoff defended his thesis untitled “Traitement de l’information appliqué aux systèmes de surveillance” in July 2005;
  - 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

- Frédéric Cao was invited speaker at the CVR'05 conference held in the Center for Vision Research in Toronto, Canada. He gave a talk on "*A contrario* methods for image comparison and video analysis". He was also invited two weeks in the Graduate Summer School, "Intelligent Extraction of Information from Graphs and High Dimensional Data" hosted by the Institute for Pure and Applied Mathematics in UCLA. He gave a talk on "*A contrario* detection of clusters, and application to shape recognition".
- N. Gengembre and P. Pérez demoed the tracking platform developed as part of the FT-RD CRE contract during the FT-RD Innovation Days held in Paris, June 2005. The same platform will be part of the demos at the 30th Irisa Anniversary Open House.
- J.-P. Le Cadre was awarded as Automatica outstanding reviewer.
- E. Mémin gave an invited talk "Estimation de mouvements fluides par imagerie" at Onera, Chatillon, at the workshop on Signal and Image Processing for Experimental Physics. He has also been invited by the AFVL (French Association of Laser Velocimetry) for a workshop on Displacements Measurement for Fluid and Solid Mechanics ("Outils pour l'analyse d'écoulements fluides à partir de séquences d'images").
- P. Pérez gave a presentation at the Ariana project's seminar (Inria Sophia Antipolis, October 2005) on "Tracking, mixtures and particles".

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, n° 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, n° 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, n° 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, n° 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, n° 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, n^o 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, n^o 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, n^o 4, April 2000, p. 573-586.

Doctoral dissertations and Habilitation theses

[11] T. BRÉHARD. *Estimation séquentielle et analyse de performances pour un problème de filtrage non linéaire partiellement observé. Application à la trajectographie par mesure d'angles*, Ph. D. Thesis, Université de Rennes 1, Mention Traitement du Signal et des Télécommunications, December 2005.

[12] G. PIRIOU. *Modélisation statistique du mouvement dans des séquences d'images pour la reconnaissance de contenus dynamiques*, Ph. D. Thesis, Université de Rennes 1, Mention Traitement du Signal et des Télécommunications, December 2005.

[13] T. VEIT. *Détection et analyse du mouvement dans des séquences d'images selon une approche probabiliste a contrario*, Ph. D. Thesis, Université de Rennes 1, Mention Traitement du Signal et des Télécommunications, December 2005.

Articles in refereed 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", vol. 14, n^o 1, 2005, p. 63-79, http://www.irisa.fr/vista/Papers/2004_ip_arnaud.pdf.

[15] F. BAVENCOFF, J.-M. VANPEPERSTRAETE, J.-P. LE CADRE. *Constrained bearings-only target motion analysis via Monte Carlo Markov chain methods*, in "IEEE Trans. on Aerospace and Electronic Systems", to appear.

[16] P. BOUTHEMY, C. HARDOUIN, G. PIRIOU, J.-F. YAO. *Mixed-state auto-models and motion texture modeling*, in "Journal of Mathematical Imaging and Vision", to appear.

[17] T. BRÉHARD, J.-P. LE CADRE. *Closed-form posterior Cramèr-Rao bound for bearings-only tracking*, in "IEEE Trans. on Aerospace and Electronic Systems", to appear.

[18] F. CAO, R. FABLET. *Automatic morphological detection of otolith nucleus*, in "Pattern Recognition Letters", to appear.

[19] F. CAO, P. MUSÉ, F. SUR. *Extracting meaningful curves from images*, in "Journal of Mathematical Imaging and Vision", vol. 22, n^o 2-3, 2005, p. 159-181.

[20] F. CAO, T. VEIT, P. BOUTHEMY. *Image comparison and motion detection by a contrario methods*, in "Computational Vision in Neural and Machine Systems", L. HARRIS, M. JENKIN (editors). , Cambridge

University Press, 2005.

[21] T. CORPETTI, D. HEITZ, G. ARROYO, E. MÉMIN, A. SANTA-CRUZ. *Fluid experimental flow estimation based on an optical-flow scheme*, in "International Journal of Experiments in Fluid", to appear.

[22] F. DAMBREVILLE, J.-P. LE CADRE. *Constrained minimax optimization of continuous search efforts for the detection of a stationary target*, in "Naval Research Logistics", to appear.

[23] R. DONATI, J.-P. LE CADRE. *Detection, target motion analysis and track association with a sensor network*, in "Journal of Information Fusion", to appear.

[24] 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", vol. 23, nº 1, 2005, p. 19-31.

[25] C. HUE, J.-P. LE CADRE, P. PÉREZ. *Posterior Cramer-Rao bounds for multi-target tracking*, in "IEEE Trans. on Aerospace and Electronic Systems", to appear.

[26] A. KOKARAM, N. REA, R. DAHYOT, M. TEKALP, P. BOUTHEMY, P. GROS, I. SEZAN. *Browsing sports video - the challenge of choice*, in "IEEE Signal Processing Magazine", to appear.

[27] I. LAPTEV. *On space-time interest points*, in "International Journal of Computer Vision", vol. 64, nº 2/3, 2005, p. 107–123.

[28] J.-P. LE CADRE. *Quelques aperçus sur l'optimisation de la répartition des efforts de recherche d'une cible*, in "Traitement du Signal, numéro spécial “Gestion intelligente des senseurs”", vol. 22, nº 4, 2005.

[29] P. MUSÉ, F. SUR, F. CAO, Y. GOUSSEAU, J.-M. MOREL. *An a contrario decision method for shape elements recognition*, in "International Journal of Computer Vision", to appear.

[30] P. MUSÉ, F. SUR, F. CAO, Y. GOUSSEAU, J.-M. MOREL. *Shape recognition based on an a contrario methodology*, in "Statistics and Analysis of Shapes", H. KRIM, A. YEZZI (editors). , to appear, Springer Verlag.

[31] N. PEYRARD, P. BOUTHEMY. *Motion-based selection of relevant video segments for video summarization*, in "Multimedia Tools and Applications", vol. 26, nº 3, 2005, p. 259-276.

[32] T. VEIT, F. CAO, P. BOUTHEMY. *An a contrario decision framework for region-based motion detection*, in "International Journal of Computer Vision", to appear.

[33] J. VERMAAK, S. GODSILL, P. PÉREZ. *Monte Carlo filtering for multi-target tracking and data association*, in "IEEE Trans. on Aerospace and Electronic Systems", vol. 41, nº 1, 2005, p. 309-332, http://www.irisa.fr/vista/Papers/2004_aes_draft_vermaak.pdf.

Publications in Conferences and Workshops

[34] E. ARNAUD, B. FAUVET, E. MÉMIN, P. BOUTHEMY. *A robust and automatic face tracker dedicated to broadcast videos*, in "Proc. IEEE Int. Conf. on Image Processing (ICIP'05), Genova, Italy", September 2005.

[35] E. ARNAUD, E. MÉMIN. *An efficient Rao-Blackwellised particle filter for object tracking*, in "Proc. IEEE Int. Conf. on Image Processing (ICIP'05), Genova, Italy", September 2005.

[36] V. AUVRAY, P. BOUTHEMY, J. LIENARD. *Estimation paramétrique multirésolution de mouvements transparents*, in "GRETSI'2005, Louvain-la-Neuve, Belgium", September 2005, http://www.irisa.fr/vista/Papers/2005_gretsi_auvray.pdf.

[37] V. AUVRAY, P. BOUTHEMY, J. LIENARD. *Multiresolution parametric estimation of transparent motions*, in "Proc. IEEE Int. Conf. on Image Processing (ICIP'05), Genova, Italy", September 2005, http://www.irisa.fr/vista/Papers/2005_icip_auvray.pdf.

[38] V. AUVRAY, J. LIENARD, P. BOUTHEMY. *Multiresolution parametric estimation of transparent motions and denoising of fluoroscopic images*, in "Int. Conf. on Medical Image Computing and Computer Assisted Intervention (MICCAI'05), Palm Springs, USA", October 2005, http://www.irisa.fr/vista/Papers/2005_auvray_miccai.pdf.

[39] F. BAVENCOFF, J.-M. VANPEPERSTRAETE, J.-P. LE CADRE. *Performance analysis of optimal data association within a linear regression framework*, in "8th International Conference on Information Fusion, Philadelphia, PA, USA", 2005.

[40] J. BOULANGER, C. KERVRANN, P. BOUTHEMY. *Adaptive spatio-temporal restoration for 4D fluoresence microscopic imaging*, in "Int. Conf. on Medical Image Computing and Computer Assisted Intervention (MICCAI'05), Palm Springs, USA", October 2005.

[41] J. BOULANGER, C. KERVRANN, P. BOUTHEMY. *An adaptive statistical method for 4d-fluorescence image sequence denoising with spatio-temporal discontinuities preserving*, in "Proc. of Imaging for Medical and Life Sciences (IMVIE2), Illkirch, France", March 2005.

[42] J. BOULANGER, C. KERVRANN, P. BOUTHEMY. *An adaptive statistical method for denoising 4D fluorescence image sequences with preservation of spatio-temporal discontinuities*, in "Proc. IEEE Int. Conf. on Image Processing (ICIP'05), Genova, Italy", September 2005.

[43] J. BOULANGER, C. KERVRANN, P. BOUTHEMY. *Approche statistique adaptative pour le filtrage de séquences d'images 3D de fluoroscopie*, in "Congrès Jeunes Chercheurs en Vision par Ordinateur (ORASIS'05), Fournols, France", May 2005.

[44] T. BRÉHARD, J.-P. LE CADRE. *Distributed target tracking for nonlinear systems: application to bearings-only tracking*, in "Proc. Int. Conf. Information Fusion (FUSION'05), Philadelphia, USA", 2005, http://www.irisa.fr/vista/Papers/2005_fusion_brehard.pdf.

[45] T. BRÉHARD, J.-P. LE CADRE. *Estimation distribuée pour un problème de filtrage non linéaire*, in "XXXVIIèmes Journées de Statistique de la SFDS, Pau, France", 2005,

http://www.irisa.fr/vista/Papers/2005_sfds_brehard.pdf.

[46] A. CUZOL, P. HELLIER, E. MÉMIN. *A novel parametric method for non-rigid image registration*, in "Proc. Information Processing in Medical Imaging (IPMI'05), Glenwood Springes, Colorado, USA", G. CHRISTENSEN, M. SONKA (editors). , LNCS, n^o 3565, July 2005, p. 456-467.

[47] A. CUZOL, E. MÉMIN. *A stochastic filter for fluid motion tracking*, in "Proc. Int. Conf. on Computer Vision (ICCV'05), Beijing, China", October 2005.

[48] A. CUZOL, E. MÉMIN. *Vortex and source particles for fluid motion estimation*, in "5th Int. Conf. on Scale-Space and PDE methods in Computer Vision (Scale-Space'05), Hofgeismar, Germany", 2005.

[49] M. P. DODIN, J.-P. LE CADRE. *Re-entry vehicle tracking, observability and theoretical bound*, in "8th International Conference on Information Fusion, Philadelphia, PA, USA", July 2005.

[50] D. HEITZ, J. CARLIER, E. MÉMIN. *Fluid dedicated optical-flow scheme*, in "Proc. Int. Symp. on Particle Image Velocimetry (PIV'05), Pasadena, CA", September 2005.

[51] I. LAPTEV, S. BELONGIE, P. PÉREZ, J. WILLS. *Periodic motion detection and segmentation via approximate sequence alignment*, in "Proc. Int. Conf. on Computer Vision (ICCV'05), Beijing, China", October 2005, http://www.irisa.fr/vista/Papers/2005_iccv_laptev.pdf.

[52] A. LEHUGER, P. LECHAT, N. LAURENT, P. PÉREZ. *Suivi de joueurs dans les séquences sportives à fort changement d'illumination : évaluation du problème et solutions*, in "Proc. Journées Compression et Représentation des Signaux Visuels (CORESA'05), Rennes, France", November 2005.

[53] N. PAPADAKIS, E. MÉMIN, F. CAO. *A variational approach for object contour tracking*, in "Proc. ICCV'05 Workshop on Variational, Geometric and Level Set Methods in Computer Vision, Beijing, China", October 2005.

[54] G. PIRIOU, P. BOUTHEMY, J.-F. YAO. *Motion content recognition in video database with mixed-state probabilistic causal models*, in "Int. Workshop on Content-Based Multimedia Indexing, CBMI'2005, Riga", June 2005.

[55] P. PÉREZ, J. VERMAAK. *Bayesian tracking with auxiliary discrete processes. Application to detection and tracking of objects with occlusions*, in "Proc. ICCV'05 Workshop on Dynamical Vision, Beijing, China", October 2005, http://www.irisa.fr/vista/Papers/2005_wdviccv_perez.pdf.

[56] P. PÉREZ, J. VERMAAK. *Visual tracking and auxiliary discrete processes*, in "Proc. Int. Symp. on Applied Stochastic Models and Data Analysis, Brest, France", May 2005.

[57] C. TILMANT, L. SARRY, T. CORPETTI, P. MOTREFI, E. GEOFFROY, J.-R. LUSSON, E. MÉMIN, J.-Y. BOIRE. *Suivi d'objets pour l'analyse du mouvement en imagerie ultrasonore*, in "Congrès Jeunes Chercheurs en Vision par Ordinateur (ORASIS'05), Fournols, France", Mai 2005.

[58] T. Veit, F. Cao, P. Bouthemy. *A maximality principle applied to a contrario motion detection*, in "Proc IEEE Int. Conf. on Image Processing (ICIP'05), Genova, Italy", September 2005.

[59] R. Venkatesh Babu, P. Pérez, P. Bouthemy. *Robust tracking with motion estimation and kernel-based color modelling*, in "Proc. IEEE Int. Conf. on Image Processing (ICIP'05), Genova, Italy", September 2005, http://www.irisa.fr/vista/Papers/2005_icip_babu.pdf.

[60] J. Vermaak, S. Maskell, M. Briers, P. Pérez. *Bayesian visual tracking with existence process*, in "Proc. IEEE Int. Conf. on Image Processing (ICIP'05), Genova, Italy", September 2005, http://www.irisa.fr/vista/Papers/2005_icip_vermaak.pdf.

[61] J. Vermaak, S. Maskell, M. Briers, P. Pérez. *Multi-target tracking and existence*, in "Proc. Int. Conf. Information Fusion (FUSION'05), Philadelphia, USA", July 2005.

[62] J. Yuan, P. Ruhnau, E. Mémin, C. Schnoerr. *Discrete orthogonal decomposition and variational fluid flow estimation*, in "5th Int. Conf. on Scale-Space and PDE methods in Computer Vision (Scale-Space'05), Hofgeismar, Germany", 2005.

Internal Reports

[63] P. Bouthemy, C. Hardouin, G. Piriou, J. Yao. *Auto-models with mixed states and analysis of motion textures*, Technical report, n° 1682, IRISA, 2005, http://www.irisa.fr/bibli/publi/pi/2005/1682/1682.html.

[64] T. Bréhard, J.-P. Le Cadre. *Closed-form posterior Cramèr-Rao bound for bearings-only tracking*, Technical report, n° 1701, IRISA, March 2005, http://www.irisa.fr/bibli/publi/pi/2005/1701/1701.html.

[65] F. Cao, P. Bouthemy. *A general criterion for image similarity detection*, Technical report, n° 1732, IRISA, 2005, http://www.inria.fr/rrrt/rr-5620.html.

[66] F. Cao, J. Delon, A. Desolneux, P. Musé, F. Sur. *A unified framework for detecting groups and application to shape recognition*, Technical report, n° 1746, IRISA, September 2005, http://www.inria.fr/rrrt/rr-5695.html.

[67] C. Kervrann, J. Boulanger. *Local adaptivity to variable smoothness for exemplar-based image denoising and representation*, Technical report, n° RR-5624, INRIA, July 2005, http://www.inria.fr/rrrt/rr-5624.html.

[68] A. Lehmann, P. Bouthemy, J.-F. Yao. *Comparison of video dynamic contents without feature matching by using rank-tests*, Technical report, n° RR-5586, INRIA, 2005, http://www.inria.fr/rrrt/rr-5586.html.

Miscellaneous

[69] C. Kervrann, J. Boulanger, P. Bouthemy. *Spatio-temporal analysis in 4D video-microscopy*, January 2005, http://www.ercim.org/publication/Ercim_News/enw60/kervrann.html, ERCIM News No.60, Special Theme: Biomedical Informatics.

Bibliography in notes

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

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

[72] V. CASELLES, L. IGUAL, O. SANDER. *An axiomatic approach to scalar data interpolation on surfaces*, in "Nümerische Matematik", to appear.

[73] V. CASELLES, J.-M. MOREL, C. SBERT. *An axiomatic approach to image interpolation*, in "IEEE Trans. on Image Processing", vol. 7, n^{o} 3, March 1998, p. 376-386.

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

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

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

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

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

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

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

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

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

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

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

[85] E. MÉMIN, P. PÉREZ. *Dense estimation and object-based segmentation of the optical flow with robust techniques*, in "IEEE Trans. on Image Processing", vol. 7, nº 5, May 1998, p. 703-719.

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

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

[88] 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.

[89] A. L. YUILLE, N. M. GRZYWACZ. *A theoretical framework for visual motion*, in "High-Level Motion Processing", T. WATANABE (editor). , MIT Press, 1998.