INRIA

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Team VISTAS

Vision Spatio-Temporelle et Apprentissage

Rennes - Bretagne-Atlantique

Theme : Vision, Perception and Multimedia Understanding

Activity
Report

2009

Table of contents

1. Team

Research Scientist
Patrick Bouthemy [DR, head of Inria Rennes Research Center, partial time, HdR]
Charles Kervrann [Inra research scientist, on Inria secondment (mise à disposition); head of team since 01/11/09]
Ivan Laptev [CR, until 01/07/09]
Hervé Jégou [CR, since 01/09/09]
Patrick Pérez [DR, head of team until 01/11/09, HdR]

Technical Staff
Florent Dutrech [2ACI DGA PEA project from 01/11/08 to 01/12/09]
Guillaume Neveu [FTRD CRE project from 01/03/08 to 01/09/09]
Rachid Benmoktar [Quaero project from 01/09/09]

PhD Student
Vijay Badrinarayanan [Cifre grant, Thomson-RD, from 01/03/06 to 01/04/09]
Emilie Dexter [DGA grant, from 01/10/06]
Sophie Blestel [MENRT grant, since 01/10/08]
Tomas Crivelli [Faculty of Engineering, University of Buenos-Aires]
Alexandre Hervieu [Inria grant, FP6-NOE Muscle, until 01/03/09]
Matthieu Fradet [Cifre grant, Thomson-RD, from 01/03/07]
Thierry Pécot [Inra from 01/10/06 to 31/03/08, Inria-Curie Institute grant from 01/04/08]
Muneeb Ullah [Inria grant, Quaero project, from 01/03/09]

Post-Doctoral Fellow
Anatole Chessel [Inria post-doc, from 01/01/08]
François Lecellier [Inria postdoc, ANR Icos-HD project, from 01/09/09]

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

2. Overall Objectives

2.1. Overall Objectives

Vista research work is concerned with various types of spatio-temporal images, (mainly video images and video-microscopy). We are investigating methods to analyze dynamic scenes, and, more generally, dynamic phenomena, within image sequences. We address the full range of problems raised by the analysis of such dynamic contents with a focus on image motion analysis issues: denoising, motion detection, motion estimation, motion-based segmentation, tracking, motion recognition and interpretation with learning. We usually rely on statistical approaches, resorting to: Markov models, Bayesian inference, robust estimation, particle filtering, learning. Application-wise, we focus our attention on two main domains: content-aware video applications and biological imaging. For that, a number of collaborations, academic and industrial, national and international, are set up.

3. Scientific Foundations

3.1. Motion estimation and motion segmentation with MRF models

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

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

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

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

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

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

where θ denotes the vector of motion model parameters.

One important step ahead in solving the motion segmentation problem was to formulate the motion segmentation problem as a statistical contextual labeling problem or in other words as a discrete Bayesian inference problem. Segmenting the moving objects is then equivalent to assigning the proper (symbolic) label (i.e., the region number) to each pixel in the image. The advantages are mainly two-fold. Determining the support of each region is then implicit and easy to handle: it merely results from extracting the connected components of pixels with the same label. Introducing spatial coherence can be straightforwardly (and locally) expressed by exploiting MRF models. Here, by motion segmentation, we mean the competitive partitioning of the image into motion-based homogeneous regions. Formally, we have to determine the hidden discrete motion variables (i.e., region numbers) $l(i)$ where i denotes a site (usually, a pixel of the image grid; it could be also an elementary block). Let $l = \{l(i), i \in S\}$. Each label $l(i)$ takes its value in the set $\Lambda = \{1, .., N_{reg}\}$ where N_{reg} is also unknown. Moreover, the motion of each region is represented by a motion model (usually, a 2D affine motion model of parameters θ which have to be conjointly estimated; we have also explored non-parametric motion modeling [47]). Let $\Theta = \{\theta_k, k = 1, .., N_{reg}\}$. The data model of relation (2) is used. The *a priori* on the motion label field (i.e., spatial coherence) is expressed by specifying a MRF model (the simplest choice is to favor the configuration of the same two labels on the two-site cliques so as to yield compact regions with regular boundaries). Adopting the Bayesian MAP criterion is then equivalent to minimizing an energy function E whose expression can be written in the general following form:

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

where $i \sim j$ designates a two-site clique. We first considered [45] 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 [56], [57]. It allowed us to avoid the alternate minimization procedure and to determine or update the number of regions through an outlier process in every region.

Specifying (simple) MRF models at a pixel level (i.e., sites are pixels and a 4- or 8-neighbor system is considered) is efficient, but remains limited to express more sophisticated properties on region geometry or to handle extended spatial interaction. Multigrid MRF models [50] is a means to address somewhat the second concern (and also to speed up the minimization process while usually supplying better results). An alternative is to first segment the image into spatial regions (based on gray level, color or texture) and to specify a MRF model on the resulting graph of adjacent regions [48]. 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 [49]. Then, robust estimators were popularized leading to the introduction of so-called auxiliary variables ζ now taking their values in $[0,1]$ [55]. 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 [54]. Depending on the followed approach, the third term of the energy $E(\mathbf{w}, \zeta)$ can be optional.

3.2. Object tracking with non-linear probabilistic filtering

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

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

$$\begin{aligned} \textbf{prediction}\ \text{step:} \qquad & p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) = \int p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})d\mathbf{x}_{t-1} \\ \textbf{filtering}\ \text{step:} \qquad & p(\mathbf{x}_t|\mathbf{y}_{1:t}) = \frac{p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t-1})}{p(\mathbf{y}_t|\mathbf{y}_{1:t-1})}, \end{aligned} \tag{5}$$

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

The tracking recursion yields closed-form expressions in only a small number of cases. The most well-known of these is the Kalman Filter (KF) for linear and Gaussian dynamic and likelihood models. For general non-linear and non-Gaussian models the tracking recursion becomes analytically intractable, and approximation techniques are required. Sequential Monte Carlo (SMC) methods [46], [52], [51], 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

$$w_t^{(n)} \propto w_{t-1}^{(n)} \frac{p(\mathbf{y}_t|\mathbf{x}_t^{(n)})p(\mathbf{x}_t^{(n)}|\mathbf{x}_{t-1}^{(n)})}{q(\mathbf{x}_t^{(n)}|\mathbf{x}_{t-1}^{(n)}, \mathbf{y}_t)}, \quad \sum_{n=1}^{N} w_t^{(n)} = 1, \tag{6}$$

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

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

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

- selection and on-line estimation of observation models with multiple data modalities: except in cases where detailed prior is available on state dynamics (e.g., in a number of TMA applications), the observation model is the most crucial modeling component. A sophisticated filtering machinery will not be able to compensate for a weak observation model (insufficiently discriminant and/or insufficiently complete). In most adverse situations, a combination of different data modalities is necessary. Such a fusion is naturally allowed by SMC, which can accommodate any kind of data model. However, there is no general means to select the best combination of features, and, even more importantly, to adapt online the parameters of the observation models associated to these features. The first problem is a difficult instance of discriminative learning with heterogeneous inputs. The second problem is one of online parameter estimation, with the additional difficulty that the estimation should be mobilized only parsimoniously in time, at instants that must be automatically determined (adaptation when the entities are momentarily invisible or simply not detected by the sensors will always cause losses of track). These problems of feature selection, online model estimation, and data fusion, have started to receive a great deal of attention in the visual tracking community, but proposed tools remain ad-hoc and restricted to specific cases.
- multiple-object tracking with data association: when tracking jointly multiple objects, data association rapidly poses combinatorial problem. Indeed, the observation model takes the form of a mixture with a large number components indexed by the set of all admissible associations (whose enumeration can be very expensive). Alternatively, the association variables can be incorporated within the state space, instead of being marginalized out. In this case, the observation model takes a simpler product form, but at the expense of a dramatic dimension increase of the space in which the estimation must be conducted.

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

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

- Content-aware video applications (Thomson, FT-RD, INA, Bertin Technologies);
- Biological imagery (Inra, Curie Institute, Biology Dpt of University of Rennes 1)

4. Application Domains

4.1. 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.

4.2. Biological imagery

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

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

Quantitative analysis of data obtained by fast 4D wide-field microscopy with deconvolution, confocal spinning-disk microscopy, Total Internal Reflectance Microscopy (TIRF), Fluorescence Recovery After Photobleaching (FRAP) combined with Green Fluorescence Protein (GFP)-tagging allows one to enlighten the role of specific proteins on HeLa human cell lines. Among these proteins, some are member of the family of Rab-GTPases that bind reversibly to specific membranes within the cells. In our study, we aim at designing computational and statistical models to understand membrane trafficking and, more precisely to better elucidate the role of Rab family proteins inside their multiprotein complexes. We mainly focus on the analysis of transport intermediates (vesicles) that deliver cellular components to appropriate places within cells. Methods have been developed for interaction estimation between Rab11 and Langerin proteins, and dynamic estimation of Rab6a and Rab6a' proteins - involved in the regulation of transport from the Golgi apparatus to the endoplasmic reticulum.

Moreover, microscopic imaging at both the light and electron microscopic level provides multiscale unique information on protein localization and interactions, and extends and enriches that obtained from molecular and biochemical techniques. The 3D reconstruction of macromolecular structures form 2D EM (Electronic Microscopy) images of vitrified biological samples (Cryo-EM) has some advantages over other imaging techniques since it has proved to be an effective technique to investigate the structure of native cells with macromolecular resolution and preserve the whole integrity of the cell. Nevertheless, the high magnification available with EM comes with a limited field of view. Also, the very low contrast of unstained and vitrified biological specimens and the need to minimize the exposure to electron radiation make the identification of specific structures a difficult and time consuming task. Therefore one needs a gentle and time efficient way to locate structures of interest, improve image contrast and remove noise for a better interpretation of the image contents. We currently investigated image segmentation methods to analyze microtubule dynamics observed in Cryo-EM in collaboration with University of Rennes 1 - UMR 6026 (D. Chrétien).

Microtubules are long tubes of 25 nanometers in diameter formed with the $\alpha\beta$-tubulin dimers and present in all eukaryotic cells. They play fundamental roles in cells life, in particular during their division since they are involved in chromosome segregation. The understanding of their structure and assembly mechanisms is of major importance not only in fundamental biological research, but also in medicine since microtubules are a major target of anti-mitotic drugs used in cancer therapy. The study of their molecular architecture is possible by cryo-electron microscopy, which enables observation of biological specimens frozen at liquid nitrogen temperature in their native state. The images obtained are projection views of the specimens perpendicular to the electron beam that have a resolution on the order of a few (tenth of) Angströms.

In the coming 3 years, we will study interactions between neighboring protofilaments at the extremity of microtubules and the biological function of open sheet.

5. Software

5.1. Motion2d software - parametric motion model estimation

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

Motion2D is a multi-platform object-oriented library to estimate 2D parametric motion models in an image sequence. It can handle several types of motion models, namely, constant (translation), affine, and quadratic models. Moreover, it includes the possibility of accounting for a global variation of illumination. The use of such motion models has been proved adequate and efficient for solving problems such as optic flow computation, motion segmentation, detection of independent moving objects, object tracking, or camera motion estimation, and in numerous application domains, such as dynamic scene analysis, video surveillance, visual servoing for robots, video coding, or video indexing. Motion2D is an extended and optimized implementation of the robust, multi-resolution and incremental estimation method (exploiting only the spatio-temporal derivatives of the image intensity function) we defined several years ago [56]. 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 project-team), Patrick Bouthemy.

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

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

Participants: Thomas Corpetti, Patrick Heas, Etienne Mémin (Fluminance project-team), Patrick Pérez.

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

5.4. VTrack software - generic interactive visual tracking platform

Participants: Guillaume Neveu, Patrick Pérez.

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

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

Participant: Ivan Laptev.

ObjectDet is an open source efficient C++ implementation of object detection that extends our previous method [53]. The software achieves object detection at the approximate rate of 10 frames per second on 320 × 240 images on a modest PC. The accuracy of the method was ranked among the top ones in The PASCAL Visual Object Classes Challenges 2006 and 2007 (VOC2006, VOC2007). The detection is achieved with a "scanning window" classifier applied to different positions and scales of the image. The underlying AdaBoost classifier is trained from histogram features computed on rectangle-annotated object images. Variations in object views can be handled by training separate classifiers for different views of the object. Different types of histogram features including Histograms of Oriented Gradient (HOG), second-order derivative histograms and color histograms are implemented and can be used in a complementary way for increased performance.

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

5.6. SAFIR-nD - image denoising software

Participant: Charles Kervrann.

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

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

Participant: Charles Kervrann.

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

5.8. HULLKGROUND software - Background subtraction by convex hull estimation

Participant: Charles Kervrann.

The HULLKGROUND software written in JAVA enables to decompose a fluorescence microscopy image sequence into two components: 1/ an image sequence showing mobile objects; 2/ an image sequence showing the slightly moving background. Each temporal signal of the sequence is processed individually and analyzed with computational geometry tools. The convex hull is estimated automatically for each pixel and subtracted to the original signal. The method is unsupervised, requires no parameter tuning and is a simplified version of the α shapes-based scale-space method.

5.9. TubuleJ software - A plugin for automatic straightening, filtering and 3D reconstruction of microtubule cryo-EM projection views

Participants: Charles Kervrann, Sophie Blestel.

The TUBULEJ software written in JAVA has been developed to analyze microtubule structures and helical structures in 2D cryo-electron microscope images. The software enables to straighten curved microtubule images by estimating automatically points locations on the microtubule axis. The local center estimation method relies on microtubule cylindrical shape analyzed in the Fourier domain. A user friendly interface is provided to filter straight fiber images by selecting manually the layer lines of interest in the Fourier domain. Third, this plugin can be used to generate a set of 2D projection views from a single microtubule projection view and a few parameters of this microtubule structure. These projection views are then back projected, by using for example the IMOD plugin of IMAGEJ, to reconstruct the 3D microtubule.

6. New Results

6.1. Image sequence processing and modeling

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

Participants: Charles Kervrann, Patrick Pérez.

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

6.1.2. Detection and segmentation of moving objects in complex scenes

Participants: Guillaume Neveu, Florent Dutrech, Patrick Pérez.

[In collaboration with A. Bugeau (UPF, Barcelona)]
Detecting individual moving objects in videos that are shot by either still or mobile cameras is an old problem, which is routinely addressed in a number of real applications such as tele-surveillance. There are, however, a number of applicative contexts where this motion analysis problem is not satisfactorily handled by existing techniques. In the context of activity analysis in dynamically cluttered environments (dynamic background, crowded scenes, etc.) for instance, the problem is the one of separating out foreground moving objects of interest from other uninteresting moving objects in the background.

We have proposed a completely automatic system to address this difficult task. It involves three main steps. First, a set of moving points is selected within a sub-grid of image pixels. A multi-cue descriptor is associated to each of these points. Clusters of points are then formed using a variable bandwidth mean shift technique with automatic bandwidth selection. Finally, segmentation of the object associated to a given cluster is performed using graph cuts. Experiments and comparisons to other motion detection methods on challenging sequences demonstrate the performance of the proposed method for video analysis in complex scenes.

6.1.3. Motion texture tracking with mixed-state Markov chains

Participants: Tomas Crivelli, Patrick Bouthemy.

[In collaboration with B. Cernuschi-Frias (Univ. Bueno Aires), G. Piriou and J.-F. Yao (IRMAR, Univ. Rennes)]
Examples of motion textures are mostly found in natural elements as fire, smoke, water, moving foliage but also in traffic and crowd scenes. Tracking this type of video contents is essential for video surveillance applications. However, standard tracking techniques fail in that motion textures are non-rigid, display highly dynamic contents, with specific statistical properties. The key characteristic of motion textures is that local motion observations depict a mixed-state nature: the null motion value appears as a discrete value with positive probability and the rest follows a continuous distribution. We thus have developed a motion texture tracking algorithm based on two main steps. First, motion values are modeled using mixed-state Markov chains which capture the main statistical (temporal) properties of mixed-state observations with only 13 parameters. This model is initially learned for the tracked content. Second, a motion texture window matching strategy is applied based on the computation of the conditional Kullback-Leibler divergence between mixed-state Markov chains. This permits to address the problem of displacement estimation, Results on complex real sequences of different nature have demonstrated an improved performance against standard methods.

6.1.4. Dynamic remote sensing

Participant: Patrick Pérez.

[In collaboration with E. Mémin and S. Gorthi, Fluminance project-team]
See Fluminance activity report.

6.1.5. Geodesic image and video editing

Participant: Patrick Pérez.

[In collaboration with A. Criminisi, T. Sharp and C. Rother, Microsoft Research Cambridge]
In this work [21], a new, unified technique to perform general edge-sensitive editing operations on n-dimensional images and videos efficiently. The first contribution is the introduction of a generalized geodesic distance transform (GGDT), based on soft masks. This provides a unified framework to address several, edgeaware editing operations. Diverse editing tasks such as denoising and non-photorealistic rendering, are all dealt with fundamentally the same, fast algorithm. Second, a new, geodesic, symmetric filter (GSF) is presented which imposes contrast-sensitive spatial smoothness into segmentation and segmentation-based editing tasks (cutout, object highlightening, colorization, panorama stiching). The effect of the filter is controlled by two intuitive, geometric parameters. In contrast to existing techniques, the GSF filter is applied to real-valued pixel likelihoods (soft masks), thanks to GGDTs and it can be used for both interactive and automatic editing tasks. Complex object topologies are dealt with effortlessly. Finally, the parallelism of GGDTs enables us to exploit modern multi-core CPU architectures as well as powerful new GPUs, thus providing great flexibility of implementation and deployment. Our technique operates on both images and videos, and generalizes naturally to n-dimensional data. The proposed algorithm has been validated via quantitative and qualitative comparisons with existing, state of the art approaches. Numerous results on a variety of image and video editing tasks further demonstrate the effectiveness of our method.

6.1.6. Aggregating local descriptors into a compact image representation

Participants: Hervé Jégou, Patrick Pérez.

[In collaboration with M. Douze and C. Schmid, Lear project-team]
We address the problem of image search on a very large scale, where three constraints have to be considered jointly: the accuracy of the search, its efficiency, and the memory usage of the representation. To address this problem, we first propose a simplification of the Fischer Kernel image representation, which is a way of aggregating local image descriptors into a vector of limited dimension. We then present an approach for coding and indexing such vectors that preserves well the accuracy of the vectorial Euclidean comparison. The evaluation shows that our approach significantly outperforms the state-of-the-art: the search accuracy is comparable to the bag-of-features approach for an image representation requiring 20 bytes of memory.

6.2. Low level visual motion analysis

6.2.1. Clustering point trajectories with various life-spans

Participants: Matthieu Fradet, Patrick Pérez.

[In collaboration with Ph. Robert, Thomson Research]
Motion-based segmentation of a sequence of images is an essential step for many applications of video analysis, including action recognition and surveillance. In this work [33], we introduce a new approach to motion segmentation operating on point trajectories. Each of these trajectories has its own start and end instants, hence its own life-span, depending on the pose and appearance changes of the object it belongs to. A set of such trajectories is obtained by tracking sparse interest points. Based on an adaptation of recently proposed J-linkage method, these trajectories are then clustered using series of affine motion models estimated between consecutive instants, and an appropriate residual that can handle trajectories with various life-spans. Our approach does not require any completion of trajectories whose life-span is shorter than the sequence of interest. We have evaluated the performance of the single cue of motion, without considering spatial prior and appearance. Using a standard test set, we validate our new algorithm and compare it to existing ones. Experimental results on a variety of challenging real sequences demonstrate the potential of our approach, even when no other image features (colours and contours in particular) are jointly exploited.

6.2.2. Multi-view synchronization of human actions and dynamic scenes

Participants: Émilie Dexter, Ivan Laptev, Patrick Pérez.

This work deals with the temporal synchronization of image sequences [31], [40], [41], [30]. Two instances of this problem are considered: (a) synchronization of human actions and (b) synchronization of dynamic scenes with view changes. To address both tasks and to reliably handle large view variations, we use self-similarity matrices which remain stable across views. We propose time-adaptive descriptors that capture the structure of these matrices while being invariant to the impact of time warps between views. Synchronizing two sequences is then performed by aligning their temporal descriptors using the Dynamic Time Warping algorithm. We have conducted quantitative comparisons between time-fixed and time-adaptive descriptors for image sequences with different frame rates. We have also demonstrated the performance of the approach on several challenging videos with large view variations, drastic independent camera motions and within-class variability of human actions.

6.2.3. Bag of salient tracklets for video detection

Participants: François Lecellier, Hervé Jégou, Patrick Pérez.

In the context of Icos-HD ANR project 8.1.3, we have started to investigate the use of point tracks, or *tracklets*, as low-level features of interest for video indexing and retrieval. Sets of such tracklets indeed capture distinctive motion information (as demonstrated in our works on multi-view video alignment 6.2.2 and view-independent action recognition 6.3.1) that should be complementary to classic static descriptors such as SIFT or SURF for retrieval tasks. In particular, when individual frame representations become highly ambiguous (e.g., when searching a given football shot in a databasis of football broadcasts), it is hoped that describing the dynamic content of the video will be of much help. To this end, we have proposed ways to characterize the saliency of tracklets and to describe them compactly for search purpose, leading to the concept of bag of salient tracklets. The experimental validation of such a description for video copy detection is currently pursued.

6.2.4. Stochastic filtering technique for the tracking of closed curves

Participant: Patrick Pérez.

[In collaboration with E. Mémin and Ch. Avenel, Fluminance project-team]
See Fluminance activity report.

6.2.5. Visual tracking by online selection of tracking filters

Participants: Vijay Badrinarayanan, Patrick Pérez.

[In collaboration with Lionel Oisel and François Le Clerc, Thomson Research]
We have developed a probabilistic graphical model that integrates multiple tracking filters for target state estimation. Based on a special form of parametric graphical prior, this model enables estimation of the target state by a linear combination of posteriors propagated by a selected set of tracking filters. A novel generic multi-cue switch tracker with a dynamically changing set of point trackers and a color based particle filter is derived from this model. Each point tracker is cast as a pseudo-random particle filter to enable tractable inference on the proposed model. It is argued by way of experiments that, in addition to the advantages contained in the fusion of multiple filters, the tracker also provides an effective way to tackle the problem of online target reference model update. Experimental results on very challenging sequences and comparative studies indicate this tracker is suited for robust position tracking over extended durations. The proposed tracker is compared, qualitatively and quantitatively, to appropriate standard schemes in the literature, to highlight the advantages and bring out the drawbacks of the proposed approach.

6.3. Dynamic event modeling, learning and recognition

6.3.1. View-independent action recognition from temporal self-similarities

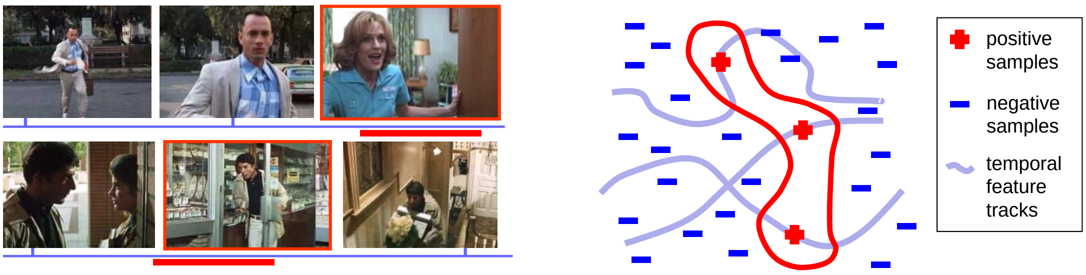
Participants: Émilie Dexter, Ivan Laptev, Patrick Pérez.

In this work [22], we address the problem of recognizing human actions under view changes. We explore self-similarities of action sequences over time and observe the striking stability of such measures across views. Building upon this key observation, we develop an action descriptor that captures the structure of temporal similarities and dissimilarities within an action sequence. Despite this temporal self-similarity descriptor not being strictly view-invariant, we provide intuition and experimental validation demonstrating its high stability under view changes. Self-similarity descriptors are also shown stable under performance variations within a class of actions, when individual speed fluctuations are ignored. If required, such fluctuations between two different instances of the same action class can be explicitly recovered with dynamic time warping, as will be demonstrated, to achieve cross-view action synchronization. More central to present work, temporal ordering of local self similarity descriptors can simply be ignored within a bag-of-features type of approach. Sufficient action discrimination is still retained this way to build a view-independent action recognition system. Interestingly, self-similarities computed from different image features possess similar properties and can be used in a complementary fashion. Our method is simple and requires neither structure recovery nor multi-view correspondence estimation. Instead, it relies on weak geometric properties and combines them with machine learning for efficient cross-view action recognition. The method is validated on three public datasets. It has similar or superior performance compared to related methods and it performs well even in extreme conditions such as when recognizing actions from top views while using side views only for training.

6.3.2. Automatic Annotation of Human Actions in Video

Participant: Ivan Laptev.

[In collaboration with O. Duchenne, J. Sivic, F. Bach and J. Ponce, Willow project-team]
Our work in [32] addresses the problem of automatic temporal annotation of realistic human actions in video using minimal manual supervision. To this end we consider two associated problems: (a) weakly-supervised learning of action models from readily available annotations, and (b) temporal localization of human actions in test videos. To avoid the prohibitive cost of manual annotation for training, we use movie scripts as a means of weak supervision. Scripts, however, provide only implicit, sometimes noisy, and imprecise information about the type and location of actions in video (cf. Figure 1(a)). We address this problem with a kernel-based discriminative clustering algorithm that locates actions in the weakly-labeled training data (cf. Figure 1(b)). Using the obtained action samples, we train temporal action detectors and apply them to locate actions in the raw video data. Our experiments demonstrate that the proposed method for weakly-supervised learning of action models leads to significant improvement in action detection. We present detection results for three action classes in four feature length movies with challenging and realistic video data.



(a) (b)

Figure 1. (a): Video clips with OpenDoor actions provided by automatic script-based annotation. Selected frames illustrate both the variability of action samples within a class as well as the imprecise localization of actions in video clips. (b): In feature space, positive samples are constrained to be located on temporal feature tracks corresponding to consequent temporal windows in video clips. Background (non-action) samples provide further constrains on the clustering.

6.3.3. Actions in Context

Participant: Ivan Laptev.

[In collaboration with M. Marszałek and C. Schmid, Lear project-team]
We exploit the context of natural dynamic scenes for human action recognition in video. Human actions are frequently constrained by the purpose and the physical properties of scenes and demonstrate high correlation with particular scene classes. For example, eating often happens in a kitchen while running is more common outdoors (cf. Figure 2). The contribution of [36] is three-fold: (a) we automatically discover relevant scene classes and their correlation with human actions, (b) we show how to learn selected scene classes from video without manual supervision and (c) we develop a joint framework for action and scene recognition and demonstrate improved recognition of both in natural video. We use movie scripts as a means of automatic supervision for training. For selected action classes we identify correlated scene classes in text and then retrieve video samples of actions and scenes for training using script-to-video alignment. Our visual models for scenes and actions are formulated within the bag-of-features framework and are combined in a joint scene-action SVM-based classifier. We report experimental results and validate the method on a new large dataset with twelve action classes and ten scene classes acquired from 69 movies.

(a) eating, kitchen (b) eating, cafe (c) running, road (d) running, street

Figure 2. Video samples from our dataset with high co-occurrences of actions and scenes and automatically assigned annotations.

6.3.4. Evaluation of local spatio-temporal features for action recognition

Participants: Muneeb Ullah, Ivan Laptev.

[In collaboration with H. Wang, A. Kläser and C. Schmid, Lear project-team]
Local space-time features have recently become a popular video representation for action recognition. Several methods for feature localization and description have been proposed in the literature and promising recognition results were demonstrated for a number of action classes. The comparison of existing methods, however, is often limited given the different experimental settings used. The purpose of this work [39] is to evaluate and compare previously proposed space-time features in a common experimental setup. In particular, we consider four different feature detectors and six local feature descriptors and use a standard bag-of-features SVM approach for action recognition. We investigate the performance of these methods on a total of 25 action classes distributed over three datasets with varying difficulty. Among interesting conclusions, we demonstrate that regular sampling of space-time features consistently outperforms all tested space-time interest point detectors for human actions in realistic settings. We also demonstrate a consistent ranking for the majority of methods over different datasets and discuss their advantages and limitations.

6.3.5. Video content recognition using trajectories

Participants: Alexandre Hervieu, Patrick Bouthemy, Jean-Pierre Le Cadre (Aspi project-team).

Content-based exploitation of video documents is of continuously increasing interest in numerous applications, e.g., for retrieving video sequences in huge TV archives, creating summaries of sports TV programs, or detecting specific actions or activities in video-surveillance. Considering 2D trajectories computed from image sequences is attractive since they capture elaborate space-time information on the viewed actions. Methods for tracking moving objects in an image sequence are now available to get reliable enough 2D trajectories in various situations. Our approach takes into account both the trajectory shape (geometrical information related to the type of motion) and the speed change of the moving object on its trajectory (dynamics-related information). Due to the trajectory features we have specified (local differential features combining curvature and motion magnitude), the designed method is invariant to translation, to rotation and to scaling while taking into account both shape and dynamics-related information on the trajectories. A novel hidden Markov model (HMM) framework is proposed which is able in particular to handle small sets of observations. Parameter setting is properly addressed. A similarity measure between the HMM is defined and exploited to tackle three dynamic video content understanding tasks : supervised recognition, clustering and detection of unexpected events. We have conducted experiments on several significative sets of real videos including sport videos. Then, hierarchical semi-Markov chains are introduced to process trajectories of several interacting moving objects. The temporal interactions between trajectories are taken into account and exploited to characterize relevant phases of the activities in the processed videos. Our method has been favorably evaluated on sets of trajectories extracted from squash and handball videos. Applications of such interaction-based models have also been extended to 3D gesture and action recognition and clustering, and temporal segmentation of actions. The results show that taking into account the interactions is of great interest for such applications.

6.4. Image processing and modeling for biological imaging

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

Participants: Charles Kervrann, Patrick Bouthemy.

[In collaboration with J. Boulanger and P. Elbau (RICAM, Austria), J.-B. Sibarita and J. Salamero (Curie Institute)]
New video-microscopy technology enables to acquire 4-D data that require the design and the development of specific image denoising methods able to preserve details and discontinuities in both the $(x - y - z)$ space dimensions and the time dimension. Images are noisy due to the weakness of the fluorescence signal in time-lapse recording. Accordingly, in collaboration with UMR 144 CNRS / Curie Institute, we have developed an original and efficient spatio-temporal filtering method for significantly increasing the signal-to-noise ratio (SNR) in noisy fluorescence microscopic image sequences where small particles have to be tracked from frame to frame. The proposed method exploits 3D+time information to improve the signal-to-noise ratio of images corrupted by mixed Poisson-Gaussian noise. A variance stabilization transform is first applied to the image-data to introduce independence between the mean and variance. This pre-processing requires the knowledge of

parameters related to the acquisition system, also estimated in our approach. In a second step, we propose an original statistical patch-based framework for noise reduction and preservation of space-time discontinuities. In the continuous setting, we have shown that the proposed estimation procedure can be interpreted as a steepest descent algorithm related to the fixed point solution corresponding to the minimization of a global energy function involving non-local terms and local image contexts described by patches. The size of each neighborhood is optimized to improve the performance of the pointwise estimator. In our experiments, the SNR is shown to be drastically improved and enhanced images can then be correctly segmented. In fluorescence video-microscopy, recent experiments demonstrated also that this method can be used as a pre-processing stage for image deconvolution, and allows us to increase the frame-rate of a factor 10 with the same SNR values. Finally, this novel approach can be used for biological studies where dynamics have to be analyzed in molecular and subcellular bio-imaging.

6.4.2. Network tomography for tracking in fluorescence microscopy imaging

Participants: Thierry Pécot, Charles Kervrann, Patrick Bouthemy.

[In collaboration with J. Boulanger (RICAM, Austria), J.-B. Sibarita and J. Salamero (Curie Institute)]
The study of protein dynamics is essential for understanding the multi-molecular complexes at subcellular levels. Green Fluorescent Protein (GFP)-tagging and time-lapse fluorescence microscopy enable to observe molecular dynamics and interactions in live cells and consequently to make progress in knowledge about protein dynamics. Original image analysis methods are then required to process challenging 2D or 3D image sequences. Tracking methods that estimate the whole trajectories of moving objects have been successfully experimented but can be applied for tracking a limited number of objects (a few dozens). To address the tracking problem of several hundreds of objects, we propose instead an original framework that provides general information about molecule transport, that is about traffic flows between origin and destination regions detected in the image sequence. Traffic estimation can be accomplished by adapting the recent advances in Network Tomography commonly used in network communications. Our estimation method is inspired from the Network Tomography (NT) concept, introduced by Vardi in network communications and further applied to video surveillance. NT-based approaches, devoted to statistical traffic analysis, simplifies the tracking process because it only requires detection of an object as it moves from one region to another and avoids the difficult data association problem. This statistical method allows us to provide a global description of traffic flows. We just need to count the number of "objects/vesicles" in different image regions at each time step. In collaboration with UMR 144 CNRS / Curie Institute, we extended the usual NT concept to non-binary routing from geodesic paths given the image sequence. Unlike previous approaches, the new formulation can be considered as a probabilistic minimal paths modeling for object tracking. We showed that the origin-destination (OD)paths are not the minimal paths between the two extremities but formed as as a set of minimal paths joining intermediate points. We also we proposed an estimation/optimization framework to derive counting measurements from image intensity (fluorescence). The traffic flow problem is also solved with additional parsimonious constraints. This approach has been developed for real image sequences and Rab proteins, known to be involved in the regulation of intracellular membrane trafficking.

6.4.3. Repetitive and transient event detection in fluorescence video-microscopy

Participant: Charles Kervrann.

[In collaboration with J. Boulanger (RICAM, Austria), A. Gidon and J. Salamero (Curie Institute)]
Endocytosis-recycling is an essential cellular trafficking process regulating the proper distribution and function of a large set of molecules, such as lipids, receptors, or adhesion transmembrane proteins. This dynamic process also participates to the homeostasis of intracellular membrane compartments. Progresses in imaging dynamics behaviors of molecules including fast video microscopy and the application of evanescent wave microscopy have allowed to image intracellular vesicular movements, exocytosis and endocytosis of fluorescent-tagged proteins. In parallel, statistical image analysis has emerged as a basic methodology in the study of many biological phenomena. However, spatio-temporal analysis of transient events occurring at different sites of the cell has not been systematically performed. In addition, more formal tests are required in testing biological hypotheses, rather than visual inspection combined with more or less manual statistical analysis. For

an unbiased quantification of repetitive and transient events, such as those observed during the trafficking of molecules traveling through the endosomal-recycling network of cells, their automatic detection become necessary. While requiring particular adjustments, our proposed approach is versatile enough, to be applicable to diverse although complementary modes of microscopy. This is illustrated for fast both video and TIRF imaging techniques in collaboration with UMR 144 CNRS / Curie Institute and RICAM (Austria). The proposed detection method can be decomposed into three main steps: i) a first pre-processing step is dedicated to the normalization of the image sequence; ii) the second step is the patch-based detection procedure to detect unusual patterns; iii) a third post-processing step allows us to cluster and count detected events in space and time. While focusing here on one particular Lectin receptor that constitutively recycles from internal compartments to the plasma membrane, it could be translated to many other studies of membrane trafficking in health and diseases such as diabetes, neurological, pigmentation or lysosomal defects.

6.4.4. Dynamic background subtraction in fluorescence video-microscopy

Participants: Charles Kervrann, Thierry Pécot, Patrick Bouthemy.

[In collaboration with A. Chessel, B. Cinquin and J. Salamero (Curie Institute)]
In collaboration with UMR 144 CNRS / Curie Institute, the main idea was to produce new descriptors able to capture spatio-temporal features of moving particles or group of particles with high variable motions. Actually, a wide range of proteins viewed using xFP probes in fluorescence video-microscopy shows a cytosolic state diffusing slowly and a membrane state eventually moving as vesicles or small tubules along the cell cytoskeleton network. Due to the property of fluorescence, the measured intensity is the sum of the contributions of these two components. The membrane and cytosolic components are then analyzed respectively by the Network Tomography approach for both the vesicle traffic estimation and the free-diffusion estimation in the cytosol. Previous parametric and non-parametric estimation methods have been tested for background subtraction. This year, we have proposed a promising and original framework based the computational geometry concept of α-shapes. The relationships with the Empirical Mode Decomposition (EMD), the curvature motion-based scale-space and some operators from mathematical morphology, ware studied. We have also investigated the Conditional Random Field modeling framework to improve the background subtraction with more flexibility. This method has been compared to several algorithms : i) gray-scale opening (a.k.a. "rolling ball"); ii) wavelet-based detection combined with image interpolation; iii) Conditional Random Field-based detection; iv) computational geometry for temporal signal analysis. The evaluation protocol, related to the actual use of image sequence decomposition, includes a qualitative evaluation on real image sequences by experts and a quantitative evaluation on simulated sequences that mimic real images.

6.4.5. 3D reconstruction in cryo-Electron Microscopy

Participants: Sophie Blestel, Charles Kervrann.

[In collaboration with D. Chrétien (UMR 6026, Rennes)]
In collaboration with UMR 6026 CNRS, we are interested in longitudinal projections of microtubules (i.e. perpendicular to the microtubule axis). Microtubules are composed of identical subunits that arrange together to form a helical lattice (see Figure 2). Because of their symmetry, most of their information can be retrieved from their Fourier spectrum. However, due to their flexibility, microtubules are generally curved and their Fourier transform can no longer be used for symmetry analysis.

We have proposed two sensitive contributions to automatically determine the local orientations and centers of short segments of microtubules in cryo-electron microscope images. Indeed, to our knowledge, the methods to determine the local centers of helices are not relevant for non centro-symmetric helices (e.g. 13-protofilament microtubules). The proposed algorithm exploits the helical symmetry of microtubules and the corresponding properties in the Fourier domain, so it can be easily extended to process other helical objects. Experimental results demonstrate that center locations are estimated with an accuracy of lower than one pixel. We have applied the algorithm to automatically straighten images of curved microtubules with odd numbers of protofilaments, and to improve 3D reconstructions of microtubules using back-projection methods.

6.4.6. *Segmentation of microtubules in cryo-Electron Tomograms*

Participants: Sophie Blestel, Charles Kervrann.

[In collaboration with D. Chrétien (UMR 6026, Rennes)]
Cryo-electron tomography allows 3D observation of biological specimens in their hydrated state. Generally, cryo-tomograms have very low signal-to-noise ratios, and conventional image segmentation methods yield poor results. To address this problem, we have considered the Conditional Random Fields (CRF) framework and we have formulated the segmentation problem as a maximum a posteriori estimation problem. Segmentation is obtained by computing the global minimum of a non-convex energy functional defined, in the discrete setting, as the sum of a fidelity term and of a regularization term. We define an original fidelity term robust to noise based on a distance between patches. Segmentation is performed section by section, with an automatic update of the reference patches for the 2 classes: ‘object’ and ‘background’. Because of the contrast anisotropy in the specimen thickness direction, the whole tomogram is segmented section by section, with an automatic update of reference patches. This method has been is evaluated on synthetic data and on cryo-electron tomograms of in vitro microtubules.

7. Contracts and Grants with Industry

7.1. Grant with Curie Institute on membrane transport/traffic estimation

Participants: Charles Kervrann, Thierry Pécot.

no. xxx, duration 11 months.

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

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

Participants: Patrick Pérez, Vijay Badrinarayanan.

no. Inria 2029, duration 36 months.

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

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

Participants: Patrick Pérez, Matthieu Fradet.

no. Inria 2210, duration 36 months.

This contract started in March 2007 is associated with the supervision of M. Fradet's thesis funded by a Cifre grant. It concerns the problem of semi-automatic object removal for film and television post-production. The first goal is, for a given static mask to be filled-in, to design a local motion analysis approach (estimation and segmentation) that allows the interpolation of motion information within the region and the effective filling of the region based on current and surrounding frames as well as measured and interpolated motion information. The second step will aim at combining previous tool with tracking tools developed in 7.2 in order to allow the removal of a moving object selected by the user in one or several key frames. Two applicative scenarios are especially targeted: removing of logos in TV broadcasts for business unit Thomson Grass Valley, and object removal in film post-production for business unit Technicolor. Part of this work lead to the filing of the following European patent: "Method for segmenting video segment" by M. Fradet, P. Pérez and Ph. Robert.

7.4. Cifre grant with Thomson-R&D on human pose estimation and action recognition

Participants: Ivan Laptev, Patrick Pérez.

duration 36 months

This contract started in December 2008 is associated with the supervision of R. Kumar's thesis funded by a Cifre grant. It addresses the problem of human pose estimation and human action recognition in still images. The work follows a novel data-driven approach aiming at exploiting the space of human images from many examples. Given many (video) samples of people spanning the "person-image" space, the goal is to explore the structure of this space and use it to pose constraints on the interpretation of new images. Estimation of human pose is useful e.g. in the context of graphics and video editing but also as a cue for interpretation of human actions. Human action recognition in still images is required for the search, indexing and automatic annotation of large image and video collections.

7.5. DGA 2ACI PEA: Detection, recognition and identification of targets in infra-red sequences

Participants: Florent Dutrecht, Patrick Pérez.

no. Inria 2920, duration: 20 months.

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

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

Participants: Guillaume Neveu, Patrick Pérez.

no. Inria 2666, duration: 22 months.

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

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

Participants: Ivan Laptev, Patrick Pérez.

duration 36 months.

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

8. Other Grants and Activities

8.1. National initiatives

8.1.1. CNRS-INRIA MINCyT: Reduced dynamical modeling

Participants: Patrick Heas, Etienne Mémin.

duration 36 months.

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

8.1.2. AIP INRA / ARC INRIA : DynaMIT project

Participant: Charles Kervrann.

no. xxx, duration 12 months.

This project, labeled within the ARC Inria program and contracted in January 2007, was supported by INRA in 2009. The Vistas team is the prime contractor of the project DYNAMIT which associates the following other groups: MIA (Mathématiques et Informatique Appliquées) Unit from Inra Jouy-en-Josas, Curie Institute ("Compartimentation et Dynamique Cellulaires" Laboratory, UMR CNRS-144 located in Paris). In this project, we develop new methods dedicated to the analysis of nD microscopy data and to the modeling of molecular and macromolecular mechanisms at the cell level (confocal spinning-disk microscopy, Fluorescence Recovery After Photobleaching microscopy, Fluorescence Lifetime Microscopy Imaging). Our main objective is then to provide computational methods and mathematical models to automatically extract, organize and model dynamic information observed in temporal series of images in multi-dimensional (nD) microscopy. The central problem addressed by this project concerns the roles played by different molecular motors in Rab dynamics and a rich set of data (mostly image sequences in video-microscopy) will support the analysis. The biologist partners use very powerful and novel molecular (RNA interference), mechanical (micro-patterning) and optical (FRAP, photoablation) tools to normalize and perturb the cell activity.

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

Participants: Patrick Pérez, Hervé Jégou.

duration 36 months.

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

8.1.4. QUAERO

Participants: Ivan Laptev, Patrick Pérez.

no. Inria 3184, duration 60 months

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

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" and is member of the Office of the GDR 2588.
- P. Pérez is the administrative head of the ECO-Sud project on Movie Restoration, whose participants are Telecom Paris Tech, Univ. Pierre and Marie Curie, ENS Cachan, Univ. of Monte Video.

8.2. Bilateral international co-operation

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

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

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

8.2.2. BFHZ-CCUFB / French-Bavarian collaboration

Participant: Charles Kervrann.

This bilateral collaboration project between Bavaria (Munich, Germany) and France is a one year project with the Department Biologie II - Ludwig-Maximilians Universitat Munich / H. Leonhardt, A. Dobay). In this study, we have proposed to evaluate the performance of already developed denoising algorithms when the time, exposure or gain is reduced during acquisition in fluorescence microscopy.

9. Dissemination

9.1. Leadership within scientific community

- *Editorial boards of journals*
  - P. Pérez is Associate Editor for the IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) and member of the editorial board of the International Journal of Computer Vision (IJCV).
  - I. Laptev is Associate Editor of Image and Vision Computing Journal (IVC).
- *Technical program committees of conferences*
  - P. Bouthemy: TPC member of MMM'2009, ICIAP'09, CBMI'09, CVIR'09, ICPR'10 .
  - C. Kervrann: TPC member of SSVM'2009, EMMCVPR'2009, ORASIS'2009, ISBI'2010, ICPR'2010.
  - I. Laptev: Area chair of CVPR'2009, TCP member of ICCV'2009, CVPR'2009, BMNC'2009.
  - P. Pérez: Area chair of ECCV'2010 and RFIA'2010, TPC member of CVPR'2009, ICCV'2009, NIPS'2009, EMMCVPR'2009, SIGGRAPH'2009, EUROGAPHICS'2009, SIGGRAPH ASIA'2009.
- *Ph.D. reviewing*
  - P. Pérez: J.-F. Boulanger (Univ. Aix-Marseille),
- *HdR reviewing*
  - P. Pérez: E. Debreuve (Univ. Nice Sophia Antipolis)
  - P. Bouthemy: P. Charbonnier (LCPC, Strasbourg)
- *Project reviewing, consultancy, administrative responsibilities*

  - P. Bouthemy is director of the Inria centre Rennes-Bretagne Atlantique and of Irisa. He is member of the Board of the scientific association AFRIF (Association Française pour la Reconnaissance et l'Interprétation des Formes). He is member of the Board of the scientific association GRETSI. He serves as a regular expert for the MRIS("Mission pour la Recherche et l'Innovation Scientifique") of the French Defense Agency (DGA).
  - C. Kervrann is member of the Scientific Council of the Applied Mathematics and Informatics Department of Inra since 2006. He is member of the AERES committee for CEMAGREF and member of the animation committee of the PIXEL microscopy platform at university of Rennes 1.
  - P. Pérez was vice president of the Inria-Rennes project-team committee ("Comité des projets"), deputy member of Inria evaluation board ("Commission d'évaluation") and member of the direction team of Irisa/Inria-Rennes ("Équipe de direction"). He conducted consultancy work for Bertin Technologies on detection and tracking of moving objects in unconstrained videos.

9.2. Teaching

- Master SISEA "Signal, Telecommunications, Images", University of Rennes , (P. Bouthemy: image sequence analysis, P. Pérez: visual tracking, C. Kervrann : geometric modeling for shapes and images).
- Master of Computer Science, Ifsic, University of Rennes 1 (P. Pérez: motion analysis).
- Master of Computer Science, ENSIMAG, (H. Jégou: multimedia databases).
- Master PIC and ENSPS Strasbourg (P. Bouthemy : image sequence analysis).
- ENSAI Rennes, 3rd year (C. Kervrann : statistical models and image analysis, P. Pérez : particle filtering and target tracking).
- ENS Cachan, Brittany, 1st year (P. Pérez: introduction to image processing and analysis)
- Master Machine, Vision ans Learning, Paris (I. Laptev)
- External thesis supervision :
  - C. Guilmart (secondment from "Corps de l'armement"', DGA-ONERA) supervised by P. Pérez;

9.3. Participation in seminars, invitations, awards

- C. Kervrann: Invited talk ("Analysis of intracellular trafficking in nD imaging at the single cell level") at the INRIA-NIH meeting (Rocquencourt, June 2009), INRIA-INRA meeting (Sophia-Antipolis, December 2009) and GIS Europia meeting (Rennes, December 2009).
- I. Laptev: Keynote at Sibgrapi 2009 Digital Video Journey, Rio de Janeiro, Brazil, October 2009; Invited talks at Workshop on Trends in Computer Vision, Prague, July 2009 and International Workshop on Video, Barcelona, May 2009; Outstanding review award at IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2009.
- P. Pérez: Invited talk at MIA'09, Paris (Mathematics and Image Analysis).

10. Bibliography

Major publications by the team in recent years

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

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

[3] J. BOULANGER, C. KERVRANN, P. BOUTHEMY, P. ELBAU, J.-B. SIBARITA, J. SALAMERO. *Patch-based non-local functional for denoising fluorescence microscopy image sequences*, in "IEEE Trans. on Medical Imaging", vol. 28, n^{o} 12, 2009.

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

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

[6] 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.

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

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

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

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

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

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

Year Publications

Doctoral Dissertations and Habilitation Theses

[13] V. BADRINARAYANAN. *Probabilistic graphical models for visual tracking of objects*, université de Rennes 1, Mention Informatique, avril 2009, http://www.irisa.fr/vista/Papers/2009_thesis_vijay.pdf, Ph. D. Thesis.

[14] E. DEXTER. *Modélisation de l'auto-similarité dans les vidéos : Application à la synchronisation de scènes et à la reconnaissance d'actions*, université de Rennes 1, Mention Informatique, décembre 2009, Ph. D. Thesis.

[15] A. HERVIEU. *Analyse de trajectoires vidéos à l'aide de modélisations markoviennes pour la reconnaissance de contenus*, université de Rennes 1, Mention Traitement du Signal et des Télécommunications, mars 2009, http://www.irisa.fr/vista/Papers/2009_these_hervieu.pdf, Ph. D. Thesis.

Articles in International Peer-Reviewed Journal

[16] V. AUVRAY, P. BOUTHEMY, J. LIENARD. *Joint motion estimation and layer segmentation in transparent image sequences - Application to noise reduction in X-ray image sequences*, in "EURASIP Journal on Advances in Signal Processing", 2009, http://www.irisa.fr/vista/Papers/2009_eurasip_auvray.pdf.

[17] J. BOULANGER, C. KERVRANN, P. BOUTHEMY. *A simulation and estimation framework for intracellular dynamics and trafficking in video-microscopy and fluorescence imagery*, in "Medical Image Analysis", vol. 13, 2009, p. 132-142, http://www.irisa.fr/vista/Papers/2009_MedIA_boulanger.pdf.

[18] J. BOULANGER, C. KERVRANN, P. BOUTHEMY, P. ELBAU, J.-B. SIBARITA, J. SALAMERO. *Patch-based non-local functional for denoising fluorescence microscopy image sequences*, in "IEEE Trans. on Medical Imaging", vol. 28, n^{o} 12, 2009, http://www.irisa.fr/vista/Papers/2009_TMI_Boulanger.pdf.

[19] A. BUGEAU, P. PÉREZ. *Detection and segmentation of moving objects in complex scenes*, in "Comput. Vis. Image Understanding", vol. 113, n^{o} 4, 2009, p. 459–476, http://www.irisa.fr/vista/Papers/2009_cviu_bugeau.pdf.

[20] P. COUPÉ, P. HELLIER, C. KERVRANN, C. BARILLOT. *NonLocal Means-based speckle filtering for ultrasound images*, in "IEEE Trans. on Image Processing", vol. 18, n^{o} 10, 2009, p. 2221-2229.

[21] A. CRIMINISI, T. SHARP, C. ROTHER, P. PÉREZ. *Geodesic Image and Video Editing*, in "ACM Trans. Graphics", 2009.

[22] I. JUNEJO, E. DEXTER, I. LAPTEV, P. PÉREZ. *View-Independent Action Recognition from Temporal Self-Similarities*, in "IEEE Trans. on Pattern Analysis and Machine Intelligence", 2009.

[23] I. LAPTEV. *Improving object detection with boosted histograms*, in "Image and Vision Computing", vol. 27, n^{o} 5, 2009, p. 535-544, http://www.irisa.fr/vista/Papers/2009_ivc_laptev.pdf.

Articles in National Peer-Reviewed Journal

[24] A. HERVIEU, P. BOUTHEMY, J.-P. LE CADRE. *Reconnaissance d'évènements vidéos par l'analyse de trajectoires à l'aide de modèles de Markov*, in "Traitement du Signal", vol. 26, n^{o} 3, 2009, p. 187-197, http://www.irisa.fr/vista/Papers/2009_TdS_hervieu.pdf.

International Peer-Reviewed Conference/Proceedings

[25] C. AVENEL, E. MEMIN, P. PÉREZ. *Tracking closed curves with non-linear stochastic filters*, in "Conf. on Scale Space and Variational Methods (SSVM'09), Voss, Norway", June 2009, p. 576-587, http://www.irisa.fr/vista/Papers/2009_ssvm_avenel.pdf.

[26] S. BLESTEL, C. KERVRANN, D. CHRÉTIEN. *A Fourier-based method for detecting curved microtubule centers: application to straightening of cryo-electron microscope images*, in "Proc. IEEE Int. Symp. on Biomedical Imaging: from nano to macro (ISBI'09), Boston, MA", June 2009, http://www.irisa.fr/vista/Papers/2009_isbi_blestel.pdf.

[27] A. CHESSEL, B. CINQUIN, S. BARDIN, J. SALAMERO, C. KERVRANN. *Computational geometry-based scale-space and modal image decomposition: application to light video-microscopy imaging*, in "Conf. on

Scale Space and Variational Methods (SSVM'09), Voss, Norway", June 2009, p. 770-781, http://www.irisa.fr/vista/Papers/2009_ssvm_chessel.pdf.

[28] A. CHESSEL, T. PÉCOT, S. BARDIN, C. KERVRANN, J. SALAMERO. *Evaluation of image sequences additive decomposition algorithms for membrane analysis in fluorescence video-microscopy*, in "Proc. IEEE Int. Symp. on Biomedical Imaging: from nano to macro (ISBI'09), Boston, MA", June 2009, http://www.irisa.fr/vista/Papers/2009_isbi_chessel.pdf.

[29] T. CRIVELLI, P. BOUTHEMY, B. CERNUSCHI-FRIAS, J. YAO. *Learning mixed-state Markov models for statistical motion texture tracking*, in "Proc. ICCV'09, Int. Workshop on Machine Learning for Vision-based Motion Analysis (MLVMA'09), Kyoto, Japan", October 2009.

[30] E. DEXTER, P. PÉREZ, I. LAPTEV. *Multi-view synchronization of human actions and dynamic scenes*, in "Proc. British Machine Vision Conference (BMVC'09), London, UK", September 2009.

[31] E. DEXTER, P. PÉREZ, I. LAPTEV, I. N. JUNEJO. *View-independent video synchronization from temporal self-similarities*, in "Proc. Int. Conf. on Computer Vision Theory and Applications (VISAPP'09), Lisboa, Portugal", vol. 2, February 2009, p. 383-391.

[32] O. DUCHENNE, I. LAPTEV, J. SIVIC, F. BACH, J. PONCE. *Automatic annotation of human actions in video*, in "Proc. Int. Conf. Comp. Vis.(ICCV'09), Kyoto, Japan", October 2009.

[33] M. FRADET, P. PÉREZ, P. ROBERT. *Clustering point trajectories with various life-spans*, in "Proc. Eur. Conf. on Visual Media Production (CVMP'09), London, UK", November 2009, http://www.irisa.fr/vista/Papers/2009_cvmp_fradet.pdf.

[34] A. HERVIEU, P. BOUTHEMY, J.-P. LE CADRE. *Trajectory-based handball video understanding*, in "ACM Int. Conf. on Image and Video Retrieval (CIVR'2009), Santorini, Greece", July 2009, p. 1-8, http://www.irisa.fr/vista/Papers/2009_civr_hervieu.pdf.

[35] C. KERVRANN, J. BOULANGER, T. PÉCOT, P. PÉREZ. *Discriminant random field and patch-based redundancy analysis for image change detection*, in "Proc. IEEE Int. Workshop on Machine Learning for Signal Processing (MLSP'09), Grenoble, Fance", September 2009, p. 1-6, http://www.irisa.fr/vista/Papers/2009_MLSPD_Kervrann.pdf.

[36] M. MARSZALEK, I. LAPTEV, C. SCHMID. *Actions in context*, in "Proc. Conf. Computer Vision and Pattern Recog. (CVPR'09), Miami Beach, Florida", June 2009, p. 2929-2936, http://www.irisa.fr/vista/Papers/2009_cvpr_marszalek.pdf.

[37] T. PÉCOT, A. CHESSEL, S. BARDIN, J. SALAMERO, P. BOUTHEMY, C. KERVRANN. *Conditional random fields for object and background estimation in fluorescence video-microscopy*, in "Proc. IEEE Int. Symp. on Biomedical Imaging: from nano to macro (ISBI'09), Boston, MA", June 2009, http://www.irisa.fr/vista/Papers/2009_isbi_pecot.pdf.

[38] H. SINGH, H. JEGOU. *Searching with Expectations*, in "Proc. of Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP'10), Dallas, TX", March 2010.

[39] H. WANG, M. ULLAH, A. KLÄSER, I. LAPTEV, C. SCHMID. *Evaluation of local spatio-temporal features for action recognition*, in "Proc. British Machine Vision Conf. (BMVC'09), London, UK", September 2009.

National Peer-Reviewed Conference/Proceedings

[40] E. DEXTER, P. PÉREZ, I. LAPTEV, I. JUNEJO. *Synchronisation de vidéos*, in "Proc. Journées Compression et Représentation des Signaux Audovisuels (CORESA'09), Toulouse, France", March 2009.

[41] E. DEXTER, P. PÉREZ, I. LAPTEV. *Synchronisation de séquences d'images Indépendamment des points de vue*, in "Proc. Congrès Jeunes Chercheurs en Vision par Ordinateur (ORASIS'09), Trégastel, France", June 2009.

[42] H. JÉGOU, M. DOUZE, C. SCHMID. *Représentation compacte des sacs de mots pour l'indexation d'images*, in "Reconnaissance des Formes et Intelligence Artificielle (RFIA'10)", January 2010.

Workshops without Proceedings

[43] S. BLESTEL, D. CHRÉTIEN, C. KERVRANN. *Segmentation of fibers in cryo-electron tomograms*, in "Colloque de la Société Francaise des Microscopies (SFmu'09), Paris, France", June 2009.

[44] P. GUICHARD, A. TASSIN, S. BLESTEL, C. KERVRANN, S. MARCO, D. CHRÉTIEN. *Daughter centriole morphogenesis and structural details of isolated human centrosomes revealed by cryo-electron tomography*, in "Colloque de la Société Francaise des Microscopies (SFmu'09), Paris, France", June 2009.

References in notes

[45] 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.

[46] 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.

[47] 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.

[48] 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.

[49] 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.

[50] 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.

[51] 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.

[52] 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.

[53] I. LAPTEV. *Improvements of object detection using boosted histograms*, in "Proc. British Machine Vision Conf. (BMVC'06), Edinburgh, U.-K.", September 2006, p. 949-958.

[54] 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.

[55] 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.

[56] 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.

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