Conditional Filters for Image Sequence Based Tracking - Application to Point Tracking

Élise Arnaud[1], Étienne Mémin[1] and Bruno Cernuschi-Frías[2]

***Abstract*— In this paper, a new conditional formulation of classical filtering methods is proposed. This formulation is dedicated to image sequence based tracking. These conditional filters allow solving systems whose measurements and state equation are estimated from the image data. In particular, the model that is considered for point tracking combines a state equation relying on the optical flow constraint and measurements provided by a matching technique. Based on this, two point trackers are derived. The first one is a linear tracker well-suited to image sequences exhibiting global dominant motion. This filter is determined through the use of a new estimator, called the conditional linear minimum variance estimator. The second one is a nonlinear tracker, implemented from a conditional particle filter. It allows tracking of points whose motion may be only locally described. These conditional trackers significantly improve results in some general situations. In particular, they allow dealing with noisy sequences, abrupt changes of trajectories, occlusions and cluttered background.**

***Index Terms*— point tracking, stochastic filtering, minimum variance estimator, particle filtering, optimal importance function, robust motion estimation, correlation measurement, gating**

I. Introduction

POINT tracking from an image sequence constitutes a basic but essential problem in computer vision. Many high level tasks depend on it, such as motion estimation, surveillance, video database management, robot vision control [1] or 3D reconstruction [2]. This problem, which consists in reconstructing a point trajectory along a given image sequence, is indeed inherently difficult. As a matter of fact, unlike structured shape tracking, no shape priors can be imposed, and the only possibility is to rely on a local characteristic of the point. More precisely, tracking a given point over time implies to assume that a local typical feature is invariant along its trajectory. Another difficulty concerns the setup of a prior dynamical model of the point motion, which is very difficult to establish without any *a priori* knowledge on the evolution law of the surrounding object. These intrinsic difficulties have brought researchers to implement local techniques based on geometric and luminance invariants, which locally characterize the gray value signal. The most used assumption is a constancy hypothesis for local photometric characteristic of the point and its neighborhood. This brightness constancy assumption along a trajectory has led to devise two different kinds of methods.

[1] IRISA, Université de Rennes 1, Campus de Beaulieu, Rennes, France
tel: +33 2 99 84 71 67 - fax: +33 2 99 84 71 71 - email: {Elise.Arnaud, Etienne.Memin}@irisa.fr
[2] LIPSIRN, Facultad de Ingenieria, Universidad de Buenos Aires, Argentina - email: bcernus@galileo.fi.uba.ar

The first ones are intuitive methods based on correlation criteria. Such techniques are used in numerous domains to track points but also to estimate motions of highly deformable media such as clouds in meteorological imagery or fluid flows in particles imagery [3]. An interesting comparative study of several similarity functions is described by Aschanden and Geggenbül in [4]. These methods remain very popular for their simplicity and efficiency. Nevertheless, in case of large geometric transformations (scaling, rotation, perspective distortion), illumination changes or occlusions, the efficiency of these methods decreases dramatically.

The second ones are defined as differential trackers, built from a differential formulation of a similarity criterion. In particular, a simple intensity conservation assumption leads to the optical flow constraint [5]. The well-known Shi-Tomasi-Kanade (STK) tracker [6] is derived from such a constraint which is expressed on a small neighborhood together with a spatial parameterization of the motion. However, these trackers remain sensitive to illumination changes. To solve this problem, the most common solution consists in including some photometric parameters of brightness and/or contrast [7]. Other adaptations have been suggested to improve result quality and to evaluate whether the feature is tracked successfully or not, such as the use of robust rejection rules [8].

In this paper, we propose to combine these two complementary formulations of the motion matching problem. In order to properly mix these two sources of information, we propose to set up their competition into a stochastic filtering modelization. Such a framework models the problem by a discrete hidden Markov chain, described by a system. This system consists of a *state equation* (also called *dynamic*), which characterizes the evolution law of the state to be estimated, and a *measurement equation* which links the observation to the state. The state of the filter can be the feature position together with additional information such as its velocity or its intensity template [9]. Stochastic filters give then procedures to estimate the distribution probability of the state conditionally to all past measurements. These filters, such as Kalman filter in the linear Gaussian case [10] or sequential Monte Carlo approximation methods in the nonlinear case [11] are well-known to improve tracker robustness to outliers and occlusions. To the best of our knowledge, these sequential Monte Carlo techniques have not been applied for point tracking problem.

In our case, it is important to note that the whole system describing the point tracking problem depends on the image sequence. Indeed, both dynamics and measurements are extracted from the image sequence at each discrete instant. They rely on the one hand on a differential method, and on the

other hand on a correlation criterion. The considered noise distributions depend also on the image data. As a consequence, all the elements of our filtering problem are estimated on the image sequence. Building a filtering model entirely on the image sequence is a solution to go round a lack of *a priori* information on the tracked feature.

One key-point of our work is therefore to propose a well-founded filtering framework which allows such a usual situation to be dealt with. The resulting filters are built following the traditional setup of stochastic filters, by considering a conditioning with respect to the image sequence data. Such a conditioning requires an adaptation of the usual estimators for tracking applications. This adaptation leads us to devise two kinds of trackers: a linear one and a nonlinear one. These two trackers will be described for the problem of point tracking, however, they will be dedicated to two different kinds of situations: the first one is dedicated to sequences exhibiting global dominant motion, and the second one is useful to track points whose motion can be only locally described. Each of them constitutes a robust and very reactive tracker. They both allow dealing with occlusion problems and abrupt changes of trajectories, in an elegant way.

The paper is organized as follows. After a short review on classical stochastic filtering, we present the motivations and the notations of the image sequence based filtering. Two general filters are derived, to solve linear and nonlinear systems. These two filters are then precised for point tracking application. The application's results are presented in the last section and compared to existing methods, including the well-known Shi-Tomasi-Kanade tracker [6].

II. Filtering problem : formulation and existing solutions

For the sake of clarity, the general principle of filtering problems is briefly introduced. We consider a discrete hidden Markov state process $\mathbf{x}_{0:n} = \{\mathbf{x}_0, \mathbf{x}_1, ..., \mathbf{x}_n\}$ of transition equation $p(\mathbf{x}_k|\mathbf{x}_{k-1})$. This probability distribution models the evolution of the state process. It is also known as *dynamic equation*. The set of observations $\mathbf{z}_{1:n} = \{\mathbf{z}_1, \mathbf{z}_1, ..., \mathbf{z}_n\}$, of marginal distribution $p(\mathbf{z}_k|\mathbf{x}_k)$, are supposed conditionally independent given the state sequence. The marginal distribution defines the *measurement equation*. At each discrete instant k, the filtering problem consists in having an accurate approximation of the posterior probability density of state $\mathbf{x}_k$ given the whole set of past and present measurements $\mathbf{z}_{1:k}$. A Bayesian recursive solution known as optimal filter is constituted by two interleaved steps:

- Assuming $p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1})$ known, the prediction step relying on the dynamic equation enables making a first approximation of the next state given all available information:

$$p(\mathbf{x}_k|\mathbf{z}_{1:k-1}) = \int p(\mathbf{x}_k|\mathbf{x}_{k-1})\ p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1})\ d\mathbf{x}_{k-1}.$$

- During the update state, the introduction of the new observation $\mathbf{z}_k$ corrects this first approximation using the measurement equation:

$$p(\mathbf{x}_k|\mathbf{z}_{1:k}) = \frac{p(\mathbf{z}_k|\mathbf{x}_k)\ p(\mathbf{x}_k|\mathbf{z}_{1:k-1})}{\int p(\mathbf{z}_k|\mathbf{x}_k)\ p(\mathbf{x}_k|\mathbf{z}_{1:k-1})\ d\mathbf{x}_k}.$$

Due to their huge dimension, a direct computation of these two sums can not be realized in a general case. Indeed, defining a computational formulation of the two sums constitutes the key point to solve in filtering problems.

In the case of linear Gaussian models, the Kalman filter [12] gives the optimal solution in terms of a recursive expression of mean and covariance of the Gaussian distribution $p(\mathbf{x}_k|\mathbf{z}_{1:k})$. Such an expression may be derived from the *minimum variance estimator*. This estimator is equivalent to the *best linear estimator* in the Gaussian case, and then corresponds to the conditional expectation of the state at time k given the set of observations $E[\mathbf{x}_k|\mathbf{z}_{1:k}]$. In the absence of Gaussian assumption, but keeping the linear properties of the model, a reasonable choice consists in relying on the *linear minimum variance estimator* of $\mathbf{x}_k$ given $\mathbf{z}_{1:k}$. The use of this estimator leads finally to the same equations as the classical Kalman filter. Nevertheless, such an estimator provides only first and second order statistics of $p(\mathbf{x}_k|\mathbf{z}_{1:k})$, and does not provide other than incomplete information on higher order moments [13].

Similarly, in the nonlinear case, the extended Kalman filter leads also to an estimation of the two first moments of the required posterior density. This non-optimal solution is derived from a local first-order linearization, which is not satisfactory facing multi-modality. Other approaches named grid-based methods propose to build a determinist mesh of the state space in order to obtain numerical estimations of the optimal filter integrals [14]. The optimal solution is reached if the state space is discrete and consists of a finite number of states. An alternative of these highly computational algorithms lies in the use of sequential Monte Carlo filters, also called particle filters [15], [16]. These approaches present the interest of not requiring to linearize the equations of the system. A representation of $p(\mathbf{x}_k|\mathbf{z}_{1:k})$ is then given in terms of a finite weighted sum of Diracs centered in elements of the state space named *particles*. The associated weights are chosen using the importance sampling principle. The swarm of weighted particles is updated recursively.

A non-exhaustive list of existing solutions for filtering problems has been briefly presented in this section. More details on algorithms and applications (particularly concerning Monte Carlo algorithms) can be found in the book edited by Doucet et al. [17].

III. Image sequence based filtering

A. Motivations

In our point of view, tracking features from image sequences may require, in some cases, to define a slightly modified framework of stochastic filtering. A first problem comes from the choice of the observation model. As a matter of fact, the measurement on which one should ideally rely is the image sequence itself. Unfortunately, images have too large dimensions and too complex structures to be used directly.

Therefore, one usually defines a digest structured observation built from the images. Different kind of data can be taken into account, such as motion [18], intensity level, color histogram [19], information from gradients [20], etc. It is important to outline that the extraction of such a measurement is usually done through a potentially highly nonlinear function with respect to image pixels.

Another source of difficulties comes from the definition of appropriate dynamic models. These dynamics, such as auto-regressive models, are usually defined *a priori*. Such dynamics are used for instance in [21], [19], for human tracking. State equations are also frequently obtained by learning, as in the Condensation algorithm. This method is used to track curves in dense visual clutter [20], or to follow a drawing action of a hand holding a pen [22]. Another example can be found in [23], where Sidenbladh et al. propose to build a database of motions to define probabilistic models in order to track 3D human motions. A severe limitation of these models arises facing the tracking of features whose trajectories exhibit abrupt changes and occlusions or simply obey too complex dynamic laws, which can hardly be learned or predicted. Indeed, a known state model is not always available [24]. This is particularly the case when tracking very general punctual entities in images of any kinds. To avoid this problem, we propose an original construction of the dynamic in this paper. We state that in such a context, one possibility consists in relying on a dynamical model extracted from the image sequence. Such a dynamic - which may be related to a spatial representation of the motion (affine, quadratic and so on) - has the advantage of introducing a contextual prior on the point motion in a simple way. Thus, it enables us to extract the velocity of the surrounding object without any knowledge on its nature (rigid, fluid or deformable).

Nevertheless, we now have to face a tracking problem for which the whole system (measurement, observation model and state model) depends on the images. For such a peculiar case, we propose here a modified formulation of classical stochastic filtering approaches. These filters are extensions of classical filters. They are named *conditional* filters as they include a conditioning with respect to the image sequence. This constitutes in some way a generalization of traditional trackers. Such *a priori*-free models but also systems built with classical dynamics can be solved in the conditional framework we propose.

Based on this idea, we propose here two different filters. The first one is a linear filter built from an estimator derived from the linear minimum variance estimator. We have called such an estimator the *conditional linear minimum variance estimator* as it includes a conditioning with respect to the image sequence. The second filter is nonlinear and implemented through a particle filter. This latter filter relies on the optimal importance sampling function. These two trackers have been applied to point tracking in image sequence.

B. Notations

For the sake of clarity let us first define the notations used in this paper. Let $\mathbf{I}_k$ denote a random variable which corresponds to an image obtained at time k. The finite sequence of variables $\{\mathbf{I}_k, k = 0, ..., n\}$ will be represented by $\mathbf{I}_{0:n}$. Knowing a realization of $\mathbf{I}_{0:k}$, the image based tracking system is modeled by the following dynamic:

$$\mathbf{x}_k = f_k^{\mathbf{I}_{0:k}}(\mathbf{x}_{k-1}, \mathbf{w}_k^{\mathbf{I}_{0:k}}),$$

associated to a observation equation defined as:

$$\mathbf{z}_k = h_k^{\mathbf{I}_{0:k}}(\mathbf{x}_k, \mathbf{v}_k^{\mathbf{I}_{0:k}}).$$

At each time k, a realization of $\mathbf{z}_k$ is obtained as the result of an estimation process applied to the image sequence $\mathbf{I}_{0:k}$. Functions $f_k^{\mathbf{I}_{0:k}}$ and $h_k^{\mathbf{I}_{0:k}}$ are assumed to be any kind of possibly nonlinear functions. These functions may be specified from $\mathbf{I}_{0:k}$. The state noise $\mathbf{w}_k^{\mathbf{I}_{0:k}}$ and the measurement noise $\mathbf{v}_k^{\mathbf{I}_{0:k}}$ may be specified as well from $\mathbf{I}_{0:k}$, and are not necessarily Gaussian. Their distributions may depend in particular on the kind of estimation processes used to define measurements and dynamics. The involved probability distributions are such that:

$$\begin{aligned} p(\mathbf{x}_k|\mathbf{x}_{0:k-1}, \mathbf{z}_{1:k-1}, \mathbf{I}_{0:n}) &= p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{I}_{0:n}), \\ p(\mathbf{z}_k|\mathbf{x}_{0:k}, \mathbf{z}_{1:k-1}, \mathbf{I}_{0:n}) &= p(\mathbf{z}_k|\mathbf{x}_k, \mathbf{I}_{0:n}). \end{aligned}$$

By analogy with the classical filtering problem, conditionally to the sequence, the Markovian assumption, as well as the conditional independence of the observations are maintained. A causal hypothesis with respect to the temporal image acquisition is added. Such an hypothesis means that the state $\mathbf{x}_k$ and the measurement $\mathbf{z}_k$ are assumed to be independent from $\mathbf{I}_{k+1:n}$. To clarify the proposed conditional model, figures 1 and 2 present the different oriented dependency graphs associated to classical filters and to conditional filters.

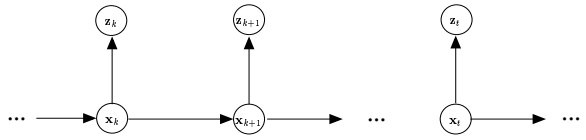


Fig. 1. Oriented dependency graph of the hidden Markov chain of processes $\mathbf{x}_{0:n}$ and $\mathbf{z}_{1:n}$ (classical filtering problem).

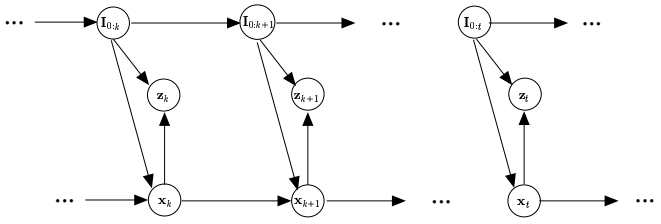


Fig. 2. Oriented dependency graph of variables $\mathbf{x}_{0:n}$, $\mathbf{z}_{1:n}$ and $\mathbf{I}_{0:n}$ associated to conditional filters.

Including a conditioning with respect to the image sequence, the optimal filter's equations can be applied to the proposed model. The posterior probability distribution of the state $\mathbf{x}_k$ given all available information, reads now $p(\mathbf{x}_k|\mathbf{z}_{1:k}, \mathbf{I}_{0:k})$. Assuming $p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1}, \mathbf{I}_{0:k-1})$ known, the two steps corresponding to a Bayesian optimal recursive solution can be

given immediately as:

$$p(\mathbf{x}_k|\mathbf{z}_{1:k-1},\mathbf{I}_{0:k}) = \int p(\mathbf{x}_k|\mathbf{x}_{k-1},\mathbf{I}_{0:k})\ p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1},\mathbf{I}_{0:k-1})\ d\mathbf{x}_{k-1},$$

$$p(\mathbf{x}_k|\mathbf{z}_{1:k},\mathbf{I}_{0:k}) = \frac{p(\mathbf{z}_k|\mathbf{x}_k,\mathbf{I}_{0:k})\ p(\mathbf{x}_k|\mathbf{z}_{1:k-1},\mathbf{I}_{0:k})}{\int p(\mathbf{z}_k|\mathbf{x}_k,\mathbf{I}_{0:k})\ p(\mathbf{x}_k|\mathbf{z}_{1:k-1},\mathbf{I}_{0:k})\ d\mathbf{x}_k}.$$

To solve this conditional tracking problem, the standard filters have to be derived to be conditional upon the image data. Two cases of this problem will be detailed. The following section presents a Conditional Linear Filter (CLF) for linear image based filtering problems. This case corresponds to systems for which functions f and h are linear with respect to the state. The resulting filter adapted to the point tracking problem will be devoted to features whose motion roughly corresponds to the global dominant motion detected in the image sequence. Section V will focus on the nonlinear case. The corresponding nonlinear filter will allow us to consider points for which the previous assumption of dominant motion dynamic is not verified. We will show in the experimental section that a combination of these two filters will enable us to track any point of any sequence respecting some kind of brightness constancy assumption.

IV. Conditional Linear Filter

The linear image based filtering problem can be modeled by the following system:

$$\left\{ \begin{array}{l} \mathbf{x}_k = A_k^{\mathbf{I}_{0:k}}\mathbf{x}_{k-1} + \mathbf{b}_k^{\mathbf{I}_{0:k}} + \mathbf{w}_k^{\mathbf{I}_{0:k}} \\ \mathbf{z}_k = H_k^{\mathbf{I}_{0:k}}\mathbf{x}_k + \mathbf{v}_k^{\mathbf{I}_{0:k}} \end{array} \right. \quad (1)$$

The index $\mathbf{I}_{0:k}$ indicates a possible dependence on the image data. Let us remind that, in our case, matrices $A_k^{\mathbf{I}_{0:k}}$, $H_k^{\mathbf{I}_{0:k}}$ and vector $\mathbf{b}_k^{\mathbf{I}_{0:k}}$ may be estimated from $\mathbf{I}_{0:k}$. Variables $\mathbf{w}_k^{\mathbf{I}_{0:k}}$, $\mathbf{v}_k^{\mathbf{I}_{0:k}}$ are supposed to be zero mean independent white noises (conditionally to the image sequence), possibly non Gaussian, of known conditional covariances denoted $Q_k^{\mathbf{I}_{0:k}}$ and $R_k^{\mathbf{I}_{0:k}}$ respectively. These covariances depend in particular on the accuracy of the estimation processes from which $A_k^{\mathbf{I}_{0:k}}$, $H_k^{\mathbf{I}_{0:k}}$ and $\mathbf{b}_k^{\mathbf{I}_{0:k}}$ have been computed. In order to tackle the problem on non-Gaussian noises, the Conditional Linear Filter is derived through an extension of the linear minimum variance estimator. This estimator, which we call the *conditional linear minimum variance estimator*, provides an estimation of the two first moments of $p(\mathbf{x}_k|\mathbf{z}_{1:k},\mathbf{I}_{0:k})$. This description is obviously sufficient to have the entire knowledge of the expected density if the linear model is Gaussian. For non-Gaussian noises, this description provides only a Gaussian approximation of the posterior density function.

A. Conditional linear minimum variance estimator

As already said, the Conditional Linear Filter is built by relying on a conditional linear minimum variance estimator. Let us first introduce this estimator and its properties.

Definition 1: Let X,Z,W be 3 jointly distributed random variables. $E^*_W[X|Z]$ denotes the best estimator of X, linear in Z, conditionally to W:

$$E^*_W[X|Z] = AZ + B$$

with A and B such that $E[\|X - AZ - B\|^2|W]$ is minimum. $E^*_W[X|Z]$ is called the conditional linear minimum variance estimator.

It must be noticed that $E^*_W[X|Z]$ is not an expectation. Denoting $\Sigma_{X,Z|W} = E[XZ^t|W] - E[X|W]E[Z|W]^t$, the following important result is obtained (see appendix I):

$$E^*_W[X|Z] = E[X|W] + \Sigma_{X,Z|W}\Sigma^{-1}_{Z,Z|W}(Z - E[Z|W]). \quad (2)$$

It can be checked that this estimator shares some similar properties to the linear minimum variance estimator. The main properties that are involved in the elaboration of the Conditional Linear Filter are listed below.

Theorem 1 (property of being unbiased): Let X,Z,W be jointly distributed random variables, then

$$E[X - E^*_W[X|Z]|W] = \mathbf{0}.$$

Theorem 2 (uncorrelated conditioning quantities): Let X, Z_1, Z_2, ..., Z_k, W be jointly distributed random variables with Z_1, Z_2, ..., Z_k uncorrelated conditionally to W, then

$$E^*_W[X|Z_{1:k}] = \sum_{i=1}^{k} E^*_W[X|Z_i] - (k-1)\ E[X|W].$$

Theorem 3 (change of conditioning variables): Let X,Z,W be jointly distributed random variables, and let $M = C\ Z + D$ (with C not singular), then

$$E^*_W[X|Z] = E^*_W[X|M].$$

Theorem 4 (orthogonality principle): Let X,Z,W be jointly distributed random variables, then

$$E[(X - E^*_W[X|Z])\ Z^t|W] = 0.$$

B. Tracking with Conditional Linear Filter

We consider a linear model of the form described in (1). To simplify the notations, the index $\mathbf{I}_{0:k}$ will be omitted in the following of this section. Let us denote $\hat{\mathbf{x}}_{k+1|k} = E^*_{\mathbf{I}_{0:k+1}}[\mathbf{x}_{k+1}|\mathbf{z}_{1:k}]$ and $\Sigma_{k+1|k}$ the associated conditional error covariance. Considering conditional expressions induced by E^*, and relying on the previously listed properties of E^*, a recursive formulation of $\hat{\mathbf{x}}_{k+1|k}$ can be found through similar manipulations to the usual Gaussian case:

$$\hat{\mathbf{x}}_{k+1|k} = A_{k+1}\ \hat{\mathbf{x}}_{k|k-1} + \mathbf{b}_{k+1} + \widetilde{K}_k\ (\mathbf{z}_k - H_k\ \hat{\mathbf{x}}_{k|k-1}),$$

where matrix $\widetilde{K}_k$ is defined using the Kalman gain K_k:

$$\begin{aligned} \widetilde{K}_k &= A_{k+1}K_k \\ &= A_{k+1}(\Sigma_{k|k-1}H_k^t)(H_k\Sigma_{k|k-1}H_k^t + R_k)^{-1}. \end{aligned}$$

A recursive expression of the conditional estimation error covariance $\Sigma_{k+1|k}$ can also be obtained as:

$$\begin{aligned} \Sigma_{k+1|k} = (A_{k+1} - \widetilde{K}_k\ H_k)\ \Sigma_{k|k-1}\ (A_{k+1} - \widetilde{K}_k\ H_k)^t \\ + Q_{k+1} + \widetilde{K}_k\ R_k\ \widetilde{K}_k^t. \end{aligned}$$

These equations can be further split to distinguish the prediction step and the update step, as in the block diagram depicted in figure 3.

Validation gate: In order to limit the computational cost, it may be useful to define a research area where the estimation process of the measurement $\mathbf{z}_k$ is applied. Such a region, called validation region or gate, is defined as an area of the measurement space where the future observation will be found with some high probability. Gates are generally used in radar tracking problems, for clutter reduction [25]. They are here defined through the use of the probability distribution $p(\mathbf{z}_k|\mathbf{z}_{1:k-1}, \mathbf{I}_{0:k})$. In image sequence based filtering, this measurement prediction region usually defines a part of the image where the future observation has to be looked for. The probability of the measurement given all past observations and the image sequence is approximated by a normal distribution (this expression is exact in case of Gaussian noises):

$$p(\mathbf{z}_k|\mathbf{z}_{1:k-1}, \mathbf{I}_{0:k}) = \mathcal{N}(\hat{\mathbf{z}}_{k|k-1}, S_k),$$

where $\hat{\mathbf{z}}_{k|k-1} = H_k \hat{\mathbf{x}}_{k|k-1}$ represents the predictive measurement and $S_k = H_k \Sigma_{k|k-1} H_k^t + R_k$ the conditional covariance given the sequence of innovations $\widetilde{\mathbf{z}}_{k|k-1} = \mathbf{z}_k - \hat{\mathbf{z}}_{k|k-1}$. An ellipsoidal probability concentration region is then defined as:

$$gate_k = \{\mathbf{z}_k : \epsilon_k = \widetilde{\mathbf{z}}_{k|k-1}^t S_k^{-1} \widetilde{\mathbf{z}}_{k|k-1} \leq \gamma\}. \qquad (3)$$

Assuming (3), the distance ϵ_k is distributed according to a Chi-square law, of p degrees of freedom, where p is the measurement vector dimension. It is important to outline that the validation gate classically depends on the estimated error covariance but also on the image sequence through R_k and H_k.

The Conditional Linear Filter is synthesized in figure 3. The double boxes represent steps processed from the image sequence. The resulting filter constitutes a tracker resembling

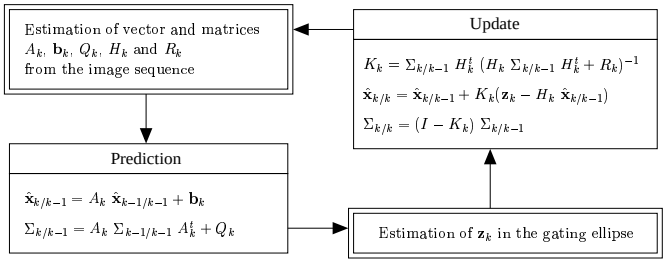


Fig. 3. Block diagram of the Conditional Linear Filter.

the Kalman filter for Gaussian linear models. It is nevertheless important to note that (i) the use of Kalman recursive equations are now well justified for this specific case through the use of a conditional minimum variance estimator and (ii) it provides a sound framework which enables a probabilistic competition between two estimation processes on the image sequence. Before precising such a filter for point tracking application, let us present a conditional filter devoted to the nonlinear case.

V. Conditional NonLinear Filter

A. General case

When we have to face a system with a nonlinear dynamic and/or a nonlinear measurement equation, it is not possible to construct an exact recursive expression of the conditional expectation of the system at time k given all available information. To overcome these computational difficulties, particle filtering techniques propose to implement recursively an approximation of the posterior density function (see [15], [16] for an extended review). Such a technique consists in propagating a swarm of N particles. Each particle $\mathbf{x}_{0:k}^{(i)}$ corresponds to a feasible trajectory of the initial system $\mathbf{x}_0$. A weight assigned to each trajectory is computed from the likelihood of observations up to k. The optimal trajectory is obtained by weighting the particles swarm.

As in the linear case, we now focus on the construction of a particle filter adapted to the general class of image sequence based models. As previously, a conditional algorithm is derived from the classical one. The possible dependence on the sequence is taken into account through a conditioning with respect to the image sequence data. We give below the corresponding equations.

Assuming the knowledge of distributions $p(\mathbf{x}_0|\mathbf{I}_0)$, $p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{I}_{0:k})$ and $p(\mathbf{z}_k|\mathbf{x}_k, \mathbf{I}_{0:k})$ $\forall k \geq 1$, we are looking for the best estimate of a point trajectory $\mathbf{x}_{0:k}$ given the image data and all the measurements. We want to estimate the conditional expectation:

$$T = E[\mathbf{x}_{0:k}|\mathbf{z}_{1:k}, \mathbf{I}_{0:k}] = \int \mathbf{x}_{0:k}\; p(\mathbf{x}_{0:k}|\mathbf{z}_{1:k}, \mathbf{I}_{0:k})\; d\mathbf{x}_{0:k}.$$

Such an integral is impossible to compute because of its dimension. The use of importance sampling allows to approximate this integral by introducing a probability distribution $\pi(\mathbf{x}_{0:k}|\mathbf{z}_{1:k}, \mathbf{I}_{0:k})$, from which one can easily sample. This distribution is called the *importance function*. Drawing a set $\{\mathbf{x}_{0:k}^{(i)}\}$ of N i.i.d samples according to the importance function, the knowledge of N associated normalized weights $\widetilde{w}_k^{(i)}$ allows approximating the conditional expectation by a finite summation:

$$\hat{T} = \frac{1}{N} \sum_{i=1}^{N} \mathbf{x}_{0:k}^{(i)}\; \widetilde{w}_k^{(i)}.$$

The non-normalized weights are given by:

$$w_k = \frac{p(\mathbf{z}_{1:k}|\mathbf{x}_{0:k}, \mathbf{I}_{0:k})\; p(\mathbf{x}_{0:k}|\mathbf{I}_{0:k})}{\pi(\mathbf{x}_{0:k}|\mathbf{z}_{1:k}, \mathbf{I}_{0:k})}.$$

In order to construct a recursive expression of the conditional expectation, a recursive expression of the importance function is assumed. This formulation also introduces the causality on image data:

$$\begin{aligned} &\pi(\mathbf{x}_{0:k}|\mathbf{z}_{1:k}, \mathbf{I}_{0:k}) \\ &= \pi(\mathbf{x}_{0:k-1}|\mathbf{z}_{1:k-1}, \mathbf{I}_{0:k-1})\; \pi(\mathbf{x}_k|\mathbf{x}_{0:k-1}, \mathbf{z}_{1:k}, \mathbf{I}_{0:k}). \end{aligned} \qquad (4)$$

Such an expression leads to a recursive equation for the weights:

$$w_k = w_{k-1}\; \frac{p(\mathbf{z}_k|\mathbf{x}_k, I_{0:k})\; p(\mathbf{x}_k|\mathbf{x}_{k-1}, I_{0:k})}{\pi(\mathbf{x}_k|\mathbf{x}_{0:k-1}, \mathbf{z}_{1:k}, \mathbf{I}_{0:k})}.$$

Nevertheless, it also induces an increase of the weight variance over time [26]. Consequently, in practice, the number of significant particles decreases dramatically over time. To limit such a degeneracy, two methods have been proposed. They are presented here in the framework of the Conditional NonLinear Filter.

A first solution consists in selecting an *optimal* importance function which minimizes the variance of the importance weights w_k conditional upon $\mathbf{x}_{0:k-1}$, $\mathbf{z}_{1:k}$ and $\mathbf{I}_{0:k}$. It is then possible to demonstrate that the optimal importance function is $p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{z}_k, \mathbf{I}_{0:k})$, and leads to a new recursive formulation of w_k:

$$w_k = w_{k-1}\ p(\mathbf{z}_k|\mathbf{x}_{k-1}, \mathbf{I}_{0:k}). \tag{5}$$

The problem with this approach is related to the fact that it requires to be able to sample from the optimal importance function $p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{z}_k, \mathbf{I}_{0:k})$, and to have an expression of $p(\mathbf{z}_k|\mathbf{x}_{k-1}, \mathbf{I}_{0:k})$. Let us remark that usually in vision applications the importance function is not known and is identified to the diffusion process (i.e. $\pi(\mathbf{x}_k|\mathbf{x}_{0:k-1}, \mathbf{z}_{1:k}) = p(\mathbf{x}_k|\mathbf{x}_{k-1})$) [11], [27]. Such a choice excludes the measurements from the diffusion step. It is thus necessary in that case to rely on an accurate dynamic model to efficiently sample particles. We will see in our point tracking application that a different choice is possible. A second solution to handle the problem of weight variance increase relies on a resampling method [28]. This method consists in removing the trajectories with weak normalized weights, and in adding copies of the trajectories associated to strong weights. Nevertheless it is important to outline that the resampling step introduces errors, since it increases the Monte Carlo variance of the estimate. As a consequence, the resampling step is necessary in practice, but should be used as rarely as possible. Obviously, these two solutions may be coupled for a better efficiency.

If the importance density only depends on $\mathbf{x}_{k-1}, \mathbf{z}_k, \mathbf{I}_{0:k}$, the first moment of $p(\mathbf{x}_k|\mathbf{z}_{1:k}, \mathbf{I}_{0:k})$ can be approximated in the same way as $E[\mathbf{x}_{0:k}|\mathbf{z}_{1:k}, \mathbf{I}_{0:k}]$. Indeed, this approximation, denoted $\hat{\mathbf{x}}_{k|k}$, can be shown to be:

$$\hat{\mathbf{x}}_{k|k} = \frac{1}{N}\sum_{i=1}^{N} \mathbf{x}_k^{(i)}\ \widetilde{w}_k^{(i)}.$$

As mentioned previously, it may be beneficial to know the expression of the optimal importance function. We will see in the next section that it is possible to infer this function for a specific class of models.

B. Conditional partial Gaussian state space models

Let us consider a conditional nonlinear system, composed of a nonlinear state equation, with an additive Gaussian noise:

$$\mathbf{x}_k = f_k^{\mathbf{I}_{0:k}}(\mathbf{x}_{k-1}) + \mathbf{w}_k^{\mathbf{I}_{0:k}}, \quad \mathbf{w}_k^{\mathbf{I}_{0:k}} \rightsquigarrow \mathcal{N}(\mathbf{0}, Q_k^{\mathbf{I}_{0:k}}) \tag{6}$$

and a linear Gaussian measurement equation:

$$\mathbf{z}_k = H_k^{\mathbf{I}_{0:k}}\mathbf{x}_k + \mathbf{v}_k^{\mathbf{I}_{0:k}}, \quad \mathbf{v}_k^{\mathbf{I}_{0:k}} \rightsquigarrow \mathcal{N}(\mathbf{0}, R_k^{\mathbf{I}_{0:k}}). \tag{7}$$

We will denote these systems *conditional partial Gaussian state space models*. For the sake of clarity, we will again omit the index $\mathbf{I}_{0:k}$. For this class of systems, the analytic expression of the optimal importance function is known. As a matter of fact, noticing that:

$$p(\mathbf{z}_k|\mathbf{x}_{k-1}, \mathbf{I}_{0:k}) = \int p(\mathbf{z}_k|\mathbf{x}_k, \mathbf{I}_{0:k})\ p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{I}_{0:k})\ d\mathbf{x}_k,$$

we deduce:

$$p(\mathbf{z}_k|\mathbf{x}_{k-1}, \mathbf{I}_{0:k}) = \mathcal{N}(H_k f_k(\mathbf{x}_{k-1}), R_k + H_k Q_k H_k^t), \tag{8}$$

which yields a simple tractable expression for the weight calculation (5). Then we have:

$$p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{z}_k, \mathbf{I}_{0:k}) = \frac{p(\mathbf{z}_k|\mathbf{x}_k, \mathbf{I}_{0:k})\ p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{I}_{0:k})}{p(\mathbf{z}_k|\mathbf{x}_{k-1}, \mathbf{I}_{0:k})}$$

and thus,

$$p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{z}_k, \mathbf{I}_{0:k}) = \mathcal{N}(\mathbf{m}_k(\mathbf{x}_{k-1}), C_k), \tag{9}$$

with

$$C_k = (Q_k^{-1} + H_k^t R_k^{-1} H_k)^{-1}$$
$$\mathbf{m}_k(\mathbf{x}_{k-1}) = C_k(Q_k^{-1}\ f_k(\mathbf{x}_{k-1}) + H_k^t R_k^{-1}\ \mathbf{z}_k).$$

In that specific case, all the expressions used in the diffusion process (9), and in the update step (8) are Gaussian. Let us remark that the unconditional version of this result is described in [16]. The Conditional NonLinear Filter corresponding to this class of specific models is therefore particularly simple to implement.

Validation gate: As in the linear case, an important issue is the definition of a validation gate corresponding to a research area for the measurement at time k. As seen previously, in the linear Gaussian case, an analytic expression of $p(\mathbf{z}_k|\mathbf{z}_{1:k-1}, \mathbf{I}_{0:k})$ may be obtained. For nonlinear models, the validation gate may be approximated by a rectangular or an ellipsoidal region, whose parameters may be complex to define. Breidt [29] suggests to use Monte Carlo simulations in order to approximate the required density but this solution appears to be time consuming. In case of conditional partial Gaussian state space model, it is possible to infer an ellipsoidal validation gate in a simple way. In order to define such a validation region, we have to compute the two first moments of $p(\mathbf{z}_k|\mathbf{z}_{1:k-1}, \mathbf{I}_{0:k})$. Empirical approximations of these quantities can be easily derived considering the specific form of the model. As a matter of fact, observing that

$$p(\mathbf{z}_k|\mathbf{z}_{1:k-1}, \mathbf{I}_{0:k})$$
$$= \int p(\mathbf{z}_k|\mathbf{x}_{k-1}, \mathbf{I}_{0:k})\ p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1}, \mathbf{I}_{0:k-1})\ d\mathbf{x}_{k-1},$$

and reminding that an approximation of $p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1}, \mathbf{I}_{0:k-1})$ is given by the weighted swarm of particles $(\mathbf{x}_{k-1}^{(i)}, \widetilde{w}_{k-1}^{(i)})$, the following approximation can be done:

$$p(\mathbf{z}_k|\mathbf{z}_{1:k-1}, \mathbf{I}_{0:k}) \simeq \sum_i \widetilde{w}_{k-1}^{(i)}\ p(\mathbf{z}_k|\mathbf{x}_{k-1}^{(i)}, \mathbf{I}_{0:k}).$$

Through expression (8), and after few simple calculations, we finally obtain a Gaussian distribution for $p(\mathbf{z}_k|\mathbf{z}_{1:k-1}, \mathbf{I}_{0:k})$:

$$p(\mathbf{z}_k|\mathbf{z}_{1:k-1}, \mathbf{I}_{0:k}) \simeq \mathcal{N}(E_{k|k-1}, V_{k|k-1}), \tag{10}$$

with

$$E_{k|k-1} = E[\mathbf{z}_k|\mathbf{z}_{1:k-1}, \mathbf{I}_{0:k}] = \sum_i \widetilde{w}_{k-1}^{(i)}\ H_k\ f_k(\mathbf{x}_{k-1}^{(i)}),$$

and

$$\begin{aligned} V_{k|k-1} &= V[\mathbf{z}_k|\mathbf{z}_{1:k-1}, \mathbf{I}_{0:k}] \\ &= \textstyle\sum_i \widetilde{w}_{k-1}^{(i)} \ [H_k \ Q_k^{(i)} \ H_k^t + R_k \\ &+ H_k \ f_k(\mathbf{x}_{k-1}^{(i)}) \ f_k^t(\mathbf{x}_{k-1}^{(i)}) \ H_k^t] \\ &- (\textstyle\sum_i \widetilde{w}_{k-1}^{(i)} \ H_k \ f_k(\mathbf{x}_{k-1}^{(i)})) \ (\textstyle\sum_i \widetilde{w}_{k-1}^{(i)} \ H_k \ f_k(\mathbf{x}_{k-1}^{(i)}))^t. \end{aligned}$$

This gives us an expression of the ellipsoidal region corresponding to our validation gate at time k:

$$gate_k = \{\mathbf{z}_k : \epsilon_k = (\mathbf{z}_k - E_{k|k-1})^t V_{k|k-1}^{-1} (\mathbf{z}_k - E_{k|k-1}) \leq \gamma\}. \tag{11}$$

In addition to a simple and optimal sampling process, being able to build a validation gate constitutes another advantage of conditional partial Gaussian state space models.

A synopsis of the resulting Conditional NonLinear Filter is depicted in figure 4.

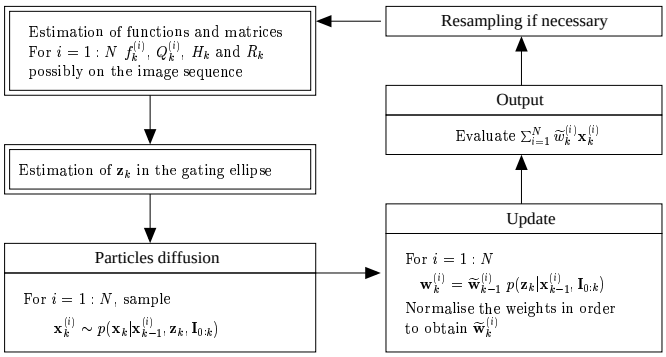


Fig. 4. Conditional NonLinear Filter, with the use of optimal importance function.

The modeling of a problem with conditional partial Gaussian state space models of the form (6-7) requires a linear measurement equation. We believe that such a choice - which consists in constraining the system to rely on a simple and rough linear measurement model - can be compensated by a sound and accurate dynamic model. At the best of our knowledge, tracking applications in computer vision, based on stochastic filtering approaches, rely on the opposite choice: a linear dynamic model is associated to a nonlinear multi-modal likelihood [19], [20], [30], [31]. We argue that a conditional partial Gaussian state space model constitutes an interesting alternative, as its properties induce a very simple filter implementation. We believe that an accurate dynamic, a pertinent estimation of the measurement noise covariance and the use of the optimal importance function allows counterbalancing a multi-modal likelihood even for tracking applications in cluttered environment. We will demonstrate these abilities in the experimental section on several real-world image sequences. The next section develops more precisely the proposed point trackers.

VI. Application to point tracking in image sequence

Let us considering a given point in the scene. The problem of tracking this feature in an image sequence can be defined as locating an estimation of the point projection in the image plane at each time. We consider the most general context, where no knowledge on the dynamic of the surrounding object is available. As said before, the solution we propose to tackle the lack of *a priori* information consists in computing the model from the image sequence and solve the system with one of the previously proposed conditional filter. In the considered tracking problem, each state $\mathbf{x}_k$ represents the location of the point projection at time k, in image $\mathbf{I}_k$, observable through the measurement $\mathbf{z}_k$. Let us point out that for the kind of system we focus on, both measurements and the dynamic equation are built from $\mathbf{I}_{0:k}$. Indeed, as motivated in the introduction, we combine a dynamic model relying on the optical flow constraint and measurements provided by a matching technique.

A. Conditional dynamic equation

The motion of a point $\mathbf{x}_{k-1}$ between the frame instants $k-1$ and k is defined through the probability distribution $p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{I}_{0:k})$. In order to be reactive to any change of speed and direction of the point, we propose to define such a state equation from a robust parametric estimation technique [32], [33]. At each time, this technique provides us the velocity of the feature of interest.

Robust motion estimation using parametric model: A robust parametric motion estimation technique enables to estimate reliably a 2D parametric model representing the dominant apparent velocity field on a given support region $\mathcal{R}$. The motion vector of a point $\mathbf{s}$, between time $k-1$ and time k is modeled as a polynomial function of the point coordinates:

$$\mathbf{u}(\mathbf{s}) = P(\mathbf{s}) \ \boldsymbol{\theta}_k$$

where $\mathbf{u}(\mathbf{s})$ denotes the estimated motion vector of pixel $\mathbf{s} = (x, y)^t$ and $\boldsymbol{\theta}_k$ the parameter vector which contains the polynomial's coefficients. $P(\mathbf{s})$ is a matrix related to the chosen parametric model whose entries depend on the spatial coordinates x and y. For example, a 6-parameter affine motion model is associated to the following matrix:

$$P(\mathbf{s}) = \begin{bmatrix} 1 & x & y & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & x & y \end{bmatrix}.$$

The parameter vector $\boldsymbol{\theta}_k$ is estimated through the minimization of a robust function $\phi_1(\boldsymbol{\theta})$:

$$\phi_1(\boldsymbol{\theta}) = \sum_{\mathbf{s} \in \mathcal{R}} \rho \ [\mathbf{I}_k(\mathbf{s} + P(\mathbf{s})\boldsymbol{\theta}) - \mathbf{I}_{k-1}(\mathbf{s})]. \tag{12}$$

ρ is a non-quadratic robust cost function allowing dealing with outliers. These outliers may be identified as points or areas that do not correspond to the estimated motion model or as regions for which the brightness consistency constraint is not valid. The minimization is achieved through a Gauss-Newton-type multi-resolution procedure, which allows handling of large magnitude motion. The principle of the incremental scheme consists in applying successive Taylor expansions of the argument of ρ around the current estimates $\hat{\boldsymbol{\theta}}$. The estimation of

the motion parameter increment $\delta\boldsymbol{\theta}$ is then obtained through the minimization of:

$$\begin{aligned}\phi_2(\delta\boldsymbol{\theta}) = \sum_{\mathbf{s}\in\mathcal{R}} \rho\ [&\mathbf{I}_k(\mathbf{s}+P(\mathbf{s})\hat{\boldsymbol{\theta}}) \\ &- \mathbf{I}_{k-1}(\mathbf{s}) + \nabla\mathbf{I}_k(\mathbf{s}+P(\mathbf{s})\hat{\boldsymbol{\theta}})^t P(\mathbf{s})\delta\boldsymbol{\theta}].\end{aligned} \tag{13}$$

The minimization is embedded within a coarse to fine strategy and performed through an iterated least squares technique at each level (see [33] for more details on the approach).

Linear or nonlinear state equation: For our point tracking application, such a robust parametric motion estimation technique allows us to define the following state equation:

$$\begin{aligned}\mathbf{x}_k &= \mathbf{x}_{k-1} + \mathbf{u}(\mathbf{x}_{k-1}) + \mathbf{w}_k \\ &= \mathbf{x}_{k-1} + P(\mathbf{x}_{k-1})\boldsymbol{\theta}_k + \mathbf{w}_k,\end{aligned} \tag{14}$$

where $\mathbf{w}_k$ is a white noise of covariance Q_k. Q_k is either fixed *a priori* or computed from the estimation residuals which are obtained during the regression procedure. It is important to point at that the motion vector $\mathbf{u}(\mathbf{x}_{k-1})$ corresponds in fact to the dominant apparent velocity on an estimation support $\mathcal{R}$ containing the point of interest $\mathbf{x}_{k-1}$. Two cases may be distinguished to choose an appropriate support region:

- When the point motion corresponds to a global dominant motion (for example, the motion of a background feature), the estimation support is fixed to the whole image grid. In that case, the motion parameter vector $\boldsymbol{\theta}_k$ does not depend on the location of $\mathbf{x}_{k-1}$ and the dynamic is linear. For an affine motion model, the state equation (14) turns to be the following linear dynamic:

$$\mathbf{x}_k = A_k\ \mathbf{x}_{k-1} + \mathbf{b}_k + \mathbf{w}_k, \tag{15}$$

  where A_k is the matrix related to rotation, divergence and shear motion, and $\mathbf{b}_k$ is a translation vector. As the noise variable $\mathbf{w}_k$ accounts for errors related to the *global* motion model, it is likely to be non-Gaussian. Nevertheless, $\mathbf{w}_k$ is assumed to be a white noise of zero mean and covariance Q_k conditionally to $\mathbf{I}_{0:k}$. Let us remark that estimating the motion parameters on the whole image grid brings a global information on the point motion. Such type of information is crucial in case of lack of local information (facing noise or occlusions).
- When the feature point follows a motion that may be only described by a local parametric model (for instance, a point on a moving object), the support $\mathcal{R}$ is set to a small domain centered at $\mathbf{x}_{k-1}$. As a consequence, the estimated parameter vector $\boldsymbol{\theta}_k$ depends on $\mathbf{x}_{k-1}$. The considered state equation becomes thus nonlinear with respect to $\mathbf{x}_{k-1}$. In that case, the noise variable $\mathbf{w}_k$ figures errors coming from a *local* motion modelization. It can be assumed to be a Gaussian white noise process, i.e. $p(\mathbf{w}_k|\mathbf{I}_{0:k}) = \mathcal{N}(0, Q_k)$. The corresponding nonlinear dynamic reads then:

$$p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{I}_{0:k}) = \mathcal{N}(\mathbf{x}_{k-1} + P(\mathbf{x}_{k-1})\boldsymbol{\theta}_k, Q_k) \tag{16}$$

B. Conditional measurement equation

Whereas the state equation defines the model of the instantaneous feature motion, the conditional measurement equation $p(\mathbf{z}_k|\mathbf{x}_k, \mathbf{I}_{0:k})$ will allow us to fix a goodness of fit criterion between the initial image and the current image. Such an information is required to overcome feature drift over time.

At time k, we assume that $\mathbf{x}_k$ is observable through a matching process whose goal is to provide, in the image $\mathbf{I}_k$, the most similar point to an initial point $\mathbf{x}_0$, in a reference template $\widetilde{\mathbf{I}}_0$. The result of this process corresponds to a correlation peak and defines the measurement $\mathbf{z}_k$ of our system. The reference template is defined as the initial image $\mathbf{I}_0$, which has been eventually updated by registration in case of large geometric and/or photometric deformations around the tracked feature. Such a reference update procedure will be further described in a specific section. Let us first precise our measurement equation.

Several matching criteria can be used to quantify the similarity between the target point and the candidate point. The conservation of the intensity pattern assumption has simply led us to consider the sum-of-squared-differences (SSD). The measurement $\mathbf{z}_k$ is achieved such as:

$$\mathbf{z}_k = \arg\min_{\mathbf{z}} \underbrace{\sum_{\mathbf{y}\in\mathcal{W}} [\widetilde{\mathbf{I}}_0(\mathbf{x}_0+\mathbf{y}) - \mathbf{I}_k(\mathbf{z}+\mathbf{y})]^2}_{r_k(\mathbf{z})}. \tag{17}$$

$r_k(\mathbf{z})$ is the residual computed on $\mathcal{W}$, a small neighborhood around $\mathbf{z}$, of size $(n \times n)$. It is assumed that this measurement carries enough pieces of information about the state of the tracked point to be able to write that $\mathbf{x}_k = \mathbf{z}_k$ to within an additional white Gaussian noise $\mathbf{v}_k$. This noise variable models a local estimation error and accounts for a confidence measure on this matching. The observation equation reads then:

$$p(\mathbf{z}_k|\mathbf{x}_k, \mathbf{I}_{0:k}) = \mathcal{N}(\mathbf{x}_k, R_k). \tag{18}$$

Estimation of measurement confidence (noise covariance matrix): A good estimation of the measurement noise covariance R_k is essential to make the tracker robust to corrupted observations. Indeed, many factors can affect the quality of the observation: the intensity pattern may undergo affine geometric deformations (scaling, rotation, translation), affine intensity changes (contrast and brightness modifications) or be occluded. To that end, we define an SSD surface, given by residuals $r_k(\mathbf{z})$ on a support $\mathcal{W}'$, of size $(n' \times n')$ around the measurement $\mathbf{z}_k$. To evaluate a confidence on the SSD result, Papanikolopoulos [10] proposes an extension of techniques exploiting the topological changes of the SSD surface [34], [35]. His method consists in fitting parabolas to the directions of the four main axes on the surface.

Our approach uses the idea of Singh and Allen [36]. It consists in transforming the SSD surface - which corresponds to an error distribution - into a response distribution:

$$\mathcal{D}_k(\mathbf{z}) = \exp(-c\ r_k(\mathbf{z})), \tag{19}$$

where c is a normalization factor. As in [37], c is fixed such as $\sum_{\mathbf{z}\in\mathcal{W}'} \mathcal{D}_k(\mathbf{z}) = 1$ through an iterative adjustment ($\sum_{\mathbf{z}\in\mathcal{W}'} \mathcal{D}_k(\mathbf{z}) - 1$ is a continuous decreasing function). We

assume that this distribution corresponds to a probability distribution of the true match location. The covariance matrix R_k associated to the measurement $\mathbf{z}_k$ is constructed from the distribution (19):

$$R_k = \begin{pmatrix} \sigma_{xx} & \sigma_{xy} \\ \sigma_{xy} & \sigma_{yy} \end{pmatrix}, \quad (20)$$

where $\sigma_{uv} = \sum_{\mathbf{z} \in \mathcal{W}'} \mathcal{D}_k(\mathbf{z})(u - u_k)(v - v_k)$, $\mathbf{z} = (x, y)^t$ and $\mathbf{z}_k = (x_k, y_k)^t$. Such a modelization defines an adaptive ellipse of uncertainty of the match location $\mathbf{z}_k$. Let us remark that the variance terms can not exceed n' (the size of the support), since the uncertainty ellipse is built on the window $\mathcal{W}'$ around $\mathbf{z}_k$. Limiting ourselves to such a process may lead to two problematic issues.

- First of all, in case of noisy image sequences, or almost uniform areas, the correlation surface may show several peaks of small magnitude that may be inferior to the noise level. In that case, differences between the magnitudes of these peaks are not significant. None of these noisy peaks should have a predominant effect on the covariance calculation. We propose to detect them and equalize their SSD surface value. Their detection relies on an approximation of the distribution of residuals $r_k(\mathbf{z})$.
  As shown in [8], residuals may assumed to be Gaussian distributed for almost identical regions. Therefore, assuming a brightness constancy assumption to within aa additional Gaussian white noise for two matched points, we have:

  $$\mathbf{I}_k(\mathbf{z}) - \mathbf{I}_0(\mathbf{x}_0) \rightsquigarrow \mathcal{N}(0, \sigma^2) \quad (21)$$

  where σ corresponds to the noise standard deviation. It ensues that for two matched areas of size n^2, $r_k(\mathbf{z})/\sigma$ is distributed as a Chi-square with n^2 degrees of freedom (we remind that n is the size of the support used to implement $r_k(\mathbf{z})$). Considering a sufficiently large estimation support, due to the Fisher approximation, we can safely assume that:

  $$\sqrt{2 * r_k(\mathbf{z})/\sigma^2} - \sqrt{2 * n^2} \rightsquigarrow \mathcal{N}(0, 1). \quad (22)$$

  In practice, if several points of the correlation surface are detected to follow this law (22), their surface values are limited to the same level (i.e. their value are set to the one of the lowest residual). We therefore assign to them the same probability of being the true match location. This process provides us a method which permits to estimate a more meaningful covariance matrix in case of correlation surfaces exhibiting several low magnitude peaks. In the following, the test of law (22) will be called residual test, and will be achieved in practice at 95%. An illustration of this improvement is presented in figure 7.
- The second problematic issue occurs when the response distribution can not be approximated by a Gaussian distribution. This is the case when the correlation surface exhibits numerous significant peaks, which are above the noise level. The covariance construction described in (20) is then not relevant anymore. This may happen in case of occlusions and particularly for highly textured areas. The corresponding $\mathcal{D}_k$ surface may be very smooth and much fitted by a uniform distribution. To overcome a mis-approximation, a Chi-square "goodness of fit" test is realized (in practice at 90%), in order to check if the response distribution is better approximated by a normal or a uniform law. In this latter case, the diagonal terms of R_k are fixed to infinity, and the off-diagonal terms are set to 0. An illustration of such a problematic case is presented in figure 6.

Measurement Procedure: Finally, the overall measurement process may be summarize as follows:

1) Estimation of the measurement $\mathbf{z}_k$ with (17).
2) Construction of the SSD surface in $\mathcal{W}'$, a neighborhood of $\mathbf{z}_k$.
3) Detection of the noisy peaks by residual tests (22) and modification of the SSD surface if some residuals belong to the image noise.
4) Construction of the response distribution from the modified SSD surface with (19).
5) Chi-square test to verify or not the goodness of fit of a uniform distribution to the response distribution.
6) Estimation of the measurement covariance error matrix R_k.

Some illustrations of the measurement noise covariance estimation are presented in figures 5, 6 and 7 for some typical cases. In these figures, we present: (a) the reference image, the feature to be matched, and the obtained observation in a second image, (b) the corresponding SSD surface between the matched points and (c) the associated response distribution. We also present the residual test, (d) the modified SSD surface and (e) the resultant response distribution on which the covariance matrix is finally estimated.

The first example (figure 5) presents an ideal case for point matching : the feature is well-characterized, on a non-noisy sequence and the luminance pattern does not undergo large deformations. In that case, the SSD surface is not affected by residual tests, and the response distribution is well-approximated by a Gaussian law. The covariance estimation is then achieved through (20). The estimated matrix is $\begin{bmatrix} 0.45 & -0.12 \\ -0.12 & 0.37 \end{bmatrix}$.

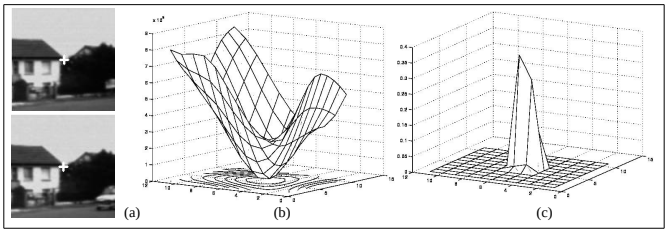


Fig. 5. Estimation of the measurement noise covariance. (noise level fixed to $\sigma = 1$) (a) top image: reference pattern and feature to be matched (the corner of the roof), bottom image: second image and obtained measurement; (b) SSD surface between the matched points; (c) corresponding response distribution. In this case, the uniform law hypothesis has been rejected. The response distribution is approximated by a 2D Gaussian law, and the covariance is estimated through (20).

An occlusion case is depicted in figure 6 to show the interest of the Chi-square "goodness of fit" test on the response

distribution. The uniform law hypothesis is here preferred. The measurement noise covariance is thus fixed to $\begin{bmatrix} \infty & 0 \\ 0 & \infty \end{bmatrix}$ as the response distribution reveals several peaks which are likely to correspond to wrong matching.

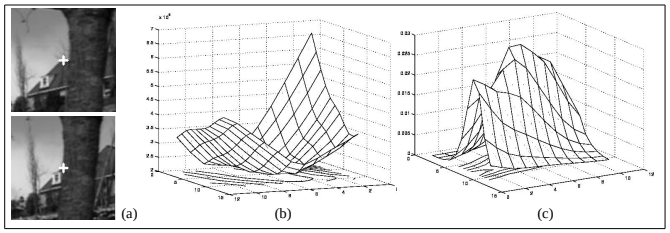


Fig. 6. Estimation of the measurement noise covariance in case of occlusion: interest of the Chi-square test. (noise level fixed to $\sigma = 1$) (a) top image: reference pattern and feature to be matched, bottom image: second image and obtained measurement; (b) SSD surface between the matched points; (c) corresponding response distribution. In this occlusion example, the uniform law hypothesis of the response distribution is preferred.

The last example shown in figure 7 demonstrates the utility of the residual tests for noisy sequences. In such a case it is clear that most of the points can not be well-matched because of the noise level. This illustration shows a miss-match on a noisy sequence. If σ is fixed to 1, the SSDs is not modified, and the resultant response distribution is approximated by a Gaussian distribution. This is presented on surface 7 (c). The corresponding covariance matrix that is estimated through (20) shows relatively small entries: $\begin{bmatrix} 2.01 & -0.95 \\ -0.95 & 4.34 \end{bmatrix}$. This result is obviously not relevant as it does not reflect the uncertainty of the measurement. On the contrary, by setting σ to 5, and by leveling the noisy peaks through the residual tests, the SSD surface is modified (see figure 7 (d)), and the uniform law hypothesis is preferred for the associated response distribution (see figure 7 (e)). This better describes the uncertaincy of the obtained measurement.

These illustrations demonstrate that the proposed tests (residual test and Chi-square test), which are added to a Gaussian modelization of the SSD surface improve significantly the results, by allowing a robust detection of occlusions and other ambiguous situations.

Reference template update procedure: The last point to be explained concerns the reference template $\widetilde{\mathbf{I}}_0$. As mentioned previously, this template around $\mathbf{x}_0$ is used as a pattern which has to be recovered in the current image. $\widetilde{\mathbf{I}}_0$ has a crucial role in the relevance of the determined observation (17). Limiting ourselves to set $\widetilde{\mathbf{I}}_0 = \mathbf{I}_0$ is too restrictive in case of long sequences, with large photometric and/or geometric deformations. The reference template has to be updated to follow the evolutions of the luminance pattern of the tracked point. To that end, one has to answer two questions which are *when* and *how* the reference has to be updated.

The first question is equivalent to the question *When are we sufficiently confident on the estimate to use its photometric pattern to update the reference ?* In the linear case, the quality of the estimate $\hat{\mathbf{x}}_{k|k}$ is simply given by the conditional error covariance $\Sigma_{k|k}$. The reference template is thus updated when its eigenvalues are below a given threshold. In the nonlinear

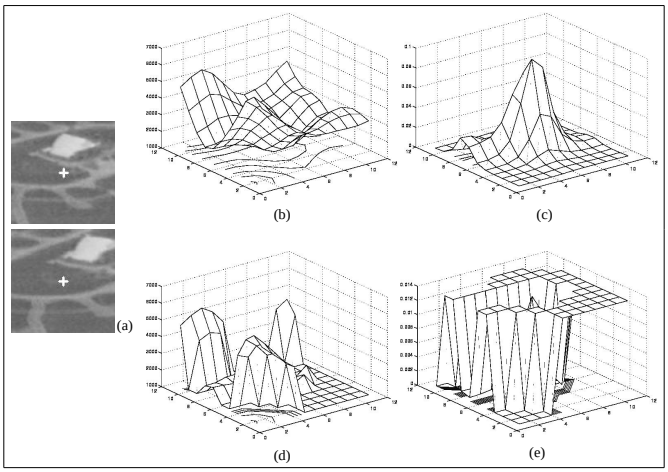


Fig. 7. Estimation of the measurement noise covariance in case of noisy sequences or uniform areas: interest of the residual tests. (a) top image: reference pattern and feature to be matched, bottom image: later image and obtained measurement; (b) SSD surface between the matched points; (c) the response distribution corresponding to $\sigma = 1$ is approximated by a Gaussian distribution; (d) modified SSD surface obtained after the residual tests with $\sigma = 5$; (e) the resultant response distribution with $\sigma = 5$ is well-approximated by a uniform distribution.

case, the procedure is the same. The difference resides in the estimation of the conditional error covariance which is computed empirically from the particle swarm.

To answer the second question, the chosen approach defines $\widetilde{\mathbf{I}}_0$ as being a part of the initial image, centered in $\mathbf{x}_0$, which can be warped to better fit to the tracked pattern. If the estimate is sufficiently accurate at time k, a motion model is estimated between the current reference $\widetilde{\mathbf{I}}_0$, centered in $\mathbf{x}_0$ and a window around $\mathbf{I}_k(\hat{\mathbf{x}}_{k|k})$ of same dimensions. Let $M_{\widetilde{\mathbf{I}}_0 \rightarrow \mathbf{I}_k}$ be this motion model and let $M_{\mathbf{I}_0 \rightarrow \widetilde{\mathbf{I}}_0}$ be the motion model which has been used to construct the current reference from $\mathbf{I}_0$. The new reference template is then built by computing a motion-compensated image. The combination of these two motion models $M_{\widetilde{\mathbf{I}}_0 \rightarrow \mathbf{I}_k} \circ M_{\mathbf{I}_0 \rightarrow \widetilde{\mathbf{I}}_0}$ is applied to the initial image $\mathbf{I}_0$ to build the new reference pattern $\widetilde{\mathbf{I}}_0$.

Figure 8 illustrates this reference update procedure. The bottom image set represents the estimates and their associated uncertainty ellipse on four images at different instants. The top images present the reference templates which are updated by successive warpings.

C. Initial steps of the point tracker

In this section, we describe the initial steps of our point tracking technique. These steps involve a point selection process and the choice between the nonlinear filter and the linear filter.

Point selection criterion: A point is considered to be tracked reliably if its neighbourhood defines a luminance pattern which carries enough information. To discard areas with unsufficient luminance gradient, we use the selection criterion proposed in [6] at the initial time. This criterion is based on the eigenvalues of the structure tensor T:

$$T(\mathbf{x}_0) = \int_{\mathcal{W}(\mathbf{x}_0)} \begin{bmatrix} \nabla \mathbf{I}_x^2 & \nabla \mathbf{I}_y \nabla \mathbf{I}_x \\ \nabla \mathbf{I}_x \nabla \mathbf{I}_y & \nabla \mathbf{I}_y^2 \end{bmatrix},$$

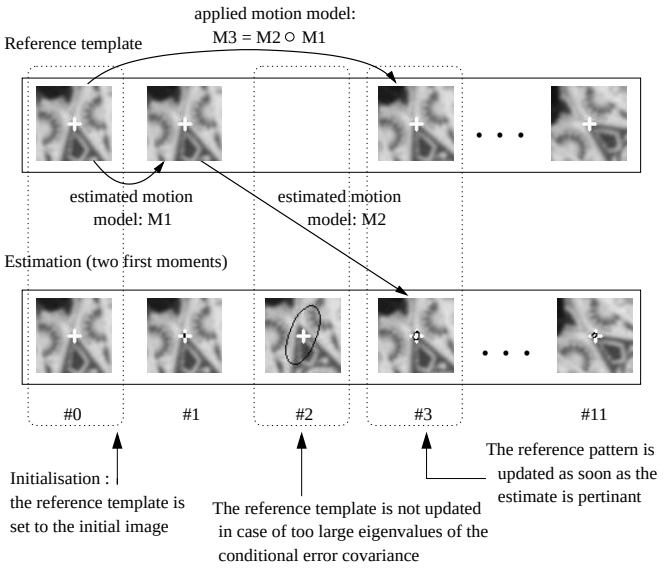


Fig. 8. Illustration of the reference update procedure.

with $[\nabla \mathbf{I}_x, \nabla \mathbf{I}_y] = [\partial \mathbf{I}_0/\partial x, \partial \mathbf{I}_0/\partial y]$. The two eigenvalues λ_1 and λ_2 give information on the intensity profile within the window $\mathcal{W}(\mathbf{x}_0)$. Too small eigenvalues are associated to constant intensity profile, whereas large values indicate a luminance pattern which can be successfully tracked. The corresponding feature is therefore accepted if $min(\lambda_1, \lambda_2) > \lambda$. Typical values of λ are within the range $[0.1, 1]$.

Filter selection - use of a motion detection map: The use of a robust motion estimation technique associated with a motion detection technique [38], [39] gives a practical way to decide whether a considered point belongs to the support of dominant motion or to a mobile area characterized by a local motion model. This motion-based segmentation is performed to partition the initial image in regions that correspond to the global motion and in regions associated to secondary motions. The procedure is applied between the image $\mathbf{I}_0$ and the registered image $\mathbf{I}_1$ at the initial time. The determination of the region boundaries are estimated through a Markov random field statistical regularization. This partition is initially used to determine the type of dynamic (linear or not) of a given point.

For points belonging to the dominant motion support the filtering problem corresponds to a linear model, with a dynamic of the form (15) and the linear observation equation (18). A synopsis of the corresponding point tracker based on the Conditional Linear Filter is described in figure 9. It is important to point out that this tracker enables us to combine global and local pieces of information on the feature point motion. When the motion of a point can be only represented by a local parameterization, the filtering problem we are dealing with is formulated through the linear likelihood (18) and the nonlinear dynamic (16). It induces the use of the Conditional NonLinear Filter as depicted on the block diagram in figure 10.

D. Remarks on computational complexity

Before presenting the experimental results of the proposed trackers, let us make some remarks about the tracker complexity. The computational cost is essentially due to the motion model estimation and to the measurement computation and its associated confidence. The filtering part is not time consuming, even for the nonlinear tracker. Indeed, the peculiar proposed model leads to a very simple and efficient algorithm. Obviously, the tracker complexity depends on the number of tracked points.

The computational cost of the dynamic parameter estimation has to be relativized. For the linear tracker, this routine is run one time at each frame instant, whatever the number of tracked points. In the nonlinear case, this routine is run N*p times, where p is the number of tracked points and N the number of particles. However, considering a simple translational model allows a significant decrease of the computational cost. Indeed, there is no necessity to use more complex model on a small support.

Concerning the observation step, a way to accelerate the computing time would be to use a correlation criterion, that can be efficiently computed in the Fourier domain. Such techniques are used in fluid imaging to estimate dense correlation field, known as Particle Image Velocimetry [40]. The measurement confidence could be also estimated in the Fourier space.

- **initialization**
  set $\hat{\mathbf{x}}_{0|0} = \mathbf{x}_0$. Set $\Sigma_{0|0}$, R_0 and $Q_{0:n}$.

for $k = 1, ..., n$

- **model (from $\mathbf{I}_{0:k}$):** compute A_k, $\mathbf{b}_k$
- **prediction:**
  $\hat{\mathbf{x}}_{k|k-1} = A_k \ \hat{\mathbf{x}}_{k-1|k-1} + \mathbf{b}_k$
  $\Sigma_{k|k-1} = A_k \ \Sigma_{k-1|k-1} \ A_k^t + Q_k$
- **measurement (from $\mathbf{I}_{0:k}$):**
  define $gate_k$ with (3),
  compute $\mathbf{z}_k$ and estimate R_k following the described procedure in §VI-B
- **update:**
  $K_k = \Sigma_{k|k-1}(\Sigma_{k|k-1} + R_k)^{-1}$
  $\hat{\mathbf{x}}_{k|k} = \hat{\mathbf{x}}_{k|k-1} + K_k \ (\mathbf{z}_k - \hat{\mathbf{x}}_{k|k-1})$
  $\Sigma_{k|k} = (I - K_k) \ \Sigma_{k|k-1}$

Fig. 9. Point tracker with Conditional Linear Filter, dedicated to points belonging to the global dominant motion.

- **initialization**
  For $i = 1, ..., N$, $\mathbf{x}_0^{(i)} \rightsquigarrow p(\mathbf{x}_0|\mathbf{I}_0)$, $w_0^{(i)} = \frac{1}{N}$. Set R_0 and $Q_{0:n}$.

for $k = 1, ..., n$

- **measurement (from $\mathbf{I}_{0:k}$):**
  define $gate_k$ with (11),
  compute $\mathbf{z}_k$ and estimate R_k following the described procedure in §VI-B
- **model (from $\mathbf{I}_{0:k}$):** for $i = 1 : N$, evaluate $\boldsymbol{\theta}_k^{(i)}$
- **prediction:** for $i = 1 : N \quad \mathbf{x}_k^{(i)} \rightsquigarrow \mathcal{N}(\mathbf{m}_k(\mathbf{x}_{k-1}^{(i)}), C_k)$
- **update:** for $i = 1 : N$
  $w_k^{(i)} = w_{k-1}^{(i)} \ p(\mathbf{z}_k|\mathbf{x}_{k-1}^{(i)}, \mathbf{I}_{0:k})$ with
  $p(\mathbf{z}_k|\mathbf{x}_{k-1}^{(i)}, \mathbf{I}_{0:k}) = \mathcal{N}(\mathbf{x}_{k-1}^{(i)} + P(\mathbf{x}_{k-1}^{(i)}) \ \boldsymbol{\theta}_{k-1}^{(i)}, R_k + Q_k)$
- **trajectory estimation:** Evaluate $\hat{\mathbf{x}}_{k|k} = \sum_{i=1}^{N} \mathbf{x}_k^{(i)} \ \widetilde{w}_k^{(i)}$
- **resampling if necessary**

Fig. 10. Point tracker with Conditional NonLinear Filter, dedicated to points whose motion can be only locally described.

VII. Experimental results

In this section, we present some experimental results on real-world sequences to demonstrate the efficiency of the two proposed point trackers, namely the Conditional Linear Filter (CLF) and the Conditional NonLinear Filter (CNLF). We compare them to the Shi-Tomasi-Kanade (STK) tracker and to a robust differential method (RDM), which corresponds to an Euler integration of our dynamic described §VI-A.

A. Results of Conditional Linear Filter

A first result of the CLF is presented on **Hangars**, a 10-frame (512 × 512 pixels) noisy sequence presenting a global chaotic motion. A comparison between CLF, STK and CNLF is presented in figure 11. In such a sequence we can remark that STK leads to poor tracking results. This is particularly true for points which may not be easily identified by characteristic luminance patterns (corner points etc.). Indeed, this is a well-known deficiency of such a tracker. On the opposite, for the CLF, the trajectories of all the points are well-recovered. This is even more noteworthy that the sequence is noisy and the motion complex, as depicted in figures 12 for a representation of point 5 and point 13 trajectories. In these figures we plot the mean over successful realizations of the CLF and CNLF. 100 trials have been run on this sequence and no failure has been observed. It is clear that for such a sequence, having a global information on top of local information is crucial. As the global motion information is not taken into account by the CNLF, the results obtained with this later are less good than thus supplied by the CLF (see figures 11, 12). Although our nonlinear tracker can be applied in every situations, the linear filter better suits to sequences exhibiting a global dominant motion for a lower computational time.

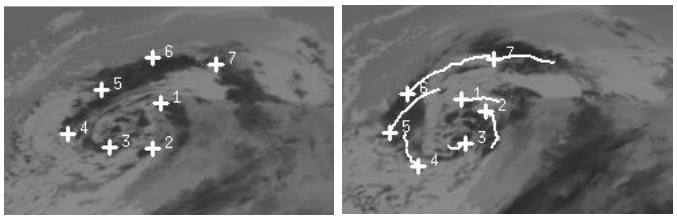
The second sequence, **Corridor**, is a 7-frame (512 × 512 pixels) sequence, which constitutes an extreme case for a global affine motion model due to depth discontinuities and large motions. The initial points are presented in figure 13(a). The complete trajectories provided by CLF, STK, RDM and CNLF are presented in figure 14. In such a sequence, it can be noticed that the STK leads to good tracking results only for two points and loses the others on frame 2. On the opposite, for the CLF, the trajectories of all the points are well-recovered. We believe that in the one hand, the global dynamic on which we rely on gives a quite good prediction for the different points (even if an affine model constitutes a crude model in that case), and on the other hand, the matching measurements enable correcting the deficiency of such a motion model. This can be checked by looking at the results of the tracker built from the dynamic (RDM). On this result, it can be observed that a dominant motion model constitutes a quite rough motion model for some points (see points 1,2,9). Another illustration of these comments is presented in figure 13(b,c) which shows the comparative trajectories of the tested algorithms, and a ground truth given by a user for points 1 and 2. In these graphics, we present the mean trajectories over successful realizations of CLF and CNLF. As it can be observed, there is a significant deviation between the ground truth and the RDM trajectory. The poor result of the STK can also be observed on the graphic 13(c). In the same way as for the sequence Hangars, the CNLF gives better results than the STK, but worse trajectories than the CLF for a higher computational-time.

B. Results of Conditional NonLinear Filter

A result of the CNLF is presented on **Caltra**, a 40-frame sequence of images (190 × 180), showing the motion of two balls fixed on a rotating rigid circle, in front of a cluttered background. Let us note that the number of used particles have been fixed to 100, to limit the computation time. Compared to STK and to RDM (fig.15), the CNLF succeeds in discriminating the balls from the wall-paper, and provides the exact trajectories. Such a result shows the ability of this tracker to deal with complex trajectories in a cluttered environment. Details of white ball trajectory are presented in figure 16(a). The obtained trajectories for the white ball with the different trackers can be observed and compared. The CNLF result accounts for the mean over the successful trajectories on 100 Monte Carlo runs (2 failures have been observed).

The figure 17 presents the results obtained by the CNLF on the sequence **Exam**. This sequence of 60 frames (512 × 512) is a medical imagery sequence of an angiography, showing a motion of contraction and dilation of vessels. These results are compared with the trajectories given by the STK. The CNLF succeeds in recovering complete trajectories whereas the STK loses half of the points.

The CNLF has also been tested on a meteorological sequence of 13 frames (276 × 396), showing the evolution of a through of low pressure. Although the SSD measurements are not very relevant in such a situation, the CNLF succeeds in recovering the vorticity motion as shown in figure 18.



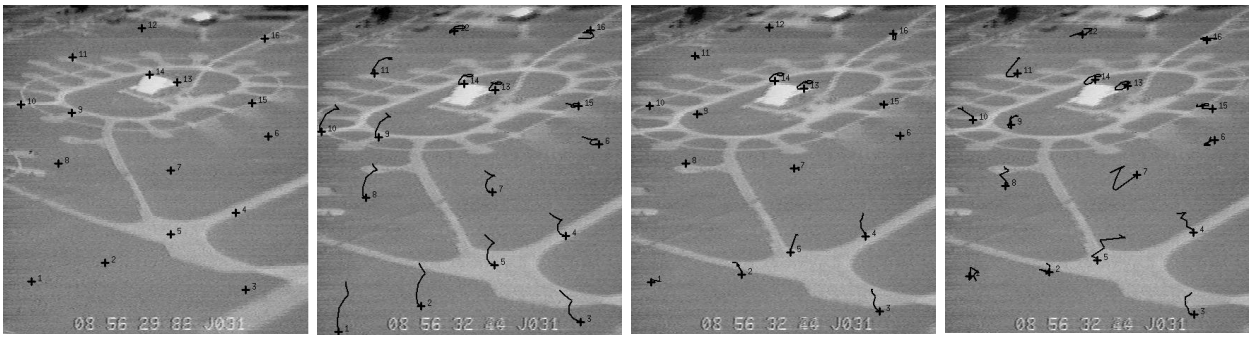
initial points on image # 0 CNLF # 12

Fig. 18. Sequence Fluid: trajectories estimated with the Conditional Non-Linear Filter.

For these three sequences, it is important to note that it would be difficult to rely on a standard linear state equation.

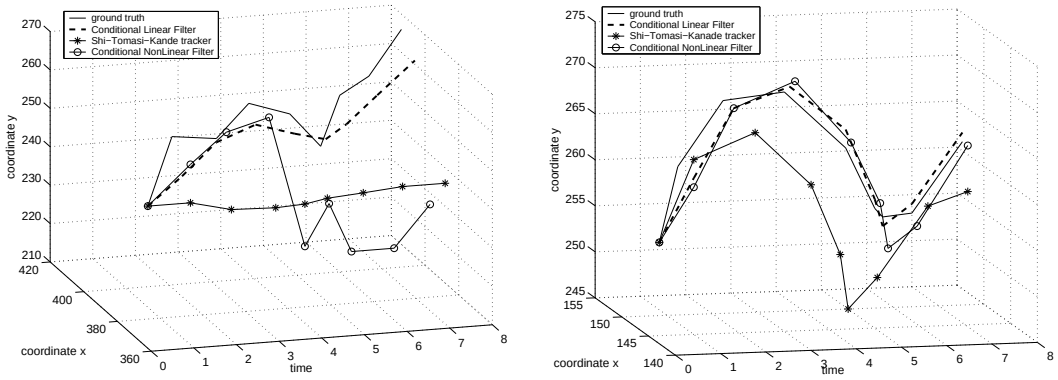
C. Robustness to large geometric deformations and occlusions

We now demonstrate the robustness of the two trackers (CLF and CNLF) to large geometric deformations and occlusions. The first sequence used for that purpose is the sequence **Minitel**. It is a 15-frame (512 × 730) sequence which presents a large camera motion. Figure 19 presents the point trajectories provided by the CLF. Without the reference update procedure described §VI-B, the large rotating deformations does not allow the achievement of a good SSD matching measurement.



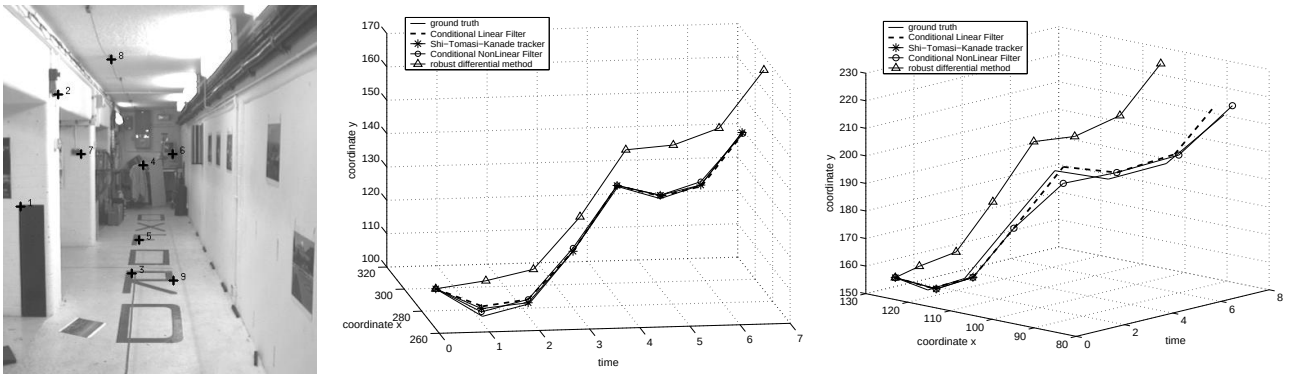
initial points on image # 0 CLF (# 9) STK (# 9) CNLF (# 9)

Fig. 11. Sequence Hangars: trajectories recovered by the Conditional Linear Filter, the Shi-Tomasi-Kanade tracker and the Conditional NonLinear Filter.



(a) point 5 (b) point 13

Fig. 12. Sequence Hangars, comparison between estimated trajectories and ground truth trajectories. The CLF and CNLF trajectories correspond to the mean trajectories over successful realizations (0 failures over 100 trials).



(a) initial points on image # 0 (b) trajectories of point 1 (c) trajectories of point 2

Fig. 13. Sequence Corridor, initial points and comparison between estimated trajectories and ground truth trajectories. The CLF and CNLF trajectories correspond to the mean trajectories over successful realizations (0 failures over 100 trials).

This procedure increases the robustness to large geometric deformations.

The last sequence, **Garden**, is a 27-frame (240 $\times$ 360) sequence. Except from the tree, this sequence presents a global translational motion. Figure 20 shows the CNLF results for the tree's points (points 2,3,4), and the CLF results for the others. Let us remark that we have intentionally chosen to track a point of the background (point 1) with the CNLF to demonstrate the filter robustness to occlusions. The mean trajectory of point 1 over 100 successful Monte-Carlo trials is given in figure 16(b). 15 failures have been observed. Following the trajectories of points moving behind the tree, we can remark that both trackers recover the point locations after they have been hidden, without specifying any occlusions scheme. Indeed, the adaptive covariance noise, estimated from the sequence, allows the conditional trackers to be resistant to occlusions.

VIII. Conclusion and perspectives

In this paper, we proposed a new formulation of stochastic filters adapted to image sequence based tracking. This framework allows considering *a priori*-free systems which

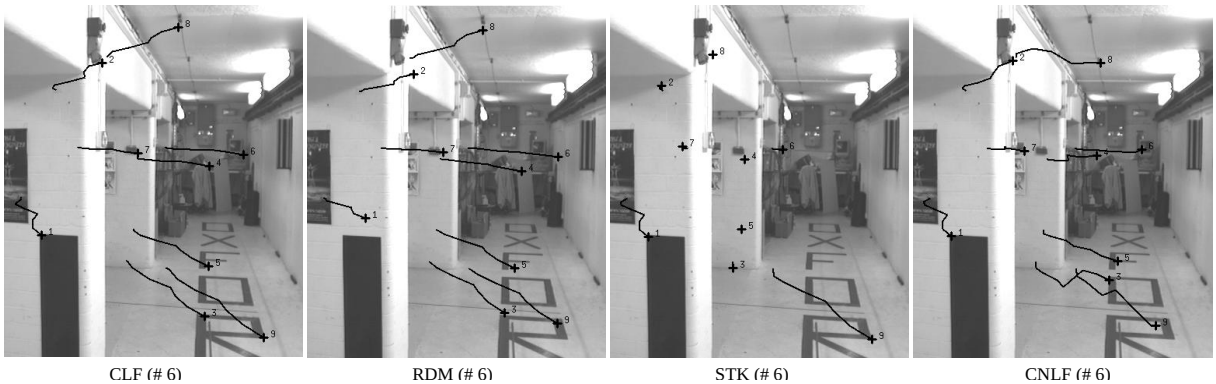


Fig. 14. Sequence Corridor: trajectories recovered by the Conditional Linear Filter, the deterministic tracker based on a robust differential method, the Shi-Tomasi-Kanade tracker and the Conditional NonLinear Filter.

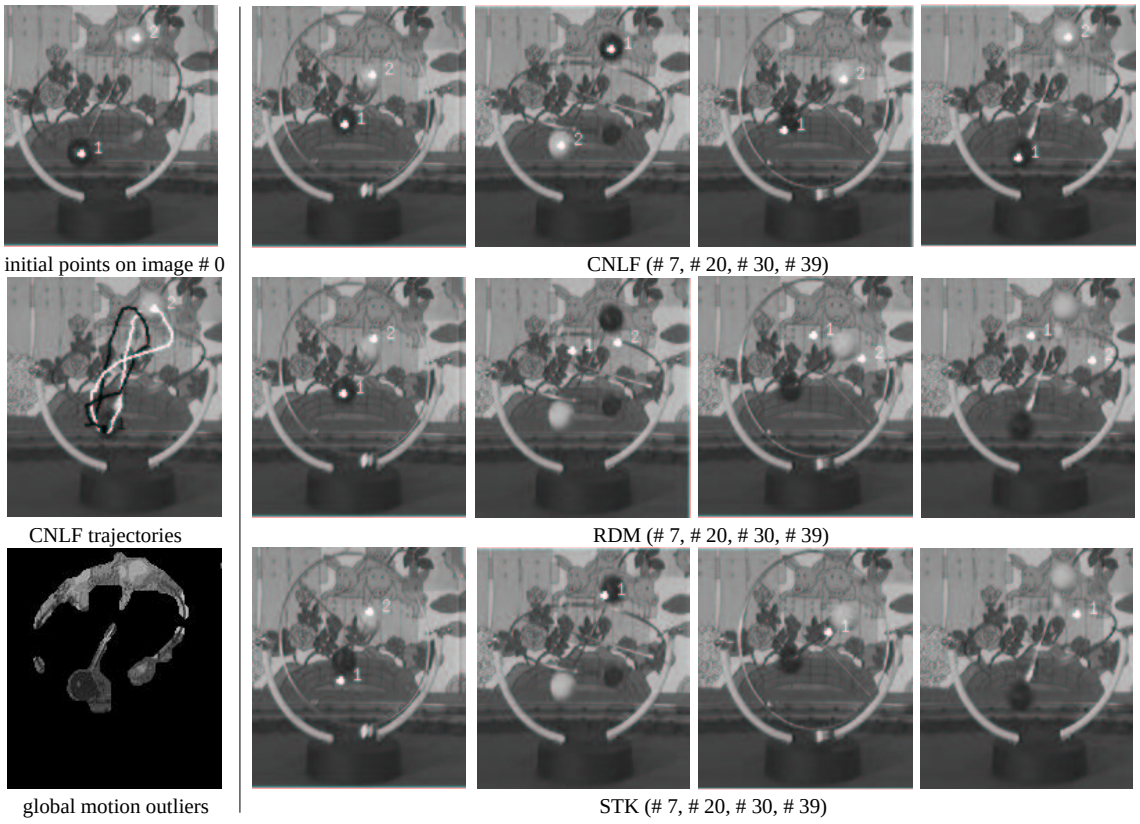
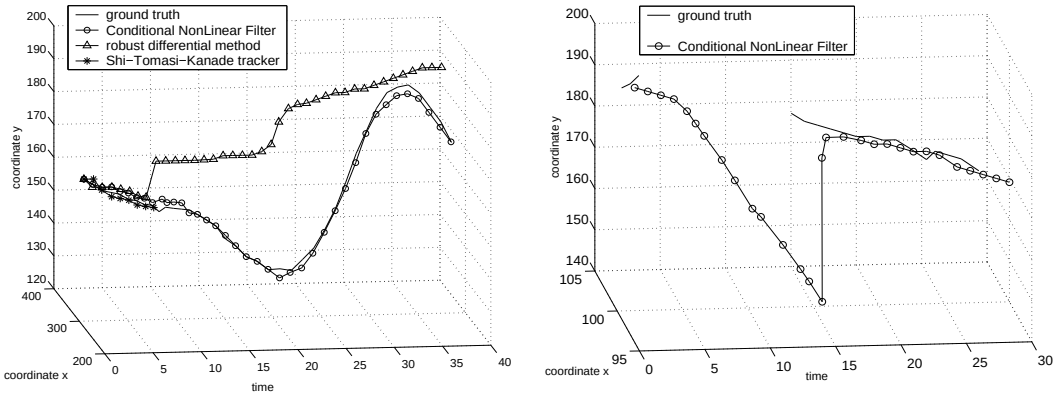


Fig. 15. Sequence Caltra: illustration of the tracking performed by the Conditional NonLinear Filter, the robust differential method, and the Shi-Tomasi-Kanade tracker.

entirely depend on the image data. In that framework, two point trackers have been described. The Conditional Linear Filter is particularly well-suited to image sequences exhibiting dominant motion situations. Indeed, it enables combining global and local pieces of information on the point motion. The Conditional NonLinear Filter is dedicated to points whose motion may be only described by a local parametric model. For both trackers, the combination of measurements provided by a matching technique and a state model relying on a motion estimation has led to very good tracking results for trajectories undergoing abrupt changes in noisy situations. Finally, an automatic computation of the measurement noise covariance leads the trackers to be robust to occlusions.

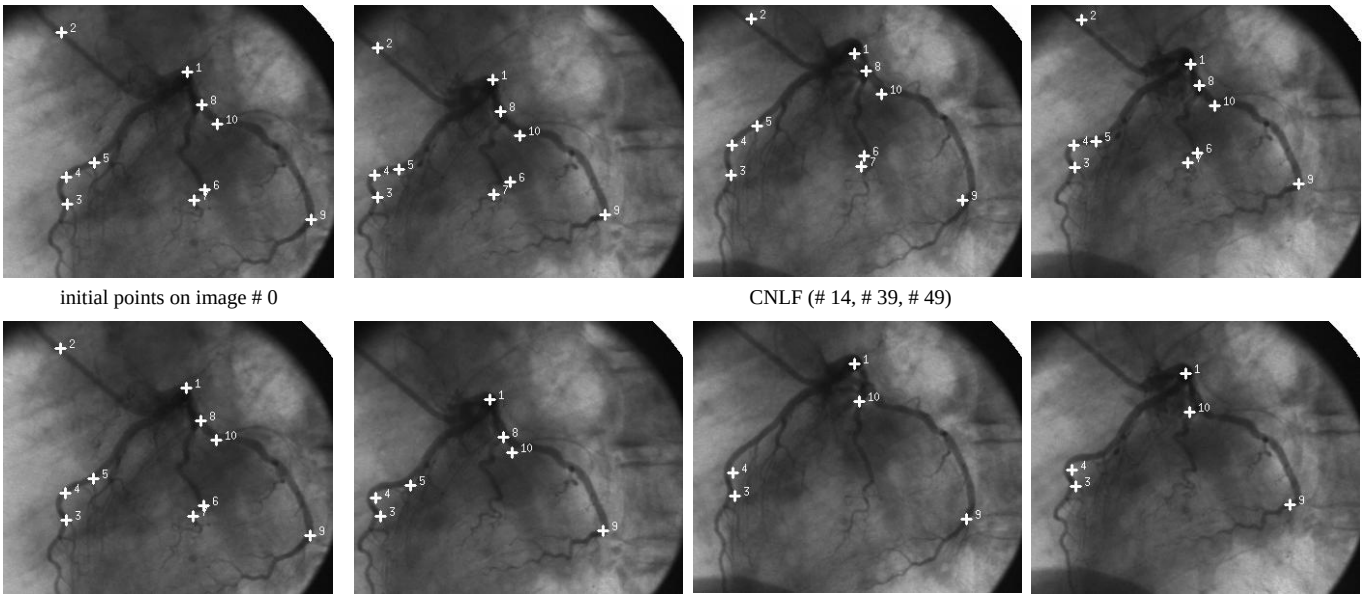
Several perspectives of this work are envisaged. In a first step, the proposed model has to be extended in order to increase the tracker robustness to false alarms. In particular, we would like to propose a likelihood allowing to consider several observations but which is still associated to a known optimal importance function. We believe that the explicit knowledge of this function is of great interest. This work is currently in progress. In a second step, it would be interesting to extend the presented methods for the tracking of complex objects. It would be also interesting to include some regularizing capabilities on the motion parameters for some specific applications. Especially, in this prospect, we plan to investigate the tracking of characteristic structures in meteorological images.



(a) Sequence Caltra, white ball trajectories

(b) Sequence Garden, point 1 trajectories

Fig. 16. Sequence Caltra and Garden, comparison of estimated trajectories and ground truth trajectories. The CNLF results account for the mean realizations over the successful trajectories on 100 Monte Carlo runs (2 failures for Caltra, 15 failures for Garden).



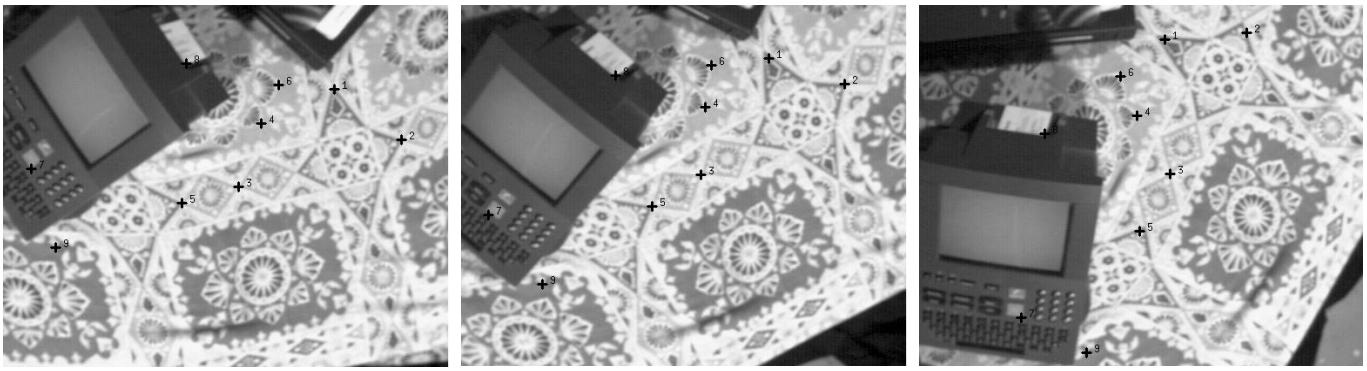
initial points on image # 0

CNLF (# 14, # 39, # 49)

initial points on image # 0

STK (# 14, # 39, # 49)

Fig. 17. Sequence Exam: trajectories recovered by the Conditional NonLinear Filter and the Shi-Tomasi-Kanade tracker.



initial points on image # 0

CLF # 6

CLF # 14

Fig. 19. Sequence Minitel: tracking result of the Conditional Linear Filter

initial points on image # 0 # 10 # 16 # 26

Fig. 20. Sequence Garden: Association of the Conditional Linear and the Conditional NonLinear Filters.

APPENDIX I
EXPRESSION OF $E^*_W[X|Y]$

Reminding that, for two arbitrary random vectors Y and W, one has:

$$\begin{aligned} E[\|Y\|^2|W] &= E[Y^tY|W] = E[tr\{YY^t\}|W] \\ &= tr\{cov(Y,Y|W)\} + E[Y|W]^tE[Y|W], \end{aligned}$$

where tr means the trace of the matrix in braces, and denoting for arbitrary random vectors X, Y and W

$$\begin{aligned} \Sigma_{X,Z|W} &\triangleq cov(X,Z|W) \\ &= E[(X-E[X|W])\ (Z-E[Z|W])^t|W] \\ &= E[XZ^t|W] - E[X|W]\ E[Z|W]^t, \end{aligned}$$

after few manipulations, one can write:

$$\begin{aligned} &E[\|X - A\ Z - B\|^2|W] \\ &= tr\{(A - \Sigma_{X,Z|W}\Sigma^{-1}_{Z,Z|W})\Sigma_{Z,Z|W} \\ &\qquad (A - \Sigma_{X,Z|W}\Sigma^{-1}_{Z,Z|W})^t\} \\ &+ \|E[X|W] - AE[Z|W] - B\|^2 \\ &+ tr\{\Sigma_{X,X|W} - \Sigma_{X,Z|W}\Sigma^{-1}_{Z,Z|W}\Sigma_{Z,X|W}\}. \end{aligned}$$

All the three terms are nonnegative.
$E[\|X - A\ Z - B\|^2|W]$ reaches its minimum for:

$$\begin{aligned} A &= \Sigma_{X,Z|W}\ \Sigma^{-1}_{Z,Z|W}, \\ B &= E[X|W] - \Sigma_{X,Z|W}\ \Sigma^{-1}_{Z,Z|W}\ E[Z|W]. \end{aligned}$$

We deduce (2), the expression of $E^*_W[X|Z]$.

REFERENCES

[1] Y. Mezouar and F. Chaumette, "Model-free optimal trajectories in the image space: Application to robot vision control," in *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, 2001.

[2] P. Sturm, "Structure and motion for dynamic scenes - the case of points moving in planes," in *IEEE Eur. Conf. on Computer Vision*, vol. 2, pp. 867–882, 2002.

[3] R. Adrian, "Particle imaging techniques for experimental fluid mechanics," *Annal Rev. Fluid Mechanism*, vol. 23, pp. 261–304, 1991.

[4] P. Aschwanden and W. Guggenbühl, "Experimental results from a comparative study on correlation-type registration algorithms," *In W. Förstner and St. Ruwiedel, editors, Robust Computer Vision*, pp. 268–289, 1992.

[5] B. Horn and B. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, pp. 185–203, 1981.

[6] J. Shi and C. Tomasi, "Good features to track," in *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pp. 593–600, 1994.

[7] H. Jin, P. Favaro, and S. Soatto, "Real-time feature tracking and outlier rejection with changes in illumination," in *IEEE Int. Conf. on Computer Vision*, vol. 1, pp. 684–689, 2001.

[8] T. Tommasini, A. Fusiello, E. Trucco, and V. Roberto, "Making good features to track better," in *IEEE Int. Conf. on Computer Vision and Pattern Recognition*, pp. 178–183, 1998.

[9] H. Nguyen, M. Worring, and R. van den Boomgaard, "Occlusion robust adaptative template tracking," in *IEEE Int. Conf. on Computer Vision*, vol. 1, pp. 678–683, 2001.

[10] N. Papanikolopoulos, P. Khosla, and T. Kanade, "Visual tracking of a moving target by a camera mounted on a robot: a combination of control and vision," *IEEE Trans. on Robotics and Automation*, vol. 9, no. 1, pp. 14–35, 1993.

[11] A. Blake and M. Isard, *Active Contours*. Springer, 1998.

[12] R. Kalman, "A new approach to linear filtering and prediction problems," *Transactions of the ASME - Journal of Basic Engineering*, pp. 35–45, 1960.

[13] Anderson and Moore, *Optimal Filtering*. Englewood Cliffs, NJ : Prentice Hall, 1979.

[14] H. W. Sorenson, *Bayesian analysis of times series and dynamic models*, j. c. spall ed., Marcel Dekker inc., ch. recursive estimation for nonlinear dynamic systems, 1988.

[15] M. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. on Signal Processing*, vol. 50, no. 2, pp. 174–188, 2002.

[16] A. Doucet, S. Godsill, and C. Andrieu, "On sequential Monte Carlo sampling methods for Bayesian filtering," *Statistics and Computing*, vol. 10, no. 3, pp. 197–208, 2000.

[17] A. Doucet, N. de Freitas, and N. Gordon, Eds., *Sequential Monte Carlo Methods in Practice*. Springer-Verlag, 2001.

[18] H. Sidenbladh, M. Black, and D. Fleet, "Stochastic tracking of 3d human figures using 2d image motion," in *IEEE Eur. Conf. on Computer Vision*, vol. 2, pp. 702–718,2000.

[19] P. Pérez, C. Hue, J. Vermaak, and M. Gangnet, "Color-based probabilistic tracking," in *IEEE Eur. Conf. on Computer Vision*, pp. 661–675, 2002.

[20] M. Isard and A. Blake, "Condensation – conditional density propagation for visual tracking," *Int. Journal of Computer Vision*, vol. 29, no. 1, pp. 5–28, 1998.

[21] M. Isard and J. MacCormick, "Bramble : A Bayesian multiple-blob tracker," in *IEEE Int. Conf. on Computer Vision*, vol. 2, pp. 34–41, 2001.

[22] M. Isard and A. Blake, "A mixed-state condensation traker with automatic model-switching," in *IEEE Int. Conf. on Computer Vision*, pp. 107–112, 1998.

[23] H. Sidenbladh, M. Black, and L. Sigal, "Implicit pobabilistic models of human motion for synthesis and tracking," in *IEEE Eur. Conf. on Computer Vision*, vol. 1, pp. 784–800, 2002.

[24] L. Torresani and C. Bregler, "Space-time tracking," in *IEEE Eur. Conf. on Computer Vision*, pp. 801–812, 2001.

[25] Y. Bar-Shalom and T. Fortmann, *Tracking and Data Association*. Accademic Press, 1988.

[26] A. Kong, J. Liu, and W. Wong, "Sequential imputations and Bayesian missing data problems," *Journal of the American Statistical Association*, vol. 89, pp. 278–288, 1994.

[27] N. Gordon, D. Salmond, and A. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *IEEE Processing-F (Radar and Signal Processing)*, vol. 140, no. 2, pp. 107–113. 1993.

[28] J.S. Liu and R. Chen, "Sequential Monte Carlo methods for dynamic systems," *Journal of the American statistical association*, vol.93, no. 443, pp.1032–1044, 1998.

[29] F. Breidt and A. Carriquiry, "Highest density gates for target traking," *IEEE Trans. on Aerospace and Electronic Systems*, vol. 36, no. 1, pp. 47–55, 2000.

[30] J. Sullivan and J. Rittscher, "Guiding random particles by deterministic search," in *IEEE Int. Conf. on Computer Vision*, vol. 1, pp. 323–330, 2001.
[31] M. Black and A. Jepson, "A probabilistic framework for matching temporal trajectories: Condensation-based recognition of gestures and expressions," in *IEEE Eur. Conf. on Computer Vision*, vol. 1, pp. 909–924, 1998.
[32] M. Black and P. Anandan, "The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields," *Computer Vision and Image Understanding*, vol. 63, no. 1, pp. 75–104, 1996.
[33] J. Odobez and P. Bouthemy, "Robust multiresolution estimation of parametric motion models," *Journal of Visual Communication and Image Representation*, vol. 6, no. 4, pp. 348–365, 1995.
[34] P. Anandan, "Measuring visual motion from image sequences," PhD dissertation, COINS TR 87-21, University of Massachusetts, 1987.
[35] L. Matthies, T. Kanade, and R. Szeliski, "Kalman filter-based algorithms for estimating depth from image sequences," *Int. Jourlal of Computer Vision*, vol. 3, no. 3, pp. 209–238, 1989.
[36] A. Singh and P. Allen, "Image-flow computation : An estimation-theoric framework and a unified perspective," *Computer Vision, Graphics, and Image Processing : Image Understanding*, vol. 56, no. 2, pp. 152–177, 1992.
[37] K. Nickels and S. Hutchinson, "Estimating uncertainty in SSD-based feature tracker," *Image and Vision Computing*, vol. 20, no. 1, pp. 47–58, 2002.
[38] M. Irani, B. Rousso, and S. Peleg, "Computing occluding and transparent motions," *Int. Journal of Computer Vision*, vol. 12, no. 1, pp. 5–16, 1994.
[39] J. Odobez and P. Bouthemy, "MRF-based motion segmentation exploiting a 2d motion model robust estimation," in *IEEE Int. Conf. on Image Processing*, vol. 3, pp. 628–632, 1995.
[40] S.P. McKenna and W.R. McGillis, "Performance of digital image velocimetry processing techniques," in *Experiments in fluids*, vol. 32, pp. 106–115, 2002.

Elise Arnaud was born in 1978. She graduated from the National Institute of Applied Sciences (INSA) of Rouen in Applied Mathematics, France, in 2001. She has been preparing a Ph. D. degree in Signal Processing and Telecommunications from the University of Rennes, France, since 2001. Her main research interests are on statistical models for tracking in image sequences.

Etienne Mémin was born in 1965. He received the Ph.D. degree in Computer Science from the University of Rennes, France, in 1993. In 2003, he received the habilitation degree in computer science from the same university. He was an assistant professor at the University of Bretagne Sud, France from 1993 to 1999. He now holds a position at the University of Rennes, France. From 2001 to 2003, he was on secondment at the CNRS. His research interests include computer vision, statistical models for image (sequence) analysis, fluid and medical images, and parallel algorithms for computer vision.

Bruno Cernuschi-Frías was born in 1952, in Montevideo, Uruguay. He is a citizen of Argentina since 1978. He received the Ph.D. degree in Electrical Engineering from Brown University, Providence, R. I., USA in 1984. He now holds a position of Full Tenured Professor in the Department of Electronics at the Faculty of Engineering, University of Buenos Aires, and is Principal Researcher of the Consejo Nacional de Investigaciones Científicas y Técnicas, (CONICET), Argentina.