Optimal importance sampling for tracking in image sequences: application to point tracking

Élise Arnaud, Etienne Mémin

▶ **To cite this version:**

Élise Arnaud, Etienne Mémin. Optimal importance sampling for tracking in image sequences: application to point tracking. IEEE European conference on computer vision, 2004, Prague, Czech Republic. inria-00306725

HAL Id: inria-00306725
https://inria.hal.science/inria-00306725

Submitted on 3 Apr 2009

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Optimal Importance Sampling for Tracking in Image Sequences: Application to Point Tracking

Elise Arnaud and Etienne Mémin

IRISA, Université de Rennes 1,
Campus de Beaulieu,
35 042 Rennes Cedex, France
{earnaud,memin}@irisa.fr

Abstract. In this paper, we propose a particle filtering approach for tracking applications in image sequences. The system we propose combines a measurement equation and a dynamic equation which both depend on the image sequence. Taking into account several possible observations, the likelihood is modeled as a linear combination of Gaussian laws. Such a model allows inferring an analytic expression of the optimal importance function used in the diffusion process of the particle filter. It also enables building a relevant approximation of a validation gate. We demonstrate the significance of this model for a point tracking application.

1 Introduction

When tracking features of any kind from image sequences, several specific problems appear. In particular, one has to face difficult and ambiguous situations generated by cluttered backgrounds, occlusions, large geometric deformations, illumination changes or noisy data. To design trackers robust to outliers and occlusions, a classical way consists in resorting to stochastic filtering techniques such as Kalman filter [13, 15] or sequential Monte Carlo approximation methods (called particle filters) [7, 10, 11, 16].

Resorting to stochastic filters consists in modeling the problem by a discrete hidden Markov state process $\mathbf{x}_{0:n} = \{\mathbf{x}_0, \mathbf{x}_1, ..., \mathbf{x}_n\}$ of transition equation $p(\mathbf{x}_k|\mathbf{x}_{k-1})$. The sequence of incomplete measurements of the state is denoted $\mathbf{z}_{1:n} = \{\mathbf{z}_1, \mathbf{z}_1, ..., \mathbf{z}_n\}$, of marginal conditional distribution $p(\mathbf{z}_k|\mathbf{x}_k)$. Stochastic filters give efficient procedures to accurately approximate the posterior probability density $p(\mathbf{x}_k|\mathbf{z}_{1:k})$. This problem may be solved exactly through a Bayesian recursive solution, named the optimal filter [10]. In the case of linear Gaussian models, the Kalman filter [1] gives the optimal solution since the distribution of interest $p(\mathbf{x}_k|\mathbf{z}_{1:k})$ is Gaussian. In the nonlinear case, an efficient approximation consists in resorting to sequential Monte Carlo techniques [4, 9]. These methods consists in approximating $p(\mathbf{x}_k|\mathbf{z}_{1:k})$ in terms of a finite weighted sum of Diracs centered in elements of the state space named particles. At each discrete instant, the particles are displaced according to a probability density function named *importance function* and the corresponding weights are updated through the likelihood.

For a given problem, a relevant expression of the *importance function* is a crucial point to achieve efficient and robust particle filters. As a matter of fact, since this function is used for the diffusion of the particle swarm, the particle repartition - or the state-space exploration – strongly depends on it. It can be demonstrated that the *optimal* importance function in the sense of a minimal weight variance criterion is the

distribution $p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{z}_k)$ [9]. As it will be demonstrated in the experimental section, the knowledge of this density improves significantly the obtained tracking results for a point tracking application.

However, the expression of $p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{z}_k)$ is totally unknown in most vision applications. In such a context, the importance function is simply fixed to the prediction density $p(\mathbf{x}_k|\mathbf{x}_{k-1})$. This constitutes a crude model which is counterbalance by a systematic re-sampling step of the particles together with sound models of highly multimodal likelihood [7, 11, 16].

In this paper, an opposite choice is proposed. We investigate simpler forms of likelihood but for which the optimal importance function may be inferred. The considered likelihood is a linear combination of Gaussian laws. In addition, such a modelization allows expressing a validation gate in a simple way. A validation gate defines a bounded research region where the measurements are looked for at each time instant.

Besides, it is interesting to focus on features for which none dynamic model can be set *a priori* or even learned. This is the case when considering the most general situation without any knowledge on the involved sequence. To tackle this situation, we propose to rely on dynamic models directly estimated from the image sequence.

For point tracking applications, such a choice is all the more interesting that any dynamic model of a feature point is very difficult to establish without any *a priori* knowledge on the evolution law of the surrounding object. As a consequence, the system we propose for point tracking depends entirely on the image data. It combines (*i*) a state equation which relies on a local polynomial velocity model, estimated from the image sequence and (*ii*) a measurement equation ensuing from a correlation surface between a reference frame and the current frame. The association of these two approaches allows dealing with trajectories undergoing abrupt changes, occlusions and cluttered background situations.

The proposed method has been applied and validated on different sequences. It has been compared to the Shi-Tomasi-Kanade tracker [17] and to a CONDENSATION-like algorithm [11].

2 Nonlinear Image Sequence Based Filtering

Classical formulation of filtering systems implies to *a priori* know the density $p(\mathbf{x}_{k+1}|\mathbf{x}_k)$, and to be able to extract, from the image sequence, an information used as a measurement of the state. However, in our point of view, feature tracking from image sequences may require in some cases to slightly modify the traditional filtering framework. These modifications are motivated by the fact that an *a priori* state model is not always available, especially during the tracking of features whose nature is not previously known. A solution to this problem may be devised relying on an estimation from the image sequences data of the target dynamics [2, 3]. In that case, it is important to distinguish *(i)* the observation data which constitute the measurements of the state from *(ii)* the data used to extract such a dynamics model. These two pieces of information are of different kinds even if they are both estimated from the image sequence – and therefore depend statistically on each other. In this unconventional situation where dynamics and measurements are both captured from the sequence, it is possible to build a proper filtering framework by considering a conditioning with respect to the image sequence data.

2.1 Image Sequence Based Filtering

Let us first fix our notations. We note $\mathbf{I}_k$ an image obtained at time k. $\mathbf{I}_{0:n}$ represents the finite sequence of random variables $\{\mathbf{I}_k, k = 0, ..., n\}$. Knowing a realization of $\mathbf{I}_{0:k}$, our tracking problem is modeled by the following dynamic and measurement equation:

$$\begin{aligned}\mathbf{x}_k &= f_k^{\mathbf{I}_{0:k}}(\mathbf{x}_{k-1}, \mathbf{w}_k^{\mathbf{I}_{0:k}}),\\ \mathbf{z}_k &= h_k^{\mathbf{I}_{0:k}}(\mathbf{x}_k, \mathbf{v}_k^{\mathbf{I}_{0:k}}).\end{aligned}$$

At each time k, a realization of $\mathbf{z}_k$ is provided by an estimation process based on 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 estimated 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 also depend on $\mathbf{I}_{0:k}$ as well, and are not necessarily Gaussian. We assume that the associated 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 formulation the Markovian assumption, as well as the conditional independence of the observations are maintained conditionally to the sequence. 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 $\mathbf{I}_{k+1:n}$. The optimal filter's equations can be applied to the proposed model. The expected posterior reads now $p(\mathbf{x}_k|\mathbf{z}_{1:k}, \mathbf{I}_{0:k})$. Supposing $p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1}, \mathbf{I}_{0:k-1})$ known, the recursive Bayesian optimal solution is:

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

To solve this conditional tracking problem, standard filters have to be derived in a conditional version. The linear version of this framework, relying on a linear minimal conditional variance estimator, is presented in [2, 3]. The nonlinear version is implemented with a particle filter and is called *Conditional NonLinear Filter*.

2.2 Conditional NonLinear Filter

Facing a system with a nonlinear dynamic and/or a nonlinear likelihood, it is not possible anymore to construct an exact recursive expression of the posterior density function of the state given all available past data. To overcome these computational difficulties, particle filtering techniques propose to implement recursively an approximation of this density (see [4, 9] for an extended review). These methods consist in approximating the posterior density by a finite weighted sum of Dirac centered on hypothesized trajectories – called particles – of the initial system $\mathbf{x}_0$:

$$p(\mathbf{x}_k|\mathbf{z}_{1:k}, \mathbf{I}_{0:k}) \approx \sum_{i=1}^{N} w_k^{(i)} \delta(\mathbf{x}_{0:k} - \mathbf{x}_{0:k}^{(i)}).$$

At each time instant (or iteration), the set of particles $\{\mathbf{x}_{0:k}^{(i)}, i = 1, ..., N\}$ is drawn from an approximation of the true distribution $p(\mathbf{x}_{0:k}|\mathbf{z}_{1:k}, \mathbf{I}_{0:k})$, called the *importance*

function and denoted $\pi(\mathbf{x}_{0:k}|\mathbf{z}_{1:k}, \mathbf{I}_{0:k})$. The closer is the approximation from the true distribution, the more efficient is the filter. The particle weights $w_k^{(i)}$ account for the deviation with regard to the unknown true distribution. The weights are updated according to importance sampling principle:

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

Choosing an importance function that recursively factorizes such as:

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

allows recursive evaluations in time of the particle weights as new measurements $\mathbf{z}_k$ become available. Such an expression implies naturally a causal assumption of the importance function w.r.t. observations and image data. The recursive weights read then:

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

Unfortunately, such a recursive assumption of the importance function induces an increase over time of the weight variance [12]. In practice, this makes the number of significant particles decrease dramatically over time. To limit such a degeneracy, two methods have been proposed (here presented in the conditional framework).

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

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

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 $p(\mathbf{z}_k|\mathbf{x}_{k-1}, \mathbf{I}_{0:k})$. In vision applications, the optimal importance function is usually not accessible. The importance function is then set to the prediction density (i.e. $\pi(\mathbf{x}_k|\mathbf{x}_{0:k-1}, \mathbf{z}_{1:k}) = p(\mathbf{x}_k|\mathbf{x}_{k-1})$). Such a choice excludes the measurements from the diffusion step.

A second solution to tackle the problem of weight variance increase relies on the use of re-sampling methods. Such methods consist in removing trajectories with weak normalized weights, and in adding copies of the trajectories associated to strong weights, as soon as the number of significant particles is too weak [9]. Obviously, these two solutions may be coupled for a better efficiency. Nevertheless it is important to outline that the resampling step introduces errors and is only the results of the discrepancy between the unknown true pdf and the importance function. As a consequence, the resampling step is necessary in practice, but should be used as rarely as possible. It can be noticed that setting the importance function to the diffusion process and resampling at each iteration leads to weight directly the particles with the likelihood. This choice has been made in the CONDENSATION algorithm [11].

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

3 Gaussian Systems and Optimal Importance Function

Filtering models for tracking in vision applications are traditionally composed of a simple dynamic and a highly multimodal and complex likelihood [3]. For such models, an evaluation of the optimal importance function is usually not accessible. In this section, we present some filtering systems relying on a class of likelihoods (eventually multimodal) for which it is possible to sample from the optimal importance function.

3.1 Gaussian System with Monomodal Likelihood

We consider first a conditional nonlinear system, composed of a nonlinear state equation, with an additive Gaussian noise, and a linear Gaussian likelihood:

$$\mathbf{x}_k = f_k^{\mathbf{I}_{0:k}}(\mathbf{x}_{k-1}) + \mathbf{w}_k^{\mathbf{I}_{0:k}}, \quad \mathbf{w}_k^{\mathbf{I}_{0:k}} \rightsquigarrow \mathcal{N}(\mathbf{w}_k^{\mathbf{I}_{0:k}}; \mathbf{0}, Q_k^{\mathbf{I}_{0:k}}) \tag{2}$$

$$\mathbf{z}_k = H_k^{\mathbf{I}_{0:k}}\mathbf{x}_k + \mathbf{v}_k^{\mathbf{I}_{0:k}}, \quad \mathbf{v}_k^{\mathbf{I}_{0:k}} \rightsquigarrow \mathcal{N}(\mathbf{v}_k^{\mathbf{I}_{0:k}}; \mathbf{0}, R_k^{\mathbf{I}_{0:k}}). \tag{3}$$

For these models the analytic expression of the optimal importance function may be inferred. As a matter of fact, noticing that:

$$p(\mathbf{z}_k|\mathbf{x}_{k-1}, \mathbf{I}_{0:k}) = \int p(\mathbf{z}_k|\mathbf{x}_k, \mathbf{I}_{0:k})\; p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{I}_{0:k})\; d\mathbf{x}_k, \tag{4}$$

we deduce:

$$\mathbf{z}_k|\mathbf{x}_{k-1}, \mathbf{I}_{0:k} \rightsquigarrow \mathcal{N}(\mathbf{z}_k; H_k f_k(\mathbf{x}_{k-1}), R_k + H_k Q_k H_k^t), \tag{5}$$

which yields a simple tractable expression for the weight calculation (1) (for the sake of clarity, the index $\mathbf{I}_{0:k}$ has been omitted). As for the optimal importance function we have:

$$p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{z}_k, \mathbf{I}_{0:k}) = p(\mathbf{z}_k|\mathbf{x}_k, \mathbf{I}_{0:k})\; p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{I}_{0:k}) / p(\mathbf{z}_k|\mathbf{x}_{k-1}, \mathbf{I}_{0:k}) \tag{6}$$

and thus,

$$\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{z}_k, \mathbf{I}_{0:k} \rightsquigarrow \mathcal{N}(\mathbf{x}_k; \mu_k, \Sigma_k), \tag{7}$$

with

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

In that particular case, all the expressions used in the diffusion process (7), and in the update step (5) are Gaussian. The filter corresponding to these models is therefore particularly simple to implement. The unconditional version of this result is described in [9].

3.2 Extension to Multimodal Likelihood

Considering only one single measurement can be too restrictive facing ambiguous situations or cluttered background. We describe here an extension of the previous monomodal case to devise a multimodal likelihood.

Let us now consider a vector of M measurements $\mathbf{z}_k = \{\mathbf{z}_{k,1}, \mathbf{z}_{k,2}, ..., \mathbf{z}_{k,M}\}$. As it is commonly done in target tracking [6] and computer vision [11], we assume that a unique measurement corresponds to a true match and that the others are due to false alarms or clutter. Noting Φ_k a random variable which takes its values in $0, ..., M$, we designate by $p(\Phi_k = m)$ the probability that measurement $\mathbf{z}_{k,m}$ corresponds to the *true* measurement at time k; $p(\Phi_k = 0)$ is the probability that none of the measurements corresponds to the *true* one. Denoting $p_{k,m} = p(\Phi_k = m|\mathbf{x}_k, \mathbf{I}_{0:k})$, and assuming that $\forall m = 1, ..., M$, the measurements $\mathbf{z}_{k,1:M}$ are independent conditionally to $\mathbf{x}_k, \mathbf{I}_{0:k}$ and $\Phi_k = m$, then the likelihood can be written as:

$$\begin{aligned} p(\mathbf{z}_k|\mathbf{x}_k, \mathbf{I}_{0:k}) = \; & p_{k,0}\; p(\mathbf{z}_k|\mathbf{x}_k, \mathbf{I}_{0:k}, \Phi_k = 0) \\ & + \sum_{m=1}^{M} \{p_{k,m}\; p(\mathbf{z}_{k,m}|\mathbf{x}_k, \mathbf{I}_{0:k}, \Phi_k = m) \prod_{j \neq m} p(\mathbf{z}_{k,j}|\mathbf{x}_k, \mathbf{I}_{0:k}, \Phi_k = m)\}. \end{aligned} \tag{8}$$

In order to devise a tractable likelihood for which an analytic expression of the optimal importance function may be derived, we make the following hypothesis. We assume that (*i*) the set of mode occurrence probabilities $\{p_{k,i}, i = 1, \cdots, M\}$ is estimated from the images at each instant ; (*ii*) the probability of having no *true* measurement is set to zero ($p_{k,0} = 0$). Such a choice differs from classical tracking assumptions [6, 11] and may be of problematic in case of occlusions. Nevertheless, as we will see it, this potential deficiency is well compensated by an efficient estimation of the measurement noise covariances. We also assume that (*iii*) considering $z_{k,m}$ as being the *true* target-originated observation, it is distributed according to a Gaussian law of mean $H_{k,m}\mathbf{x}_k$ and covariance $R_{k,m}$. As a last hypothesis (*iv*), we assume that the false alarms are uniformly distributed over a measurement region (also called gate) at time k. The total area of the validation gate V_k will be denoted $|V_k|$.

All these assumptions lead to an observation model which can be written as a linear combination of Gaussian laws:

$$\begin{aligned} \mathbf{z}_k|\mathbf{x}_k, \mathbf{I}_{0:k} \rightsquigarrow & \sum_{m=1}^{M} \left[p_{k,m}\, \mathcal{N}(\mathbf{z}_{k,m}; H_{k,m}\, \mathbf{x}_k, R_{k,m}) \prod_{j \neq m} \frac{\mathbb{1}_{\mathbf{z}_{k,j} \in V_k}}{|V_k|} \right] \\ & = \frac{1}{|V_k|^{M-1}} \sum_{m=1}^{M} p_{k,m}\, \mathcal{N}(\mathbf{z}_{k,m}; H_{k,m}\, \mathbf{x}_k, R_{k,m}). \end{aligned} \tag{9}$$

In the same way as for the monomodal measurement equation (§3.1), it is possible for such a likelihood associated to a Gaussian state equation of form (2) to know the optimal importance function. Let us remind that in our case the considered diffusion process requires to evaluate $p(\mathbf{z}_k|\mathbf{x}_{k-1}, \mathbf{I}_{0:k})$ and to sample from $p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{z}_k, \mathbf{I}_{0:k})$. Applying identity (4), with expression (9), the density used for the weight recursion reads:

$$\mathbf{z}_k|\mathbf{x}_{k-1}, \mathbf{I}_{0:k} \rightsquigarrow \frac{1}{|V_k|^{M-1}} \sum_{m=1}^{M} p_{k,m} \mathcal{N}(\mathbf{z}_{k,m}; H_{k,m} f_k(\mathbf{x}_{k-1}), H_{k,m} Q_k H_{k,m}^t + R_{k,m}). \tag{10}$$

The optimal importance function is deduced using identity (6) and expression (10):

$$\mathbf{x}_k|\mathbf{x}_{k-1},\mathbf{z}_k,\mathbf{I}_{0:k} \rightsquigarrow \frac{\mathcal{N}(\mathbf{x}_k; f_k(\mathbf{x}_{k-1}), Q_k) \sum_{m=1}^{M} p_{k,m} \mathcal{N}(\mathbf{x}_k; H_{k,m}^{-1}\mathbf{z}_k, H_{k,m}^{-1} R_{k,m} H_{k,m}^{-1t})}{|V_k|^{M-1}\, p(\mathbf{z}_k|\mathbf{x}_{k-1},\mathbf{I}_{0:k})}$$

Through Gaussian identities this expression reads as a Gaussian mixture of the form:

$$\mathbf{x}_k|\mathbf{x}_{k-1},\mathbf{z}_k,\mathbf{I}_{0:k} \rightsquigarrow \sum_{m=1}^{M} p_{k,m} \frac{\alpha_{k,m}}{S}\, \mathcal{N}(\mathbf{x}_k; \mu_{k,m}, \Sigma_{k,m}) \tag{11}$$

with

$$\begin{cases} \Sigma_{k,m} = (Q_k^{-1} + H_{k,m}^t R_{k,m}^{-1} H_{k,m})^{-1} \\ \mu_{k,m} = \Sigma_{k,m}\, (Q_k^{-1}\, f_k(\mathbf{x}_{k-1}) + H_{k,m}^t R_{k,m}^{-1}\, \mathbf{z}_{k,m}) \\ S = |V_k|^{M-1}\, p(\mathbf{z}_k|\mathbf{x}_{k-1},\mathbf{I}_{0:k}) \\ \alpha_{k,m} = \dfrac{|\Sigma_{k,m}|^{\frac{1}{2}}}{2\pi |R_{k,m}|^{\frac{1}{2}} |Q_k|^{\frac{1}{2}}} \exp(-\dfrac{1}{2}(\|f_k(\mathbf{x}_{k-1})\|^2_{Q_k^{-1}} + \|\mathbf{z}_{k,m}\|^2_{R_{k,m}^{-1}} - \|\mu_{k,m}\|^2_{\Sigma_{k,m}^{-1}})) \end{cases}$$

Let us point out that the proposed systems lead to a simple implementation as the involved distributions are all combinations of Gaussian laws. In addition, as described in the next subsection, such systems allow to define a relevant validation gate for the measurements.

3.3 Validation Gate

When tracking in cluttered environment, an important issue resides in the definition of a region delimiting the space where future observations are likely to occur [6]. Such a region is called *validation region* or *gate*. Selecting a too small gate size may lead to miss the target-originated measurement, whereas selecting a too large size is computationally expensive and increases the probability of selecting false observations.

In our framework, the validation gate is defined through the use of the probability distribution $p(\mathbf{z}_k|\mathbf{z}_{1:k-1},\mathbf{I}_{0:k})$. For linear Gaussian systems, an analytic expression of this distribution may be obtained. This leads to an ellipsoidal probability concentration region. For nonlinear models, the validation gate can be approximated by a rectangular or an ellipsoidal region, whose parameters are usually complex to define. Breidt [8] proposes to use Monte Carlo simulations in order to approximate the density $p(\mathbf{z}_k|\mathbf{z}_{1:k-1},\mathbf{I}_{0:k})$, but this solution appears to be time consuming. For the systems we propose, it is possible to approximate efficiently this density by a Gaussian mixture. The corresponding validation gate V_k consists in an union of ellipses. Observing that:

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

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 $(\mathbf{x}_{k-1}^{(i)}, w_{k-1}^{(i)})$, the following approximation can be done:

$$p(\mathbf{z}_k|\mathbf{z}_{1:k-1},\mathbf{I}_{0:k}) \simeq \sum_i w_{k-1}^{(i)}\; p(\mathbf{z}_k|\mathbf{x}_{k-1}^{(i)},\mathbf{I}_{0:k}). \tag{12}$$

Introducing expression (10) in (12) leads to an expression of $p(\mathbf{z}_k|\mathbf{z}_{1:k-1}, \mathbf{I}_{0:k})$ as a combination of $N \times M$ Gaussian distributions (N is the number of particles). As considering $N \times M$ ellipses is computationally expensive, we approximate the density by a sum of M Gaussian laws. We then finally obtain an approximation of V_k as an ellipse union $V_k = \bigcup_{m=1:M} \Psi_m = \{\epsilon_m : (\epsilon_m - \xi_{k,m})^t C_{k,m}^{-1} (\epsilon_m - \xi_{k,m}) \leq \gamma_m\}$ with the first and second moments defined as:

$$\begin{cases} \xi_{k,m} = \sum_i w_{k-1}^{(i)} \; H_{k,m} \; f_k(\mathbf{x}_{k-1}^{(i)}) \\ C_{k,m} = \sum_i w_{k-1}^{(i)} [H_{k,m} Q_k H_{k,m}^t + R_{k,m}^{(i)} + H_{k,m} f_k(\mathbf{x}_{k-1}^{(i)}) f_k^t(\mathbf{x}_{k-1}^{(i)}) H_{k,m}^t] - \\ (\sum_i w_{k-1}^{(i)} H_{k,m} \; f_k(\mathbf{x}_{k-1}^{(i)}))(\sum_i w_{k-1}^{(i)} H_{k,m} \; f_k(\mathbf{x}_{k-1}^{(i)}))^t. \end{cases}$$

The parameter γ_m is chosen in practice as the 99th percentile of the probability for $\mathbf{z}_{k,m}$ to be the *true* target-originated measurement.

In addition to a simple and optimal sampling process, the possibility to build a relevant approximation of a validation gate constitutes another advantage of the Gaussian models we propose. In order to demonstrate experimentally their significance, these systems have been applied to a point tracking application.

4 Application to Point Tracking

The objective of point tracking consists in reconstructing the 2D point trajectory along the image sequence. To that purpose, it is necessary to make some conservation assumptions on some information related to the feature point. These hypotheses may concern the point motion, or a photometric/geometric invariance in a neighborhood of the point.

The usual assumption of luminance pattern conservation along a trajectory has led to devise two kinds of methods. The first ones are intuitive methods based on correlation [5]. The second ones are defined as differential trackers, built on a differential formulation of a similarity criterion. In particular, the well-known Shi-Tomasi-Kanade tracker [17] belongs to this latter class.

In this paper, the proposed approach for point tracking is also built on the basis of luminance pattern consistency. In this application, each state $\mathbf{x}_k$ represents the location of the point projection at time k, in image $\mathbf{I}_k$. In order to benefit from the advantages of the two class of method, we propose to combine a dynamic relying on a differential method and measurements based on a correlation criterion. The system we focus on is therefore composed of measurements and dynamic equations which both depend on $\mathbf{I}_{0:k}$. The noise covariance considered at each time is also automatically estimated on the image sequence. To properly handle such a system, the point tracker is built from the filtering framework presented in § 2.

4.1 Likelihood

At time k, we assume that $\mathbf{x}_k$ is observable through a matching process whose goal is to provide the most similar points to $\mathbf{x}_0$ from images $\mathbf{I}_0$ and $\mathbf{I}_k$. The result of this process is the measurement vector $\mathbf{z}_k$. Each observation $z_{k,m}$ corresponds to a correlation peak. The number of correlation peaks (or components of $\mathbf{z}_k$) is fixed to a given number. Several matching criteria can be used to quantify the similarity between two points.

The consistency assumption of a luminance pattern has simply led to consider the sum-of-squared-differences criterion.

As in [18] the correlation surface, denoted $r_k(x, y)$ and computed over the validation gate V_k, is converted into a response distribution: $\mathcal{D}_k \triangleq \exp(-c\, r_k(x, y))$, where c is a normalizing factor, fixed such as $\int_{V_k} \mathcal{D}_k = 1$. This distribution is assumed to represent the probability distribution associated to the matching process. The relative height of the different peaks defines the probability $p_{k,m}$ of the different measurements $z_{k,m}$. The covariance matrices $R_{k,m}$ are estimated from the response distribution on local supports centered around each observation. A Chi-Square "'goodness of fit" test is realized, in order to check if this distribution is locally better approximated by a Gaussian or by a uniform law [3]. An approximation by a Gaussian distribution indicates a clear discrimination of the measurement, and $R_{k,m}$ is therefore set to the local covariance of the distribution. At the opposite, an approximation by a uniform distribution indicates an unclear peak detection on the response distribution. This may be due to an absence of correlation in presence of occlusions or noisy situations. In this case, the diagonal terms of $R_{k,m}$ are fixed to infinity, and the off-diagonal terms are set to 0. Finally, in this application, matrices $H_{k,m}$ are set to identity.

4.2 Dynamic Equation

As we wish to manage situations where no *a priori* knowledge on the dynamic of the surrounding object is available, and in order to be reactive to any unpredictable change of speed and direction of the feature point, the dynamic we consider is estimated from $\mathbf{I}_{0:k}$. The state equation describes the motion of a point $\mathbf{x}_{k-1}$ between images $k-1$ and k, and allows a prediction of $\mathbf{x}_k$. A robust parametric motion estimation technique [14] is used to estimate reliably a 2D parametric model representing the dominant apparent velocity field on a given support $\mathcal{R}$. The use of such a method on an appropriate local support around $\mathbf{x}_{k-1}$ provides an estimate of the motion vector at the point $\mathbf{x}_{k-1}$ from images $\mathbf{I}_{k-1}$ and $\mathbf{I}_k$. As $\mathcal{R}$ is a local domain centered at $\mathbf{x}_{k-1}$, the estimated parameter vector depends in a nonlinear way on $\mathbf{x}_{k-1}$. The noise variable $\mathbf{w}_k$ accounts for errors related to the local motion model. It is assumed to follow a zero mean Gaussian distribution of fixed covariance .

5 Experimental Results

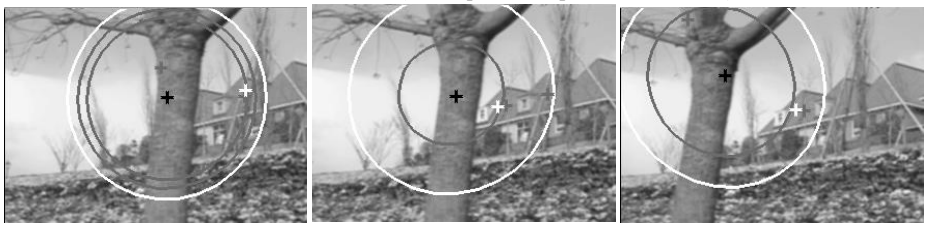
In this section, we present some experimental results on four different sequences to demonstrate the efficiency of the proposed point tracker.

The first result is presented to demonstrate the interest of the optimal importance function. To that purpose, we have chosen to study an occlusion case, on the **Garden** sequence. This sequence shows a garden and a house occluded by a tree. Let us focus on a peculiar feature point located on the top of a house roof. This point is visible in the two first images and stays hidden from frame #3 to frame #15. Two algorithms have been tested for the tracking of this point. Both of them rely on the same filtering system (the one described in section § 3.2). The first one is the method we propose (namely the Conditional NonLinear Filter (CNLF), with the use of the optimal importance function), whereas the second one is a CONDENDATION-like algorithm, for which the considered importance function is identified to the diffusion process. Figure 1 presents the obtained results. The use of the optimal importance function allows us to recover the actual point

location after a long occlusion. This shows clearly the benefit that can be obtained when taking into account the measurement in the diffusion process.

initial conditions frame #0

(a) Conditional NonLinear Filter (with optimal importance function) #0,13,18,30



(b) CONDENSATION-like algorithm (without optimal importance function) #13,18,30

Fig. 1. Interest of the optimal interest function in case of occlusion. Tracking results obtained with (a) the Conditional NonLinear Filter, with the use of optimal importance function and (b) a CONDENSATION-like algorithm, without the use of optimal importance function. For both algorithms, the considered filtering system is the one described in §3.2. Black crosses present the estimates. White and gray crosses corresponds to the observations, and the ellipses to their associated validation gates. The white crosses show the measurement of highest probability.

The second sequence, **Corridor**, constitutes a very difficult situation, since it combines large geometric deformations, high contrast, and ambiguities. The initial points and the final tracking results provided by the Shi-Tomasi-Kanade (STK) tracker, and the CNLF are presented in figure 2. In such a sequence, it can be noticed that the STK leads to good tracking results only for a small number of points. On the opposite, for the CNLF, the trajectories of all the feature points are well-recovered. Let us point out that for this sequence, considering one or several observations per point leads nearly to the same results. Another result of the CNLF, with a multimodal likelihood, is presented on the sequence **Caltra**. This sequence shows the motion of two balls, fixed on a rotating rigid circle, on a cluttered background. Compared to STK (fig.3), 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.

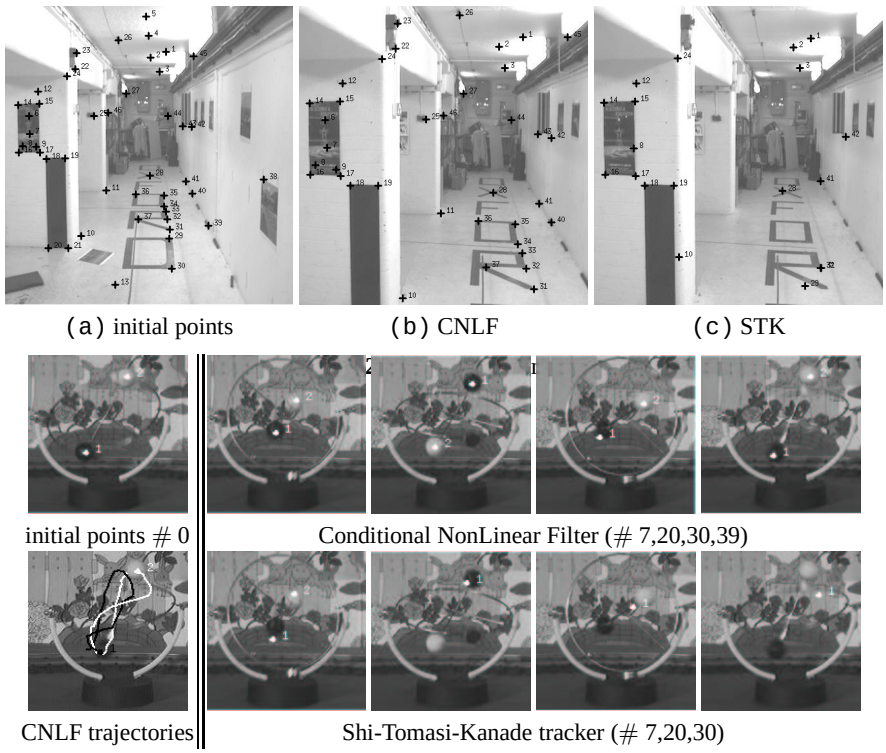


Fig. 3. Caltra sequence

The last result on the **hand** sequence demonstrates that considering several observations improves the tracking results in case of ambiguous situations. This sequence presents finger motions of one hand. Figure 4 illustrates the results obtained with the CNLF, considering a monomodal likelihood (a) and a multimodal likelihood (b). As it can be observed, considering only one correlation peak per point leads here to mistake the different fingers. This confusing situations are solved by taking into account several (here, 3) observations.

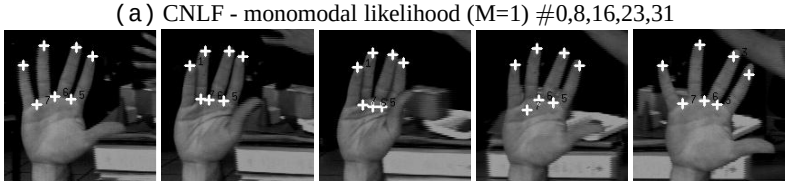
6 Conclusion

In this paper, we proposed a Conditional NonLinear Filter for point tracking in image sequence. This tracker has the particularity of dealing with *a priori*-free systems, which entirely depend on the image data. In that framework, a new filtering system has been described. To be robust to cluttered background, we have proposed a ‘¡peculiar class of multimodal likelihood. Unlike usual systems used in vision applications within non linear stochastic filtering framework, we deal with system which allows an exact estimate of the optimal importance function. The knowledge of the optimal function enables to include naturally measurements into the diffusion process and authorizes to build a relevant approximation of a validation gate. Such a framework, applied to a point tracking application, enables to significantly improve the result of traditional trackers. The resulting point tracker has been shown to be robust to occlusions and complex trajectories.

References

1. B.D.O. Anderson and J.B. Moore. *Optimal Filtering.* Englewood Cliffs, 1979.

(a) CNLF - monomodal likelihood (M=1) #0,8,16,23,31



(b) CNLF - multimodal likelihood (M=3) #0,8,16,23,31

Fig. 4. Hand sequence: Conditional NonLinear Filter results. (a) Only one observation is considered per point and the involved system is the one described in §3.1; (b) 3 observations are considered per point and the involved system is the one described in §3.2

2. E. Arnaud, E. Mémin and B. Cernuschi-Frías. A robust stochastic filter for point tracking in image sequences. *ACCV*, 2004.
3. E. Arnaud, E. Mémin and B. Cernuschi-Frías. Conditional filters for image sequence based tracking - application to point tracker. accepted for publication *IEEE trans. on Im. Proc*, 2004.
4. M.S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking. *TSP*, 50(2), 2002.
5. 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*, p. 268–289, 1992.
6. Y. Bar-Shalom and T.E. Fortmann. *Tracking and Data Association.* Accademic Press, 1988.
7. M.J. Black and A.D. Jepson. A probabilistic framework for matching temporal trajectories: Condensation-based recognition of gestures and expressions. In *ECCV*, p. 909–924, 1998.
8. F.J. Breidt and A.L. Carriquiry. Highest density gates for target traking. *IEEE Trans. on Aerospace and Electronic Systems*, 36(1):47–55, 2000.
9. A. Doucet, S. Godsill, and C. Andrieu. On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, 10(3):197–208, 2000.
10. N.J. Gordon, D.J. Salmond, and A.F.M. Smith. Novel approach to non-linear/non-Gaussian Bayesian state estimation. *IEEE Processing-F (Radar and Signal Processing)*, 140(2), 1993.
11. M. Isard and A. Blake. Condensation – conditional density propagation for visual tracking. *IJCV*, 29(1):5–28, 1998.
12. A. Kong, J.S. Liu, and W.H. Wong. Sequential imputations and Bayesian missing data problems. *Journal of the American Statistical Association*, 89(425):278–288, 1994.
13. F. Meyer and P. Bouthemy. Region-based tracking using affine motion models in long image sequences. *CVGIP:IU*, 60(2):119–140, 1994.
14. J.-M. Odobez, P. Bouthemy. Robust multiresolution estimation of parametric motion models. *Journ. of Vis. Com. and Im. Repr.*, 6(4):348-365, 1995.
15. N.P. Papanikolopoulos, P.K. 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*, 9(1):14–35, 1993.
16. P. Pérez, C. Hue, J. Vermaak, and M. Gangnet. Color-based probabilistic tracking. In *ECCV*, p. 661–675, 2002.
17. J. Shi and C. Tomasi. Good features to track. In *CVPR*, p. 593–600, 1994.
18. A. Singh and P. Allen. Image-flow computation : An estimation-theoric framework and a unified perspective. *CVGIP: IU*, 56(2):152–177, 1992.