International Journal of Computer Vision 74(1), 75–102, 2007

DOI: 10.1007/s11263-006-0003-2

Partial Linear Gaussian Models for Tracking in Image Sequences Using Sequential Monte Carlo Methods

ELISE ARNAUD*
Inria Rhône-Alpes/Universite Joseph Fourier, 655, avenue de l'Europe, 38330 Montbonnot, France
elise.arnaud@inrialpes.fr

ETIENNE MÉMIN
IRISA/Univ. Rennes I, Campus de Beaulieu, 35042 Rennes Cedex, France
memin@irisa.fr

Received October 6, 2005; Accepted August 15, 2006

First online version published in December, 2006

Abstract. The recent development of Sequential Monte Carlo methods (also called particle filters) has enabled the definition of efficient algorithms for tracking applications in image sequences. The efficiency of these approaches depends on the quality of the state-space exploration, which may be inefficient due to a crude choice of the function used to sample in the associated probability space. A careful study of this issue led us to consider the modeling of the tracked dynamic system with partial linear Gaussian models. Such models are characterized by a non linear dynamic equation, a linear measurement equation and additive Gaussian noises. They allow inferring an analytic expression of the optimal importance function used in the diffusion process of the particle filter, and enable building a relevant approximation of a validation gate. Despite of these potential advantages partial linear Gaussian models have not been investigated. The aim of this paper is therefore to demonstrate that such models can be of real interest facing difficult usual issues such as occlusions, ambiguities due to cluttered backgrounds and large state space. Three instances of these models are proposed. After a theoretical analysis, their significance is demonstrated by their performance for tracking points and planar objects in challenging real-world image sequences.

Keywords: sequential Monte Carlo methods, optimal importance function, Rao-Blackwellization, validation gate, point tracking, planar structure tracking

1. Introduction

Visual tracking has been extensively studied in computer vision due to its huge number of applications. Among them, one can find surveillance, medical imaging, video compression or augmented reality. Recently, the use of sequential Monte Carlo methods (also known as particle filters) has led to the development of very efficient tracking techniques (Arnaud et al., 2005; Isard and Blake, 1998; Pérez et al., 2004). The popularity of these stochastic filtering approaches can be explained by their simplicity, easiness of implementation, constraint-free model and robustness to difficult situations. Indeed, several specific problems may appear when tracking features of any kind from image sequences. In particular, one has to face difficult and ambiguous situations generated by cluttered backgrounds, occlusions, large geometric deformations, illumination changes or noisy data.

Resorting to stochastic filters consists in modeling the dynamic system to be tracked as a discrete hidden Markov state process. The goal is to estimate the value of the random Markovian process—also called *state process* and denoted $\mathbf{x}_{0:n} = \{\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_n\}$—from realizations of the observation process that are obtained at each instant. The set of measurements are denoted $\mathbf{z}_{1:n} = \{\mathbf{z}_1, \mathbf{z}_1, \ldots, \mathbf{z}_n\}$. The system is described by (a) the distribution of the state process at initial time $p(\mathbf{x}_0)$, (b) a probability distribution modeling the evolution

*Correspondence author.

of the state process $p(\mathbf{x}_k \mid \mathbf{x}_{k-1})$—this distribution represents a *dynamic equation*, also called *evolution law*, and is usually set *a priori*—and (c) a likelihood (representing the *measurement equation*) $p(\mathbf{z}_k \mid \mathbf{x}_k)$ that links the observation to the state. In this framework, the *posterior* distribution, i.e. the law of the state process knowing the set of observations, carries the whole information on the process to be estimated. More precisely, as tracking is a causal problem, the distribution of interest is the law of the state given the set of past and present observations $p(\mathbf{x}_k \mid \mathbf{z}_{1:k})$, known as *filtering distribution*. The problem of recursively estimating this distribution may be solved exactly through a Bayesian recursive solution, named the optimal filter (Gordon et al., 1993). This solution requires to compute integrals of huge dimension. In the case of linear Gaussian models, the Kalman filter (Anderson and Moore, 1979) gives the optimal solution since the distribution of interest $p(\mathbf{x}_k \mid \mathbf{z}_{1:k})$ is Gaussian. In the nonlinear case, an efficient approximation consists in resorting to sequential Monte Carlo techniques (Arulampalam et al., 2002; Doucet et al., 2000; Gordon et al., 1993; Liu and Chen, 1998). These methods consist in approximating $p(\mathbf{x}_k \mid \mathbf{z}_{1:k})$ 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 using the system's equations.

For a given problem, a relevant expression of the *importance function* is crucial in order to achieve an efficient and robust particle filter. 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 $p(\mathbf{x}_k \mid \mathbf{x}_{k-1}, \mathbf{z}_k)$ (Doucet et al., 2000). As it will be demonstrated, the knowledge of this density leads to consider the *optimal particle filter* whose implementation is very simple and that improves significantly the tracking results. However, because the densities involved are rarely available in practice, such an efficient sampling is not used in most vision applications. This is all the more true as most tracking applications in computer vision rely on sound models of highly multimodal likelihood associated with simple dynamic equations (Isard and Blake, 1998; Pérez et al., 2004; Wu and Huang, 2004). In such a context, the importance function is usually simply fixed to the evolution law $p(\mathbf{x}_k \mid \mathbf{x}_{k-1})$. This constitutes a crude model which is counterbalanced by a systematic selection of the particles of large weight.

In this paper, we investigate the opposite choice: we will deeply study the use of *partial linear Gaussian models* (Arnaud et al., 2005; Doucet et al., 2000) for the formalization of the dynamic system. These models are characterized by a non linear dynamic equation, a linear measurement equation and additive Gaussian noises. Such a choice—which consists in constraining the system to rely on a simple and rough linear measurement model—may appear, at first sight, to be too restrictive. However, we argue that a partial linear Gaussian model is an interesting alternative that should be considered for 2D visual tracking. These models allow the use of the optimal importance function for the diffusion step. The resulting algorithms are simple and efficient as the particles explore the state space in an optimal way. When a pertinent estimation of the measurement noise covariance is possible, this optimal particle's repartition enables the trackers to be resistant to occlusions and ambiguities, without defining specific frameworks for these difficult situations. In addition, this setup allows expressing in a simple way a validation gate that defines a bounded search region where the measurements are looked for at each new image.

As for the limitations of the use of these partial linear Gaussian models, they are mainly of two types. First, the definition of a linear measurement equation is required. This can be simply done as soon as some feature candidates may be detected. Secondly it implies to have an informative dynamic equation to counterbalance the roughly modeled likelihood. In this paper, we propose a solution to the problem of building such a dynamic in the most general case, when there is a lack of *a priori* information on the dynamic of the tracked object. We propose to tackle this problem by defining the dynamic equation from instantaneous on-line motion estimation.

The remainder of the paper is organized as follow. In Section 2, we recall the filtering problem formulation and make an overview of the existing solutions with a specific attention to sequential Monte Carlo methods. In Section 3, we will focus on the use of these methods in the framework of tracking in image sequences. A brief state of the art of the possible models of the dynamic system will be given. Then in the three following sections, the partial linear Gaussian models will be studied through three instances of increasing complexity. These three cases will allow us to show how the use of such models may be of real interest facing occlusions (Section 4), ambiguities (Section 5) and large state space (Section 6) The last section will add some remarks on the computational complexity.

To illustrate our work, a feature point tracker and a planar object tracker have been designed. They both rely on systems that are entirely estimated on the image sequences. In particular, the model that is considered for point tracking combines a dynamic equation relying on an instantaneous motion estimated with the optical flow constraint and measurements provided by a matching technique. The defined model for the object tracking is a direct extension of the latter, including a geometric

constraint. They both are built on a partial linear Gaussian formalization leading to consider optimal particle filters. Some results on real-world sequences will demonstrate the efficiency of this framework facing difficult situations. Let us finally remark that this paper follows a work presented in Arnaud and Mémin (2004) and Arnaud et al. (2005).

2. Filtering Problem and Sequential Monte Carlo Methods

In this section, the general principal of filtering problems is briefly introduced. A quick overview of the different algorithms is given, stressing the different possible choices of importance function.

Let us consider a discrete hidden Markov state process $\mathbf{x}_{0:n} = \{\mathbf{x}_0, \mathbf{x}_1, \ldots, \mathbf{x}_n\}$ of transition equation $p(\mathbf{x}_k \mid \mathbf{x}_{k-1})$. The set of observations $\mathbf{z}_{1:n} = \{\mathbf{z}_1, \mathbf{z}_1, \ldots, \mathbf{z}_n\}$, of marginal distribution $p(\mathbf{z}_k \mid \mathbf{x}_k)$, are supposed conditionally independent given the state sequence. 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 $\mathbf{z}_{1:k}$. The distribution of interest $p(\mathbf{x}_k \mid \mathbf{z}_{1:k})$ is called the *filtering distribution*. The Bayesian recursive solution known as *optimal filter* consists in two steps:

prediction step: $p(\mathbf{x}_k \mid \mathbf{z}_{1:k-1})$

$$= \int p(\mathbf{x}_k \mid \mathbf{x}_{k-1})\; p(\mathbf{x}_{k-1} \mid \mathbf{z}_{1:k-1})\, d\mathbf{x}_{k-1} \qquad (1)$$

filtering step: $p(\mathbf{x}_k \mid \mathbf{z}_{1:k})$

$$= \frac{p(\mathbf{z}_k \mid \mathbf{x}_k)\; p(\mathbf{x}_k \mid \mathbf{z}_{1:k-1})}{\int p(\mathbf{z}_k \mid \mathbf{x}_k)\; p(\mathbf{x}_k \mid \mathbf{z}_{1:k-1})\, d\mathbf{x}_k}. \qquad (2)$$

Once an approximation of the filtering distribution is known, an estimate of the state can be obtained through the use of *maximum a posteriori* or *minimum mean square error* estimates.

The tracking recursion yields closed-form expressions only for specific cases. The most well-known case is the Kalman filter (Anderson and Moore, 1979) for linear Gaussian models. Non optimal extensions of the Kalman filter, based on a Gaussian approximation of the filtering distribution (Extended Kalman filter, Unscented Kalman filter (Wan and van der Merwe, 2000)), have been devised for non linear systems. In the general multimodal case, such an approximation is not satisfactory. For general non-linear and non-Gaussian models, the recent development of sequential Monte Carlo approaches (Arulampalam et al., 2002; Doucet et al., 2001; Gordon et al., 1993; Liu and Chen, 1998) has lead to new efficient algorithms. Sequential Monte Carlo methods have been widely used for a large variety of applications. Among them, one can find bearings-only tracking, positioning and navigation (Gustafsson et al., 2002), speech modeling and enhancement (Vermaak et al., 2002), surface-based registration (Ma and Ellis, 2004), mobile robotics (Kwok et al., 2004) or tracking in image sequences (Arnaud et al., 2005; Isard and Blake, 1998; Pérez et al., 2004).

The idea behind particle filtering is very simple. These techniques propose to implement recursively an approximation of the sought density $p(\mathbf{x}_k \mid \mathbf{z}_{1:k})$. This approximation consists in a finite weighted sum of N Diracs centered on hypothesized locations in the state space—called particles—of the initial system $\mathbf{x}_0$. At each particle $\mathbf{x}_k^{(i)}$ $(i = 1 : N)$ is assigned a weight $w_k^{(i)}$ describing its relevance. This approximation can be formulated with the following expression:

$$p(\mathbf{x}_k \mid \mathbf{z}_{1:k}) \approx \sum_{i=1:N} w_k^{(i)} \delta_{\mathbf{x}_k^{(i)}}(\mathbf{x}_k).00 \qquad (3)$$

Assuming that the approximation of $p(\mathbf{x}_{k-1} \mid \mathbf{z}_{1:k-1})$ is known, the recursive implementation of the filtering distribution is done by propagating the swarm of weighted particles $\{\mathbf{x}_{k-1}^{(i)}, w_{k-1}^{(i)}\}_{i=1:N}$. At each time instant (or iteration), the set of new particles $\{\mathbf{x}_k^{(i)}\}_{i=1:N}$ is drawn from an approximation of the true distribution $p(\mathbf{x}_k \mid \mathbf{z}_{1:k})$, called the *importance function* and denoted $\pi(\mathbf{x}_k \mid \mathbf{x}_{0:k-1}^{(i)}, \mathbf{z}_{1:k})$. The closer the approximation to the true distribution, the more efficient the filter. The importance weights $w_k^{(i)}$ account for the deviation w.r.t. the unknown true distribution. To maintain a consistent sample, the importance weights are updated according to a recursive evaluation as the new measurement $\mathbf{z}_k$ becomes available:

$$w_k^{(i)} \propto w_{k-1}^{(i)} \frac{p\left(\mathbf{z}_k \mid \mathbf{x}_k^{(i)}\right)\, p\left(\mathbf{x}_k^{(i)} \mid \mathbf{x}_{k-1}^{(i)}\right)}{\pi\left(\mathbf{x}_k^{(i)} \mid \mathbf{x}_{0:k-1}^{(i)}, \mathbf{z}_{1:k}\right)}, \quad \sum_{i=1:N} w_k^{(i)} = 1. \qquad (4)$$

Limiting ourself to these two steps for updating the swarm of particles induces an increase over time of the weight variance. In practice, this *degeneracy problem* makes the number of significant particles decreases dramatically over time implying an impoverishment of the estimate (Kong et al., 1994). From time to time, it is thus necessary to perform a resampling step. This procedure aims at removing particles with weak normalized weights, and multiplying particles associated to strong weights, as soon as the number of significant particles is too small. Consequently, resampled particles tend to be concentrated in areas where important features exist. Various resampling strategies have been described in the literature but will not be developed in this paper. More details can be found in Doucet et al. (2000). These three main steps (sampling/calculation of the importance

weights/resampling) constitute the general framework of particle filtering. Historically, the first proposed particle filter including a resampling step has been built with the following rules: (a) to set the importance function to the evolution law, i.e. $\pi(\mathbf{x}_k \mid \mathbf{x}^{(i)}_{0:k-1}, \mathbf{z}_{1:k}) = p(\mathbf{x}_k \mid \mathbf{x}^{(i)}_{k-1})$ and (b) to proceed the resampling step at each iteration. This scheme corresponds to the Bootstrap filter (Gordon et al., 1993) (also called the CONDENSATION algorithm in the computer vision community (Isard and Blake, 1998)).

It is clear that adding a resampling procedure improves the quality of the estimates by reducing the degeneracy problem. However, unnecessary resampling may introduce its own challenge as samples with higher probability may be oversampled, and regions corresponding to secondary modes of the filtering distribution may be not well explored. Another strategy whose aim is to reduce the degeneracy problem consists in using an *optimal importance function* which minimizes the variance of the weights conditioned upon $\mathbf{x}_{0:k-1}$ and $\mathbf{z}_{1:k}$. It is then possible to prove that choosing:

$$\pi\left(\mathbf{x}_k \mid \mathbf{x}^{(i)}_{0:k-1}, \mathbf{z}_{1:k}\right) = p\left(\mathbf{x}_k \mid \mathbf{x}^{(i)}_{k-1}, \mathbf{z}_k\right) \tag{5}$$

corresponds to this optimal choice (Doucet et al., 2000). With this distribution, the recursive formulation of w_k becomes:

$$w^{(i)}_k \propto w^{(i)}_{k-1}\, p\left(\mathbf{z}_k \mid \mathbf{x}^{(i)}_{k-1}\right), \quad \sum_{i=1:N} w^{(i)}_k = 1. \tag{6}$$

The resulting particle filter is called the *Optimal particle filter*. This algorithm requires to be able to sample from $p(\mathbf{x}_k \mid \mathbf{x}^{(i)}_{k-1}, \mathbf{z}_k)$ and to evaluate $p(\mathbf{z}_k \mid \mathbf{x}^{(i)}_{k-1})$ up to a proportionality constant, which is rarely possible in practice. However, analytic expressions of these two required densities can be written for particular models, including *partial linear Gaussian models* (Doucet et al., 2000). These systems are characterized by a linear measurement equation and additive Gaussian noises. To the best of our knowledge, such systems have not been used for vision applications apart from a point tracking application in one of our previous work (Arnaud et al., 2005). This kind of models will be deeply studied in this paper.

The choice of the importance function is of crucial importance for the quality of the particle filter estimates. As a consequence, the goal of more recent approaches is to design efficient importance functions approximating as closely as possible the optimal one, and to guide the particles in high likelihood areas. These approaches aim also at introducing the measurements into the sampling step. Let us cite for example the *Auxiliary particle filter* (Pitt and Shephard, 2001), *Extended particle filter* (Doucet et al., 2000), *Unscented particle filter* (van der Merwe et al., 2000), and the hybrid filters combining a particle filter and Monte Carlo Markov Chain methods (Godsill and Clapp, 2001; Musso et al., 2001).

3. Tracking in Image Sequences Using Sequential Monte Carlo methods

In this section, we focus on the use of sequential Monte Carlo methods for tracking in image sequences. Different possible approaches to model the evolution law and the likelihood are given. In particular, we point at the fact that the usual model is based on a highly nonlinear likelihood, and a simple linear evolution law. We introduce the partial non linear Gaussian models, that rely on the opposite choice, and can be solved with the Optimal particle filter. To the best of our knowledge, these models have not been applied for tracking in image sequences, apart in one of our previous work (Arnaud and Mémin, 2004; Arnaud et al., 2005).

As said before, visual tracking with sequential Monte Carlo methods requires to model the problem by a system composed of an evolution law $p(\mathbf{x}_k \mid \mathbf{x}_{k-1})$ and a likelihood $p(\mathbf{z}_k \mid \mathbf{x}_k)$. Facing a given application, the choice of the algorithm depends on the system's characteristics (linear or not, Gaussian or not, etc.).

The dynamic model accounts for the *prior* on the state evolution. A precise knowledge of the tracked feature's nature allows defining an accurate evolution law. When a physical law of the feature motion is known, this model can be directly inferred. A typical example is the use of the Navier-Stokes equation as the evolution law for the tracking of fluid structures (Cuzol and Mémin, 2005). When this is not the case, an *a priori* evolution model can be obtained by learning. A large number of contributions on this subject have been given by the *Visual Dynamics Research Group* (University of Oxford) (Blake et al., 1999; North and Blake, 1998). In particular, such dynamics have been widely used in the CONDENSATION algorithm framework. This type of evolution laws can be of great interest for tracking features whose motion may be described by a limited number of classes (e.g. tracking of a bouncing ball, of a drawing hand, etc.) (Isard and Blake, 1998; Rittscher and Blake, 1999). When the number of classes characterizing the feature's evolution is too high, an alternative consists in using a database of examples describing the steps of the different possible evolutions of the state (Sidenbladh et al., 2002). 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, it is clear that a known state model is not always available. This is particularly the case when tracking very general entities in images of any kind. The most used strategy

consists then in adopting a very weak model for the state evolution, typically auto-regressive models (Black and Fleet, 1999; Pérez et al., 2002, 2004).

Such crude dynamic laws are usually counterbalanced by sound models of likelihood, evaluating the relevance of an observation given the feature state. The nature of the observation is directly linked to a representation of the tracked feature. This representation corresponds to a set of cues that are obtained from the image sequence and that enables detecting or identifying the object of interest. The choice of a good representation is of crucial importance. Indeed, it has to be persistent along the sequence not to loose the object, and at the same time discriminant to avoid ambiguity situations. The involved cues are usually edge information (Isard and Blake, 1998; Wu and Huang, 2004), color distributions (Comaniciu et al., 2003; Pérez et al., 2002, 2004; Wu and Huang, 2004), motion (Vermaak et al., 2002; Pérez et al., 2004), and appearance-based model (Jepson et al., 2001; Nguyen et al., 2001; Sullivan and Rittscher, 2001). Recent works have widely studied the interest of fusing different cues to enhance the representation quality (Pérez et al., 2004; Wu and Huang, 2004). Schematically, a reference representation is associated to the feature of interest. This reference model is then compared to similar models from candidate image regions. The likelihood accounts then for a distance measurement between the representations. Such likelihood are usually highly multimodal and non-linear.

A direct consequence of using these sound measurement models is that the implemented filter can not be the optimal particle filter. Indeed, let us remind that using the optimal importance function requires to be able to sample from $p(\mathbf{x}_k \mid \mathbf{x}_{k-1}^{(i)}, \mathbf{z}_k)$ and to evaluate $p(\mathbf{z}_k \mid \mathbf{x}_{k-1}^{(i)}) = \int p(\mathbf{z}_k \mid \mathbf{x}_k)\ p(\mathbf{x}_k \mid \mathbf{x}_{k-1}^{(i)})\ d\mathbf{x}_k$ up to a proportionality constant. An analytic form if this integral will not be available in the general case. Using the dynamic model as the importance function is then a common assumption, leading to implement the Bootstrap filter. This algorithm has the advantage of being very simple but it may be inefficient because the last available measurement is not taken into account in the particle diffusion. The state space may be badly explored as the particles do not optimally cover the areas of high likelihood.

Different approaches have been proposed to address these issues, including the use of previously cited algorithms. For example, the unscented particle filter has been applied to face tracking (Rui and Chen, 2001), and a hybrid filter has been adapted for human tracking (Sminchisescu and Triggs, 2002). Other methods, specific to the computer vision community, have been developed to enhance the quality of the state space exploration. They all aim at introducing some image based data to guide the particles. For instance, one approach is to build the importance function as a mixture law of the dynamic model and Gaussian laws centered in some positions of the state space that have been selected by a detection module (Okuma et al., 2004; Pérez et al., 2004; Vermaak et al., 2002). Other possible strategies aim at fusing a deterministic evolution of the particles with a probabilistic sampling (Sullivan and Rittscher, 2001), or fusing observations of several image cues during the sampling step (Pérez et al., 2004; Wu and Huang, 2004).

However, as pointed out in Doucet et al. (2000), analytic evaluation of the densities $p(\mathbf{x}_k \mid \mathbf{x}_{k-1}, \mathbf{z}_k)$ and $p(\mathbf{z}_k \mid \mathbf{x}_{k-1})$ (required for the use of the optimal particle filter) is possible considering partial linear Gaussian models. As said before, these models are characterized by a linear measurement equation. A linear measurement model can be defined if candidate features may be detected and if these observations constitute components of the state vector. The likelihood does not account any more for a distance between the reference representation and the candidate representations, but for a distance between the measurement and the candidates. As a consequence, the relevance of this measurement model strongly depends on the candidate detection module. In that case, it may be interesting to define a region of interest where to run the candidate detection module. Such a region is called a *validation gate*, and is defined as an area of the measurement space where the next observation has a high probability to appear. Moreover, to counterbalance a possible deficiency of the candidate detection, an informative evolution law has to be considered. But, as said before, defining such a dynamic equation is difficult when facing the tracking of general entities whose evolution is not known in advance. To avoid this problem, an original construction of the dynamic has been proposed in Arnaud et al. (2005). In such a context, one possibility consists in relying on a dynamic model extracted from the image sequence. Such a data-driven 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 state evolution in a simple way. Let us remark that this solution is valid facing a 2D tracking problem. For example, the developed strategy, that consists in relying on an instantaneous estimated motion can not be directly extended to a 3D tracking problem, as a 3D velocity estimation would be required.

4. Dealing with Occlusions with a Simple Model

In this section, we study the simplest instance of partial linear Gaussian models. This analysis is done in a methodological point a view: the associated optimal particle filter and validation gate are detailed. To validate the fact that such a model handles nicely the occlusion problem, a point tracker is designed. This study has already been described in Arnaud and Mémin

(2004), but is recalled here to facilitate the reading of the following.

Let us consider the simplest version of partial linear Gaussian models composed of a non linear dynamic equation with an additive Gaussian noise, and a linear Gaussian measurement equation:

$$\mathbf{x}_k = f_{k\,|\,\mathbf{I}}\,(\mathbf{x}_{k-1}) + \mathbf{w}_k, \quad \mathbf{w}_k \sim \mathcal{N}(\mathbf{w}_k; 0,\, Q_{k\,|\,\mathbf{I}}) \tag{7}$$

$$\mathbf{z}_k = H_{k\,|\,\mathbf{I}}\,\mathbf{x}_k + \mathbf{v}_k, \quad \mathbf{v}_k \sim \mathcal{N}(\mathbf{v}_k; 0,\, R_{k\,|\,\mathbf{I}}). \tag{8}$$

The state noise $\mathbf{w}_k$ and the measurement noise $\mathbf{v}_k$ are supposed to be independent. The subscript $|\mathbf{I}$ on the state function $f_{k\,|\,\mathbf{I}}$, the measurement matrix $H_{k\,|\,\mathbf{I}}$, and the noise covariances $Q_{k\,|\,\mathbf{I}}$ and $R_{k\,|\,\mathbf{I}}$ indicates that these matrices and the function may be estimated from the image sequence.[1] This system can be written as:

$$\begin{cases} p(\mathbf{x}_k \,|\, \mathbf{x}_{k-1}) = \mathcal{N}(\mathbf{x}_k;\, f_{k\,|\,\mathbf{I}}\,(\mathbf{x}_{k-1}),\;\; Q_{k\,|\,\mathbf{I}}) \\ p(\mathbf{z}_k \,|\, \mathbf{x}_k) = \mathcal{N}(\mathbf{z}_k;\, H_{k\,|\,\mathbf{I}}\,\mathbf{x}_k,\;\; R_{k\,|\,\mathbf{I}}). \end{cases} \tag{9}$$

As introduced in Section 2, this system enables the use of the optimal particle filter. As a matter of fact, noticing that:

$$p(\mathbf{z}_k \,|\, \mathbf{x}_{k-1}) = \int p(\mathbf{z}_k \,|\, \mathbf{x}_k)\; p(\mathbf{x}_k \,|\, \mathbf{x}_{k-1})\; d\mathbf{x}_k \tag{10}$$

and

$$p(\mathbf{x}_k \,|\, \mathbf{x}_{k-1}, \mathbf{z}_k) = \frac{p(\mathbf{z}_k \,|\, \mathbf{x}_k)\; p(\mathbf{x}_k \,|\, \mathbf{x}_{k-1})}{p(\mathbf{z}_k \,|\, \mathbf{x}_{k-1})}, \tag{11}$$

we obtain the exact expression for the distributions required in the optimal particle filter (Doucet et al., 2000):

$$p\big(\mathbf{z}_k \,|\, \mathbf{x}_{k-1}^{(i)}\big) = \mathcal{N}\big(\mathbf{z}_k; H_{k\,|\,\mathbf{I}}\; f_{k\,|\,\mathbf{I}}\,\big(\mathbf{x}_{k-1}^{(i)}\big), R_{k\,|\,\mathbf{I}} + H_{k\,|\,\mathbf{I}}\; Q_{k\,|\,\mathbf{I}}^{(i)}\; H_{k\,|\,\mathbf{I}}^t\big) \tag{12}$$

$$p\big(\mathbf{x}_k \,|\, \mathbf{x}_{k-1}^{(i)}, \mathbf{z}_k\big) = \mathcal{N}(\mathbf{x}_k; \mathbf{m}_{k\,|\,\mathbf{I}},\;\; \Sigma_{k\,|\,\mathbf{I}}) \tag{13}$$

$$\text{where} \quad \Sigma_{k\,|\,\mathbf{I}} = \big({Q_{k\,|\,\mathbf{I}}^{(i)}}^{-1} + H_{k\,|\,\mathbf{I}}^t\; R_{k\,|\,\mathbf{I}}^{-1}\; H_{k\,|\,\mathbf{I}}\big)^{-1} \tag{14}$$

$$\mathbf{m}_{k\,|\,\mathbf{I}} = \Sigma_{k\,|\,\mathbf{I}}\,\big({Q_{k\,|\,\mathbf{I}}^{(i)}}^{-1}\; f_{k\,|\,\mathbf{I}}\,\big(\mathbf{x}_{k-1}^{(i)}\big) + H_{k\,|\,\mathbf{I}}^t\; R_{k\,|\,\mathbf{I}}^{-1}\; \mathbf{z}_k\big). \tag{15}$$

Therefore, the densities involved in the diffusion process and in the update step are Gaussian. These two stages are therefore particularly simple to implement. Let us note that the index (i) on matrix $Q_{k\,|\,\mathbf{I}}$ is referred to the case where the dynamic equation is estimated on line for each particle.

4.1. Validation Gate

This simple model is valid when only one measurement is observed at each iteration. In practice, several measurements may be available. Each of these observations may have been generated by the hidden state or may be false alarms. Taking into account a false alarm in the update of the filtering density may dramatically decrease the quality of the estimate, and may cause the loss of the tracked entity. To solve the problem of distinguishing the state-originated measurement from false alarms, a validation gate need to be introduced. Such a region is defined as an area of the measurement space where the future observation will be found with some high probability. At each iteration, it is thus defined through the probability distribution $p(\mathbf{z}_k \,|\, \mathbf{z}_{1:k-1})$. Gates are generally used in radar tracking problems, for clutter reduction (Bar-Shalom and Li, 1995). 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. Selecting a too small gate size may lead to miss the state-originated measurement, whereas selecting a too large size is computationally expensive and increases the probability of selecting a false observation.

For linear Gaussian systems, an analytic expression of the Gaussian distribution $p(\mathbf{z}_k \,|\, \mathbf{z}_{1:k-1})$ may be obtained. This enables considering an exact 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 and Carriquiry (2000) proposes to use Monte Carlo simulations in order to approximate the density $p(\mathbf{z}_k \,|\, \mathbf{z}_{1:k-1})$, but this solution appears to be time consuming.

In case of partial linear Gaussian models, it is possible to infer an ellipsoidal validation gate in a simple way. As a matter of fact, from an approximation of the filtering density $p(\mathbf{x}_{k-1} \,|\, \mathbf{z}_{1:k-1})$ with a swarm of particles $\{\mathbf{x}_{k-1}^{(i)}, w_{k-1}^{(i)}\}_{i=1\ldots N}$, we can write:

$$\begin{aligned} p(\mathbf{z}_k \,|\, \mathbf{z}_{1:k-1}) &= \int p(\mathbf{z}_k \,|\, \mathbf{x}_{k-1})\; p(\mathbf{x}_{k-1} \,|\, \mathbf{z}_{1:k-1})\; d\mathbf{x}_{k-1} \\ &\simeq \sum_{i=1}^{N} w_{k-1}^{(i)}\; p\big(\mathbf{z}_k \,|\, \mathbf{x}_{k-1}^{(i)}\big) \end{aligned} \tag{16}$$

This equation may be exploited as soon as an analytical expression of $p(\mathbf{z}_k \,|\, \mathbf{x}_{k-1}^{(i)})$ is known. Considering the simplest version of partial linear Gaussian models (9), we have:

$$p\big(\mathbf{z}_k \,|\, \mathbf{x}_{k-1}^{(i)}\big) = \mathcal{N}\big(\mathbf{z}_k; H_{k\,|\,\mathbf{I}}\; f_{k\,|\,\mathbf{I}}\,\big(\mathbf{x}_{k-1}^{(i)}\big), R_{k\,|\,\mathbf{I}} + H_{k\,|\,\mathbf{I}}\; Q_{k\,|\,\mathbf{I}}^{(i)}\; H_{k\,|\,\mathbf{I}}^t\big). \tag{17}$$

Observing Eqs. (16) and (17), it appears that the exact validation gate is composed of the union of N ellipses, where each ellipse is associated to a particle. Obtaining such a validation gate is computationally too expensive for a large amount of particles. We propose here to approximate this region by a unique ellipse as illustrated by Fig. 1.

To this end, we have to compute the two first moments of $p(\mathbf{z}_k \mid \mathbf{z}_{1:k-1})$ denoted $\mathbf{m}_k^{vg}$ and Σ_k^{vg}. Empirical approximations of these quantities can be easily derived using (16–17). The expression of the first moment can be written as (the notation $\mathbb{E}$ denotes the expectation):

$$\begin{aligned}\mathbf{m}_k^{vg} &= \mathbb{E}[\mathbf{z}_k \mid \mathbf{z}_{1:k-1}] \simeq \sum_{i=1}^{N} w_{k-1}^{(i)} \int \mathbf{z}_k \; p\big(\mathbf{z}_k \,\big|\, \mathbf{x}_{k-1}^{(i)}\big) \\ &= \sum_{i=1}^{N} w_{k-1}^{(i)} \; H_{k\,|\,\mathbf{I}} \; f_{k\,|\,\mathbf{I}}\left(\mathbf{x}_{k-1}^{(i)}\right), \end{aligned} \tag{18}$$

Using the relation between the expectation and the variance $var[\mathbf{a}] = \mathbb{E}[\|\mathbf{a}\|^2] - \|\mathbb{E}[\mathbf{a}]\|^2$ as for the second moment, we obtain:

$$\begin{aligned}\Sigma_k^{vg} &= var[\mathbf{z}_k \mid \mathbf{z}_{1:k-1}] \qquad (19)\\ &\simeq \sum_{i=1}^{N} w_{k-1}^{(i)} \Big[R_{k\,|\,\mathbf{I}} + H_{k\,|\,\mathbf{I}} \; Q_{k\,|\,\mathbf{I}}^{(i)} \; H_{k\,|\,\mathbf{I}}^{t} \\ &\quad + \left\| H_{k\,|\,\mathbf{I}} \; f_{k\,|\,\mathbf{I}}\left(\mathbf{x}_{k-1}^{(i)}\right)\right\|^2\Big] \\ &\quad - \left\| \sum_{i=1}^{N} w_{k-1}^{(i)} \; H_{k\,|\,\mathbf{I}} \; f_{k\,|\,\mathbf{I}}\left(\mathbf{x}_{k-1}^{(i)}\right)\right\|^2, \qquad (20)\end{aligned}$$

and we have the following approximation:

$$p(\mathbf{z}_k \mid \mathbf{z}_{1:k-1}) \simeq \mathcal{N}\big(\mathbf{z}_k; \mathbf{m}_k^{vg}, \Sigma_k^{vg}\big). \tag{21}$$

This gives us an expression of the ellipsoidal region corresponding to the validation gate $\mathcal{V}_k$:

$$\mathcal{V}_k = \big\{\mathbf{z}_k : \big(\mathbf{z}_k - \mathbf{m}_k^{vg}\big)^t \; \Sigma_k^{vg\;-1} \; \big(\mathbf{z}_k - \mathbf{m}_k^{vg}\big) \le \gamma\big\}. \tag{22}$$

The parameter γ is chosen in practice such as the probability of finding the state-originated observation is equal to 0.99. It is important to outline that the expression of the validation gate depends on the error covariance matrices $R_{k\,|\,\mathbf{I}}$ and $Q_{k\,|\,\mathbf{I}}^{(i)}$.

4.2. *How Does the Model React to an Occlusion of the Tracked Feature ?*

Even if the model studied is very simple, its good properties enable us to deal with occlusions in a simple and elegant manner. This is all the more true if an accurate estimation of the measurement noise covariance matrix $R_{k\,|\,\mathbf{I}}$ is possible. In that case, if the feature of interest is occluded, the observed measurement is detected as a false alarm and the values of $R_{k\,|\,\mathbf{I}}$ are significantly increased. The particle propagation depends thus mainly on the dynamic model. But as soon as a pertinent observation is detected, the small values of the matrix $R_{k\,|\,\mathbf{I}}$ attract the particles in the neighborhood of the measurement, i.e. in the area of highest likelihood. This automatic scheme for recovering the location of the feature after an occlusion is due to the particle sampling with the optimal importance function $p(\mathbf{x}_k \mid \mathbf{x}_{k-1}^{(i)}, \mathbf{z}_k)$, as well as to the automatic increasing of the validation gate when false alarms are detected.

In order to demonstrate experimentally the significance of partial linear Gaussian model, we have formalized a point tracking application with such a setting. We present in the next section the basic model we considered. Although remaining in a partial linear Gaussian setting, this model will be enriched further to deal with clutter noise and ambiguities or to handle the tracking of point clouds.

4.3. *Application to Point Tracking*

As many others (Aschwanden and Guggenbül, 1992; Shi and Tomasi, 1994), our point tracker is designed on the basis of luminance pattern consistency. Here each state $\mathbf{x}_k$ represents the location of the point projection at time k, in image $\mathbf{I}_k$. The basic partial linear Gaussian model we propose combines a dynamic relying on a differential optical-flow method and measurements obtained using a correlation criterion. The system we focus on is therefore composed of measurements and dynamic equations which both depend on $\mathbf{I}_{0:k}$.

4.3.1. Dynamic Model and State Noise Covariance Matrix Estimation. The motion of a point $\mathbf{x}_{k-1}$ between the frame instants $k-1$ and k is defined through the probability distribution $p(\mathbf{x}_k \mid \mathbf{x}_{k-1})$. To be reactive to any change of speed and direction of the point, we propose to define the state equation as:

$$\mathbf{x}_k = \mathbf{x}_{k-1} + \mathbf{u}_k(\mathbf{x}_{k-1}) + \mathbf{w}_k, \tag{23}$$

where $\mathbf{w}_k$ is assumed to be a the zero-mean Gaussian white noise of covariance $Q_{k\,|\,\mathbf{I}}$. This noise covariance matrix describes the relevance of the defined model and $\mathbf{u}_k(\mathbf{s})$ denotes the motion vector associated to a given pixel $\mathbf{s} = (x, y)^t$. This motion vector is estimated from a robust parametric technique (Odobez and Bouthemy, 1995). Such a technique enables us to reliably compute a 2D parametric model representing the dominant image motion within the considered fixed support $\mathcal{R}$. More precisely, the motion vector of a point $\mathbf{s} \in \mathcal{R}$ between the

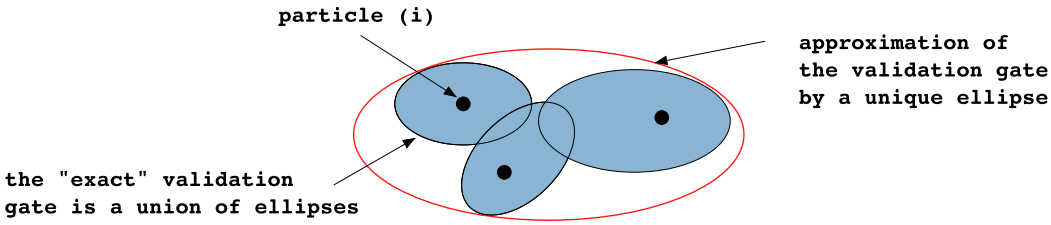


Figure 1. Validation gate for model (9) ; example with three particles. To each particle $\mathbf{x}_{k-1}^{(i)}$ is associated a blue ellipse that corresponds to $p(\mathbf{z}_k \mid \mathbf{x}_{k-1}^{(i)})$. The exact validation gate is the union of these ellipses. It is approximated by one unique ellipse, represented in red.

images $\mathbf{I}_{k-1}$ and $\mathbf{I}_k$ is modeled as a polynomial function of the point coordinates:

$$\mathbf{u}_k(\mathbf{s}) = P(\mathbf{s})\, \boldsymbol{\theta}_{k\,|\,\mathbf{I}}(\mathbf{s}), \tag{24}$$

where $\boldsymbol{\theta}_{k\,|\,\mathbf{I}}(\mathbf{s})$ denotes the parameter vector composed of the unknown polynomial's coefficients. $P(\mathbf{s})$ is a known matrix related to the chosen parametric model and whose entries depend on the spatial coordinates x and y. For example, a 6-parameter affine motion model is associated to the following matrix:

$$P(\mathbf{s}) = \begin{bmatrix} 1 & x & y & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & x & y \end{bmatrix}. \tag{25}$$

The parameter vector $\boldsymbol{\theta}_{k\,|\,\mathbf{I}}(\mathbf{s})$ is estimated through the minimization of a function defined using the brightness consistency assumption:

$$\boldsymbol{\theta}_{k\,|\,\mathbf{I}}(\mathbf{s}) = \arg\min_{\boldsymbol{\theta}} \int_{\mathcal{R}} \rho\left(\nabla \mathbf{I}_k^t(\mathbf{s}) P(\mathbf{s}) \boldsymbol{\theta} + \frac{\partial \mathbf{I}_k}{\partial t}(\mathbf{s}) \right) d\mathbf{s}, \tag{26}$$

where $\mathcal{R}$ is the estimation support and ρ is a robust cost function allowing to cope with outliers (occlusion area, secondary motions, etc.) (Black and Anandan, 1996; Odobez and Bouthemy, 1995). The minimization is achieved through a Gauss-Newton-type multiresolution procedure that allows handling large magnitude motions. Choosing an affine motion model and defining the support $\mathcal{R}$ as a small neighborhood around $\mathbf{x}_{k-1}$ provides a robust way to compute an instantaneous motion vector $\mathbf{u}_k(\mathbf{x}_{k-1})$ of the tracked feature $\mathbf{x}_{k-1}$. Our non linear dynamic model is thus specified on-line thanks to this instantaneous motion vector and reads finally:

$$\mathbf{x}_k = \mathbf{x}_{k-1} + P(\mathbf{x}_{k-1})\, \boldsymbol{\theta}_{k\,|\,\mathbf{I}}(\mathbf{x}_{k-1}) + \mathbf{w}_k. \tag{27}$$

Let us note that the non linearity of such a dynamic is due to the dependence of $\boldsymbol{\theta}_{k\,|\,\mathbf{I}}$ to $\mathbf{x}_{k-1}$ through the choice of the estimation support $\mathcal{R}$ as a region centered in $\mathbf{x}_{k-1}$. This model may be written as:

$$p(\mathbf{x}_k \mid \mathbf{x}_{k-1}) = \mathcal{N}(\mathbf{x}_k; \mathbf{x}_{k-1} + P(\mathbf{x}_{k-1})\boldsymbol{\theta}_{k\,|\,\mathbf{I}}(\mathbf{x}_{k-1}),\, Q_{k\,|\,\mathbf{I}}). \tag{28}$$

On top on providing the parameters describing the motion on the fixed support $\mathcal{R}$, the use of a *robust* estimation procedure provides the number of outliers and inliers involved in the estimation of $\boldsymbol{\theta}_{k\,|\,\mathbf{I}}$ within $\mathcal{R}$ (Black and Anandan, 1996; Odobez and Bouthemy, 1995). This information can be used to evaluate the parameters estimation quality and to infer the noise covariance matrix $Q_{k\,|\,\mathbf{I}}$. Indeed, as said before, this matrix carries information on the relevance of the defined state model, and indirectly describes the estimation accuracy of the motion $\mathbf{u}_k(\mathbf{x}_{k-1})$. Let us denote σ_k the ratio between the number of outliers and the number of inliers involved in the estimation of $\boldsymbol{\theta}_{k\,|\,\mathbf{I}}$. The analysis of σ_k along the tracking process (in time) can be exploited as a motion estimation quality indicator. This gives us a criterion to fix the dynamic noise covariance $Q_{k\,|\,\mathbf{I}} = \alpha \times Id$ (Id is the identity matrix): a small value is assigned to the variable α if σ_k is close to 1. At the opposite a very large one is set if this ratio decreases. The switch decisions are done automatically, studying the evolution of σ_k in time. More precisely, the switch decisions depend on the increasing/decreasing jumps of σ_k and are supplied by a statistical test (Page-Hinkley test (1971)).

4.3.2. Measurement Model and Measurement Noise Covariance Matrix Estimation. Whereas the evolution law defines the model of the instantaneous feature motion, the measurement equation $p(\mathbf{z}_k \mid \mathbf{x}_k)$ will allow us to fix a goodness of fit criterion between a reference point and possible candidates in 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 the initial one $\mathbf{x}_0$ 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 of the point's neighborhood. In case of

large geometric and/or photometric deformations around the tracked feature this reference image is eventually updated by registration (see Arnaud et al. (2005) for more details). 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) on a window $\mathcal{W}(\mathbf{x}_0)$ (of fixed size) centered in $\mathbf{x}_0$. At first sight, the use of the SSD can seem not relevant, compared to other correlation criterion (ZSSD, ZNSSD, CC, ZNCC (Aschwanden and Guggenbül, 1992)) that take into account the usual difficulties occurring in natural video like changing of viewpoints, lighting condition, etc. However, this choice is justified by the possibility of estimating a measure of quality of the SSD result. This procedure, further explained below, takes into account the image noise through a statistical test that could not be proceed for other criterion (Arnaud et al., 2005). The measurement $\mathbf{z}_k$ is achieved such as:

$$\mathbf{z}_k = \arg\min_{\mathbf{z}\in\mathcal{V}_k} \underbrace{\sum_{\mathbf{m}\in\mathcal{W}(\mathbf{x}_0)} [\widetilde{\mathbf{I}}_0(\mathbf{m}) - \mathbf{I}_k(\mathbf{z}+\mathbf{m})]^2}_{r_k(\mathbf{z})}, \quad (29)$$

where $\mathcal{V}_k$ is the validation gate (see Section 4.1 p. 10) and $r_k(\mathbf{z})$ is the SSD residual. It is assumed that this measurement carries enough pieces of information about the state of the tracked point to be able to write that $\mathbf{x}_k = \mathbf{z}_k$ up to an additional white Gaussian noise $\mathbf{v}_k$. This noise variable models a local estimation error and accounts for a confidence information on this matching. The observation equation reads then:

$$p(\mathbf{z}_k \mid \mathbf{x}_k) = \mathcal{N}(\mathbf{z}_k; \mathbf{x}_k, R_{k\mid\mathbf{I}}). \quad (30)$$

A good estimation of the matrix $R_{k\mid\mathbf{I}}$ is essential to make the tracker robust to corrupted observations due for instance to large geometric or photometric deformations, occlusions, etc. To that end, an SSD surface is defined around the correlation peak. This surface corresponds to an error distribution. Two treatments are applied to this surface in order to `(a)` suppress the SSD values due to the image noise using a statistical test and `(b)` enforce the shape of the surface using an exponential function. A statistical Chi-square "goodness of fit" test is then performed on the modified surface in order to check whether such a surface is better approximated by a Gaussian or an uniform law. In this latter case, the associated covariance is fixed to $\infty \times Id$. In the Gaussian case, it is set to the empirical covariance calculated on the modified surface. For an interested reader, more details are available in Arnaud et al. (2005). A description of the final tracker is given by Algorithm 1.

Algorithm 1 feature point tracker based on the simplest version of partial linear Gaussian models

- **initialization:** for $i = 1 : N$, generate $\mathbf{x}_0^{(i)} \sim p(\mathbf{x}_0)$, and fix $w_0^{(i)} = 1/N$ for $k = 1, 2, \ldots$
- **estimations on the image sequence:**
  1. detection of $\mathbf{z}_k$ in the validation gate $\mathcal{V}_k$ with the SSD using $\widetilde{\mathbf{I}}_0$ and estimation of $R_{k\mid\mathbf{I}}$
  2. for $i = 1 : N$, estimation of $\boldsymbol{\theta}_{k\mid\mathbf{I}}(\mathbf{x}_{k-1}^{(i)})$ and $Q_{k\mid\mathbf{I}}^{(i)}$ using a robust parametric estimation technique on a small neighborhood around $\mathbf{x}_{k-1}^{(i)}$
- **sequential importance sampling:**
  1. sampling: for $i = 1 : N$, generate $\mathbf{x}_k^{(i)} \sim p(\mathbf{x}_k \mid \mathbf{x}_{k-1}^{(i)}, \mathbf{z}_k)$ with

$$\begin{aligned} p\big(\mathbf{x}_k \mid \mathbf{x}_{k-1}^{(i)}, \mathbf{z}_k\big) &= \mathcal{N}(\mathbf{x}_k; \mathbf{m}_{k\mid\mathbf{I}},\ \Sigma_{k\mid\mathbf{I}}) \\ \Sigma_{k\mid\mathbf{I}} &= \big(Q_{k\mid\mathbf{I}}^{(i)\,-1} + R_{k\mid\mathbf{I}}^{-1}\big)^{-1} \\ \mathbf{m}_{k\mid\mathbf{I}} &= \Sigma_{k\mid\mathbf{I}}\big(Q_{k\mid\mathbf{I}}^{(i)-1}\big[\mathbf{x}_{k-1}^{(i)} + P\big(\mathbf{x}_{k-1}^{(i)}\big) \\ &\qquad \boldsymbol{\theta}_{k\mid\mathbf{I}}\big(\mathbf{x}_{k-1}^{(i)}\big)\big] + R_{k\mid\mathbf{I}}^{-1}\,\mathbf{z}_k\big). \end{aligned}$$

  2. calculation of importance weights: for $i = 1 : N$, calculate $w_k^{(i)} = p(\mathbf{z}_k \mid \mathbf{x}_k^{(i)})$, with

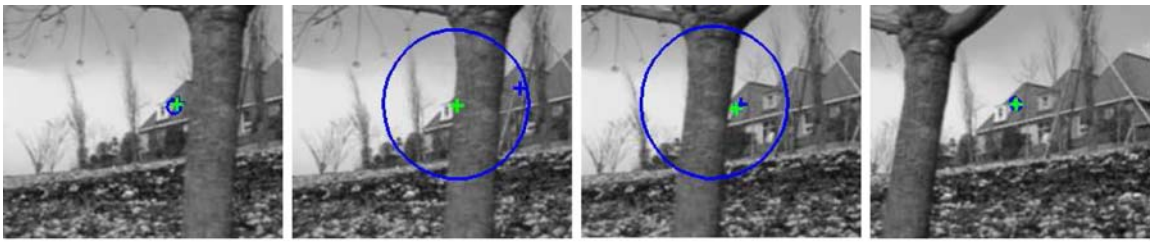
$$\begin{aligned} p\big(\mathbf{z}_k \mid \mathbf{x}_k^{(i)}\big) = \mathcal{N}\big(\mathbf{z}_k; \mathbf{x}_{k-1}^{(i)} + P\big(\mathbf{x}_{k-1}^{(i)}\big)\,\boldsymbol{\theta}_{k\mid\mathbf{I}}\big(\mathbf{x}_{k-1}^{(i)}\big), \\ R_{k\mid\mathbf{I}} + Q_{k\mid\mathbf{I}}^{(i)}\big) \quad \sum_{i=1:N} w_k^{(i)} = 1. \end{aligned}$$

- **estimation of the feature position**
- **eventual update of the reference image** $\widetilde{\mathbf{I}}_0$
- **resampling if necessary**

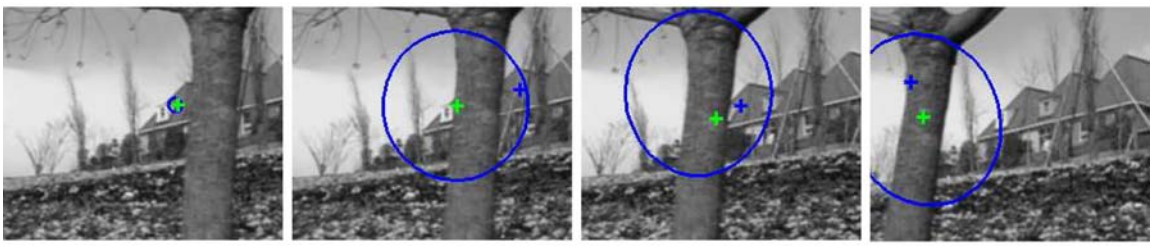
4.4. Experimental Results

In this section, we present some experimental results demonstrating the efficiency of the proposed point tracker that rely on the simplest instance of partial linear Gaussian models.

Comparison with the CONDENSATION *Algorithm.* The goal of the first result presented is to enhance the interest of diffusing the particles with 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 specific feature point located on the top of a house roof. This point is visible in the first two images and is hidden by the tree 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 4.3), but have different particle diffusion processes. The first one is the method we propose,



Proposed tracker, using the optimal importance function (frames 2,6,16,30)



CONDENSATION algorithm, without optimal importance function (frames 2,6,16,30)

Figure 2. Garden sequence—Interest of the optimal importance function and of the validation gate—the green cross indicates the state estimate, the blue cross corresponds to the observation and the ellipse to the validation gate.

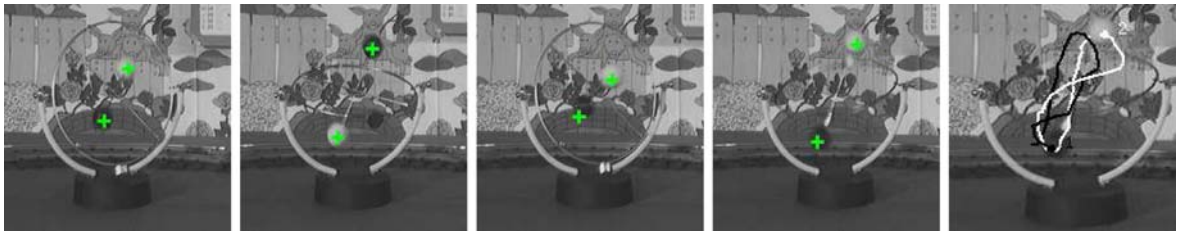


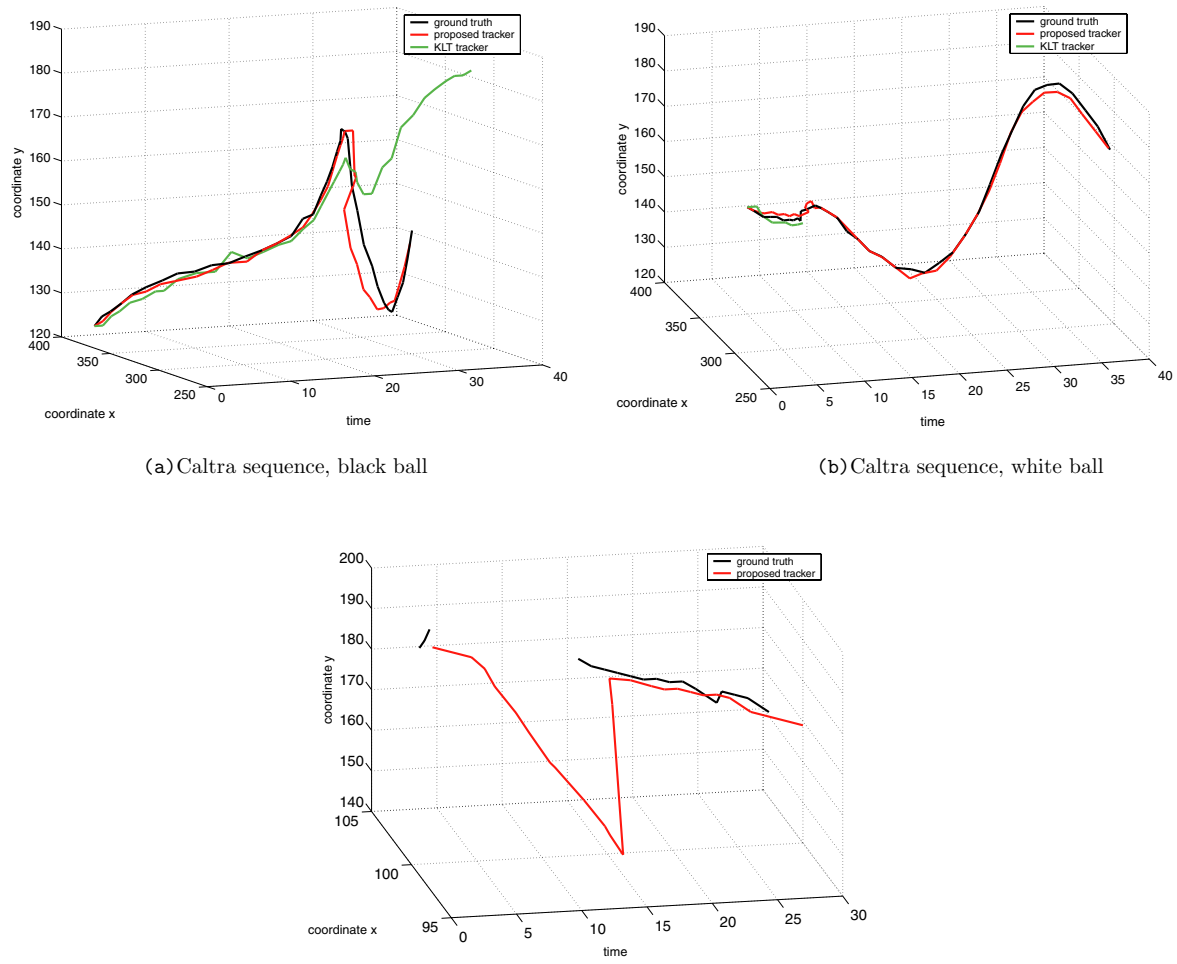
Figure 3. Caltra sequence—obtained result on a real-world sequence (frames 8,21,31,40 and obtained trajectories)—average trajectories over the successful trajectories on 100 realizations.

based on a partial linear Gaussian model. It allows the exploration of the state space with the optimal importance sampling by taking into account the new measurement in an optimal way. The second one is the CONDENSATION algorithm, for which the considered importance function is identified to the evolution law. Figure 2 presents the obtained results. The use of the optimal importance function allows us to recover the actual point location after a long occlusion. On the contrary, the CONDENSATION algorithm is not efficient for such situations, when the evolution density and the likelihood do not overlap. This figure illustrates also the importance of our validation gate: the research region of the next observation is considerably reduced as soon as the the confidence on the state estimation is high.

Validation on a Real-World Sequence. A result of the proposed tracker 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. As it can be noticed on Fig. 3 the tracker succeeds in discriminating the balls from the wallpaper, and provides the exact trajectories. Such a result shows the ability of this tracker to deal with complex trajectories in a cluttered environment. The trajectories account for the average realization over the successful trajectories on 100 realizations (Only 2 failures have been observed).

In order to validate our algorithm, a comparative study has been done. In particular, the proposed point tracker has been compared with the KLT tracker (Shi and Tomasi, 1994) and with ground truth trajectories extracted manually. The results are displayed for the black and white balls of the caltra sequence (Fig. 4(a–b)), and for the point on the roof for the garden sequence (Fig. 4(c)). On these figures, the trajectories can be observed and compared. Other comparative results are available in Arnaud et al. (2005) showing that our algorithm is robust and avoids any target's drift along the image sequence.

In Case of Strong Similarities Between Several Image Patches. The last result illustrates the limitation of the



(a) Caltra sequence, black ball

(b) Caltra sequence, white ball

(c) Garden sequence, roof point

Figure 4. Comparison of estimated trajectories and ground truth trajectories. The results obtained with our tracker account for the mean realizations over the successful trajectories on 100 Monte Carlo runs.

simplest version of partial linear Gaussian models. This result has been obtained on the 31-frames sequence **Concorde** and corresponds to the average trajectories over 100 realizations. This sequence presents the road traffic on a roundabout located in Paris. This sequence exhibits a lot of ambiguities due to photometric similarities of several cars. It presents also illumination changes when the cars go by the shadow area in the right part of the images. The tracking results are presented in Fig. 5. From this figure, it can be observed that the tracker fails in recovering the complete trajectory of the upper truck and of several cars (see for instance the merging of the two car trajectories located in the bottom part of the images or the mis-estimation of the car trajectories located in the shadow area after image #22). These failures can be explained by the fact that measurements corresponding to the highest correlation peaks are in some cases false alarms. They do not correspond to the right observation. As it will be explained in the next section, this situation can be solved by considering a more elaborated partial linear Gaussian model that consider several measurements per state.

5. Dealing with Ambiguities (Multi-Hypothesis Tracking)

In this section, we study the instance of partial linear Gaussian models that we designed for the case of multi-hypothesis tracking. As before, a methodological analysis is presented (algorithm, validation gate). One of the main contribution of this section is the multi-hypothesis model of the likelihood, which has the particularity of being a simple mixture of Gaussian laws. The original corresponding filter is applied and validated for point tracking in case of ambiguities.

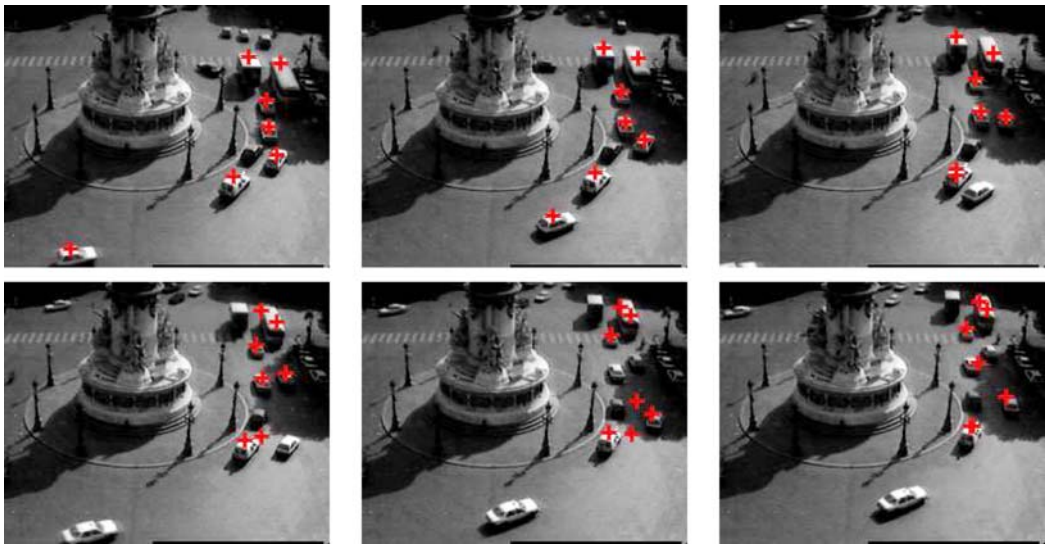


Figure 5. Concorde sequence (frames 1,10,17,22,27,30)—average trajectories over 100 realizations.

As shown in the latter result (Fig. 5), considering only one single observation may induce a mis-estimation of the state. In particular, in case of noisy images with potential cluttered background, the measurement detection module may be disturbed by the generated ambiguities and may produce false alarms. In addition to the use of a validation gate, a way to overcome this problem is to consider several measurements (hypothesis) simultaneously per state—Let us note that all the considered observations have to belong to the validation gate in order to be the possible state-originated measurement. The resulting problem of *data association* (i.e. determining the state-originated measurement from the false alarms) has to be solved jointly with the tracking problem. [2] A probabilistic formulation of the possible associations can be expressed through the following association hypotheses:

H1: A measurement can be generated by the feature of interest or be a false alarm,
H2: At a given instant, the feature can generate zero or only one measurement.

At time k, Let us note $\mathbf{z}_k$ the vector gathering the M_k measurements-hypothesis $\{\mathbf{z}_{k,m}\}_{m=1:M_k}$. Let also $\Phi_k \in \{0 : M_k\}$ denote the random variable describing the association events:

$$\Phi_k = \begin{cases} j \text{ if the hypothesis } \mathbf{z}_{k,j} \text{ corresponds to the state-originated measurement,} \\ 0 \text{ if the } M_k \text{ measurements-hypothesis are false alarms.} \end{cases} \tag{31}$$

The combination of the assumptions **H1** and **H2** expresses the fact that the possible associations described by the variable Φ_k are exhaustive and mutually exclusive. Using the total probability theorem, the likelihood of the model can be written as a mixture law:

$$\begin{aligned} p(\mathbf{z}_k \mid \mathbf{x}_k) &= \sum_{j=0}^{M_k} p(\Phi_k = j \mid \mathbf{x}_k)\, p(\mathbf{z}_k \mid \Phi_k = j, \mathbf{x}_k) \\ &= \sum_{j=0}^{M_k} \beta_{k,j}\, p(\mathbf{z}_k \mid \Phi_k = j, \mathbf{x}_k) \end{aligned} \tag{32}$$

$$\text{with } \beta_{k,j} \triangleq p(\Phi_k = j \mid \mathbf{x}_k) \quad j = 0 : M_k. \tag{33}$$

In order to derive analytical expressions of these two terms several assumptions have to be done.

First of all, we assume that (a) the measurements-hypothesis are conditionally independent:

$$p(\mathbf{z}_k \mid \Phi_k = j, \mathbf{x}_k) = \prod_{m=1}^{M_k} p(\mathbf{z}_{k,m} \mid \Phi_k = j, \mathbf{x}_k). \tag{34}$$

We also assume that, (b) considering the measurement-hypothesis $\mathbf{z}_{k,m}$ as being the state-originated observation, its conditional likelihood is a Gaussian law of mean $h_{k\,|\,\mathbf{I},m}(\mathbf{x}_k)$ and covariance $R_{k\,|\,\mathbf{I},m}$, and (c) false alarms are uniformly distributed over the validation gate at time k. Consequently, we have:

$$\begin{aligned} &p(\mathbf{z}_{k,m} \mid \Phi_k = j, \mathbf{x}_k) \\ &\quad = \begin{cases} \mathcal{N}(\mathbf{z}_{k,m}; h_{k\,|\,\mathbf{I},m}(\mathbf{x}_k), R_{k\,|\,\mathbf{I},m}) & j = m \\ V_k^{-1} & j = 0 : M_k; j \neq m, \end{cases} \end{aligned} \tag{35}$$

where V_k is the total area of the validation gate. From (32), (34) and (35), the likelihood reads as a mixture of

uniform and Gaussian laws:

$$p(\mathbf{z}_k \mid \mathbf{x}_k) = \sum_{j=0}^{M_k} \left[\beta_{k,j} \prod_{m=1}^{M_k} p(\mathbf{z}_{k,m} \mid \Phi_k = j, \mathbf{x}_k) \right]$$
$$= \beta_{k,0} V_k^{-M_k} + \sum_{j=1}^{M_k} \left[\beta_{k,j} V_k^{1-M_k} \mathcal{N}(\mathbf{z}_{k,j}; h_{k|\mathbf{I},j}(\mathbf{x}_k), R_{k|\mathbf{I},j}) \right]. \quad (36)$$

This model to deal with clutter has been introduced in Gordon (1997) and presented for several problems like bearings only tracking (Marrs et al., 2002), or tracking in image sequences (Isard and Blake, 1998). Since in our case the measurements provided by a detection module are linearly linked to the state this expression becomes:

$$p(\mathbf{z}_k \mid \mathbf{x}_k) = \beta_{k,0} V_k^{-M_k} + \sum_{j=1}^{M_k} \left[\beta_{k,j} \; V_k^{1-M_k} \right.$$
$$\left. \mathcal{N}(\mathbf{z}_{k,j}; H_{k \mid \mathbf{I},j} \, \mathbf{x}_k, R_{k \mid \mathbf{I},j}) \right]. \quad (37)$$

As the process $\mathbf{x}_k$ is unknown, a direct estimation of $\beta_{k,j} \triangleq p(\Phi_k = j \mid \mathbf{x}_k)$ is obviously impossible. A first solution (Isard and Blake, 1998) consists in approximating $p(\Phi_k = j \mid \mathbf{x}_k)$ by the *a priori* law $p(\Phi_k = j)$ and to assume the equi-probability of the associations. This leads to the following estimation:

$$\hat{\beta}_{k,j} = \frac{1 - \beta_{k,0}}{M_k} \quad \forall j = 1 : M_k, \quad (38)$$

where $\beta_{k,0}$ is a parameter to be fixed. Recalling that $\beta_{k,0}$ represents the probability of having no target-originated measurements, we can interpret $\beta_{k,0}$ as being the probability of having the target occluded. Such a probability is indeed very difficult to predict. A second solution (Marrs et al., 2002) is to approximate $\beta_{k,j}$ by $p(\Phi_k = j \mid \mathbf{z}_{1:k})$ and to consider the common model of false alarms proposed by Bar-Shalom and Li (1995) that supposes the number of false alarms distributed according to a Poisson law. The resulting expressions of $\beta_{k,j}$ are complex and still involve a parameter similar to $\beta_{k,0}$.

In addition to the difficulty of choosing a relevant value for this parameter $\beta_{k,0}$, these solutions do not allow the use of the importance optimal function. As a matter of fact, the uniform term of the likelihood prevents to get the analytical expressions of the distributions $p(\mathbf{z}_k \mid \mathbf{x}_{k-1})$ and $p(\mathbf{x}_k \mid \mathbf{x}_{k-1}, \mathbf{z}_k)$. We propose thus to set to zero the probability of having no state-originated measurement (i.e. $\beta_{k,0} = 0$). Such a choice differs from the classical tracking assumptions. It may sound strange and potentially problematic in case of occlusions. Nevertheless, we believe that such a deficiency can in practice be well compensated by an efficient estimation of the measurement noise covariances $R_{k \mid \mathbf{I},j}$. As a matter of fact, facing an occlusion, if these covariance matrices are set to $\infty \times Id$, none of the observations will be taken into account. This is equivalent to set the probability of having no state-originated measurement to a positive value. As for the set of mode occurrence probabilities $\beta_{k,j}$, $j \neq 0$, we propose to estimate them at each instant from the images. This will be further described in Section 5.2. The final likelihood that we consider consists thus in a mixture of Gaussian laws:

$$p(\mathbf{z}_k \mid \mathbf{x}_k) = \; V_k^{1-M_k} \; \sum_{j=1}^{M_k} \beta_{k,j} \, \mathcal{N}(\mathbf{z}_{k,j}; H_{k \mid \mathbf{I},j} \mathbf{x}_k, R_{k \mid \mathbf{I},j}). \quad (39)$$

We can remark that this model, associated to the same evolution model considered before (Eq. (7)) is a direct multimodal extension of the model proposed in case of a unique observation (9). It is also a partial linear Gaussian model which can be handled efficiently with the optimal particle filter. Let us remind that the diffusion process considered requires to evaluate $p(\mathbf{z}_k \mid \mathbf{x}_{k-1})$ and to sample from $p(\mathbf{x}_k \mid \mathbf{x}_{k-1}, \mathbf{z}_k)$. Applying the identity (10), the density used for the weight recursion is a mixture of Gaussians and reads:

$$p\big(\mathbf{z}_k \mid \mathbf{x}_{k-1}^{(i)}\big) = V_k^{1-M_k} \sum_{j=1}^{M_k} \beta_{k,j} \mathcal{N}\big(\mathbf{z}_{k,j}; H_{k \mid \mathbf{I},j} \, f_{k \mid \mathbf{I}}\big(\mathbf{x}_{k-1}^{(i)}\big), R_{k \mid \mathbf{I},j} + H_{k \mid \mathbf{I},j} \; Q_{k \mid \mathbf{I}}^{(i)} \; H_{k \mid \mathbf{I},j}^t\big). \quad (40)$$

The optimal importance function is deduced using equality (11) (proof available in annexe (Section 9)):

$$p\big(\mathbf{x}_k \mid \mathbf{x}_{k-1}^{(i)}, \mathbf{z}_k\big) = \sum_{j=1}^{M_k} \beta_{k,j} \frac{\alpha_{k,j}}{S_k} \mathcal{N}(\mathbf{x}_k; \mathbf{m}_{k \mid \mathbf{I},j}, \Sigma_{k \mid \mathbf{I},j}), \quad (41)$$

with

$$S_k = \sum_{j=1}^{M_k} \beta_{k,j} \, \mathcal{N}\big(\mathbf{z}_{k,j}; H_{k \mid \mathbf{I},j} \; f_{k \mid \mathbf{I}} \left(\mathbf{x}_{k-1}^{(i)}\right), R_{k \mid \mathbf{I},j} + H_{k \mid \mathbf{I},j} \; Q_{k \mid \mathbf{I}}^{(i)} \; H_{k \mid \mathbf{I},j}^t\big), \quad (42)$$

$$\Sigma_{k \mid \mathbf{I},j} = \left({Q_{k \mid \mathbf{I}}^{(i)}}^{-1} + H_{k \mid \mathbf{I},j}^t \; R_{k \mid \mathbf{I},j}^{-1} \; H_{k \mid \mathbf{I},j} \right)^{-1}, \quad (43)$$

$$\mathbf{m}_{k \mid \mathbf{I},j} = \Sigma_{k \mid \mathbf{I},j} \big({Q_{k \mid \mathbf{I}}^{(i)}}^{-1} f_{k \mid \mathbf{I}} \left(\mathbf{x}_{k-1}^{(i)}\right) + H_{k \mid \mathbf{I},j}^t R_{k \mid \mathbf{I},j}^{-1} \mathbf{z}_{k,j} \big), \quad (44)$$

$$\alpha_{k,j} = C \exp\left(-\frac{1}{2} \big[\left\| f_{k \mid \mathbf{I}} \left(\mathbf{x}_{k-1}^{(i)}\right) \right\|^2_{{Q_{k \mid \mathbf{I}}^{(i)}}^{-1}} + \|\mathbf{z}_{k,j}\|^2_{R_{k \mid \mathbf{I},j}^{-1}} - \|\mathbf{m}_{k \mid \mathbf{I},j}\|^2_{\Sigma_{k \mid \mathbf{I},j}^{-1}} \big] \right), \quad (45)$$

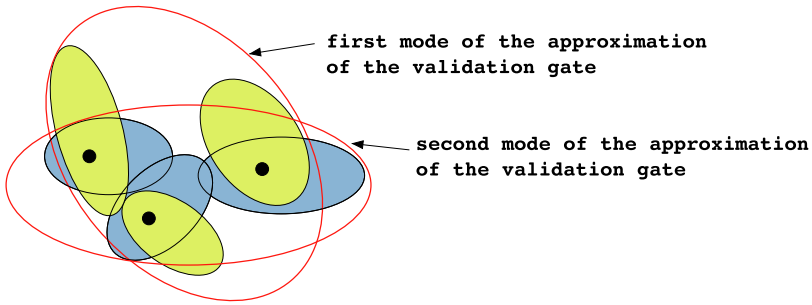


Figure 6. Validation gate for the model dedicated to the ambiguities; example with three particles and two measurements. To each particle $\mathbf{x}_{k-1}^{(i)}$ is associated a blue ellipse that corresponds to $p(\mathbf{z}_{k,1} \mid \mathbf{x}_{k-1}^{(i)})$ (linked to the first observation) and a yellow ellipse that corresponds to $p(\mathbf{z}_{k,2} \mid \mathbf{x}_{k-1}^{(i)})$ (linked to the second observation). The exact validation gate is the union of these 3×2 ellipses. It is approximated by two ellipses, represented in red. These two red ellipses correspond to an approximation of the union of the ellipses respectively blue and yellow, and are associated to the observations respectively 1 and 2.

$$\text{and where } C = (2\pi)^{-\frac{n_z}{2}} \mid \Sigma_{k \mid \mathbf{I},j} \mid^{\frac{1}{2}} \left| Q_{k \mid \mathbf{I}}^{(i)} \right|^{-\frac{1}{2}} \mid R_{k \mid \mathbf{I},j} \mid^{-\frac{1}{2}}. \tag{46}$$

In the last expression n_z denotes the dimension of the measurement vector $\mathbf{z}_k$. It can be easily proved that the expression (41) is a Gaussian mixture (the idea is to prove that the sum of the coefficients $\sum_{j=1}^{M_k} \beta_{k,j} \frac{\alpha_{k,j}}{S_k}$ is equal to 1). The implementation of the optimal particle filter through expressions (40–41) remains simple as the involved distributions are all mixtures of Gaussian laws. In practice, for each particle, the implementation of the sampling step is done as follows: first, a Gaussian law of (41) is selected (random draw over the Gaussians weights), and then the new state of the particle is drawn from it. As the selected Gaussian law is associated to one measurement-hypothesis, then the weight of the new particle is calculated, knowing the chosen measurement-hypothesis.

The partial linear Gaussian system presented in this section is particularly interesting as it permits both to consider several observations and to diffuse the particles with the optimal importance function. As it will be shown in the experimental section, such a system enables to solve the ambiguous situations generated by a cluttered background.

5.1. Validation Gate

The multimodal partial linear Gaussian model proposed here enables us to take into account several measurements-hypothesis. These measurements all belong to the validation gate area. As in the previous simple case, a relevant validation gate needs to be defined at each iteration through the distribution $p(\mathbf{z}_k \mid \mathbf{z}_{1:k-1})$. The same approach is applied here to get an approximation of this distribution. Using (16) and (40), we have:

$$p(\mathbf{z}_k \mid \mathbf{z}_{1:k-1}) \simeq V_k^{1-M_k} \sum_{i=1}^{N} w_{k-1}^{(i)} \left[\sum_{j=1}^{M_k} \beta_{k,j} \mathcal{N}\big(\mathbf{z}_{k,j}; H_{k \mid \mathbf{I},j}\ f_{k \mid \mathbf{I}}\left(\mathbf{x}_{k-1}^{(i)}\right), R_{k \mid \mathbf{I},j} + H_{k \mid \mathbf{I},j}\ Q_{k \mid \mathbf{I}}^{(i)}\ H_{k \mid \mathbf{I},j}^t\big) \right] \tag{47}$$

It can be immediately deduced that the exact validation gate is composed of the union of $N \times M_k$ ellipses, where N is the number of particles and M_k the number of available observations at time k. This expression is computationally too expensive and has to be simplified. In the same spirit as before and as illustrated in Fig. 6, the approximation we propose is to define the validation area $\mathcal{V}_k$ by M_k ellipses (or modes). Each ellipse Ψ_j corresponds to observation j.

$$\mathcal{V}_k = \bigcup_{j=1}^{M_k} \Psi_j; \Psi_j = \{\mathbf{z}_{k,j} : \left(\mathbf{z}_{k,j} - \mathbf{m}_{k,j}^{vg}\right)^t \Sigma_{k,j}^{vg\ -1} \left(\mathbf{z}_{k,j} - \mathbf{m}_{k,j}^{vg}\right) \leq \gamma_j\}. \tag{48}$$

The value γ_j is fixed in practice as the 99th percentile of the probability for $\mathbf{z}_{k,j}$ to be the *true* state-originated measurement. The expressions of the parameters $\mathbf{m}_{k,j}^{valid}$ and $\Sigma_{k,j}^{vg}$ can be deduced from the simple monomodal case:

$$\mathbf{m}_{k,j}^{vg} \simeq \sum_{i=1}^{N} w_{k-1}^{(i)}\ H_{k \mid \mathbf{I},j}\ f_{k \mid \mathbf{I}}\left(\mathbf{x}_{k-1}^{(i)}\right), \tag{49}$$

$$\Sigma_{k,j}^{vg} \simeq \sum_{i=1}^{N} w_{k-1}^{(i)} \Big[R_{k \mid \mathbf{I},j} + H_{k \mid \mathbf{I},j} Q_{k \mid \mathbf{I}}^{(i)} H_{k \mid \mathbf{I},j}^t + \left\| H_{k \mid \mathbf{I},j}\ f_{k \mid \mathbf{I}}\left(\mathbf{x}_{k-1}^{(i)}\right) \right\|^2 \Big] - \left\| \sum_{i=1}^{N} w_{k-1}^{(i)} H_{k \mid \mathbf{I},j} f_{k \mid \mathbf{I}}\left(\mathbf{x}_{k-1}^{(i)}\right) \right\|^2 \tag{50}$$

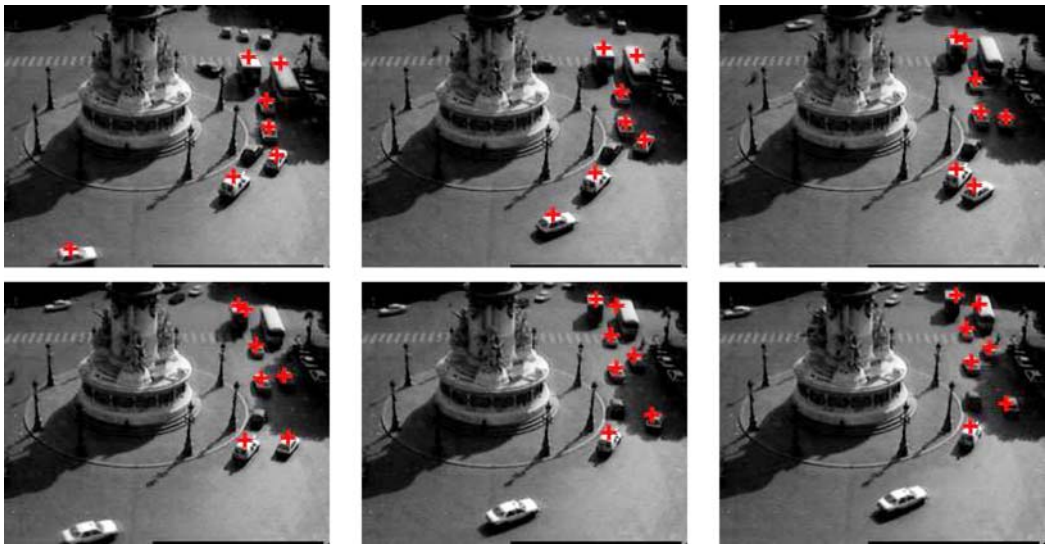


Figure 7. Concorde sequence—tracker dedicated to ambiguities, with three measurements per tracked feature (frames 1, 10, 17, 22, 27, 30)—average trajectories over 100 realizations.

5.2. *Application to Point Tracking*

To validate the relevance of the multimodal partial linear Gaussian model in presence of ambiguities, we have extended the monomodal point tracker previously described. This new tracker relies on:

- The same on-line evolution law:

$$p(\mathbf{x}_k \mid \mathbf{x}_{k-1}) = \mathcal{N}(\mathbf{x}_k; \mathbf{x}_{k-1} + P(\mathbf{x}_{k-1})\boldsymbol{\theta}_{k\mid\mathbf{I}}(\mathbf{x}_{k-1}),\ Q_{k\mid\mathbf{I}}). \tag{51}$$

  The motion parameter and the noise covariance matrix $Q_{k\mid\mathbf{I}}$ are estimated in the same way as before (i.e. from a local robust least squares process and a Page-Hinkley statistical test see Section 4.3.1)
- A vector of M_k observations $\mathbf{z}_k = \{\mathbf{z}_{k,m}\}_{m=1:M_k}$ of the state defined as being the points of the current image that are the most similar to the point of interest in a reference template. These measurements are the result of a matching process and correspond to the highest correlation peaks. They all belong to the validation area. The expression of the likelihood is:

$$p(\mathbf{z}_k \mid \mathbf{x}_k) \propto \sum_{j=1}^{M_k} \beta_{k,j}\, \mathcal{N}(\mathbf{z}_{k,j}\ ;\ \mathbf{x}_k,\ R_{k\mid\mathbf{I},j}). \tag{52}$$

  The coefficient $\beta_{k,j}$ denotes the probability that the observation j is the target-originated observation at time k. In our application, these coefficients are chosen proportionally to the correlation score of the associated measurement, such as $\sum_{j=1}^{M_k} \beta_{k,j} = 1$.[3] The effect of that procedure is to give more weight to the measurements with high correlation. However, in the sampling step of the algorithm, the measurement-hypothesis that will be favored will not necessarily be the one corresponding to the highest correlation score. Indeed, the choice of the favored direction takes into account the value of $\{\beta_{k,j}\}_{j=1:M_k}$ but also the coherence between the measurements-hypothesis and the dynamic model (see Eq. (41)).
  Each measurement noise covariance matrix $R_{k\mid\mathbf{I},j}$ is estimated on line. These estimations are done following the scheme presented for the previous simple tracker, i.e. by modeling the correlation surface as a probability distribution of the true match location.

The whole method embedded into an optimal particle filter implementation is briefly described in Algorithm 2.

5.3. *Experimental Results*

Experimental results are presented in order to enhance the interesting features of our multimodal point tracker. The used real-world sequences exhibit numerous photometric ambiguities.

The first result concerns the sequence **Concorde** shown previously and for which the monomodal point tracker failed (cf. Fig. 5). As it can be noticed from Fig. 7, considering three correlation hypothesis per feature improves significantly the result quality. The whole trajectories are now successfully recovered. Maintaining several correlation modes enables in particular to recover the right target even if a mis-estimation has been momentar-

Algorithm 2 feature point tracker based on a partial linear Gaussian model dedicated to ambiguities

- **initialization:** for $i = 1 : N$, generate $\mathbf{x}_0^{(i)} \sim p(\mathbf{x}_0)$, and fix $w_0^{(i)} = 1/N$ for $k = 1, 2, \ldots$
- **estimations on the image sequence:**
  1. detection of $\{\mathbf{z}_{k,j}\}_{j=1:M_k}$ with SSD criterion in the validation gate $\mathcal{V}_k$ using $\widetilde{\mathbf{I}}_0$, estimation of $\{R_{k\,|\,\mathbf{I},j}\}_{j=1:M_k}$ and $\{\beta_{k,j}\}_{j=1:M_k}$
  2. for $i = 1 : N$, estimation of $\boldsymbol{\theta}_{k\,|\,\mathbf{I}}(\mathbf{x}_{k-1}^{(i)})$ and $Q_{k\,|\,\mathbf{I}}^{(i)}$ using a robust parametric estimation technique on a small neighborhood around $\mathbf{x}_{k-1}^{(i)}$
- **sequential importance sampling:**
  1. sampling: for $i = 1 : N$, generate $\mathbf{x}_k^{(i)} \sim p(\mathbf{x}_k \,|\, \mathbf{x}_{k-1}^{(i)}, \mathbf{z}_k)$ with

$$p\big(\mathbf{x}_k \,|\, \mathbf{x}_{k-1}^{(i)}, \mathbf{z}_k\big) = \sum_{j=1}^{M_k} \beta_{k,j} \frac{\alpha_{k,j}}{S_k} \mathcal{N}(\mathbf{x}_k; \mathbf{m}_{k\,|\,\mathbf{I},j}, \Sigma_{k\,|\,\mathbf{I},j})$$

$$\Sigma_{k\,|\,\mathbf{I},j} = \big(Q_{k\,|\,\mathbf{I}}^{(i)-1} + R_{k\,|\,\mathbf{I},j}^{-1}\big)^{-1}$$

$$\mathbf{m}_{k\,|\,\mathbf{I},j} = \Sigma_{k\,|\,\mathbf{I},j}\big(Q_{k\,|\,\mathbf{I}}^{(i)-1}\big[\mathbf{x}_{k-1}^{(i)} + P\big(\mathbf{x}_{k-1}^{(i)}\big) \boldsymbol{\theta}_{k\,|\,\mathbf{I}}\big(\mathbf{x}_{k-1}^{(i)}\big)\big] + R_{k\,|\,\mathbf{I},j}^{-1}\mathbf{z}_{k,j}\big)$$

     The expressions of $\alpha_{k,j}$ and S_k correspond respectively to (45) and (42).

  2. calculation of importance weights: for $i = 1 : N$, calculate $w_k^{(i)} = p(\mathbf{z}_k \,|\, \mathbf{x}_k^{(i)})$ with

$$p\big(\mathbf{z}_k|\mathbf{x}_k^{(i)}\big) \propto \sum_{j=1}^{M_k} \beta_{k,j} \mathcal{N}\big(\mathbf{z}_{k,j}; \mathbf{x}_{k-1}^{(i)} + P\big(\mathbf{x}_{k-1}^{(i)}\big)\boldsymbol{\theta}_{k|\mathbf{I}}\big(\mathbf{x}_{k-1}^{(i)}\big), R_{k\,|\,\mathbf{I},j} + Q_{k\,|\,\mathbf{I}}^{(i)}\big) \quad \sum_{i=1:N} w_k^{(i)} = 1.$$

- **estimation of the feature position**
- **eventual update of the reference image** $\widetilde{\mathbf{I}}_0$
- **resampling if necessary**

ily done (such a behavior can be noticed for the point located on the bus in the upper part of the image).

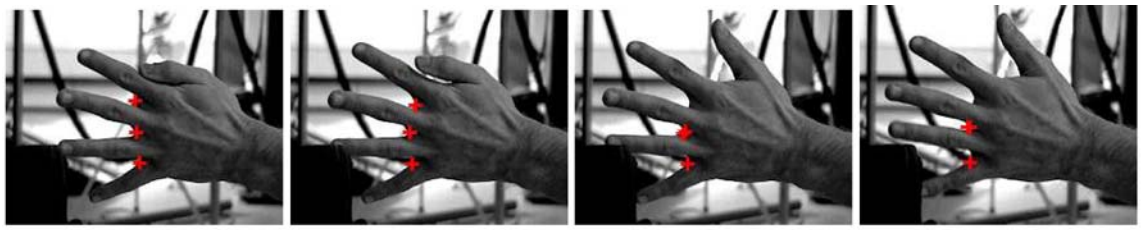
A second comparison is given on the 20-frame sequence **Hand**. Three target points have been fixed on finger intersections. These points are quite close to each other and correspond to highly similar photometric patterns. The multimodal and monomodal trackers have been run both on this sequence. The results are presented in Figs. 8(a) and (c) respectively. It can be noticed in Fig. 8(a) that the results provided by the monomodal tracker are not satisfactory. The feature located between the forefinger and the middle finger is wrongly tracked (cf. the 2 last images). The tracking of this point is studied in more details in Fig. 8(b). On this illustration are represented the particles (in yellow), the point estimate (in red) and the measurement (in blue). The tracking failure is clearly due to a false alarm generated by a high correlation score. This high correlation score causes the setting of small values in the noise covariance matrix and induces finally the diffusion of the particles around this wrong observation.

On the contrary, the results obtained with the extended version of the tracker considering three measurements per feature are correct (cf. Fig. 8(c)): the tracking is not perturbed anymore by false alarms. As in Fig. 8(b), Fig. 8(d) focuses on the tracking of the point between the index and middle fingers. On this illustration are represented the particles (in yellow), the point estimate (in red), the measurement of highest correlation score (in blue) and the two secondary observations (in green). Paying attention to image 15, it can be noticed that the main measurement is located on a wrong position whereas it is a secondary observation associated to a lower correlation peak which actually corresponds to the state. In that case, let us remind that the optimal importance function is a mixture of Gaussian laws whose coefficients depend on the correlation peaks but also on the distance between these measurements and the prediction. Such a weighting is naturally handled within a diffusion based on the optimal importance function permits here to estimate the correct feature location.

6. Dealing with Large State Space Dimension

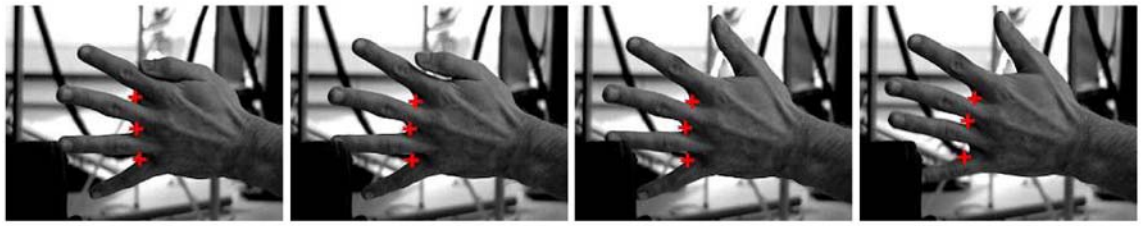
In this section, we extend the simple partial linear Gaussian model to the case of large state space. The methodological study shows that one may have access to the Optimal Rao-Blackwellized particle filter, whose sampling exploration is done efficiently, as well as to a validation gate. This new algorithm is applied to object tracking. A model for tracking objects described by a set of points is proposed. The original resulting tracker is shown to be resistant to partial/complete occlusions, object deformations and motion blur.

The two previous parts have demonstrated the interest of alternative tracking system based on an informative evolution model, a simple likelihood and a sound estimation of the measurement noise covariance matrices. These partial linear Gaussian models have been successfully validated on a feature point tracking application. However, as the measurement model considered assumes a linear link between observations provided by a detection module and the state, one may object that such a setting is limited to the tracking of low dimensional features and can hardly be extended to state vector of larger dimension and associated only to local observations. The



(a) simple tracker, with only one measurement per tracked feature (frames 1, 8, 16, 19) – average trajectories over 100 realizations

(b) Tracking result of the point situated between the forefinger and the middle finger. Use of the simple tracker, considering only one measurement. This observation is represented in blue, the feature location estimate in red and the set of particles in yellow (100 particles have been used) – frames 2, 7, 15, 19



(c) tracker dedicated to ambiguities, with three measurements per tracked feature (frames 1, 8, 16, 19) – average trajectories over 100 realizations

(d) Tracking result of the point located between the forefinger and the middle finger. Use of the tracker dedicated to ambiguities, considering three measurements. The first observation (corresponding to the highest correlation score) is represented in blue, the two secondary measurements are in green. The feature location estimate in red and the set of particles in yellow (100 particles have been used) – frames 2, 7, 15, 19

Figure 8. Hand sequence—(a–b) simple tracker—(c–d) tracker dedicated to ambiguities.

goal of this section is to contradict such assertion and to show that a partial linear Gaussian model can be advantageously implemented for tracking in a state space of large dimension. To this aim we will apply such a framework to a problem of object tracking.

For the sake of generality, we assume here that the target object can be represented as a collection of D descriptors (such as feature points, sift points, edge portions, etc.). To overcome the difficulty of devising a global object detector in the image, we assume that only local and independent observations for each descriptor are available. Let us note $\mathbf{x}_k$ the state vector composed by the complete set of characteristics (locations, direction, etc.) of all the descriptors; $\mathbf{z}_k$ is the measurement vector which gathers the whole set of descriptors measurements (one per descriptor involved). In order to use a partial linear Gaussian model, a linear measurement equation is assumed:

$$\mathbf{z}_k = H_{k\,|\,\mathbf{I}}\,\mathbf{x}_k + \mathbf{v}_k, \tag{53}$$

where $H_{k\,|\,\mathbf{I}}$ is the measurement matrix and $\mathbf{v}_k$ is a white noise of zero mean and of covariance $R_{k\,|\,\mathbf{I}}$. As for the construction of the object evolution law, the ideal case is to consider a dynamic equation that describes both the spatial coherence of the set of descriptors and its dynamic evolution. To that end, in order to take benefit from the available information on the object motion and on its geometric structure, the descriptor set is divided in two subsets $\mathbf{r}_k$ (the *motion* descriptors) and $\mathbf{y}_k$ (the *geometric* descriptors). Two different components of the object evolution law are attached to this decomposition. The first one, associated to the motion descriptors, encodes the rigid motion of a reference frame connected to the object. The second one describes the object deformations. Let us note that similar assumptions coupling a rigid motion component (rotation and translation) with a blend-shape component are widely used in structure-from-motion tracking (Brand, 2001; Torresani and Bregler, 2002). The dynamic model we consider may be written:

$$\mathbf{x}_k = \begin{bmatrix} \mathbf{r}_k \\ \mathbf{y}_k \end{bmatrix} = \begin{bmatrix} f_{k\,|\,\mathbf{I}}\,(\mathbf{r}_{k-1}) \\ g_{k\,|\,\mathbf{I},\mathbf{r}}\,(\mathbf{y}_{k-1}) \end{bmatrix} + \begin{bmatrix} \mathbf{w}_k^{\mathbf{r}} \\ \mathbf{w}_k^{\mathbf{y}} \end{bmatrix}, \tag{54}$$

where $\mathbf{w}_k^{\mathbf{r}}$ and $\mathbf{w}_k^{\mathbf{y}}$ are white noises of zero mean and covariance matrices respectively $Q_{k\,|\,\mathbf{I}}^{\mathbf{r}}$ and $Q_{k\,|\,\mathbf{I}}^{\mathbf{y}}$. As previously, in order to be reactive to any unpredictable changes of speed and direction of the object, we rely on an online dynamic model for the reference frame rigid motion. Even if it describes a rigid motion, the function $f_{k\,|\,\mathbf{I}}$ may be non linear w.r.t. components $\mathbf{r}_k$ as its estimation may depend in a non linear way on the location of the associated descriptors. As for the function $g_{k\,|\,\mathbf{I},\mathbf{r}}$ (whose definition may depend on $\mathbf{r}_k$), it links the descriptors of subset $\mathbf{y}_k$. It is devoted to the description of the *a priori* geometric deformation of the object between two successive instants. Several parametric constraints may be used depending on the object tracked. Let us remark that the function $g_{k\,|\,\mathbf{I},\mathbf{r}}$ is linear if $\mathbf{y}_k$ can be described in a linear interpolation scheme—eventually knowing $\mathbf{r}_k$ (for ex. $\mathbf{y}_k$ is described using piecewise linear splines with $\mathbf{r}_k$ as control point, or using blend shape decomposition).

There is no particular constraint on the state space dimension to implement partial linear Gaussian models (53–54) through sequential Monte Carlo methods. Nevertheless, as large dimensional state space is well-known to be the Achille's heel of nonlinear Bayesian filters, any procedures allowing dimension reduction is welcome. Among them, Rao-Blackwellization allows reducing the sampling space when some state components can be optimally estimated conditionally to the others. The idea is to estimate the truly nonlinear/non-Gaussian part of the state using a particle filter, while the remaining conditional state components can be estimated analytically using the Kalman filter, a hidden Markov model filter, or any other finite-dimensional optimal filter (Doucet et al., 2000). The resulting algorithms are called *Rao-Blackwellized particle filters*. They have been successfully applied for bearings only tracking and navigation (Gustafsson et al., 2002), digital communication (Chen and Liu, 2000) and recently in computer vision (Khan et al., 2004).

Rao-Blackwellized filters assume that the evolution law of the state $\mathbf{x}_k = [\mathbf{r}_k, \mathbf{y}_k]^t$, is such that component $\mathbf{r}_k$ is independent of component $\mathbf{y}_k$ conditionally to $\mathbf{r}_{k-1}$:

$$\begin{aligned} p(\mathbf{x}_k \,|\, \mathbf{x}_{k-1}) &= p(\mathbf{r}_k, \mathbf{y}_k \,|\, \mathbf{r}_{k-1}, \mathbf{y}_{k-1}) \\ &= p(\mathbf{r}_k \,|\, \mathbf{r}_{k-1})\; p(\mathbf{y}_k \,|\, \mathbf{r}_k, \mathbf{y}_{k-1}). \end{aligned} \tag{55}$$

Assuming in addition that the density $p(\mathbf{y}_k \,|\, \mathbf{r}_k, \mathbf{z}_{1:k})$ can be optimally estimated, as the objective distribution reads:

$$p(\mathbf{x}_k \,|\, \mathbf{z}_{1:k}) = p(\mathbf{r}_k, \mathbf{y}_k \,|\, \mathbf{z}_{1:k}) = p(\mathbf{y}_k \,|\, \mathbf{r}_k, \mathbf{z}_{1:k})\; p(\mathbf{r}_k \,|\, \mathbf{z}_{1:k}), \tag{56}$$

the only remaining difficulty relies in the estimation of $p(\mathbf{r}_k \,|\, \mathbf{z}_{1:k})$. Compared to the original filtering density this new density lives in a space of reduced dimension. Given a particle swarm $\{\mathbf{r}_k^{(i)}, w_k^{(i)}\}_{i=1:N}$ and a particle filter approximation of $p(\mathbf{r}_k \,|\, \mathbf{z}_{1:k})$:

$$p(\mathbf{r}_k \,|\, \mathbf{z}_{1:k}) \simeq \sum_{i=1}^{N} w_k^{(i)}\; \delta_{\mathbf{r}_k^{(i)}}(\mathbf{r}_k), \tag{57}$$

the marginal density $p(\mathbf{y}_k \mid \mathbf{z}_{1:k})$ can then be approximated by the following mixture of laws:

$$p(\mathbf{y}_k \mid \mathbf{z}_{1:k}) = \int p(\mathbf{y}_k \mid \mathbf{r}_k, \mathbf{z}_{1:k})\; p(\mathbf{r}_k \mid \mathbf{z}_{1:k})\; d\mathbf{r}_k \simeq \sum_{i=1}^{N} w_k^{(i)}\; p\big(\mathbf{y}_k \mid \mathbf{r}_k^{(i)}, \mathbf{z}_{1:k}\big). \tag{58}$$

Theoretical results deduced from the Rao-Blackwell theorem (Casella and Robert, 1996) demonstrate the interest of this marginalization approach. In particular, considering two estimates of $\mathbb{E}[\mathbf{y}_k \mid \mathbf{z}_{1:k}]$, namely the standard estimate:

$$\hat{E}_1[\mathbf{y}_k \mid \mathbf{z}_{1:k}] = \frac{1}{N}\sum_{i=1}^{N} \mathbf{y}_k^{(i)} \quad \text{where} \quad \big(\mathbf{y}_k^{(i)}, \mathbf{r}_k^{(i)}\big) \sim p(\mathbf{y}_k, \mathbf{r}_k \mid \mathbf{z}_{1:k}) \tag{59}$$

and the Rao-Blackwellized estimate :

$$\hat{E}_2[\mathbf{y}_k \mid \mathbf{z}_{1:k}] = \frac{1}{N}\sum_{i=1}^{N} \mathbb{E}\big[\mathbf{y}_k \big| \mathbf{r}_k^{(i)}, \mathbf{z}_{1:k}\big] \quad \text{where} \quad \mathbf{r}_k^{(i)} \sim p(\mathbf{r}_k \mid \mathbf{z}_{1:k}) \tag{60}$$

it can be shown that (Doucet et al., 2000):

$$var[\hat{E}_2[\mathbf{y}_k \mid \mathbf{z}_{1:k}]] \leq var[\hat{E}_1[\mathbf{y}_k \mid \mathbf{z}_{1:k}]]. \tag{61}$$

This result proves the better efficiency of the Rao-Blackwellized estimate. Let us remark that the superiority of the Rao-Blackwell estimate will also be validated from an experimental point of view Section 6.3. For interested reader, other theoretical results are given in Doucet et al. (2000). The resulting Rao-Blackwellized particle filter consists finally in three main steps. First, an importance sampling stage is performed to generate the swarm $\{\mathbf{r}_k^{(i)}, w_k^{(i)}\}_{i=1:N}$ which approximates $p(\mathbf{r}_k \mid \mathbf{z}_{1:k})$. For each particle $\mathbf{r}_k^{(i)}$, the density $p(\mathbf{y}_k \mid \mathbf{r}_k^{(i)}, \mathbf{z}_{1:k})$ is estimated. A final resampling step is eventually required.

Considering the system (53–54), if one wishes to use the Kalman filter to compute optimally the density $p(\mathbf{y}_k \mid \mathbf{r}_k, \mathbf{z}_{1:k})$ the dynamic equation of the geometric descriptors has to be linear. We will therefore consider the following partial linear Gaussian model (which is a particular case of model (53–54)):

$$\begin{cases} \mathbf{r}_k = f_{k\mid\mathbf{I}}(\mathbf{r}_{k-1}) + \mathbf{w}_k^{\mathbf{r}} \\ \mathbf{y}_k = G_{k\mid\mathbf{I},\mathbf{r}}\mathbf{y}_{k-1} + \mathbf{w}_k^{\mathbf{y}} \\ \mathbf{z}_k = \begin{bmatrix} \mathbf{z}_k^{\mathbf{r}} \\ \mathbf{z}_k^{\mathbf{y}} \end{bmatrix} = H_{k\mid\mathbf{I}}\mathbf{x}_k + \mathbf{v}_k = \begin{bmatrix} M_{k\mid\mathbf{I}} & 0 \\ 0 & N_{k\mid\mathbf{I}} \end{bmatrix} \begin{bmatrix} \mathbf{r}_k \\ \mathbf{y}_k \end{bmatrix} + \begin{bmatrix} \mathbf{v}_k^{\mathbf{r}} \\ \mathbf{v}_k^{\mathbf{y}} \end{bmatrix}, \end{cases} \tag{62}$$

where $G_{k\mid\mathbf{I},\mathbf{r}}$ is a matrix whose entries may depend on the current estimation of $\mathbf{r}_k$. We assume that $\mathbf{v}_k^{\mathbf{r}}$ and $\mathbf{v}_k^{\mathbf{y}}$ are white noises of zero mean and covariances respectively $R_{k\mid\mathbf{I}}^{\mathbf{r}}$ and $R_{k\mid\mathbf{I}}^{\mathbf{y}}$.

As for the estimation of $p(\mathbf{r}_k \mid \mathbf{z}_{1:k})$, this model allows the exploration of the state space using the optimal importance function. Indeed, similarly to the case presented in Section 4, the two densities required for the optimal sampling, $p(\mathbf{r}_k \mid \mathbf{r}_{k-1}, \mathbf{z}_k)$ and $p(\mathbf{z}_k \mid \mathbf{r}_{k-1})$, are Gaussian. This leads to consider the *Optimal Rao-Blackwellized Particle filter* that corresponds to a bank of Kalman filters (one filter per particle) since the distributions $p(\mathbf{y}_k \mid \mathbf{r}_k^{(i)}, \mathbf{z}_{1:k}) = \mathcal{N}(\mathbf{y}_k; \hat{\mathbf{y}}_{k\mid k}, \Sigma_{k\mid k})$ are estimated through the Kalman filter's recursive expressions:

$$\hat{\mathbf{y}}_{k\mid k-1} = G_{k\mid\mathbf{I},\mathbf{r}}\hat{\mathbf{y}}_{k-1\mid k-1} \tag{63}$$

$$\Sigma_{k\mid k-1} = G_{k\mid\mathbf{I},\mathbf{r}}\Sigma_{k-1\mid k-1}\; G_{k\mid\mathbf{I},\mathbf{r}}^t + Q_k^{\mathbf{y}} \tag{64}$$

$$K_k = \Sigma_{k\mid k-1}N_{k\mid\mathbf{I}}^t\big(N_{k\mid\mathbf{I}}\Sigma_{k\mid k-1}N_{k\mid\mathbf{I}}^t + R_{k\mid\mathbf{I}}\big)^{-1} \tag{65}$$

$$\hat{\mathbf{y}}_{k\mid k} = \hat{\mathbf{y}}_{k\mid k-1} + K_k\big[\mathbf{z}_k^{\mathbf{y}} - N_{k\mid\mathbf{I}}^t\hat{\mathbf{y}}_{k\mid k-1}\big] \tag{66}$$

$$\Sigma_{k\mid k} = \Sigma_{k\mid k-1} - K_k\;\Sigma_{k\mid k-1}. \tag{67}$$

6.1. Validation Gate

Following the same schemes as before, a validation gate may be still considered for this kind of partial linear Gaussian model using:

$$p(\mathbf{z}_k \mid \mathbf{z}_{1:k-1}) = \int p(\mathbf{z}_k \mid \mathbf{x}_{k-1})\; p(\mathbf{x}_{k-1} \mid \mathbf{z}_{1:k-1})\; d\mathbf{x}_{k-1} \simeq \sum_{i=1}^{N} w_{k-1}^{(i)}\; p\big(\mathbf{z}_k \mid \mathbf{x}_{k-1}^{(i)}\big), \tag{68}$$

where in that case the particle swarm $\{\mathbf{x}_{k-1}^{(i)}, w_{k-1}^{(i)}\}_{i=1\ldots N} = \{\mathbf{r}_{k-1}^{(i)}, \mathbf{y}_{k-1}^{(i)}, w_{k-1}^{(i)}\}_{i=1\ldots N}$ has been obtained with the optimal Rao-Blackwellized particle filter. Applying

$$p(\mathbf{z}_k \mid \mathbf{x}_{k-1}) = \int p(\mathbf{z}_k \mid \mathbf{x}_k)\; p(\mathbf{x}_k \mid \mathbf{x}_{k-1})\; d\mathbf{x}_k, \tag{69}$$

we can remark that for the model under concern (62), we have:

$$p\big(\mathbf{z}_k \mid \mathbf{x}_{k-1}^{(i)}\big) = \mathcal{N}\left(\mathbf{z}_k; \begin{bmatrix} M_{k\mid\mathbf{I}}\; f_{k\mid\mathbf{I}}\left(\mathbf{r}_{k-1}^{(i)}\right) \\ N_{k\mid\mathbf{I}}\; G_{k\mid\mathbf{I},\mathbf{r}}\; \mathbf{y}_{k-1}^{(i)} \end{bmatrix}, \begin{bmatrix} R_{k\mid\mathbf{I}}^{\mathbf{y}} + M_{k\mid\mathbf{I}}Q_{k\mid\mathbf{I}}^{\mathbf{y}\;(\mathbf{i})}M_{k\mid\mathbf{I}}^t & 0 \\ 0 & R_{k\mid\mathbf{I}}^{\mathbf{r}} + N_{k\mid\mathbf{I}}Q_{k\mid\mathbf{I}}^{\mathbf{r}\;(\mathbf{i})}\; N_{k\mid\mathbf{I}}^t \end{bmatrix}\right). \tag{70}$$

As we have chosen to consider only local and independent measurements, the matrices involved in the measurement model, i.e. $M_{k\,|\,\mathbf{I}}$, $N_{k\,|\,\mathbf{I}}$, $R^{\mathbf{y}}_{k\,|\,\mathbf{I}}$ and $R^{\mathbf{r}}_{k\,|\,\mathbf{I}}$ are block-diagonal (where each block corresponds to one descriptor). Making the additional weak assumption that the state noise covariance matrices $Q^{\mathbf{y}}_{k\,|\,\mathbf{I}}$ and $Q^{\mathbf{r}}_{k\,|\,\mathbf{I}}$ are block-diagonal, the covariance matrix of the Gaussian law $p(\mathbf{z}_k \,|\, \mathbf{x}^{(i)}_{k-1})$ (Eq. (70)) is also block diagonal. Each block of this matrix is associated to one descriptor of the object representation. As a consequence, it is possible to consider a global validation gate composed of a collection of ellipsoidal regions, each of these regions being associated to one descriptor. The parameters of each of these local validation gates are computed following the same procedure as the one described in Section 4.1 for the simplest partial linear Gaussian model.

If $\mathbf{z}_{k,d}$ denotes the observation associated to descriptor d (noted $\mathbf{x}_{k,d}$) at time k, and $\mathbf{m}^{(i)}_{k,d}$ and $\Sigma^{(i)}_{k,d}$ the two first moments of the Gaussian density $p(\mathbf{z}_{k,d} \,|\, \mathbf{x}^{(i)}_{k-1,d})$ (whose expression may be directly deduced from Eq. (70)), the global validation gate is defined as:

$$\mathcal{V}_k = \bigcup_{d=1}^{D} \Psi_d, \tag{71}$$

with each ellipse Ψ_d (corresponding to the descriptor d) is given by:

$$\Psi_d = \left\{ \mathbf{z}_{k,d} : \left(\mathbf{z}_{k,d} - \mathbf{m}^{vg}_{k,d}\right)^t {\Sigma^{vg}_{k,d}}^{-1} \left(\mathbf{z}_{k,d} - \mathbf{m}^{vg}_{k,d}\right) \leq \gamma_d \right\}. \tag{72}$$

The parameter γ_d is chosen in practice such as the probability of finding the real measurement of the descriptor d is equal to 0.99. The expressions of the parameters $\mathbf{m}^{vg}_{k,d}$ and $\Sigma^{vg}_{k,d}$ are:

$$\mathbf{m}^{vg}_{k,d} \simeq \sum_{i=1}^{N} w^{(i)}_{k-1}\, \mathbf{m}^{(i)}_{k,d}, \tag{73}$$

$$\Sigma^{vg}_{k,d} \simeq \sum_{i=1}^{N} w^{(i)}_{k-1} \left[\Sigma^{(i)}_{k,d} + \left\| \mathbf{m}^{(i)}_{k,d} \right\|^2 \right] - \left\| \sum_{i=1}^{N} w^{(i)}_{k-1} \mathbf{m}^{(i)}_{k,d} \right\|^2. \tag{74}$$

6.2. *Application to Object Tracking Relying on a Point Cloud Description*

In this section, we propose to apply the optimal Rao-Blackwellized particle filter to the tracking of an object described by a set of feature points. This tracking problem has been mainly investigated for structure from motion problems (Brand, 2001; Torresani and Bregler, 2002). These approaches rely on factorization methods and assume as a hard constraint that a sufficient number of correspondences are known for the whole sequence. Alternatively, trackers for groups of feature points in video shots has been defined through epipolar fits and RANSAC estimator (Sivic et al., 2004). All these studies rely on deterministic successive instantaneous estimation processes such as optic flow, correlation from frame to frame or wide baseline matching. Indeed, they aim more at estimating a global structure of the point cloud than at tracking features in a robust way. By robust we mean: robust to image sequence noise (such additive noise, blur and motion blur), free of drift, robust to occlusions or to failures of the measurement process and robust to model's imprecisions. Stochastic filters enable naturally such an estimation. In addition, as previously explained, these approaches provide a prediction of the image region to focus on. The recursive aspect of these filters (online estimation) constitutes also a major difference with batch structure from motion techniques based on factorization which require all point correspondences in the sequence. However, to the best of our knowledge, no study has been done to propose an efficient tracker allowing partial or complete occlusions of the target and being robust to problematic sources of noise such as motion blur.

In the following paragraphs we describe all the different components of the tracking system of the form (62) devoted to point cloud tracking. We detail successively the initialization process, the estimation of the different ingredients of the dynamic model (i.e. the *motion* function $f_{k\,|\,\mathbf{I}}$ and the *geometric* matrix $G_{k\,|\,\mathbf{I},\mathbf{r}}$), and the elements related to the measurement model.

6.2.1. Selection of the Feature Points. The target point cloud is automatically determined using the Harris point detector (Harris and Stephens, 1988) on the first image of the sequence. This detector relies on the assumption that characteristic points in a region of interest are associated with maxima of the local autocorrelation function. Compared to other detectors, it has been shown to provide feature points of good quality for a tracking application (Tissainayagam and Suter, 2004). Relying on the assumption that the more discriminant the neighborhood of the feature, the more reliable the tracking of this feature, we set the motion points to features associated to the highest maxima. These points are used to define a reference frame related to the object. The other points constitute the geometric subset.

6.2.2. Motion Function. Following the same idea as for the previous tracker we wish to rely on a dynamic model describing as accurately as possible the global motion of the point cloud. To this aim we still rely on a motion model estimated online through a robust estimation process. The

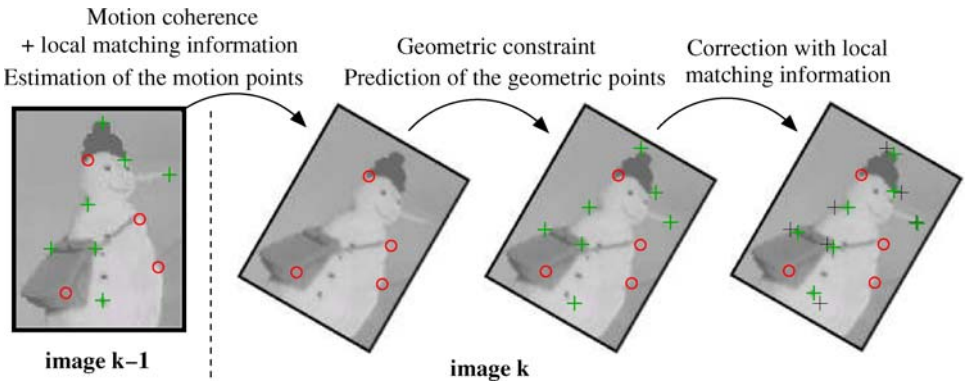


Figure 9. Illustration of the steps to be achieved for each particle—red: motion points ; green: geometric points.

resulting evolution law of the set of motion points $\mathbf{r}_k$ reads:

$$p(\mathbf{r}_k \mid \mathbf{r}_{k-1}) = \mathcal{N}\big(\mathbf{r}_k; \mathbf{r}_{k-1} + P(\mathbf{r}_{k-1})\boldsymbol{\theta}_{k\mid\mathbf{I}}(\mathbf{r}_{k-1}),\, Q^{\mathbf{r}}_{k\mid\mathbf{I}}\big). \tag{75}$$

The estimation support on which the motion parameter vector $\boldsymbol{\theta}_{k\mid\mathbf{I}}$ is estimated is a square window region including the whole set of motion points. The noise covariance matrix $Q^{\mathbf{r}}_{k\mid\mathbf{I}}$ is estimated as before.

6.2.3. Geometric Matrix. The geometric matrix $G_{k\mid\mathbf{I},\mathbf{r}}$ depends on the motion subset. It can be evaluated for any given configuration of the reference frame constituted by the motion points. This matrix describes the evolution of the geometric points between two frame instants. Several constraints may be used. Two cases have been considered here to distinguish the tracking of a planar object from the tracking of a non planar object. In both cases, the evolution law of the geometric subset $\mathbf{y}_k$ reads:

$$p(\mathbf{y}_k \mid \mathbf{r}_k, \mathbf{y}_{k-1}) = \mathcal{N}\big(\mathbf{y}_k\ ;\ G_{k\mid\mathbf{I},\mathbf{r}}\,\mathbf{y}_{k-1},\, Q^{\mathbf{y}}_{k\mid\mathbf{I}}\big), \tag{76}$$

where the noise covariance matrix $Q^{\mathbf{y}}_{k\mid\mathbf{I}}$ is fixed and set to $\sigma^2 \times Id$.

- *The Tracked Object has a Non Planar Structure.* For the tracking of a non planar object, the simplest and more general solution is to set the dynamic model of the geometric points to a constant position model in the coordinate reference frames defined by the motion points. In that case $G_{k\mid\mathbf{I},\mathbf{r}}$ is thus a matrix describing a change of coordinate system. Associated to high covariance value such a model defines just a weak dynamic model which *a priori* penalizes configurations resulting from large deformations between two frame instants.
- *The Tracked Object has a Planar Structure.* In this case the projective coordinates of the point cloud at distinct instants are linearly linked by an homography matrix. This matrix can be estimated from at least four matched motion points (Faugeras and Lustman, 1988). Assuming therefore that $\mathbf{r}_k$ is composed of at least four feature points, an homography matrix may be estimated for each configuration (i.e. for each particles) of the motion points.

Let us remark that other models could be considered. The only limitation, related to the linearity of $G_{k\mid\mathbf{I},\mathbf{r}}$, is that $\mathbf{y}_k$ has to be described using a linear interpolation scheme. Such a description can handle a large number of cases. For example, articulated motion can be described by piecewise linear splines where main articulations are control points (here the control points are $\mathbf{r}_k$, estimated with a local or global motion model). We can also think of deformable objects described by a mesh of points $\mathbf{y}_k$ that may be estimated as a linear combination of possible configurations (blend shape description). In that case the set $\mathbf{r}_k$ may describe the rigid motion.

6.2.4. Measurement Model. In order to avoid the difficulty of defining a global detection process for the set of points, we propose to rely on local independent measurements of the different features. The considered observation vector $\mathbf{z}_k$ gathers all the correlation peaks of each point composing $\mathbf{x}_k$. In the same way as in the case of a single point tracking, this vector is determined through a matching technique between the current image and a target representation $\widetilde{\mathbf{I}}_0$. This representation which constitutes a reference image of the object is composed by a collection of local image patches centered at the initial point locations in the first image of the sequence. In the same manner as done previously, this reference may be updated by warping the collection of image patches. The matching criterion used to quantify the similarity

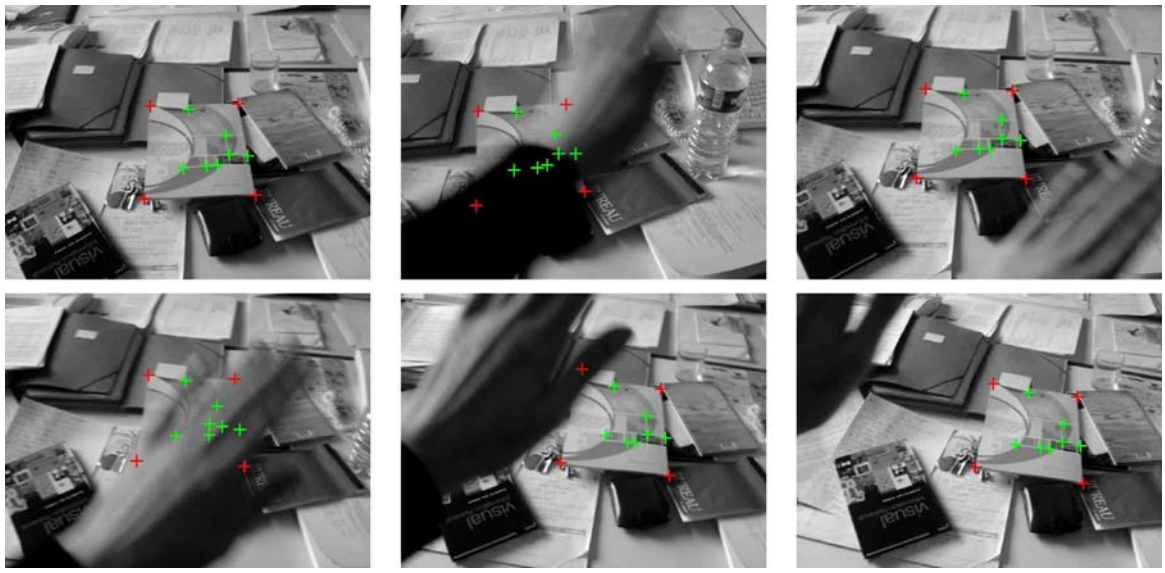


Figure 10. Desk sequence—dealing with complete occlusions (frames 1, 16, 37, 41, 47, 53)—average trajectory over 100 runs.

between two points is the SSD criterion. The resulting measurement equation reads:

$$p(\mathbf{z}_k \mid \mathbf{x}_k) = \mathcal{N}(\mathbf{z}_k; \mathbf{x}_k, R_{k\mid\mathbf{I}}) \tag{77}$$

The noise covariance $R_{k\mid\mathbf{I}}$ is a block-diagonal matrix. For each point, the associated block accounts for an accuracy measurement of the matching procedure. The estimation of these blocks follows the procedure detailed in Section 4.3.2. Such online estimations of the noise covariance matrices improve greatly the tracker robustness to occlusions.

The final object tracker is based on the model composed of Eqs. (75)–(77). This model is a partial linear Gaussian model and its structure allows using the optimal Rao-Blackwellized particle filter. The steps of the filter are summarized in Algorithm 3, and illustrated for a single particle in Fig. 9.

6.3. *Experimental Results*

In this section experimental results obtained on real-world sequences are presented. To highlight the tracker efficiency, we have chosen to present some results on difficult situations that would make deterministic trackers fail (i.e. motion blur, complete/partial occlusions, partial exit of the object out the image frame). On these results, the red crosses denote the motion points and the green crosses account for the geometric points. All the used sequences have been shot with a hand-held-camera and present therefore global chaotic motions generating motion blur on some images.

Algorithm 3 Object tracker

- **initialization:**
  for $i = 1 : N$, generate $\mathbf{r}_0^{(i)} \sim p(\mathbf{r}_0)$, set $w_0^{(i)} = 1/N$, $\hat{\mathbf{y}}_{0\mid 0}^{(i)} = \hat{\mathbf{y}}_0$ and $\Sigma_{0\mid 0}^{(i)} = \Sigma_0$ or $k = 1, 2, \ldots$
- **estimations on the image sequence:**
  1. detection of $\mathbf{z}_k$ using SSD criteria in $\mathcal{V}_k$, and estimation of $R_{k\mid\mathbf{I}}^{\mathbf{r}}$ and $R_{k\mid\mathbf{I}}^{\mathbf{y}}$
  2. for $i = 1 \ldots N$ estimation of $\boldsymbol{\theta}_{k\mid\mathbf{I}}(\mathbf{r}_{k-1}^{(i)})$ and $Q_{k\mid\mathbf{I}}^{\mathbf{r}\,(i)}$ using a robust parametric estimation technique on an image region including $\mathbf{r}_{k-1}^{(i)}$
- **sequential importance sampling:**
  1. sampling: for $i = 1 : N$, generate $\mathbf{r}_k^{(i)} \sim p(\mathbf{r}_k \mid \mathbf{r}_{k-1}^{(i)}, \mathbf{z}_k)$
  2. calculation of normalized importance weights: for $i = 1 \ldots N$, calculate $w_k^{(i)} = p(\mathbf{z}_k \mid \mathbf{r}_k^{(i)})$
- **estimations on the image sequence:**
  for each particle $\mathbf{r}_k^{(i)}$, estimation of the geometric matrix $G_{k\mid\mathbf{I},\mathbf{r}}^{(i)}$
- **bank of Kalman filters:**
  for each particle $\mathbf{r}_k^{(i)}$, estimation of the parameters $\hat{\mathbf{y}}_{k\mid k}^{(i)}$, $\Sigma_{k\mid k}^{(i)}$ of the Gaussian law $p(\mathbf{y}_k \mid \mathbf{r}_k^{(i)}, \mathbf{z}_{1:k})$ using the Kalman filter (Eqs. (63—67))
- **estimation of the point cloud position**
- **resampling if necessary**

Tracking of Planar Objects. The first results (Figs. 10 and 11) present the tracking of a planar structure and have been obtained using an homography model for the

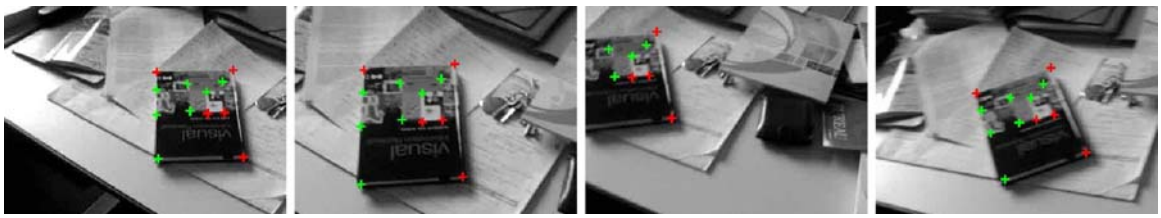


Figure 11. Book sequence—dealing with partial exit out of the image, large amplitude motion and motion blur (frames 1, 48, 87, 121)—average trajectory over 100 realizations.

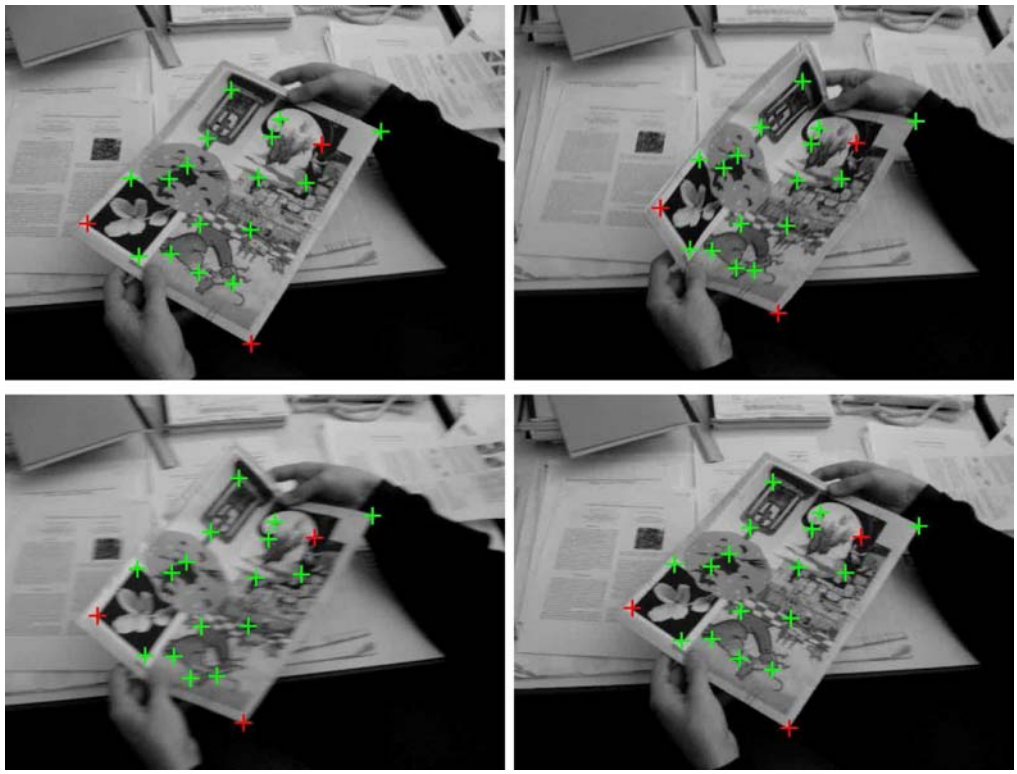


Figure 12. Paper sequence—dealing with object deformations (frames 1, 26, 39, 50)—average trajectory over 100 realizations.

geometric constraint. The results obtained on the **Desk** sequence are illustrated on Fig. 10. It shows a successful tracking of a book lying on an overstuffed desk that constitutes a cluttered background. As it can be observed on frames 16 and 41, this object is completely occluded twice. Although the tracker is naturally disturbed during the occlusions (as it can be observed by the trajectories of few points of the cloud), it recovers well the respective locations of all the points at the occlusion end. The second results shows a successful tracking of another book lying on a desk. This is a long sequence of 121 frames exhibiting long range motion, motion blur and a partial exit of the target.

Tracking of Non Planar Objects. The two last results are dedicated to the tracking of deformable and non planar structures. They have been obtained setting the geometric constraint as a simple constant position model in a coordinate reference defined by the motion subset. The **Paper** sequence shows a textured planar object undergoing deformations. As it is illustrated in Fig. 12, the trajectories of each feature of the cloud have been successfully recovered in spite of the several possible ambiguities. Finally, our algorithm has been tested on the challenging **Tiger** sequence. The obtained results are presented on Fig. 13. The non rigid deformations of the walking tiger are very important. Strong similarities can also be observed between the different feature appearances, and these latter may change a lot along the image sequence. However, the tracker succeeds in recovering the right trajectory on the point cloud. In order to validate the interest of the Rao-Blackwellization version of the particle fil-

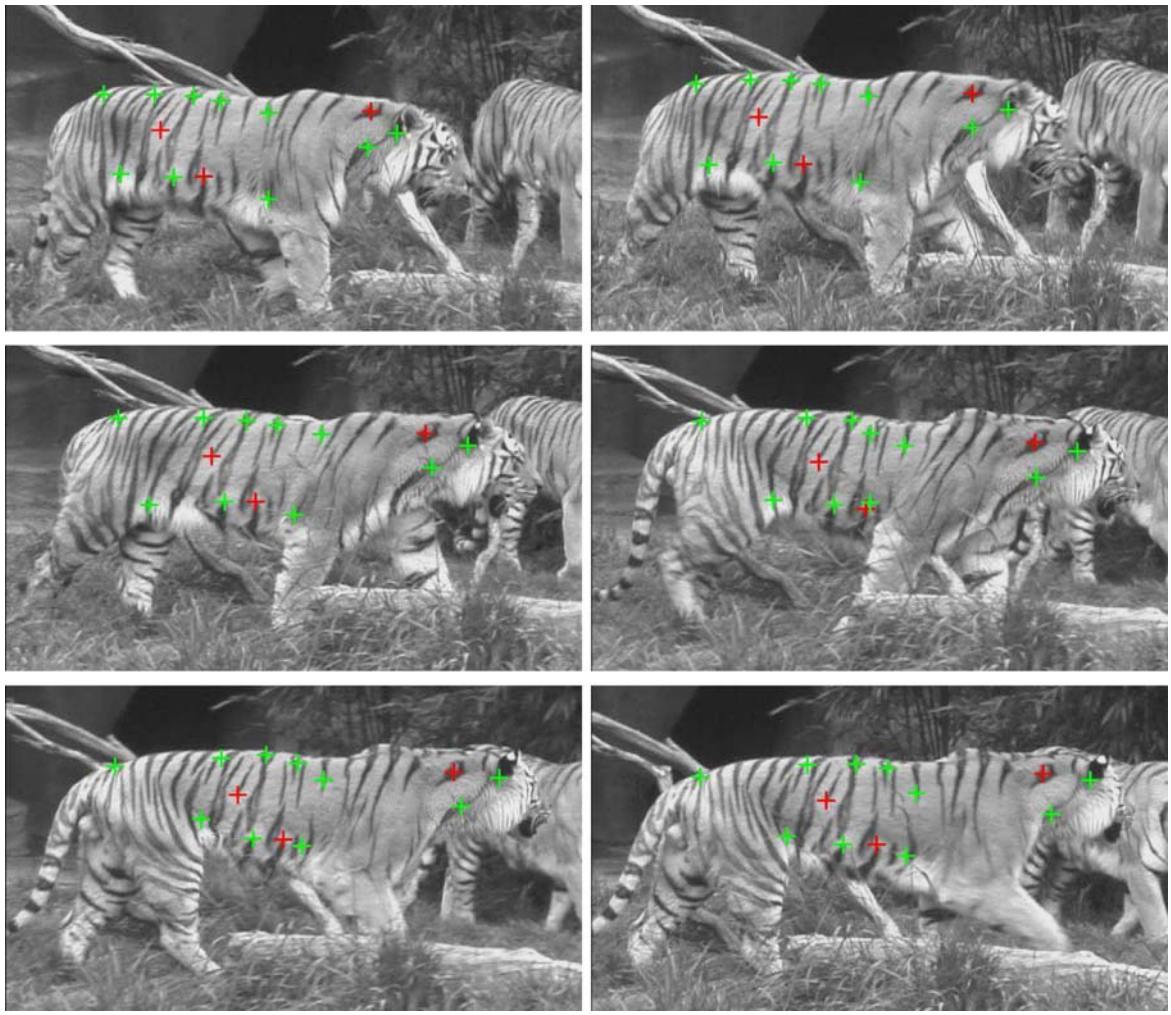


Figure 13. Tiger sequence—frames 1, 6, 13, 20, 26, 30—average trajectory over 100 realizations—Rao-Blackwellized optimal particle filter.

ter, Fig. 14 presents the result obtained with and without the Rao-Blackwellization (simple optimal particle filter). Comparing the results shows clearly the efficiency of the Rao-Blackwell estimate.

7. Remarks on Computational Complexity

Before concluding, let us make some remarks about the trackers complexity. Concerning the two point trackers (Algorithms 1 and 2), the computational cost is essentially due to the motion model estimation, to the measurement computation and to its associated confidence measure. The computation of the dynamic model update has to be relativized. Indeed, the robust motion estimation process can be implemented very efficiently. The version we use works in real time. Nevertheless, since the dynamic model has to be calculated for each particle, it is true that such an image-based model costs more than usual auto-regressive models. 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. The measurement confidence could be also estimated in the Fourier space. The filtering part is not time consuming as the peculiar proposed models lead to very simple and efficient algorithms (involving only Gaussian densities). However, the complexity of the multi-hypothesis tracker slightly increases with respect to the simple one. Obviously, this increase depend on the number of considered measurement-hypothesis.

As for the object tracker (Algorithm 3), the same remarks as before account for the model calculations. However, since the filter is composed of a set of particles, for each of them one need to compute a Kalman filter, the whole procedure is time-consuming.

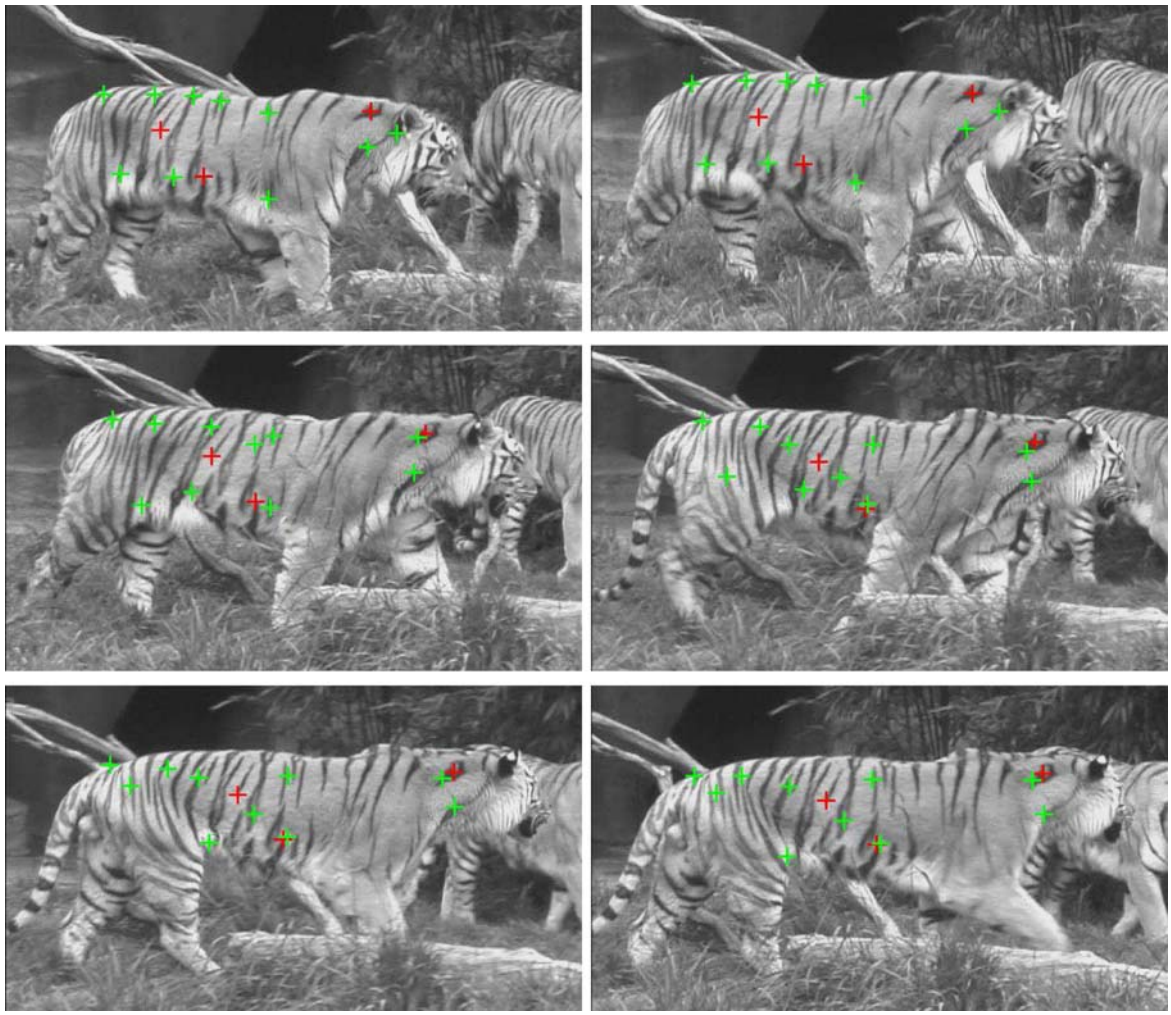


Figure 14. Tiger sequence—frames 1, 6, 13, 20, 26, 30—average trajectory over 100 realizations—optimal particle filter.

8. Conclusion

The work presented in this paper takes place in the framework of Sequential Monte Carlo approaches for tracking in image sequences. In particular, the use of partial linear Gaussian systems to model the unknown dynamic model has been deeply investigated. Such systems have been shown to describe the problem in an opposite manner as the traditional tracking models since they associate a rough linear measurement model with a non linear dynamic equation. They have the great advantage to allow an exact expression of the optimal importance function. The knowledge of this optimal function enables us to include naturally the measurements into the state space exploration process and authorizes to build a relevant approximation of a validation gate. Three instances of increasing complexity have been introduced, allowing to cope with occlusions, ambiguities and large state space.

These models have been first successfully validated on a point tracking application. The obtained results have demonstrated the ability of the proposed method to deal with occlusions, motion blur, complex trajectories and clutter without defining any specific scheme. The *a priori*-free system we have considered has been entirely defined on the image data. This latter has finally been extended and validated for a planar object tracking application.

9. Annexe

Let us consider the following multimodal partial linear Gaussian model (studied in Section 5):

$$\begin{cases} p(\mathbf{x}_k \mid \mathbf{x}_{k-1}) = \mathcal{N}(\mathbf{x}_k; f_{k\mid\mathbf{I}}(\mathbf{x}_{k-1}),\ Q_{k\mid\mathbf{I}}) \\ p(\mathbf{z}_k \mid \mathbf{x}_k) = V_k^{1-M_k} \sum_{j=1}^{M_k} \beta_{k,j}\, \mathcal{N}(\mathbf{z}_{k,j}; \\ \qquad\qquad H_{k\mid\mathbf{I},j}\, \mathbf{x}_k,\, R_{k\mid\mathbf{I},j}). \end{cases} \tag{78}$$

we have

$$p\big(\mathbf{z}_k \mid \mathbf{x}_{k-1}^{(i)}\big) = \int p(\mathbf{z}_k \mid \mathbf{x}_k) p\big(\mathbf{x}_k \mid \mathbf{x}_{k-1}^{(i)}\big)\, d\mathbf{x}_k$$
$$= V_k^{1-M_k} \sum_{j=1}^{M_k} \beta_{k,j} \mathcal{N}\big(\mathbf{z}_{k,j}; H_{k\mid\mathbf{I},j}\ f_{k\mid\mathbf{I}}\big(\mathbf{x}_{k-1}^{(i)}\big),$$
$$R_{k\mid\mathbf{I},j} + H_{k\mid\mathbf{I},j}\ Q_{k\mid\mathbf{I}}^{(i)}\ H_{k\mid\mathbf{I},j}^t\big). \tag{79}$$

Remarking that:

$$p(\mathbf{x}_k \mid \mathbf{x}_{k-1}, \mathbf{z}_k) = \frac{p(\mathbf{z}_k \mid \mathbf{x}_k)\ p(\mathbf{x}_k \mid \mathbf{x}_{k-1})}{p(\mathbf{z}_k \mid \mathbf{x}_{k-1})}, \tag{80}$$

and using the expressions of model (78), and Eq. (79) one can write:

$$p\big(\mathbf{x}_k \mid \mathbf{x}_{k-1}^{(i)}, \mathbf{z}_k\big) = \frac{V_k^{1-M_k} \mathcal{N}(\mathbf{x}_k; f_{k\mid\mathbf{I}}(\mathbf{x}_{k-1}^{(i)}), Q_{k\mid\mathbf{I}}^{(i)}) \sum_{j=1}^{M_k} \beta_{k,j} \mathcal{N}(\mathbf{z}_{k,j}; H_{k\mid\mathbf{I},j}\mathbf{x}_k, R_{k\mid\mathbf{I},j})}{V_k^{1-M_k} \sum_{j=1}^{M_k} \beta_{k,j} \mathcal{N}(\mathbf{z}_{k,j}; H_{k\mid\mathbf{I},j} f_{k\mid\mathbf{I}}(\mathbf{x}_{k-1}^{(i)}), R_{k\mid\mathbf{I},j} + H_{k\mid\mathbf{I},j} Q_{k\mid\mathbf{I}}^{(i)} H_{k\mid\mathbf{I},j}^t)}. \tag{81}$$

The denominator can be replaced by a constant S_k, independent of $\mathbf{x}_k$, such as:

$$S_k = \sum_{j=1}^{M_k} \beta_{k,j} \mathcal{N}\big(\mathbf{z}_{k,j}; H_{k\mid\mathbf{I},j} f_{k\mid\mathbf{I}}\big(\mathbf{x}_{k-1}^{(i)}\big), R_{k\mid\mathbf{I},j}$$
$$+ H_{k\mid\mathbf{I},j} Q_{k\mid\mathbf{I}}^{(i)} H_{k\mid\mathbf{I},j}^t\big), \tag{82}$$

then, we have:

$$p\big(\mathbf{x}_k \mid \mathbf{x}_{k-1}^{(i)}, \mathbf{z}_k\big) = \frac{1}{S_k} \sum_{j=1}^{M_k} \beta_{k,j} \mathcal{N}(\mathbf{z}_{k,j}; H_{k\mid\mathbf{I},j}\mathbf{x}_k, R_{k\mid\mathbf{I},j})$$
$$\mathcal{N}\big(\mathbf{x}_k; f_{k\mid\mathbf{I}}\big(\mathbf{x}_{k-1}^{(i)}\big), Q_{k\mid\mathbf{I}}^{(i)}\big). \tag{83}$$

Let us introduce the variable N:

$$N = (\mathbf{x}_k - f_{k\mid\mathbf{I}}(\mathbf{x}_{k-1}))^t Q_{k\mid\mathbf{I}}^{-1}(\mathbf{x}_k - f_{k\mid\mathbf{I}}(\mathbf{x}_{k-1}))$$
$$+ (\mathbf{z}_{k,j} - H_{k\mid\mathbf{I},j}\mathbf{x}_k)^t R_{k\mid\mathbf{I},j}^{-1}(\mathbf{z}_{k,j} - H_{k\mid\mathbf{I},j}\mathbf{x}_k). \tag{84}$$

By noting that:

$$N = \mathbf{x}_k^t\big(Q_{k\mid\mathbf{I}}^{-1} + H_{k\mid\mathbf{I},j}^t R_{k\mid\mathbf{I},j}^{-1} H_{k\mid\mathbf{I},j}\big)\mathbf{x}_k \tag{85}$$
$$- 2\mathbf{x}_k^t\big(Q_{k\mid\mathbf{I}}^{-1} f_{k\mid\mathbf{I}}(\mathbf{x}_{k-1}) + H_{k\mid\mathbf{I},j}^t R_{k\mid\mathbf{I},j}^{-1}\mathbf{z}_{k,j}\big)$$
$$+ (f_{k\mid\mathbf{I}}(\mathbf{x}_{k-1}))^t Q_{k\mid\mathbf{I}}^{-1} f_{k\mid\mathbf{I}}(\mathbf{x}_{k-1}) + \mathbf{z}_{k,j}^t R_{k\mid\mathbf{I},j}^{-1}\mathbf{z}_{k,j}$$
$$= (\mathbf{x}_k - \mathbf{m}_{k\mid\mathbf{I},j})^t \Sigma_{k\mid\mathbf{I},j}^{-1}(\mathbf{x}_k - \mathbf{m}_{k\mid\mathbf{I},j}) + K \tag{86}$$

where

$$\Sigma_{k\mid\mathbf{I},j} = \big(Q_{k\mid\mathbf{I}}^{-1} + H_{k\mid\mathbf{I},j}^t R_{k\mid\mathbf{I},j}^{-1} H_{k\mid\mathbf{I},j}\big)^{-1} \tag{87}$$
$$\mathbf{m}_{k\mid\mathbf{I},j} = \Sigma_{k\mid\mathbf{I},j}\big(Q_{k\mid\mathbf{I}}^{-1} f_{k\mid\mathbf{I}}(\mathbf{x}_{k-1}) + H_{k\mid\mathbf{I},j}^t R_{k\mid\mathbf{I},j}^{-1}\mathbf{z}_{k,j}\big) \tag{88}$$
$$K = (f_{k\mid\mathbf{I}}(\mathbf{x}_{k-1}))^t Q_{k\mid\mathbf{I}}^{-1} f_{k\mid\mathbf{I}}(\mathbf{x}_{k-1}) + \mathbf{z}_{k,j}^t R_{k\mid\mathbf{I},j}^{-1}\mathbf{z}_{k,j}$$
$$+ \mathbf{m}_{k\mid\mathbf{I},j}^t \Sigma_{k\mid\mathbf{I},j}^{-1}\mathbf{m}_{k\mid\mathbf{I},j}, \tag{89}$$

the optimal importance function reads:

$$p\big(\mathbf{x}_k \mid \mathbf{x}_{k-1}^{(i)}, \mathbf{z}_k\big) = \sum_{j=1}^{M_k} \beta_{k,j} \frac{\alpha_{k,j}}{S_k} \mathcal{N}(\mathbf{x}_k; \mathbf{m}_{k\mid\mathbf{I},j}, \Sigma_{k\mid\mathbf{I},j}) \tag{90}$$

with

$$\alpha_{k,j} = C \exp\Bigg(-\frac{1}{2}\Big[\big\| f_{k\mid\mathbf{I}}\big(\mathbf{x}_{k-1}^{(i)}\big)\big\|_{Q_{k\mid\mathbf{I}}^{(i)-1}}^2 + \|\mathbf{z}_{k,j}\|_{R_{k\mid\mathbf{I},j}^{-1}}^2$$
$$- \|\mathbf{m}_{k\mid\mathbf{I},j}\|_{\Sigma_{k\mid\mathbf{I},j}^{-1}}^2\Big]\Bigg) \tag{91}$$

$$\text{and} \quad C = (2\pi)^{-\frac{n_z}{2}} \mid \Sigma_{k\mid\mathbf{I},j} \mid^{\frac{1}{2}} \mid Q_{k\mid\mathbf{I}}^{(i)} \mid^{-\frac{1}{2}} \mid R_{k\mid\mathbf{I},j} \mid^{-\frac{1}{2}}, \tag{92}$$

where n_z denotes the dimension of the measurement vector $\mathbf{z}_k$.

Notes

1. Let us note that all the system's ingredients may depend on the image data. Ideally one should consider a conditioning upon the image sequence data $\mathbf{I}_{0:k}$ as in Arnaud et al. (2005) and rely therefore on the probability distributions $p(\mathbf{z}_k \mid \mathbf{x}_k, \mathbf{I}_{0:k})$ and $p(\mathbf{x}_k \mid \mathbf{x}_{k-1}, \mathbf{I}_{0:k})$. In that case, the new filtering density is $p(\mathbf{x}_k \mid \mathbf{z}_{1:k}, \mathbf{I}_{0:k})$. Such a conditioning is omitted along the whole paper for sake of clarity.
2. Let us remark that the problem of data association is all the more difficult in the case of multi-target tracking. As we focus here on single feature tracking application, we will not address this issue. For more details, interested readers may refer to Bar-Shalom and Li (1995) and Vermaak et al. (2005).
3. Let us remark that in this work, the value of $\{\beta_{k,j}\}_{j=1:M_k}$ is estimated without taking into account the covariance matrices $\{R_{k\mid\mathbf{I},j}\}_{j=1:M_k}$, contrary to an optimal EM strategy (Bilmes, 1997). However, we think that our method has the advantage to be simple and fast. In the next future, we plan to compare it with the optimal strategy.

Acknowledgments

The authors would like to thank B. Cernuschi-Frias, C. Basso, A. Doucet, F. Le Gland and P. Pérez for fruitful discussions and suggestions on this work.

References

Anderson, B.D.O. and Moore, J.B. 1979. *Optimal Filtering*. Prentice Hall, Englewood Cliffs, NJ.

Arnaud, E. and Mémin, E. 2004. Optimal importance sampling for tracking in image sequences: application to point tracking. In *European Conference on Computer Vision*, Vol. 3, pp. 302–314.

Arnaud, E., Mémin, E., and Cernuschi-Frías, B. 2005. Conditional filters for image sequence based tracking—application to point tracking. *IEEE Transactions of Image Processing*, 14(1):63–79.

Arulampalam, M.S., Maskell, S., Gordon, N., and Clapp, T. 2002. A tutorial on particles filters for online nonlinear/non-Gaussian Bayesian tracking. *IEEE Transactions on Signal Processing*, 50(2):174–18.

Aschwanden, P. and Guggenbül, W. 1992. Experimental results from a comparative study on correlation type registration algorithms. In *Robust Computer Vision: Quality of Vision Algorithms*, Förstner and Ruwiedel, (Ed.), Wichmann, pp. 268–282.

Bar-Shalom, Y. and Li, X.-R. 1995. *Multitarget-Multisensor Tracking: Principles and Techniques*. Artech House Publishers.

Bilmes, J. 1997. A gentle tutorial on the em algorithm and its application to parameter estimation for gaussian mixture and hidden markov models. Technical report, University of Berkeley.

Black, M. and Anandan, P. 1996. The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. *Computer Vision and Image Understanding*, 63(1):75–104.

Black, M. and Fleet, D. 1999. Probabilistic detection and tracking of motion discontinuities. In *International Conference on Computer Vision*, Vol. 2, pp. 551–558.

Blake, A., North, B., and Isard, M. 1999. Learning multi-class dynamics. *Advances in Neural Information Processing Systems*, 11:389–395.

Brand, M. 2001. Morphable 3d models from video. In *Conference on Computer Vision and Pattern Recognition*.

Breidt, F.J. and Carriquiry, A.L. 2000. Highest density gates for target tracking. *IEEE Transactions on Aerospace and Electronic Systems*, 36:46–55.

Casella, G. and Robert, C. 1996. Rao-Blackwelisation of sampling schemes. *Biometrika*, 83(1):81–94.

Chen, R. and Liu, J.S. 2000. Mixture Kalman filters. *Journal of the Royal Statistical Society, Serie B*, 62(3):493–508.

Comaniciu, D., Ramesh, V., and Meer, P. 2003. Kernel-based object tracking. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(5):564–574.

Cuzol, A. and Mémin, E. 2005. A stochastic filter for fluid motion tracking. In *International Conference on Computer Vision*.

Doucet, A., de Freitas, N., and Gordon, N. (Eds.) 2001. *Sequential Monte Carlo Methods in Practice*. Springer-Verlag, New York. Series Statistics for Engineering and Information Science.

Doucet, A., de Freitas, N., Murphy, K., and Russell, S. 2000. Rao-Blackwellised particle filtering for dynamic Bayesian networks. In *Conference on Uncertaincy in Artificial Intelligence*.

Doucet, A., Godsill, S., and Andrieu, C. 2000. On sequential Monte Carlo sampling methods for Bayesian filtering. *Statistics and Computing*, 10(3):197–208.

Faugeras, O. and Lustman, F. 1988. Motion and structure from motion in a piecewise planar environment. Technical report, INRIA, Institut National de Recherche en Informatique et Automatique.

Godsill, S. and Clapp, T. 2001. Improvement strategies for Monte Carlo particle filters. In *(Doucet et al., 2001)*, chapter 7. Springer-Verlag.

Gordon, N. 1997. A hybrid Bootstrap filter for target tracking in clutter. *IEEE Transactions in Aerospace and Electronic Systems*, 33(1):353–358.

Gordon, N., Salmond, D., and Smith, A. 1993. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEEE Proceedings-F*, 140(2):107–113.

Gustafsson, F., Gunnarsson, F., Bergman, N., Forssell, U., Jansson, J., Karlsson, R., and Nordlund, P.-J. 2002. Particle filters for positioning, navigation and tracking. *IEEE Transactions on Signal Processing*, 50(2):425–437.

Harris, C. and Stephens, M. 1988. A combined corner and edge detector. In *Proceedings of the Alvey Vision Conference*, pp. 147–151.

Hinkley, D. 1971. Inference about the change-point from cumulative sum tests. *Biometrika*, 58(509–523).

Isard, M. and Blake, A. 1998. Condensation—conditional density propagation for visual tracking. *International Journal of Computer Vision*, 29(1):5–28.

Isard, M. and Blake, A. 1998. A mixed-state CONDENSATION tracker with automatic model-switching. In *International Conference on Computer Vision*, pp. 107–112.

Jepson, A., Fleet, D., and El-Maraghi, T. 2001. Robust online appearance models for visual tracking. In *Conference on Computer Vision and Pattern Recognition*, Vol. 1, pp. 415–422.

Khan, Z., Balch, T., and Dellaert, F. 2004. A Rao-Blackwellized particle filter for eigentracking. In *Conference on Computer Vision and Pattern Recognition*, Vol. 2, pp. 980–986.

Kong, A., Liu, J.S., and Wong, W.H. 1994. Sequential imputations and Bayesian missing data problems. *Journal of the American Statistical Association*, 89:278–288.

Kwok, C., Fox, D., and Meilă, M. 2004. Real-time particle filter. *Proceedings of the IEEE (issue on State Estimation)* 92(2).

Liu, J.S. and Chen, R. 1998. Sequential Monte Carlo methods for dynamic systems. *Journal of the American Statistical Association*, 93:1032–1044.

Ma, B. and Ellis, R.E. 2004. Surface-based registration with a particle filter. In *Procedings of Medical Image Computing and Computer-Assisted Intervention*.

Marrs, A., Maskell, S., and Bar-Shalom, Y. 2002. Expected likelihood for tracking in clutter with particle filters. In *SPIE Conference on Signal and Data Processing of Small Targets*.

Musso, C., Oudjane, N., and Le Gland, F. 2001. Improving regularised particle filters. In *(Doucet et al., 2001)*, chapter 12. Springer-Verlag.

Nguyen, H.T., Worring, M., and van den Boomgaard, R. 2001. Occlusion robust adaptative template tracking. In *International Conference on Computer Vision*, Vol. 1, pp. 678–683.

North, B. and Blake, A. 1998. Learning dynamical models using Expectation-Maximisation. In *International Conference on Computer Vision*.

Odobez, J.-M. and Bouthemy, P. 1995. Robust multiresolution estimation of parametric motion models. *Journal of Visual Communication and Image Representation*, 6(4):348–365.

Okuma, K., Taleghani, A., De Freitas, N., Little, J., and Lowe, D. 2004. A boosted particle filter: multitarget detection and tracking. In *European Conference on Computer Vision*.

Pitt, M. and Shephard, M. 2001. Auxiliary variable based particle filters. In *(Doucet et al., 2001)*, chapter 13. Springer-Verlag.

Pérez, P., Hue, C., and Gangnet, M. 2002. Color-based probabilistic tracking. In *European Conference on Computer Vision*, pp. 661–675.

Pérez, P., Vermaak, J., and Blake, A. 2004. Data fusion for visual tracking with particles. In *Proceedings of the IEEE (Issue on State Estimation)*, 92(3):495–513.

Rittscher, J. and Blake, A. 1999. Classification of human body motion. In *International Conference on Computer Vision*, pp. 634–639.

Rui, Y. and Chen, Y. 2001. Better proposal distributions: object tracking using unscented particle filter. In *Conference on Computer Vision and Pattern Recognition*, Vol. 2, pp. 786–793.

Shi, J. and Tomasi, C. 1994. Good features to track. In *Conference on Computer Vision and Pattern Recognition*, pp. 593–600.

Sidenbladh, H., Black, M., and Sigal, L. 2002. Implicit probabilistic models of human motion for synthesis and tracking. In *European Conference on Computer Vision*, Vol. 1, pp. 784–800.

Sivic, J., Schaffalizky, F., and Zisserman, A. 2004. Object lovel grouping for video shots. In *European Conference on Computer Vision*, pp. 85–98.

Sminchisescu, C. and Triggs, B. 2002. Hyperdynamic importance sampling. In *European Conference on Computer Vision*.

Sullivan, J. and Rittscher, J. 2001. Guiding random particles by deterministic search. In *International Conference on Computer Vision*, pp. 323–330.

Tissainayagam, P. and Suter, D. 2004. Assessing the performance of corner detectors for point feature tracking applications. *Image and Vision Computing*, 22(8):663–679.

Torresani, L. and Bregler, C. 2002. Space-time tracking. In *European Conference on Computer Vision*.

van der Merwe, R., Doucet, A., de Freitas, N., and Wan, E. 2000. The unscented particle filter. Technical report CUED/F-INFENG/TR 380, University of Cambridge, Engineering Department.

Vermaak, J., Andrieu, C., Doucet, A., and Godsill, S.J. 2002. Particle methods for Bayesian modeling and enhancement of speech signals. *IEEE Transactions on Speech and Audio Processing*, 10(3):173–185.

Vermaak, J., Godsill, S.J., and Pérez, P. 2005. Monte Carlo filtering for multi-target tracking and data association. *IEEE Transations on Aerospace and Electronic Systems* 41(1):309–332.

Vermaak, J., Pérez, P., Gangnet, M., and Blake, A. 2002. Towards improved observation models for visual tracking: selective adaptation. In *European Conference on Computer Vision*, Vol. 1, pp. 645–660.

Wan, E.A. and van der Merwe, R. 2000. The unscented kalman filter for nonlinear estimation. In *IEEE Symposium on Adaptive Systems for Signal Processing, Communication and Control*.

Wu, Y. and Huang, T. 2004. Robust vision tracking by integrating multiple cues based on co-inference learning. *International Journal of Computer Vision*, 58(1):55–71.