An efficient Rao-Blackwellized particle filter for object tracking

Élise Arnaud, Etienne Mémin

AN EFFICIENT RAO-BLACKWELLIZED PARTICLE FILTER FOR OBJECT TRACKING

Elise Arnaud, Etienne Mémin

IRISA, Université de Rennes 1,
Campus Beaulieu, Rennes, France

ABSTRACT

In this paper we present a technique for the tracking of textured almost planar object. The target is modeled as a noisy planar cloud of points. The tracking is led with an appropriate non linear stochastic filter. The particular system that we devised is conditionally Gaussian and can be efficiently implemented through variance reduction principle known as Rao-Blackwellisation. Our model allows also to melt a correlation measurements with dynamic model estimated from the images. Such a cooperation within a stochastic filtering framework allows the tracker to be robust to occlusions and target's unpredictable changes of speed and direction. We demonstrate the efficiency of the tracker on different types of real world sequences.

1. INTRODUCTION

Recently, the use of sequential Monte Carlo methods (also knows as particle filters) has led to the development of very efficient visual tracking techniques [8, 3, 1, 2]. The popularity of particle filters can be explained by their simplicity, their robustness to difficult situations – such as occlusions, noise, large displacements, etc. – their easiness of implementation. These techniques relies on a discrete hidden Markov chain modeling defined by a *dynamic equation* characterizing the evolution law of the unknown state (the target), and a *measurement equation* which links some available observations of the target to the state. These techniques provide an approximation of the distribution of the target state given the sequence of past and current measurements and allows as a consequence a probabilistic competition between known *a priori* information on the target and different measurements processes.

However, this kind of techniques may be unsuccessful for the tracking of objects of great dimension since sampling in high dimensional spaces is generally inefficient. In some cases, this problem can be overcame if the system has a "tractable substructure", which can be analytically marginalized out, conditionally on some part of the state. Such a methodology, leading to **Rao-Blackwellized particle filters** (RBPFs), enables reducing significantly the size of the state space over which we need to sample [3]. Rao-Blackwellized particle filters have been successfully applied for bearings only tracking and navigation, digital communication and recently in computer vision [4].

In this paper, we investigate the use of a RBPF for object tracking. To that aim, we propose to describe the object of interest through a set of feature points. The state to be estimated consists in the concatenation of all the points location. The modeling we suggest is composed of : (a) local measurement equations for each point based on a matching process and (b) a dynamic law that describes the spatial coherence of the point cloud, using motion information and geometric constraint. The resulting tracker is robust to acquisition noise, and allows partial or complete occlusions as well as deformations of the target. .

2. RAO-BLACKWELLIZED PARTICLE FILTER

2.1. Particle filtering principle

For sake of clarity, the general principle of filtering problems is briefly introduced. We consider here a discrete hidden Markov state process $\mathbf{x}_{0:n} = \{\mathbf{x}_0, \mathbf{x}_1, ..., \mathbf{x}_n\}$ of transition equation $p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{I}_{0:k})$ which represents the evolving law of the state (the dynamic equation). The conditioning on the image variable sequence, denoted here as $\mathbf{I}_{0:k}$, indicates a possible dependency of the state and the image [6]. The set of measurements $\mathbf{z}_{1:n} = \{\mathbf{z}_1, \mathbf{z}_1, ..., \mathbf{z}_n\}$, of marginal distribution $p(\mathbf{z}_k|\mathbf{x}_k, \mathbf{I}_{0:k})$ 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 $p(\mathbf{x}_{0:k}|\mathbf{z}_{1:k}, \mathbf{I}_{0:k})$. In the case of linear Gaussian models, the Kalman filter [7] gives the optimal solution[1]. If this system appears to be nonlinear and/or non Gaussian, this distribution can not be analytically computed. In such a case, an efficient approximation consists in resorting to particle filters.

These techniques propose to implement recursively an approximation of the sought density (see [8, 3] for an extended review). This approximation consists in a finite weighted sum of Dirac centered on hypothesized trajectories – called particles – of the initial system $\mathbf{x}_0$:

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

---

[1]Let us note that when the dynamic equation depends on the image data, the use of the kalman filter can be justified to recover a Gaussian approximation of the pdf through a peculiar estimator we named conditional minimum variance estimator [6]

At each time instant (or iteration), the set of particles $\{\mathbf{x}_{0:k}^{(i)}, i = 1, ..., N\}$ is drawn from an approximation of the true distribution $p(\mathbf{x}_{0:k}|\mathbf{z}_{1:k}, \mathbf{I}_{0:k})$, called the *importance function* and denoted $\pi(\mathbf{x}_{0:k}|\mathbf{z}_{1:k}, \mathbf{I}_{0:k})$. The closer the approximation from the true distribution, the more efficient the filter. The particle weights $w_k^{(i)}$ account for the deviation w.r.t. the unknown true distribution. The weights are updated according to importance sampling principle:

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

Choosing an importance function that recursively factorizes such as: $\pi(\mathbf{x}_{0:k}|\mathbf{z}_{1:k}, \mathbf{I}_{0:k}) = \pi(\mathbf{x}_{0:k-1}|\mathbf{z}_{1:k-1}, \mathbf{I}_{0:k-1}) \times \pi(\mathbf{x}_k|\mathbf{x}_{0:k-1}, \mathbf{z}_{1:k}, \mathbf{I}_{0:k})$ allows recursive evaluations in time of the particle weights as the new measurement $\mathbf{z}_k$ becomes available. Such an expression implies naturally a causal assumption of the importance function w.r.t. observations and image data. The recursive weights read then:

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

Unfortunately, such a recursive assumption of the importance function induces an increase over time of the weight variance. In practice, this makes the number of significant particles decreases dramatically over time. To limit such a degeneracy, two methods have been proposed.

A first solution consists in using an *optimal* importance function which minimizes the variance of the weights conditioned upon $\mathbf{x}_{0:k-1}$, $\mathbf{z}_{1:k}$ and $\mathbf{I}_{0:k}$ in our case. It is then possible to prove that $p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{z}_k, \mathbf{I}_{0:k})$ is this optimal distribution. With this distribution, the recursive formulation of w_k becomes: $w_k^{(i)} = w_{k-1}^{(i)}\, p(\mathbf{z}_k|\mathbf{x}_{k-1}^{(i)}, \mathbf{I}_{0:k})$. The optimal importance function is usually not accessible. In vision applications, this function is often set to the prediction density, i.e. $\pi(\mathbf{x}_k|\mathbf{x}_{0:k-1}, \mathbf{z}_{1:k}, \mathbf{I}_{0:k}) = p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{I}_{0:k})$. Such a choice has the drawback to exclude the measurements from the particle diffusion step. For conditional Gaussian systems such as the one we are dealing with, the importance function is analytically known [6, 9]. This constitutes one of the main advantages of the system we are considering. The optimal importance function provides a simple and natural mechanism to include the measurements into the sampling process. Such an option is for instance very beneficial in case of occlusions [6].

A second solution to tackle the problem of weight variance increase relies on the use of re-sampling methods. Such methods remove trajectories with weak normalized weights, and add copies of the trajectories associated to strong weights, as soon as the number of significant particles is too small. Obviously, these two solutions may be coupled for a better efficiency.

2.2. Rao-Blackwellized particle filter

As a general principle, particle filters can be applied whatever the size of the state space. However, it is crucial to take advantage of the system's properties in order to reduce the dimension of the sampling space. When some state vector components can be estimated conditionally to the others with an optimal filter (such as the Kalman filter), it is possible to apply the Rao-Blackwellization principle of variance reduction.

Let us assume that $\mathbf{x}_k$ can be decomposed into two groups of components $\mathbf{r}_k$ and $\mathbf{y}_k$, such as:
$$\begin{aligned} p(\mathbf{x}_k|\mathbf{x}_{k-1}, \mathbf{I}_{0:k}) &= p(\mathbf{r}_k, \mathbf{y}_k|\mathbf{r}_{k-1}, \mathbf{y}_{k-1}, \mathbf{I}_{0:k}) \\ &= p(\mathbf{r}_k|\mathbf{r}_{k-1}, \mathbf{I}_{0:k})\, p(\mathbf{y}_k|\mathbf{r}_{k-1:k}, \mathbf{y}_{k-1}, \mathbf{I}_{0:k}). \end{aligned}$$

Let us also assume that the posterior density $p(\mathbf{y}_{0:k}|\mathbf{r}_{0:k}, \mathbf{z}_{1:k}, \mathbf{I}_{0:k})$ can be optimally estimated. As the objective distribution reads :

$$p(\mathbf{x}_{0:k}|\mathbf{z}_{1:k}, \mathbf{I}_{0:k}) = p(\mathbf{y}_{0:k}|\mathbf{r}_{0:k}, \mathbf{z}_{1:k}, \mathbf{I}_{0:k})\, p(\mathbf{r}_{0:k}|\mathbf{z}_{1:k}, \mathbf{I}_{0:k}),$$

the only remaining difficulty is the estimation of $p(\mathbf{r}_{0:k}|\mathbf{z}_{1:k}, \mathbf{I}_{0:k})$. Compared with the objective density this density lives in a space of reduced dimension and is assumed to be approximated through particle filtering. Given a particle swarm $\{\mathbf{r}_{0:k}^{(i)}, w_k^{(i)}\}_{i=1:N}$ approximating $p(\mathbf{r}_{0:k}|\mathbf{z}_{1:k}, \mathbf{I}_{0:k})$, we have the following expressions:

$$p(\mathbf{y}_{0:k}|\mathbf{z}_{1:k}, \mathbf{I}_{0:k}) \simeq \sum_{i=1:N} w_k^{(i)}\, p(\mathbf{y}_{0:k}|\mathbf{r}_{0:k}^{(i)}, \mathbf{z}_{1:k}, \mathbf{I}_{0:k}),$$

$$p(\mathbf{y}_k|\mathbf{z}_{1:k}, \mathbf{I}_{0:k}) \simeq \sum_{i=1:N} w_k^{(i)}\, p(\mathbf{y}_k|\mathbf{r}_{0:k}^{(i)}, \mathbf{z}_{1:k}, \mathbf{I}_{0:k}).$$

The resulting algorithm, named the Rao-Blackwellized particle filter (RBPF), consists finally in three main steps. First, an importance sampling stage is performed to generate the swarm $\{\mathbf{r}_{0:k}^{(i)}, w_k^{(i)}\}_{i=1:N}$. For each particle $\mathbf{r}_{0:k}^{(i)}$, an approximation of $p(\mathbf{y}_{0:k}|\mathbf{r}_{0:k}^{(i)}, \mathbf{z}_{1:k}, \mathbf{I}_{0:k})$ is estimated with an optimal filter. In order to avoid particle degeneracy a final re-sampling step is generally required.

2.3. Conditionally Gaussian systems

Let us now consider the particular case of a conditionally Gaussian system:
$$\begin{cases} \mathbf{r}_k = f_k(\mathbf{r}_{k-1}) + \mathbf{w}_k^{\mathbf{r}} \\ \mathbf{y}_k = H_k\, \mathbf{y}_{k-1} + \mathbf{w}_k^{\mathbf{y}} \\ \mathbf{z}_k^{\mathbf{r}} = G_k\, \mathbf{r}_k + \mathbf{v}_k^{\mathbf{r}} \\ \mathbf{z}_k^{\mathbf{y}} = K_k\, \mathbf{y}_k + \mathbf{v}_k^{\mathbf{y}} \end{cases} \tag{1}$$

The state to be estimated is $\mathbf{x}_k = [\mathbf{r}_k, \mathbf{y}_k]^t$, the measurements are denoted $\mathbf{z}_k = [\mathbf{z}_k^{\mathbf{r}}, \mathbf{z}_k^{\mathbf{y}}]^t$ and f_k is a nonlinear function which possibly depends on the image sequence. Matrices H_k, G_k, and K_k may also be specified from the image data. $\mathbf{w}_k^{\mathbf{r}}, \mathbf{w}_k^{\mathbf{y}}, \mathbf{v}_k^{\mathbf{r}}, \mathbf{v}_k^{\mathbf{y}}$ are assumed to be independent zero-mean Gaussian noises of covariances $Q_k^{\mathbf{r}}, Q_k^{\mathbf{y}}, R_k^{\mathbf{r}}$ and $R_k^{\mathbf{y}}$. It is important to notice that the matrices H_k and K_k are available only when the estimation of $\mathbf{r}_{0:k-1}$ is known.

Such a system has very interesting properties. First of all, the distribution $p(\mathbf{y}_k|\mathbf{r}_{0:k}, \mathbf{z}_{1:k}, \mathbf{I}_{0:k})$ is Gaussian and is entirely described by its two first moments. The RBPF is in that case equivalent to a Kalman filter bank. Moreover, as for the estimation of $p(\mathbf{r}_{0:k}|\mathbf{z}_{1:k}, \mathbf{I}_{0:k})$, this system enables using the optimal importance function for the sampling of particles $\mathbf{r}_{0:k}^{(i)}$. Indeed, it can be shown that the two required densities $p(\mathbf{r}_k|\mathbf{r}_{k-1}, \mathbf{z}_k^{\mathbf{r}}, \mathbf{I}_{0:k})$ and $p(\mathbf{z}_k^{\mathbf{r}}|\mathbf{r}_{k-1}, \mathbf{I}_{0:k})$ are Gaussian of mean $\mu_k = \Sigma_k \left(Q_k^{\mathbf{r}\,-1} f_k(\mathbf{r}_{k-1}) + G_k^t R_k^{\mathbf{r}\,-1} \mathbf{z}_k^{\mathbf{r}} \right)$ and of covariance $\Sigma_k = (Q_k^{\mathbf{r}\,-1} + G_k^t R_k^{\mathbf{r}\,-1} G_k)^{-1}$ for the first pdf and of mean $\mu_k = G_k f_k(\mathbf{r}_{k-1})$ and of covariance $(R_k^{\mathbf{r}} + G_k Q_k^{\mathbf{r}} H_k^{\mathbf{r}\,t})$ for the second.

3. MODEL FOR POINT SET TRACKING

Back to our application, the aim of this section is to describe the model we propose to track a cloud of features points. In the considered tracking problem, the state $\mathbf{x}_k$ represents the concatenation of all the points location at time k, in image $\mathbf{I}_k$, observable through the measurement $\mathbf{z}_k$.

3.1. A dynamic equation in two steps

In order to benefit from the information available on the object motion and on its geometric structure, the point set is separated in two subsets $\mathbf{r}_k$ (called the *reference points*) and $\mathbf{y}_k$ (called the *attached points*) . Each of these subsets will capture one of these two pieces of information. The dynamic equation describing the evolution of $\mathbf{x}_k = [\mathbf{r}_k, \mathbf{y}_k]^t$ is then built in two stages.

• The first stage concerns the displacement of the reference points $\mathbf{r}_k$. In order to be reactive to any unpredictable change of speed and direction, the dynamic we consider is based on an instantaneous motion vector $\omega_k(\mathbf{r}_k)$ estimated on line from $\mathbf{I}_{0:k}$. A robust parametric motion estimation technique [10] is used to estimate reliably a 2D parametric model representing the dominant apparent velocity field on a given support $\mathcal{R}$. Assuming an affine motion, we have $\omega_k(\mathbf{r}_k) = P(\mathbf{r}_k)\boldsymbol{\theta}_k$, where $P(\mathbf{r}_k)$ is a matrix whose entries depend on the locations of $\mathbf{r}_k$, and $\boldsymbol{\theta}_k$ is the motion parameter vector of dimension 6. The use of such a method on an appropriate local support around $\mathbf{r}_{k-1}$ provides an estimate of the motion vector for each motion point. As $\mathcal{R}$ is a domain depending on $\mathbf{r}_{k-1}$, the estimated parameter vector $\boldsymbol{\theta}_k$ depends in a nonlinear way on $\mathbf{r}_{k-1}$. The resulting state equation is:

$$\mathbf{r}_k = \mathbf{r}_{k-1} + P(\mathbf{r}_{k-1})\,\boldsymbol{\theta}_k + \mathbf{w}_k^{\mathbf{r}}. \qquad (2)$$

The noise variable $\mathbf{w}_k^{\mathbf{r}}$ accounts for errors related to the local motion model. It is assumed to follow a zero mean Gaussian distribution of fixed covariance $Q_k^{\mathbf{r}}$.

• The second stage concerns the evolution of the attached points $\mathbf{y}_k$. At this level, an estimation of $\mathbf{r}_{k-1}$ and $\mathbf{r}_k$ is available. This enables us to calculate a matrix that links, in a geometric point of view, the two point subsets. Several

constraints may be employed such as homographic or affine constraint. The only limitation related to the use of a RBPF is that the resulting dynamic equation of the attached points must be linear :

$$\mathbf{y}_k = H_k(\mathbf{r}_{k-1:k})\,\mathbf{y}_{k-1} + \mathbf{w}_k^{\mathbf{y}}. \qquad (3)$$

Matrix H_k depends on $\mathbf{r}_{k-1:k}$. It is estimated once an estimate of the reference points location is known. The noise variable $\mathbf{w}_k^{\mathbf{y}}$ accounts for errors related to the homography matrix estimation. It is assumed to follow a zero mean Gaussian distribution of fixed covariance $Q_k^{\mathbf{y}}$.

3.2. Measurement equation

The proposed dynamic equation introduces a spatial and geometric coherence of the the cloud points. In a complementary way, to improve the tracker robustness to occlusion, a local measurement model is considered for each feature of the cloud. At time k, we assume that each feature point is observable through a matching process. We then define the observation vector $\mathbf{z}_k$ as the vector gathering all the correlation peaks of each point composing $\mathbf{x}_k$. This vector is determined through a correlation technique between the current image and a target representation. This representation which constitutes a reference image of the object is composed by a collection of local image patches centered on the initial point locations in the first image of the sequence. Several correlation criteria may be used to quantify the similarity between two points. The consistency assumption of a luminance pattern has simply led us to consider the sum-of-squared-differences (SSD) criterion. The resulting measurement equation we consider reads:

$$[\mathbf{z}_k^{\mathbf{r}}, \mathbf{z}_k^{\mathbf{y}}]^t = [\mathbf{r}_k, \mathbf{y}_k]^t + [\mathbf{v}_k^{\mathbf{r}}, \mathbf{v}_k^{\mathbf{y}}]^t \qquad (4)$$

The noise variables $\mathbf{v}_k^{\mathbf{r}}$, $\mathbf{v}_k^{\mathbf{y}}$ are assumed to be distributed according to zero mean Gaussian distributions with covariances $R_k^{\mathbf{r}}$ and $R_k^{\mathbf{y}}$. These matrices are block-diagonal matrices. For each point, the associated block accounts for an accuracy measurement of the matching procedure. More details concerning this estimation are given in [6]. The used accuracy criterion consists roughly in representing the SSD surface as two possible pdfs corresponding either to an occlusion situation or to a standard situation where a noisy measurement is available. Statistical tests and on-line parameter estimations enable us to set the distribution of interest. Such an on-line estimations of the noise covariance matrices improve greatly the tracker robustness to occlusions.

3.3. Final algorithm

Our points cloud tracking problem is finally modeled by a system composed of equations (2), (3), (4). This system is clearly conditionally Gaussian of the form (1). It is solved by an efficient Rao-Blackwellized particle filter whose particles are sampling using the optimal importance function.

| initial point cloud, # 1 | # 9 | # 12 | # 14 | # 18 |
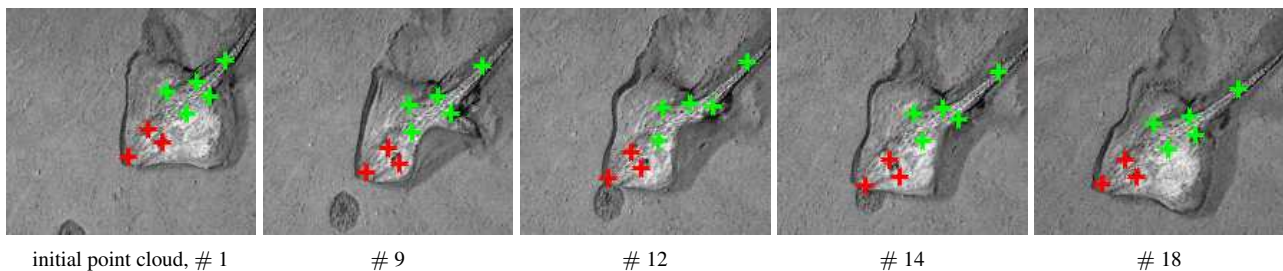
Fig. 1. Fish sequence : use of an affine constraint for deformable object tracking

4. EXPERIMENTAL RESULTS

In this section, we present some experimental results on two real-world sequences to demonstrate the efficiency of the proposed object tracker. On these results, the red crosses show the *reference points* and the green crosses account for the *attached points*.

The first sequence shows a skate evolving on sand. The skate is characterized by its large deformations. As shown on fig. 1, the fish trajectory is well reconstructed, despite the absence of high photometric contrasts. For this sequence, an affine model has been use as geometric constraint. The second result (fig. 2) presents the tracking of a book on a cluttered desk. As we focus on this sequence on a planar object tracking, we used an homography model for the geometric constraint. As can be observed on this sequence, the target undergoes a complex trajectories and is in the same time occluded twice. in spite of these difficulties our tracker has been able to recover successfully the trajectory of the points cloud.

initial point cloud, # 0	# 15
# 22	# 39
# 44	# 59
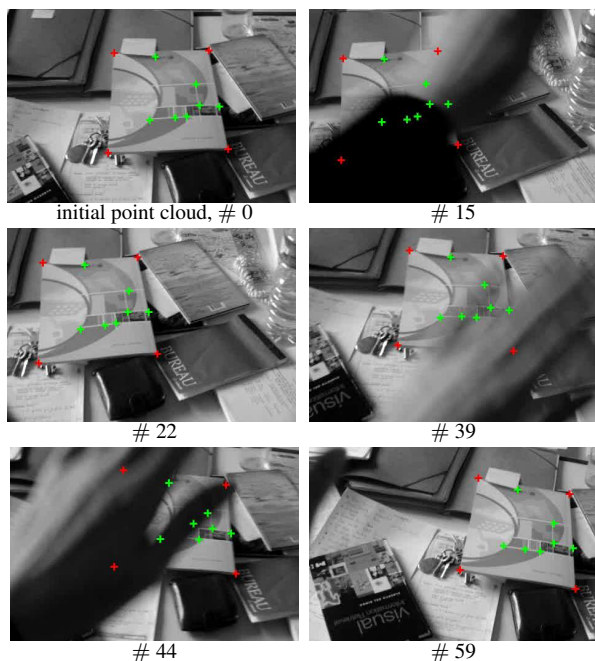
Fig. 2. Office sequence : use of an homographic constraint for planar object tracking

5. CONCLUSION

In this paper we have proposed a technique for the tracking of textured objects. The technique relies on an efficient particle filter. The tracker is robust to occlusion and to complex unpredictable trajectories. The method is devoted to the tracking of objects that can be described through a collection of points. A part of these points must also exhibits a linear relation with respect to the others.

6. REFERENCES

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

[2] P. Pérez, Jaco Vermaak, and Andrew Blake, "Data fusion for visual tracking with particles," *Proceedings of the* IEEE *(issue on State Estimation)*, vol. 92, no. 3, pp. 495–513, 2004.

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

[4] Z. Khan, T. Balch, and F. Dellaert, "a Rao-Blackwellized particle filter for eigentracking," in *Conference on Computer Vision and Pattern Recognition*, 2004, vol. 2, pp. 980–986.

[5] C. Harris and M. Stephens, "A combined corner detection and edge detector," in *Proceedings of the Alvey Vision Conference*, 1988, pp. 147–151.

[6] E. Arnaud, E. Mémin, and B. Cernuschi-Frías, "Conditional filters for image sequence based tracking - application to point tracking," IEEE *Transactions of Image Processing*, 2005.

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

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

[9] E. Arnaud and E. Mémin, "Optimal importance sampling for tracking in image sequences: application to point tracking," in *European Conference on Computer Vision*, 2004.

[10] J.-M. Odobez and P. Bouthemy, "Robust multiresolution estimation of parametric motion models," *Journal of Visual Communication and Image Representation*, vol. 6, no. 4, pp. 348–365, 1995.