ROBUST VISUAL TRACKING VIA MCMC-BASED PARTICLE FILTERING

D-N. Truong Cong[1], *F. Septier*[1], *C. Garnier*[1], *L. Khoudour*[2], *Y. Delignon*[1]

[1]Institut TELECOM / Telecom Lille1 / LAGIS UMR CNRS 8219, France
[2]CETE du Sud-Ouest: Centre d'Etudes Techniques de l'Equipement, France

ABSTRACT

We present in this paper a new visual tracking framework based on the MCMC-based particle algorithm. Firstly, in order to obtain a more informative likelihood, we propose to combine the color-based observation model with a detection confidence density obtained from the Histograms of Oriented Gradients (HOG) descriptor. The MCMC-based particle algorithm is then employed to estimate the posterior distribution of the target state to solve the tracking problem. The global system has been tested on different real datasets. Experimental results demonstrate the robustness of the proposed system in several difficult scenarios.

***Index Terms*—** Visual tracking, particle filtering, MCMC, HOG.

1. INTRODUCTION

Visual tracking is one of the most fundamental tasks in many computer vision applications, such as intelligent visual surveillance, human-computer interaction, traffic monitoring or video indexing. Among the numerous tracking methods proposed in the literature [1], particle filtering (PF), which was first introduced by Isard and Blake [2], has obtained considerable success in various kinds of visual tracking problems. Such a method recursively approximates the posterior probability density function with a set of weighted random sampled particles evolving in the state space. The obtained estimates can be set arbitrarily close to the optimal solution (in the Bayesian sense) at the expense of computational complexity.

Many particle filtering-based visual trackers have been proposed in the literature. Most of them attempt to reinforce observation models, which must be robust to occlusions, pose changes, camera viewpoints, and environment variations. The most common approach consists in constructing an adaptive appearance-based model, such as color histograms [3, 4], mixture of Gaussians [5, 6], or multiple features fusion [7]. Some systems try to improve the dissimilarity measure between observation models in order to improve the tracking performance [6, 8]. Other contributions include the use of affine transformations in the state space model [9, 10, 11]. The authors show that, by defining the state equations on the two-dimensional affine group, the tracking robustness is enhanced considerably in difficult scenarios.

Some other approaches have also been proposed in the literature to improve the traditional importance sampling step in the particle filters [12, 13, 14]. Indeed, due to their sampling mechanism, particle filters tend to be inefficient when applied to high dimensional problems such as target tracking. A notable contribution is the use of Markov Chain Monte Carlo (MCMC) in a sequential setting. In [13], Khan et al. replace the traditional importance sampling step in the particle filter with a MCMC sampling step to obtain a more efficient MCMC-based multi-target filter. However, the computational demand of the proposed algorithm can become excessive as the number of particles increases owing to the direct Monte Carlo computation of the prediction density at each time step. To avoid this numerical integration, the authors in [14, 15] propose an alternative sequential MCMC algorithm.

In this paper, we address the tracking problem in complex scenes using a single uncalibrated camera. We present a visual tracking framework based on the MCMC-based particle algorithm presented by Septier et al. [15, 16] with some improvements. In order to reinforce the likelihood measurement, we propose to combine a color-based observation model with additional probabilistic information obtained from the Histograms of Oriented Gradients (HOG) detector. This approach is different from [7] which fuses the information of color histogram and HOG descriptor as a single human feature to track. The MCMC-based particle algorithm is then employed to estimate the posterior distribution of the target state to solve the tracking problem.

The outline of the paper is as follows: in Section 2, we present the proposed method for object tracking. Section 3 presents performance results of the system on several real datasets. Finally, in Section 4, conclusions and important short-term perspectives are given.

2. VISUAL TRACKING PROBLEM FORMULATION

In this paper, we focus on the problem of single object visual tracking. The aim is to estimate the conditional probability $p\left(\mathrm{X}_k|Z_{0:k}\right)$ of the target state X_k at time k given the sequence of observations $Z_{0:k} = (Z_0, \ldots, Z_k)$. This posterior probability $p\left(\mathrm{X}_k|Z_{0:k}\right)$, known as the filtering distribution, can be expressed recursively using the Bayes filter equation:

$$p\left(\mathrm{X}_k\,|Z_{0:k}\right) \propto \int p\left(Z_k\,|\mathrm{X}_k\right) p\left(\mathrm{X}_k\,|\mathrm{X}_{k-1}\right) p\left(\mathrm{X}_{k-1}\,|Z_{0:k-1}\right) d\mathrm{X}_{k-1} \quad (1)$$

where the dynamic model $p\left(\mathrm{X}_k\,|\mathrm{X}_{k-1}\right)$ governs the temporal evolution of the state X_k given the previous state X_{k-1}, and the observation likelihood model $p\left(Z_k\,|\mathrm{X}_k\right)$ measures the fitting accuracy of the observation data Z_k given the state X_k.

In the following subsections, we first define the dynamic and observation models used in our system, and then describe in detail the proposed MCMC-based tracking algorithm.

2.1. Dynamic and observation models

Given the state vector $\mathrm{X}_k = [\mathbf{c}_k, \mathbf{r}_k]^T$, where $\mathbf{c}_k = [x_k, y_k]$ are the coordinates of the object centroid and $\mathbf{r}_k = [r_k^x, r_k^y]$ are the width

This work was supported by the DISCOVER project 2011-2012 funded by the Institut Telecom.

and height of the object, the state evolution is defined as:

$$\begin{cases} \Sigma_k \sim \mathcal{IW}\left(\Sigma_k \,|n, \Sigma_{k-1}\right) \\ \mathrm{X}_k \sim \mathcal{N}\left(\mathrm{X}_k \,|\mathrm{X}_{k-1}, \Sigma_k\right) \end{cases} \tag{2}$$

where the state vector X_k has a Gaussian distribution with mean vector X_{k-1} and covariance matrix Σ_k. The covariance matrix Σ_k, which defines the region of uncertainty around the current state, follows an inverse Wishart distribution with n degrees of freedom and scale matrix Σ_{k-1}. As in [17], the covariance matrix is modeled as a dynamic random variable in order to adapt to motion changing.

To evaluate how likely a candidate region represents the target, we define the likelihood model by combining two sources of information: the color-based similarities estimated by using a region-based color histogram coupled with the Earth Mover's Distance (EMD), and a detection confidence density obtained from the intermediate output of the HOG-based detector. Each term is described in detail below.

Color-based model. The region of interest is first horizontally decomposed in p equidistant bands. Fixing the number of bands rather than their size allows obtaining invariance whatever the scale of the region. The histogram of each color component is then computed in each band. The region-based color histogram is thus composed of $n = p \times b \times c$ values, where p is the total number of equal parts of the region, b is the number of histogram bins, and c is the number of color components. The advantages of such a model are the consideration of the spatial information and the simple estimation.

The similarity between the candidate and the reference models is estimated by using Earth Mover's Distance (EMD) [18], which computes the matching cost between two histograms of each band and each color component. The similarity is thus defined as:

$$L_C = \exp\left(-\sum_{i=1}^{p}\sum_{j=1}^{c} EMD\left(h_{ij}, h'_{ij}\right)/\sigma^2\right) \tag{3}$$

where h_{ij} and h'_{ij} denote the histograms of the color component j and the corresponding band i of two models, σ is a predefined parameter.

Detection confidence. The second term of the likelihood model is based on the detection confidence density built up by using the HOG detector [19]. This intermediate information, obtained before applying non-maximum suppression, has already been integrated in the likelihood model in [20] in combination with the final results of the HOG detector. Here we only exploit the raw output obtained in the Support Vector Machine (SVM) classification step to define a confidence map. Given a pixel k and all the sliding-windows w_i to which the pixel k belongs, the detection confidence score of the pixel k is given by:

$$p_d\left(k\right) = \sum_{\forall w_i \ni k} \exp\left[\alpha\left(e\left(w_i\right) - e_{\max}\right)\right] \tag{4}$$

where $e\left(w_i\right)$ is the distance obtained from the SVM output of window w_i, $e_{\max}$ is a parameter for normalizing the SVM classification output, and α is a parameter modifying the discrimination level between the presence and non-presence of people in an image.

Once the detection confidence scores of all pixels of an image have been calculated, they are normalized by their sum in order to obtain a spatial distribution. Figure 1 presents an example of the detection confidence map for a given frame.

Given a candidate state X_k^i, the detection confidence-based likelihood L_H is defined as the sum of the detection confidence scores of all pixels belonging to the region r defined by X_k^i. The final likelihood of the candidate state X_k^i is thus defined as the product of these two terms:

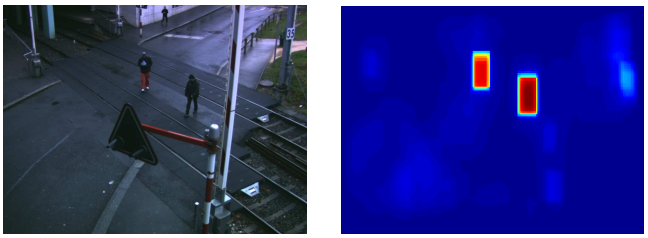

Fig. 1. Detection confidence map (left to right: original image, confidence map).

$$p\left(Z_k \left| \mathrm{X}_k^i \right.\right) \propto L_C \times L_H \tag{5}$$

2.2. Visual tracking via MCMC-based particle filter

2.2.1. *Overview*

As mentioned above, our tracking framework is based on the MCMC-based particle algorithm proposed in [15, 16] which considers the general joint posterior distribution of S_k and S_{k-1} as the target distribution:

$$p\left(\mathrm{S}_k, \mathrm{S}_{k-1} \,|Z_{0:k}\right) \propto p\left(\mathrm{S}_k \,|\mathrm{S}_{k-1}\right) p\left(Z_k \,|\mathrm{S}_k\right) p\left(\mathrm{S}_{k-1} \,|Z_{0:k-1}\right) \tag{6}$$

The MCMC procedure is used to make inference from this joint distribution. The posterior distribution $p\left(\mathrm{S}_{k-1} \,|Z_{0:k-1}\right)$ at time $k-1$ is first approximated by an empirical distribution based on the current particle set $\hat{p}\left(\mathrm{S}_{k-1} \,|Z_{0:k-1}\right)$:

$$p\left(\mathrm{S}_{k-1} \,|Z_{0:k-1}\right) \approx \frac{1}{N_p}\sum_{p=1}^{N_p} \delta\left(\mathrm{S}_{k-1} - \mathrm{S}_{k-1}^p\right) \tag{7}$$

where N_p is the number of particles used in the algorithm and p is the particle index. Then, after many joint draws from Eq.(6) using an appropriate MCMC scheme, the converged MCMC output for S_k can be extracted to give an updated marginalized particle approximation of $p\left(\mathrm{S}_k \,|Z_{0:k}\right)$. In this way, sequential inference can be achieved.

2.2.2. *MCMC-based particle filter*

In our problem, the state of interest is $\mathrm{S}_k = \{\mathrm{X}_k, \Sigma_k\}$ and the target distribution is the joint posterior distribution $p\left(\mathrm{X}_k, \mathrm{X}_{k-1}, \Sigma_k, \Sigma_{k-1} \,|Z_{0:k}\right)$. At each MCMC iteration, a joint Metropolis-Hastings (MH) proposal step is first carried out for jointly updating $\{\mathrm{X}_k, \mathrm{X}_{k-1}, \Sigma_k, \Sigma_{k-1}\}$. Then, X_k and Σ_k are updated individually by using the refinement Metropolis-within-Gibbs step. These two steps are repeated $(N_b + N_p)$ times, where N_b is the burn-in period length and N_p is the number of particles. The detail of the m-th iteration of the proposed algorithm at time k is described in the following:

(i) Joint MH proposal step. For the joint proposal step, we simulate a sample $\{\mathrm{X}_k^*, \mathrm{X}_{k-1}^*, \Sigma_k^*, \Sigma_{k-1}^*\}$ from the proposal function $q_1\left(\mathrm{X}_k, \mathrm{X}_{k-1}, \Sigma_k, \Sigma_{k-1} | \, Z_{0:k}\right)$ given by:

$$q_1\left(\mathrm{X}_k, \mathrm{X}_{k-1}, \Sigma_k, \Sigma_{k-1} | \, Z_{0:k}\right) = q_{11}\left(\mathrm{X}_k \,|\mathrm{X}_{k-1}, \Sigma_k\right) q_{12}\left(\Sigma_k \,|\Sigma_{k-1}\right) q_{13}\left(\mathrm{X}_{k-1}, \Sigma_{k-1} \,|Z_{0:k-1}\right) \tag{8}$$

where $q_{11}\left(\mathrm{X}_k \left| \mathrm{X}_{k-1}, \Sigma_k \right.\right)$ and $q_{12}\left(\Sigma_k \left| \Sigma_{k-1} \right.\right)$ are the state evolution defined by Eq. (2), $q_{13}\left(\mathrm{X}_{k-1}, \Sigma_{k-1} \left| Z_{0:k-1} \right.\right)$ is the particle approximation of the posterior distribution $p\left(\mathrm{X}_{k-1}, \Sigma_{k-1} \left| Z_{0:k-1} \right.\right)$ obtained at the previous time step $k-1$.

Thus, suppose that at time $k-1$, there are N_p samples $\left\{\mathrm{X}_{k-1}^j, \Sigma_{k-1}^j\right\}_{j=1}^{N_p}$ drawn from the previous filtering density $p\left(\mathrm{X}_{k-1}, \Sigma_{k-1} \left| Z_{0:k-1} \right.\right)$. We first randomly select a joint sample $\left\{\mathrm{X}_{k-1}^*, \Sigma_{k-1}^*\right\}$ from this set, then given $\left\{\mathrm{X}_{k-1}^*, \Sigma_{k-1}^*\right\}$, we sample from the dynamic model defined by Eq. (2) to obtain $\left\{\mathrm{X}_k^*, \Sigma_k^*\right\}$.

The proposed candidate is accepted with probability $\rho_1 = \min\left(1, \dfrac{p\left(Z_k \left| \mathrm{X}_k^* \right.\right)}{p\left(Z_k \left| \mathrm{X}_k^{m-1} \right.\right)}\right)$.

(ii) Refinement step. For the individual refinement steps, we propose to sample the two main components of the state vector X_k and the covariance matrix Σ_k successively. The coordinates of the object centroid, $\mathbf{c}_k$, is refined with the proposal function $q_2\left(\mathbf{c}_k \left| \mathbf{c}_{k-1}^m, \mathbf{r}_k^m, \mathbf{r}_{k-1}^m, \Sigma_k^m \right.\right)$ given by:

$$\mathbf{c}_k^* \sim q_2\left(\mathbf{c}_k \left| \mathbf{c}_{k-1}^m, \mathbf{r}_k^m, \mathbf{r}_{k-1}^m, \Sigma_k^m \right.\right) = \mathcal{N}\left(\mathbf{c}_k \left| \tilde{\mu}_k^m, \tilde{\Sigma}_k^m \right.\right) \quad (9)$$

with mean vector $\tilde{\mu}_k^m = \mathbf{c}_{k-1}^m + \Sigma_{k(12)}^m \left(\Sigma_{k(22)}^m\right)^{-1}\left(\mathbf{r}_k^m - \mathbf{r}_{k-1}^m\right)$ and covariance matrix $\tilde{\Sigma}_k^m = \Sigma_{k(11)}^m - \Sigma_{k(12)}^m \left(\Sigma_{k(22)}^m\right)^{-1}\Sigma_{k(21)}^m$ given $\Sigma_k^m = \begin{bmatrix} \Sigma_{k(11)}^m & \Sigma_{k(12)}^m \\ \Sigma_{k(21)}^m & \Sigma_{k(22)}^m \end{bmatrix}$

The acceptation probability is $\rho_2 = \min\left(1, \frac{p\left(Z_k \left| \mathbf{c}_k^*, \mathbf{r}_k^m \right.\right)}{p\left(Z_k \left| \mathrm{X}_k^m \right.\right)}\right)$.

The refinement procedure for the size of the target $\mathbf{r}_k$ is carried out similarly with the following proposal function:

$$\mathbf{r}_k^* \sim q_3\left(\mathbf{r}_k \left| \mathbf{r}_{k-1}^m, \mathbf{c}_k^m, \mathbf{c}_{k-1}^m, \Sigma_k^m \right.\right) = N\left(\mathbf{r}_k \left| \hat{\mu}_k^m, \hat{\Sigma}_k^m \right.\right) \quad (10)$$

where $\hat{\mu}_k^m = \mathbf{r}_{k-1}^m + \Sigma_{k(21)}^m \left(\Sigma_{k(11)}^m\right)^{-1}\left(\mathbf{c}_k^m - \mathbf{c}_{k-1}^m\right)$ and $\hat{\Sigma}_k^m = \Sigma_{k(22)}^m - \Sigma_{k(21)}^m \left(\Sigma_{k(11)}^m\right)^{-1}\Sigma_{k(12)}^m$.

The acceptation probability is $\rho_3 = \min\left(1, \frac{p\left(Z_k \left| \mathbf{c}_k^m, \mathbf{r}_k^* \right.\right)}{p\left(Z_k \left| \mathrm{X}_k^m \right.\right)}\right)$.

The final refining step regarding the matrix covariance Σ_k is based on the proposal function $q_4\left(\Sigma_k \left| \Sigma_k^m, \Sigma_{k-1}^m \right.\right)$ defined as:

$$\Sigma_k^* \sim q_4\left(\Sigma_k \left| \Sigma_k^m, \Sigma_{k-1}^m \right.\right) = \mathcal{IW}\left(\Sigma_k \left| n, \Sigma_{k-1}^m \right.\right) \quad (11)$$

The acceptance probability is $\rho_4 = \min\left(1, \dfrac{p\left(\mathrm{X}_k^m \left| \mathrm{X}_{k-1}^m, \Sigma_k^* \right.\right)}{p\left(\mathrm{X}_k^m \left| \mathrm{X}_{k-1}^m, \Sigma_k^m \right.\right)}\right)$.

3. EXPERIMENTAL RESULTS

In this section, we demonstrate the performance of our tracking algorithm on challenging scenarios extracted from PETS'06 dataset [21] and from our own dataset captured at a level crossing (named LC dataset). For each experimentation, the initialization is manual and the MCMC-based particle filters are run with $N_p = 1000$ particles and a burning period of 500 iterations. The color-based model is formed by concatenating 4 bands with 8-bin histograms for each R, G, B channel. We also compare our method to the well-known PF-based tracker. For a fair comparison, both methods adopt the same dynamical model and observation model as described in Section 2.1.

For the PETS'06 dataset, the tracking results shown in Figure 2 appear to be satisfactory with regard to the presence of occlusion. The scenario extracted from the LC dataset is more challenging because of the presence of partial occlusion and the large variation in the target motion. Figure 3 visually compares the results obtained by the PF-based tracker and the proposed algorithm. Although both trackers never lose the target, the tracking accuracy of our tracker is consistently better than that of PF-based tracker.

Fig. 2. Tracking results obtained from the PETS'06 dataset (occlusion occurred at frame 1070 between two people).

In order to perform a quantitative analysis of the proposed approach, we have manually segmented the targets in two sequences illustrated in Figures 2 and 3. The centroid and the size of these segmentations are used as the ground truth data. The performances of the tracking system are evaluated by calculating the difference in pixel between the ground truth data and the target state vector estimated by the trackers.

Figure 4 shows the per-frame tracking errors for the sequence shown in Figure 3 obtained by using four trackers: conventional PF and MCMC-based particle using the color-based observation model and the proposed observation model (i.e. the color descriptor coupled with the HOG-based confidence map). Table 1 resumes the error means in terms of centroid and size of the target of both sequences in Figures 2 and 3. One can notice that the MCMC method improves considerably the tracking performance in comparison with the PF-based approach. Moreover, the use of the proposed HOG-based confidence map in the observation model leads to better results thanks to the additional information provided in the likelihood.

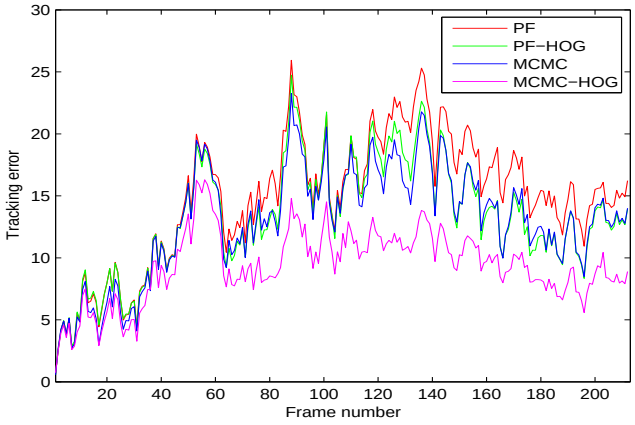

Fig. 4. Per-frame tracking errors for the sequence shown in Figure 3 obtained by different methods.

4. CONCLUSION

In this paper, we have presented a new visual tracking framework based on the MCMC-based particle algorithm. We have also proposed to combine the color-based observation model with the detection confidence density obtained from the HOG descriptor in order

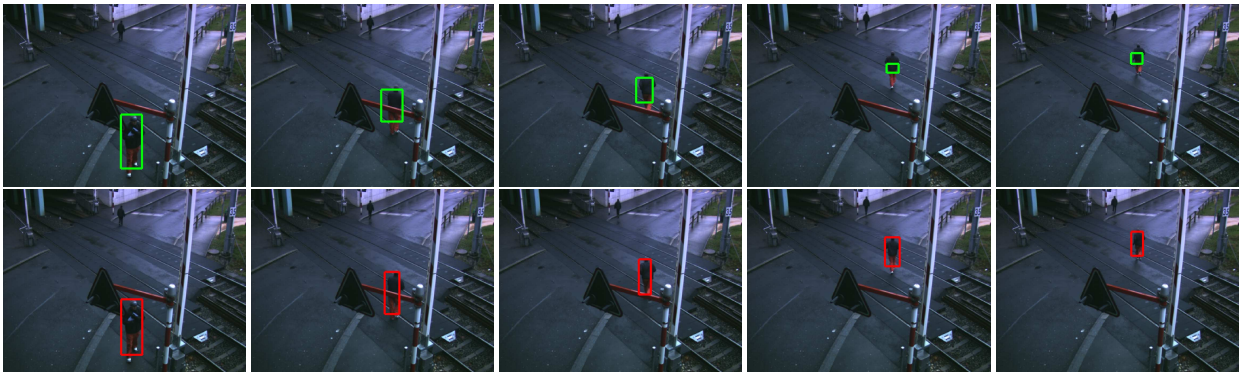

Fig. 3. Tracking results obtained from a sequence captured at a level crossing (first row: results obtained by the PF-based tracker, second row: results of the proposed algorithm) (occlusion occurs at frame 40 and the person starts to run at frame 98).

Table 1. Mean of tracking errors for different methods.

	LC		PETS'06	
	Center	Size	Center	Size
PF	7.81	14.75	4.34	5.7
PF-HOG	6.45	11.22	3.8	5.34
MCMC	6.79	11.98	3.47	5.49
MCMC-HOG	5.06	7.39	3.15	5.09

to increase the robustness of the tracker. The global system has been tested on different real datasets. Experimental results show that our system outperforms the conventional particle filter, even when it uses the HOG-based confidence density.

Several perspectives are envisaged to improve the performance of the system. Firstly, we aim to include an automatic update of the target reference histogram during the tracking. We also aim to extend our tracking framework to multiple object tracking in more realistic scenarios.

5. REFERENCES

[1] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," *ACM Computing Surveys*, vol. 38, no. 4, pp. 1–45, 2006.

[2] M. Isard and A. Blake, "Condensation-conditional density propagation for visual tracking," *Int. J. Comput. Vision*, vol. 29, no. 1, pp. 5–28, 1998.

[3] P. Pérez, C. Hue, J. Vermaak, and M. Gangnet, "Color-based probabilistic tracking," in *ECCV*. 2002, pp. 661–675, Springer.

[4] K. Nummiaro, E. Koller-Meier, and L. Van Gool, "An adaptive color-based particle filter," *Image Vision Comput.*, vol. 21, no. 1, pp. 99–110, 2003.

[5] S.K. Zhou, R. Chellappa, and B. Moghaddam, "Visual tracking and recognition using appearance-adaptive models in particle filters," *IEEE Trans. on Image Process.*, vol. 13, no. 11, pp. 1491–1506, 2004.

[6] H. Wang, D. Suter, K. Schindler, and C. Shen, "Adaptive object tracking based on an effective appearance filter," *PAMI*, vol. 29, no. 9, pp. 1661–1667, 2007.

[7] L. Jin, J. Cheng, and H. Huang, "Human tracking in the complicated background by particle filter using color-histogram and hog," in *ISPCS*, 2010, pp. 1–4.

[8] A. Yao, G. Wang, X. Lin, and X. Chai, "An incremental bhattacharyya dissimilarity measure for particle filtering," *Pattern Recognit.*, vol. 43, no. 4, pp. 1244–1256, 2010.

[9] X. Li, W. Hu, Z. Zhang, X. Zhang, and G. Luo, "Robust visual tracking based on incremental tensor subspace learning," in *ICCV*, 2007.

[10] D.A. Ross, J. Lim, R.S. Lin, and M.H. Yang, "Incremental learning for robust visual tracking," *Int. J. Comput. Vision*, vol. 77, no. 1, pp. 125–141, 2008.

[11] J. Kwon, K.M. Lee, and F.C. Park, "Visual tracking via geometric particle filtering on the affine group with optimal importance functions," in *CVPR*, 2009, pp. 991–998.

[12] W.R. Gilks and C. Berzuini, "Following a moving target–monte carlo inference for dynamic bayesian models," *Journal of the Royal Statistical Society*, vol. 63, no. 1, pp. 127–146, 2001.

[13] Z. Khan, T. Balch, and F. Dellaert, "MCMC-based particle filtering for tracking a variable number of interacting targets," *PAMI*, vol. 27, pp. 1805–1918, 2005.

[14] S.K. Pang, J. Li, and S.J. Godsill, "Models and algorithms for detection and tracking of coordinated groups," in *IEEE Aerospace Conference*. IEEE, 2008, pp. 1–17.

[15] F. Septier, A. Carmi, SK Pang, and SJ Godsill, "Multiple object tracking using evolutionary and hybrid MCMC-based particle algorithms," in *15th IFAC Symposium on System Identification*, 2009.

[16] F. Septier, S.K. Pang, A. Carmi, and S. Godsill, "On MCMC-based particle methods for bayesian filtering: Application to multitarget tracking," in *3rd IEEE International Workshop CAMSAP*, 2009, pp. 360–363.

[17] J. Vermaak, ND Lawrence, and P. Perez, "Variational inference for visual tracking," in *CVPR*, 2003.

[18] Y. Rubner, J. Puzicha, C. Tomasi, and J.M. Buhmann, "Empirical evaluation of dissimilarity measures for color and texture," *CVIU*, vol. 84, no. 1, pp. 25–43, 2001.

[19] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *CVPR*, 2005, pp. 886–893.

[20] M.D. Breitenstein, F. Reichlin, B. Leibe, E. Koller-Meier, and L. Van Gool, "Online multi-person tracking-by-detection from a single, uncalibrated camera," *PAMI*, 2010.

[21] "PETS 2006," http://www.cvg.rdg.ac.uk/PETS2006/data.html.