USING VISUAL SALIENCY FOR OBJECT TRACKING WITH PARTICLE FILTERS

Désiré Sidibé, David Fofi, and Fabrice Mériaudeau

LE2I Laboratory - UMR CNRS 5158, Université de Bourgogne
12 Rue de la Fonderie, 71200, Le Creusot, France
email: dro-desire.sidibe@u-bourgogne.fr

ABSTRACT

This paper presents a robust tracking method based on the integration of visual saliency information into the particle filter framework. While particle filter has been successfully used for tracking non-rigid objects, it shows poor performances in the presence of large illumination variation, occlusions and when the target object and background have similar color distributions. We show that considering saliency information significantly improves the performance of particle filter based tracking. In particular, the proposed method is robust against occlusion and large illumination variation while requiring a reduced number of particles. Experimental results demonstrate the efficiency and effectiveness of our approach.

1. INTRODUCTION

Tracking moving objects is an important task in many computer vision applications including video surveillance, smart rooms, mobile robotics, augmented reality and video compression. Despite many effort, it is still a challenging problem due to the presence of noise, changes of illumination, cluttered background and occlusions that introduce uncertainty in the estimation of the object's state.

The main objective of tracking is to roughly predict and estimate the location of a target object in each frame of a sequence. Many methods have been developed and can be divided into two groups: deterministic methods and stochastic methods [12]. Methods of the former group iteratively search for the local maxima of a similarity measure between a template of the target and the current image. The Kanade-Lucas-Tomasi tracker [7] and the mean-shift tracker [3] are examples of this category of methods. In contrast, methods of the latter group use the state space representation of the moving object to model its underlying dynamics. The tracking problem can then be viewed as a Bayesian inference problem. In the case of a linear dynamic model with Gaussian noise, Kalman filter provides an optimal solution. However, for non-linear and non-Gaussian cases, it is impossible to find analytic solutions. Over the last decade, particles filters, also known as condensation or sequential Monte Carlo methods, have proved to be very efficient for object tracking [4, 5, 9]. Different types of features can be used within the particle filter framework. Color distribution [9, 8] is robust against noise and partial occlusion, but becomes ineffective in the presence of illumination changes, or when the background and the target have similar colors. Edges or contour features [5] are more robust to illumination variations, but are sensitive to clutter and are computationally expensive. For better performances, one can combine color and edge features as in [11, 12].

When looking at a scene, humans tend to focus on regions that are visually salient, i.e. that are more conspicuous in contrast with respect to their neighborhood [6, 10]. Salient regions detection has been used in many applications including image retrieval, image segmentation and object recognition. Most of the existing detection methods are based on a low-level approach and use different features such as color, intensity and orientation. In general, separate feature maps are created for each of the features considered and then combined to obtain the final saliency map. One representative method is the work of Itti *et al.* [6] who employes color, intensity and orientation maps with a histogram entropy thresolding analysis. Recently, Achanta *et al.* [1] introduce a frequency-based method which exploits color and luminance features. Their method is easy to implement, fast and provide full resolution saliency maps.

In this paper, we integrate visual saliency information into the particle filters framework for object tracking. In particular, we show how to combine color and saliency distributions to increase the robustness to large illumination variations and to similar background color. Visual saliency has also been used for tracking by Zhang *et al.* [14], but their method is based on the detection of salient objects using both static and motion features.

The paper is organized as follows. An overview of particle filtering based tracking is given in Section 2. In Section 3, the proposed method is described, explaining the visual saliency detection and its combination with color for tracking. Experimental results and discussion are shown in Section 4 and, finally, concluding remarks are given in Section 5.

2. PARTICLE FILTERING OVERVIEW

This section briefly introduces the particle filter method for tracking. For more details and theoretical proofs, the reader is invited to refer to [4, 2]. A particle filter is a sequential Monte Carlo method, which recursively approximates the posterior distribution using a finite set of weighted samples $\{x_t^i, w_t^i\}_{i=1,\dots,N}$. Each sample x_t^i represents a hypothetical state of the target with a corresponding importance weight w_t^i. Given all observations up to time t, $Z_t = \{z_0, z_1, \dots, z_t\}$, the goal is to estimate the state of the target object x_t, i.e. to find the posterior distribution $p(x_t|Z_t)$. Let's assume that the system is governed by the following state-space representation:

$$\left\{ \begin{array}{r} x_t = f(x_{t-1}) + v_{t-1} \\ z_t = h(x_t) + n_t \end{array} \right. , \tag{1}$$

where f and h are respectively the system transition and the measurement functions, and v_{t-1} and n_t are the system and measurement noises.

The particle filter, like any sequential Bayesian technique, uses a prediction and correction strategy. The prediction stage uses the system transition model to predict the posterior at time t as:

$$p(x_t|Z_{t-1}) = \int p(x_t|x_{t-1})p(x_{t-1}|Z_{t-1})dx_{t-1}. \tag{2}$$

The correction step uses the available observation at time t, z_t, to update the posterior using Bayes' rule:

$$p(x_t|Z_t) = \frac{p(z_t|x_t)p(x_t|Z_{t-1})}{p(z_t|Z_{t-1})}. \tag{3}$$

2.1 Color Distribution Model

Different types of features can be used to measure the observation likelihood of the samples. Among them, color distribution is robust against noise and partial occlusion, and fast to compute [9, 8]. Usually, color distributions are represented by histograms in the RGB or HSV color space. The color distribution $p(\mathbf{x}) = \{p_u(\mathbf{x})\}_{u=1,\dots,m}$ of a region centered at location $\mathbf{x}$ is given by:

$$p_u(\mathbf{x}) = C\sum_{i=1}^{N_p} k(\left\|\frac{\mathbf{x}_i - \mathbf{x}}{h}\right\|^2)\delta[b(\mathbf{x}_i) - u], \tag{4}$$

where C is a normalizer, δ is the Kronecker function, k is a kernel with bandwith h, N_p is the number of pixels in the region and $b(\mathbf{x}_i)$ is a function that assigns one of the m-bins to a given color at location $\mathbf{x}_i$. The kernel k is used to consider spatial information by lowering the contribution of farther pixels.

Several distances can be defined to compute the similarity between color distributions such as the KL-distance or histogram intersection. Here, we adopt the popular Bhattacharyya coefficient as a similarity measure [3]. If we denote $p^* = \{p_u^*(\mathbf{x}_0)\}_{u=1,\dots,m}$ as the reference color model of the object and $p(\mathbf{x}_t)$ as a candidate color model, then the distance between p^* and $p(\mathbf{x}_t)$ is defined by:

$$\rho[p^*, p(\mathbf{x}_t)] = \left[1 - \sum_{u=1}^{m}\sqrt{p_u^*(\mathbf{x}_0)p_u(\mathbf{x}_t)}\right]^{\frac{1}{2}} \tag{5}$$

Each sample x_t^i is assigned an importance weight which corresponds to the likelihood that x_t^i is the true location of the object. In the case of the bootstrap filter [4], the weights are given by the observation likelihood:

$$w_t^i = p(z_t|x_t^i) \propto e^{-\lambda\rho[p^*,p(x_t^i)]^2}, \tag{6}$$

where $\lambda = 20$ in our experiments as in [9].

3. VISUAL SALIENCY BASED TRACKING

In particle filtering based tracking, one has to resolve the contradiction between robustness and tracking speed. In fact, a large number of particles, or samples, leads to more robust results at the price of high computational load and slow tracking speed. Moreover, the features distributions have to be evaluated for each of the samples. Based on this considerations, we use the saliency detection method of Achanta *et al.* [1] in our work. This method is computationally efficient while providing full resolution saliency maps.

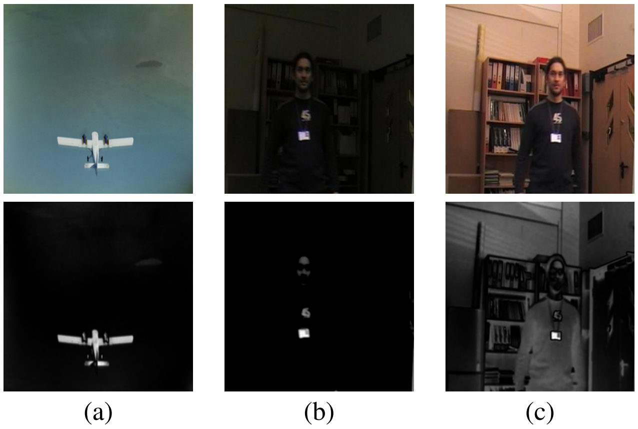

(a) (b) (c)

Figure 1: Saliency detection example. Top row shows original images and bottom row shows corresponding saliency maps.

3.1 Saliency Detection

The saliency detection method fully described in [1] is based on color and luminance features. For each pixel of the image, we compute the degree of saliency with respect to its neighborhood as the Euclidean distance between the pixel vector and the average vector of the input image in the *Lab* color space. Formally, the input image I is first converted to *CIELab* color space I^*. From I^*, one computes the mean image feature $I_\mu^* = [L_\mu, a_\mu, b_\mu]^T$ and a Gaussian blurred image I_σ^* using a 5×5 separable binomial kernel. The saliency at a pixel location (x, y) is then given by:

$$S(x,y) = \left\|I_\mu^* - I_\sigma^*(x,y)\right\|, \tag{7}$$

where $\|\|$ is the L_2 norm.

The method emphasizes the largest salient objects and generates sharper and well-defined boundaries of salient objects.

Some saliency detection results are shown in Figure 1. As it can be seen, regions that stand out relative to their neighbors are detected as been salient parts of the image. However, in some situations, the object of interest might be detected as being less salient. This is depicted in Figure 1-c, where pixels belonging to the face of the person have lower saliency values compared to background pixels. Therefore, saliency information has to be carefully combined with color information to achieve good tracking results. The next subsection explains how we combine these two information.

3.2 Using Saliency for Tracking

Based on the saliency detection method described in Section 3.1, we define a saliency distribution for a region of the image in a similar way to the color distribution, i.e. using equation (4). The similarity between two saliency distributions is measured by the Bhattacharyya distance.

Figure 2 shows examples of similarity measures between color and saliency distributions. As we can notice, in cases where the object of interest and the background have similar colors, color feature is not enough to identify the object. In the given example, the distance between the color distributions of the reference model in Figure 2-a and the candidate model in Figure 2-b is $\rho_{12} = 0.4983$ while the distance between the color distributions of the reference model and the candidate model in Figure 2-c is $\rho_{13} = 0.4933$. It is thus, hard to distinguish the correct location of the object. Using saliency distributions, the distances are respectively, $\rho'_{12} = 0.1963$ and $\rho'_{13} = 0.3620$, showing the distinctiveness of visual saliency.

Despite the distinctiveness of saliency feature, in some situations, the object of interest might be detected as being less salient than the background, as mentioned in Section 3.1. Therefore, in order to improve the robustness of the tracker, we combine both color and saliency features, automatically weighting their respective contribution to the likelihood function.

More precisely, given N samples $\{x_t^i\}_{i=1,\dots,N}$ at time t, let ρ_c^i be the distance between the reference and the i-th candidate color distributions, and let ρ_s^i be the distance between the reference and the i-th candidate saliency distributions. Then each sample x_t^i is assigned an importance weight given by:

$$w_t^i \propto (1-\alpha_t)e^{-\lambda(\rho_c^i)^2} + \alpha_t e^{-\lambda(\rho_s^i)^2}. \tag{8}$$

The weighting parameter α_t is evaluated at every time t using the following formula:

$$\alpha_t = \frac{\overline{\rho_c}}{\overline{\rho_c} + \overline{\rho_s}}, \tag{9}$$

where $\overline{\rho_c}$ is the mean value of $\{\rho_c^i\}_{i=1,\dots,N}$.

By employing a time varying weighting parameter, the tracker can adaptively give more importance to one feature or the other based on the color and saliency distributions of every frame of the sequence. Thus, we can deal with large illumination variations and similar background color.

3.3 Particle Filter Tracking

To implement the particle filter, one has to define the state vector and the dynamic model of the system. We define the state as $\mathbf{x} = [x, y, s_x, s_y]^T$, where (x, y) is the location of the target object, s_x and s_y are the scales in the x and y directions. In the prediction stage of the particle filter, the samples are propagated through a dynamic model. We use a first order auto-regressive (AR) process for simplicity:

$$\mathbf{x}_t = A\mathbf{x}_{t-1} + \mathbf{v}_{t-1}, \tag{10}$$

where $\mathbf{v}_{t-1}$ is a multivariate Gaussian random noise and A defines the deterministic system model. A constant velocity model is usually used for the dynamic model.

In the update stage, the observation likelihood for each sample, i.e., the weights for each sample, are estimated using equation (8). In practice, to avoid degeneracy, i.e. all but one particle having negligible weights after a certain number of recursive steps, a bootstrap resampling is performed. The resampling step is also designed to handle sample impoverishment, i.e. particles that have high weights are statistically selected more often than others [2].

4. EXPERIMENTS

To evaluate the performance of the proposed tracking method, we applied it to different sequences showing confusing background color and large illumination variations. The image size of the sequences is 320 × 240. We use the HSV color space with a 8 × 8 × 8 bins histogram to represent the color distribution, and we employ a 16 bins histogram for saliency distribution. The HSV color space is used instead of RGB because it is less sensitive to lighting conditions. In all experiments, the tracked object is manually initialized in the first frame.

To show the robustness of our method against similar background color, we use the *Plane* sequence of 300 frames and compare the basic color based particle filter with the saliency based one. In both cases, we use 100 particles for tracking. The top row of Figure 3 shows the results of the color histogram based tracker. The tracker totally loses the target after several frames. The bottom row in Figure 3 shows the tracking results using our approach. We can see that the tracker robustly follows the target despite confusing background color.

In particle filter tracking, the robustness and the tracking speed are proportional to the number of samples used. A large number of samples provides more robust results at the price of high computational load and slow tracking speed. Experiments show that the proposed tracker can successfully track the target with a reduced number of particles. For example, our method robustly follows the target in the *Plane* sequence with as few as only 20 particles, which significantly reduces the computational time.

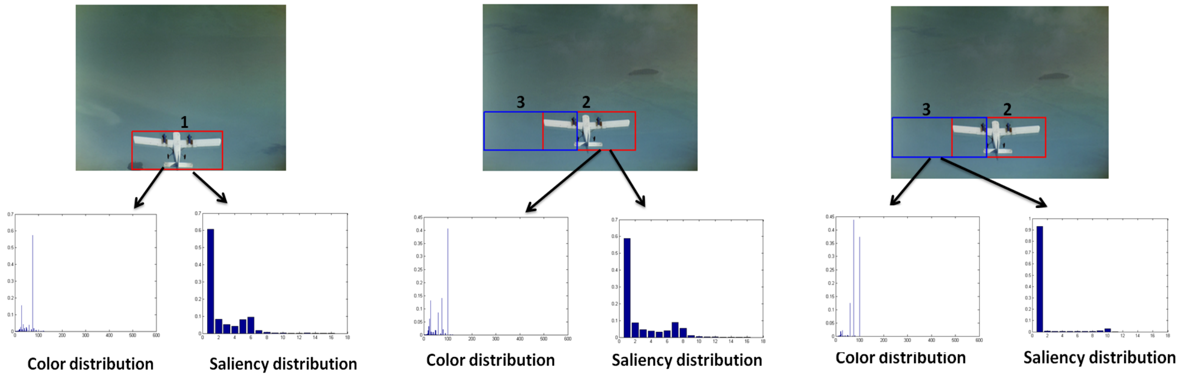
In the second experiment, the *Face* sequence is used to evaluate the performance of the proposed tracker against severe and sudden illumination changes. The results are shown in Figure 4. It is important to point out that because of the poor lighting environment, the background color distribution is similar to that of the face. This makes the color based tracker to lose the target. Furthermore, the color based tracker can hardly adapt to a sudden illumination change. On the contrary, the proposed saliency based tracker performs extremely well in this situation. Note that the tracker is also robust against occlusion since it can recover the target object even if it is fully oclluded as shown by frames 612 and 679 in Figure 4.

For a quantitative evaluation of the tracking methods, we use the spatial overlap metric defined in [13]. Let S_{GT}^i and S_T^i be, respectively, the ground truth and the estimated bounding box of the object in the i-th frame of a sequence. The spatial overlap is defined as:

$$\zeta_i = \frac{\text{Area}(S_{GT}^i \cap S_T^i)}{\text{Area}(S_{GT}^i \cup S_T^i)} \tag{11}$$

The object in frame k is accurately estimated if $\zeta_k > T$, where the threshold T is set to 0.25 in our experiments.

The tracking performances are given in Table 1. For each method, the accuracy is defined as the ratio between the number of frames the object location is accurately estimated and the total number of frames in the sequence. Our saliency based tracker outperforms the color based tracker for all three sequences, providing excellent results for the *Plane* and *Face* sequences. In the case of the *Walk* sequence, the tracking accuracy is limited due to the fact that the walking person (the target) is hardly distinguishable from the background. However, using visual saliency improves the tracking results as shown in Figure 5.



(a) (b) (c)

Figure 2: Similarity measure between color and saliency distributions. (a) Reference color and saliency distributions. (b) First color and saliency candidate distributions. (c) Second color and saliency candidate distributions. Using color feature, the distances between the color distributions are respectively, $\rho_{12} = 0.4983$ and $\rho_{13} = 0.4933$. Using saliency distributions, the distances are $\rho'_{12} = 0.1963$ and $\rho'_{13} = 0.3620$.

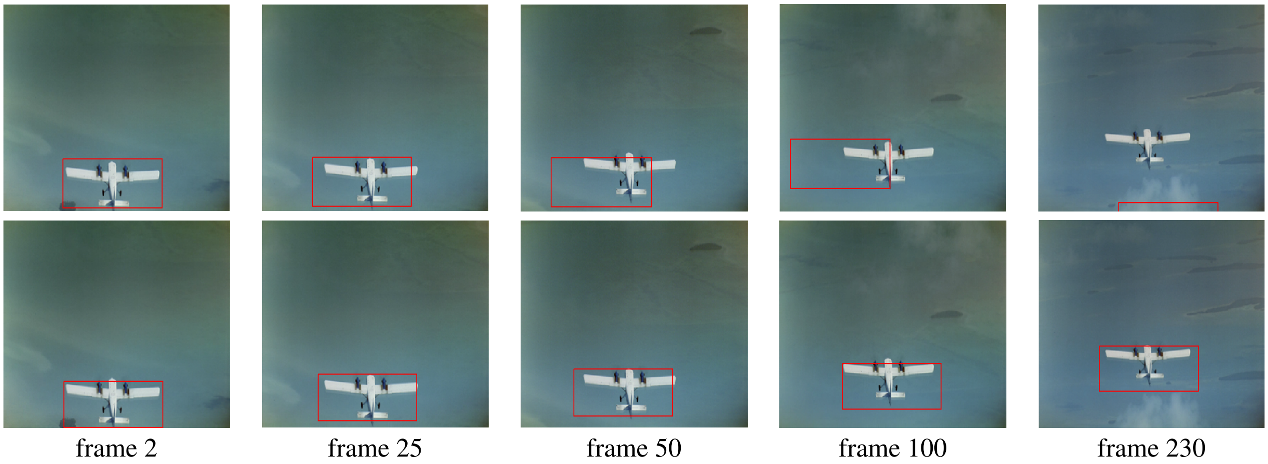


Figure 3: Tracking results using the *Plane* sequence. Top row shows results with the color based tracker and bottom row shows results with the proposed method.

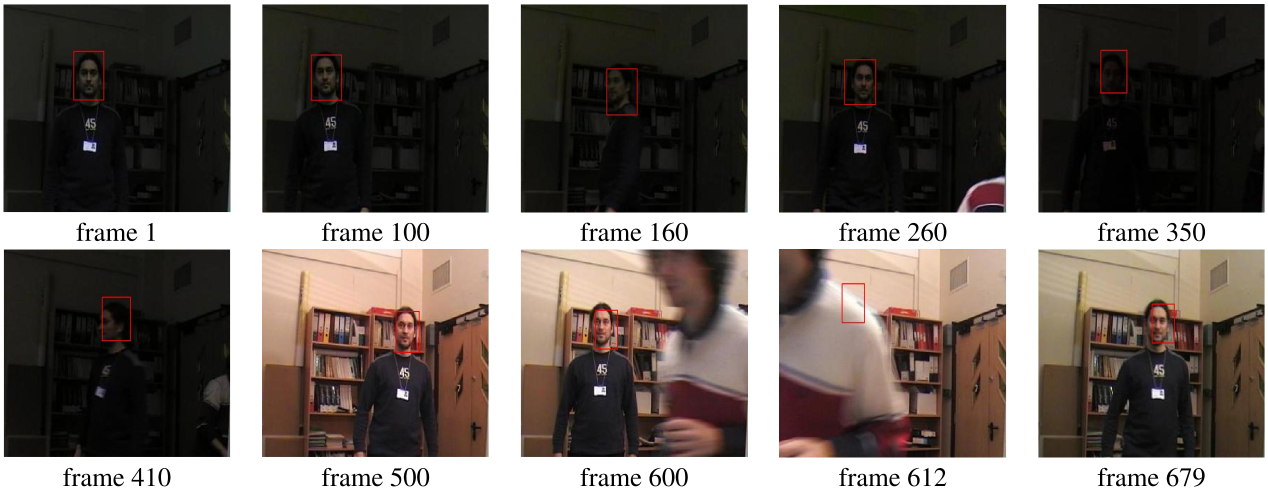
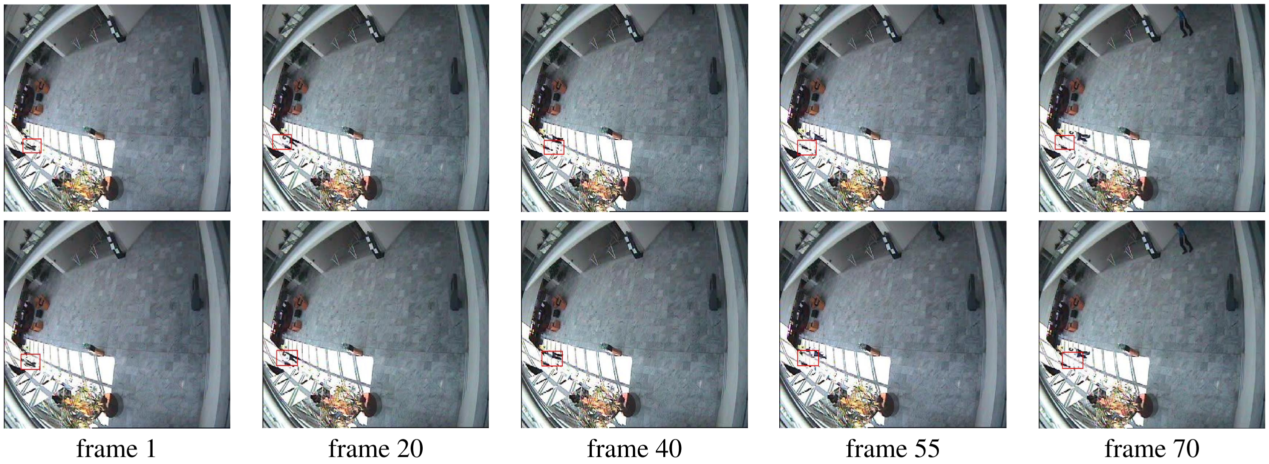


Figure 4: Tracking results in the presence of severe illumination change and occlusion. The target face is consistently and robustly tracked by the proposed method despite poor lighting condition and occlusion.



frame 1 frame 20 frame 40 frame 55 frame 70

Figure 5: Tracking results using the *Walk* sequence. Top row shows results with the color based tracker and bottom row shows results with the proposed method.

Table 1: Tracking Performance Scores

	correct/total		Accuracy (%)	
	C	C-S	C	C-S
Plane sequence	201/300	300/300	67	100
Face sequence	429/680	671/680	63	98
Wall sequence	21/70	48/70	30	68

C = color based tracker
C-S = color and saliency based tracker.

5. CONCLUSIONS

In this paper a robust tracking method is proposed. It is based on the integration of visual saliency information into the particle filter framework. We have shown how to effectively combine color and saliency information in order to make the tracker robust against occlusion, confusing background color and large illumination variation. Experiments with different sequences show that the proposed tracking method outperforms the established color based tracker, while requiring a reduced number of particles. A direction of future work would be an extension for multi-object tracking, incorporating shape and texture features.

REFERENCES

[1] R. Achanta, S. Hemami, F. Estrada, and S. Süsstrunck. Frequency-tuned Salient Region Detection. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, pages 1597–1604, 2009.

[2] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp. A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking. *IEEE Transactions on Signal Processing*, 5(2:174–188, 2002.

[3] D. Comaniciu, V. Ramesh, and P. Meer. Real-Time Tracking of Non-Rigid Objects using Mean Shift. In *Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition*, pages 142–149, 2000.

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

[5] M. Isard and A. Blake. Condensation - Conditional Density Propagation for Visual Tracking. *International Journal of Computer Vision*, 28(1):5–28, 1998.

[6] L. Itti, C. Koch, and E. Niebur. A Model of Saliency Based Visual Attention for Rapid Scene Analysis. *IEEE Transactions on Pattern Analysis and Machine Itelligence*, 20(11):1254–1259, 1998.

[7] B. D. Lucas and T. Kanade. An Iterative Image Registration Technique With an Application to Stereo Vision. In *Proceedings of International Joint Conference on Artificial Intelligence*, pages 674–859, 1981.

[8] K. Nummiaro, E. Koller-Meier, and L. V. Gool. An Adaptive Color-based Particle Filter . *Image and Vision Computing*, 21(1):99–110, 2003.

[9] P. Pérez, C. Hue, J. Vermaak, and J. Gangnet. Color-Based Probabilistic Tracking. In *Proceedings of the European Conference on Computer Vision*, pages 661–675, 2002.

[10] J. Tsotsos, S. Culhane, W. Wai, Y. Lai, N. Davis, and F. Nuflo. Modeling Visual Attention via Selective Tuning. *Artificial Intelligence*, 78(1-2):507–545, 1995.

[11] Y. Wu. Robust Visual Tracking by Integrating Multiple Cues Based on Co-inference Learning. *International Journal of Computer Vision*, 58(1):55–71, 2004.

[12] C. Yang, R. Duraiswami, and L. Davis. Fast Multiple Object Tracking via a Hierarchical Particle Filter. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 53–60, 2002.

[13] F. Yin, D. Makris, and S. Velastin. Performance Evaluation of Object Tracking Algorithms. In *Proceedings of the IEEE International Workshop on Performance Evaluation of Tracking and Surveillance*, 2007.

[14] G. Zhang, Z. Yuan, N. Zheng, X. Sheng, and T. Liu. Visual Saliency Based Object Tracking. In *Proceedings of the Asian Conference on Computer Vision*, 2009.