Probabilistic Detection-based Particle Filter for Multi-target Tracking

Xuan Song, Jinshi Cui, Hongbin Zha and Huijing Zhao
Key Laboratory of Machine Perception (Ministry of Education),
Peking University, China
{songxuan,cjs,zha,zhaohj}cis.pku.edu.cn

Abstract

In this paper, we present a Probabilistic Detection-based Particle Filter (PD-PF) for tracking a variable number of interacting targets. When the objects do not interact with each other, our method performs like the deterministic detection-base methods. When the objects are in close proximity, the interactions and occlusions are modelled by a mixed proposal constructed by probabilistic detections and information from dynamic models. Specially, prior of detection-reliability minimizes the influence of non-detection or false alarm in the tracking. Moreover, we run independent PD-PF for each target, such that particles are sampled in a small state space, thus our method not only obtains a better approximation of posterior than joint particle filter or independent particle filter when interactions occur, but also has an acceptable computational complexity. Different evaluations demonstrate the validity and efficiency of the proposed method.

1 Introduction

Tracking multiple targets which are similar in appearance becomes significantly more challenging when the targets interact. The goal of this work is to obtain a better approximation of posterior than previous methods when interactions and occlusions occur. Moreover, most detection-based tracking algorithms rely on the performance of detection methods, how to minimize the influence of non-detection or false alarm is another focus point in this paper.

Over the last couple of years, a large number of algorithms for multi-target tracking have been presented [1-9]. Some of them [1,2] are obtained by finding the best match according to some similarity function in the neighborhood of the predicted position. However, these methods use appearance based tracking methods, thus tracking multiple targets which are similar in appearance becomes very difficult. On the other hand, some algorithms [3-5] use detection based data association framework. But detection driven tracking schemes greatly rely on the performance of the detection algorithms. Only observations at the high responding detections are considered as potential measurements. This leads to that false alarm and non-detection which in turn significantly influence the performance of the tracker.

BMVC 2008 doi:10.5244/C.22.23

Recently, some tracking algorithms which incorporate the information of dynamic models and detections have been proposed [6,7]. The combination of detections and dynamic models leads to fewer failures than either one on its own, as well as addressing both detection and consistent track formation in the same framework. However, these methods can be seem as deterministic detection-based particle filter, which means one deterministic detection has main contribution for each target in the sampling and filter procedure. Hence, the main problems of deterministic detection-based particle filter are: 1) lose the information of interactions among detections; 2) non-detection or false alarms influence tracking performance greatly.

In this paper, we propose a Probabilistic Detection-based Particle Filter (PD-PF) for multi-target tracking. In our method, we incorporate possible probabilistic detections and information from dynamic model to construct a mixed proposal for particle filter, which quite effectively models interactions and occlusions among targets. Compared to the previous methods, the main advantages of our method are: 1) although our approach runs independent particle filter for each target, the mixed proposal constructed by probabilistic detections and dynamics exactly models the interactions and occlusions among targets. Thus, our method not only obtains a better approximation of posterior for each target, but also can maintain correct tracking when the interactions and occlusions occur; 2) because probabilistic detections include prior model of detection-reliability, influence of false alarm and non-detection can be minimized. Experimental results demonstrate the validity and efficiency of the proposed method.

2 Probabilistic Detection

Compared to the deterministic detection, "probabilistic detection" in our method should contain two aspects of implication. Firstly, some interacting detections have different contributions for each target in the tracking. Obviously, some of these detections are constructive for the tracking, and some of these have negative influences. However, these detections just model the interactions and occlusions among targets by competing in a mixed proposal. Hence, each detection should have a probability which reflects its contribution in the tracking. Secondly, because of the uncertainty of detected method, non-detection and false alarm frequently take place. Thus, each detection should have a prior of detection-reliability which minimizes the influence of uncertain detections. In this section, we give a comprehensive discussion about probabilistic model of detections.

2.1 Possible Interacting Detections

Each target should have one or more possible detections. We use the statistical distance [10] to find possible detections for the target. We assume $\mathbf{d} = (X, Y)$ is the predicted position of target, $\mathbf{d}^* = (X^*, Y^*)$ the position of detection, the distance function should be

$$\frac{(X^* - X)^2}{(G\sigma_X)^2} + \frac{(Y^* - Y)^2}{(G\sigma_Y)^2} \leq 1\,, \tag{1}$$

where G is the gate value, σ_X and σ_Y the covariance.

An example is shown in Fig 1, for the predicted target 1 (yellow rectangle) in frame t+1, there are four possible detections (green point) in the distance function (red ellipse). We suppose that the four detections all have influence on target 1 in the tracking. Although some of them do not associate with this target, they have some influence on the target 1. We can use these interacting detections to model the interactions among targets. On the other hand, if a target cannot find possible detections or a detection associates with no targets, it should be a disappearing target or a new target.

In the next two subsections, we will give details about how to evaluate the contribution of these possible detections in the tracking.

Frame t Frame t+1

Figure 1**:** Example of possible detections. For the target 1 (yellow rectangle) in the frame t, we obtain a predicted position in frame t+1. With the help of distance function (red ellipse), we find four possible detections (green point) for the target 1.

2.2 Prior Model of Detections

Because of the limitation of detected approach, non-detection or false alarms frequently take place. An example of uncertain detections is shown in Fig 2. To minimize the influence of uncertain detections, a probabilistic model of detections should be considered. For the Adaboost detection [11], non-detection or false detection frequently occur in the overlap region of targets. Hence, we can use the interactions among targets to depict this prior of detection-reliability, a prior probability P_{prior} should be

$$P_{prior} = P_{ada} \times P_{Interact}, \tag{2}$$

where P_{ada} is the prior probability of incorrect detections which is depending on the classification of Adaboost. For instance, the detector can continue to give false alarms persistently in some textured area, we use this prior to depict this condition. $P_{Interact}$ is the penalty function to depict the interacting region of targets.

We can use undirected graph model to construct this penalty function. Given an undirected graph (V, E), where the nodes V represent random variables and the edges E specify a neighborhood system. The joint probability over the random variables is then factored as a product of local potentials function f at each node.

Here, if some targets interact with each other, we compute their joint field distribution as

$$P_{Interact} \propto \prod_{i \in L} f_i(x_{C_i}) \quad , \tag{3}$$

where C_i is the subset of nodes associated with the ith factor of cliques, f_i the factored potentials function of cliques, L the index set of cliques. In our application,

Non-detection of one player False alarm

Figure 2: Example of uncertain detections.

we follow [8], $f_i(x_{C_i})$ is means of Gibbs distribution

$$f_i(x_{C_i}) \propto \exp(-\lambda_i g(x_{C_i})) \quad , \tag{4}$$

where λ_i is the factor of specific clique, $g(x_{C_i})$ a penalty function which is depending on the number of pixels overlap between interacting targets.

An example of graph model for a particular configuration is shown in Fig 3. For each target, we use a circle to represent "region of influence", if some targets have the overlapping region, we assume that there are interactions among them, a joint field distribution should be computed. For instance, target 1, target 2 and target 3 interact with each other, we can compute their joint distribution as

$$P_{Interact}^{x_1,x_2,x_3} \propto f_1(x_1,x_2) \times f_1(x_1,x_3) \times f_1(x_2,x_3) \times f_2(x_1,x_2,x_3) \, . \tag{5}$$

Hence, for a detection which is in this region, its prior probability of uncertain detections $P_{prior}^{\mathbf{d}_i}$ should be

$$P_{prior}^{\mathbf{d}_i} = P_{ada} \times P_{Interact}^{x_1,x_2,x_3}(\mathbf{d}_i) \, , \tag{6}$$

where $\mathbf{d}_i$ is the location of this detection.

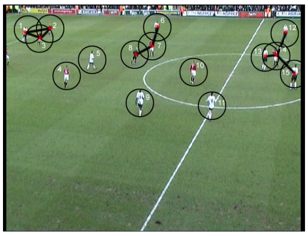


Figure 3: Example of graph model for a particular configuration. In this figure, there are three interacting regions, which have different probability distribution. Detections in these regions have higher probability to be incorrect detections.

2.3 Likelihood Function

The likelihood function should reflect the positive contribution of detections in the tracking. Therefore, it should be the similarity between targets and detections. For instance, detections which are associated to the targets should be more helpful in the tracking. Let $\mathbf{x}_{t,k}$ denote the state of specific target k in frame t, and $\mathbf{y}_t$ the observations. A likelihood function should be

$$P(\mathbf{y}_t \mid \mathbf{x}_{k,t}) \propto P_{motion}(\mathbf{y}_t \mid \mathbf{x}_{k,t}) \times P_{color}(\mathbf{y}_t \mid \mathbf{x}_{k,t}) \, , \tag{7}$$

where $P_{motion}(\mathbf{y}_t \mid \mathbf{x}_{k,t})$ is the similarity of motion between targets and detections that is

$$P_{motion}(\mathbf{y}_t \mid \mathbf{x}_{k,t}) = \frac{1}{\sqrt{2\pi / \boldsymbol{\sigma}}} \exp\left(-\alpha \frac{(\mathbf{d}^*_{k,t} - \mathbf{d}_{k,t})^2}{\boldsymbol{\sigma}^2} \right), \tag{8}$$

where $\mathbf{d}^*_{k,t}$ is the position of detection, $\mathbf{d}_{k,t}$ the predicted position of target, $\boldsymbol{\sigma}$ the covariance, and α the weights ratio which adjusts the weight of $P_{motion}(\mathbf{y}_t \mid \mathbf{x}_{k,t})$ in the likelihood.

On the other hand, $P_{color}(\mathbf{y}_t \mid \mathbf{x}_{k,t})$ is the similarity of color between targets and detections. In our approach, we adopt a multi-color model [1] based on Hue-Saturation-Value color histogram. The target or detection is represented by an N-bin color histogram extracted from the region $R(\mathbf{d}_{k,t})$ centered at location $\mathbf{d}_{k,t}$ of camera image. It is denoted as $Q(\mathbf{d}_{k,t}) = \{q(n; \mathbf{d}_{k,t})\}_{n=1,\dots,N}$, where

$$q(n; \mathbf{d}_{k,t}) = K \sum\nolimits_{\mathbf{k} \in R(\mathbf{d}_{k,t})} \delta(b(\mathbf{k}) - n), \tag{9}$$

where δ is the Kronecker delta function, K a normalization constant ensuring $\sum_{n=1}^{N} q(n; \mathbf{d}_{k,t}) = 1$, $b(\mathbf{k}) \in \{1, \dots, N\}$ is the bin index associated with color vector at pixel location $\mathbf{k}$.

The color likelihood must favor candidate color histograms $Q(\mathbf{d}_{k,t})$ close to the reference histogram $Q^*(\mathbf{d}^*_{k,t})$. A distance D on the HSV color distributions which is derived from the Bhattacharyya similarity coefficient [1] is chosen, and defined as

$$D(q^*, q(\mathbf{d}_{k,t})) = \left[1 - \sum_{n=1}^{N} \sqrt{q^*(n; \mathbf{d}^*_{k,t}) q(n; \mathbf{d}_{k,t})} \right]^{\frac{1}{2}}, \tag{10}$$

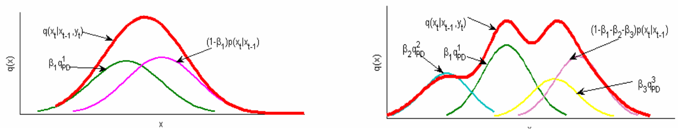
Therefore, the color likelihood is then evaluated as

$$p_{color}(\mathbf{y}_t \mid \mathbf{x}_{k,t}) \propto e^{-(1-\alpha) D^2 (\mathbf{d}_{k,t}, \mathbf{d}^*_{k,t})}, \tag{11}$$

2.4 Probabilistic Detections

Now, we can evaluate the contribution of these detections for each target in the tracking. We can use association probability to depict the contribution of detection. In order to consider all the interacting situations between targets in a joint space, we use joint association probability which is motivated by Joint Probabilistic Data Asscciation (JPDA) [4,10,12]. We define an association event $\boldsymbol{\theta}$ expressed as a vector with dimension n_d. Here n_d is the number of detections. Each $\boldsymbol{\theta}$ uniquely determines how each detection is assigned to a specific target. The vector can be drawn from a set of numbers as $\{0, 1, 2, \dots\}$ and $\boldsymbol{\theta}(j) = k$ means detection j is from target k. Therefore, for each event $\boldsymbol{\theta}$, the joint likelihood is

$$L(\boldsymbol{\theta}) = \prod_{j=1}^{n_d} P(\mathbf{y}_t^j \mid \mathbf{x}_{k,t}^{\boldsymbol{\theta}(j)}), \tag{12}$$



(a) (b)

Figure 4: Mixed proposal distribution. (a)When target did not interact with each other, the proposal had one detection which was similar to deterministic detection-based method. (b)When target interacted with each other, the proposal had more than one detections. These detections and dynamics competed by (13). Particles were sampled in multiple peak proposal distribution.

Hence the joint association probability β_{kj} is

$$\beta_{kj} = \kappa \sum_{\boldsymbol{\theta} \in \Theta_{kj}} L(\boldsymbol{\theta}) \times P_{prior}^{\mathbf{d}j} , \tag{13}$$

where Θ_{kj} is the set of joint association events that include all the case detection j being from target k. κ is a normalization factor ensuring that β_{kj} sums up to one over all $\boldsymbol{\theta}$. Therefore, we call $\beta_{kj} q_{PD}^{j}(\mathbf{x}_t \mid \mathbf{x}_{t-1}, \mathbf{y}_t)$ probabilistic detection, where $q_{PD}^{j}(\mathbf{x}_t \mid \mathbf{x}_{t-1}, \mathbf{y}_t)$ is the information of detections which have influence on the target k. In section 3, we will embed probabilistic detections into the particle filter framework.

3 Probabilistic Detection-based Particle Filter

Particle filter has been a successful numerical approximation technique for Bayesian sequential estimation. Starting with a weighted set of samples $\left\{ w_t^n, \mathbf{x}_t^n \right\}_{n=1}^{N}$ approximately distributed according to $p(\mathbf{x}_{t-1} \mid \mathbf{y}_{t-1})$, new samples are generated from a suitably designed proposal distribution, which may depend on the old state and new measurements. To maintain a consistent sample the new importance weights are set to

$$w_t^{(n)} \propto w_{t-1}^{(n)} \frac{p(\mathbf{y}_t \mid \mathbf{x}_t^{(n)}) p(\mathbf{x}_t^{(n)} \mid \mathbf{x}_{t-1}^{(n)})}{q(\mathbf{x}_t^{(n)} \mid \mathbf{x}_{t-1}^{n}, \mathbf{y}_t)} \quad , \quad \sum_{n=1}^{N} w_t^{(n)} = 1 . \tag{14}$$

From time to time it is necessary to resample the particles to avoid degeneracy of the importance weights. The resample procedure essentially multiplies particles with high importance weights, and discards those with low importance weights.

One of the crucial design issues in particle filter is the choice of the proposal distribution $q(\mathbf{x}_t \mid \mathbf{x}_{t-1}, \mathbf{y}_t)$. In our method, we run an independent particle filter for each target and incorporate the probabilistic detections and dynamic model to construct proposal distribution

$$q_k(\mathbf{x}_t \mid \mathbf{x}_{t-1}, \mathbf{y}_t) = \sum_{j=1}^{n_o} \beta_{kj} q_{PD}^{j}(\mathbf{x}_t \mid \mathbf{x}_{t-1}, \mathbf{y}_t) + (1 - \sum_{j=1}^{n_o} \beta_{kj}) \times p_k(\mathbf{x}_t \mid \mathbf{x}_{t-1}) . \tag{15}$$

When a target does not interact with others, this mixed proposal has one detection or a high weighted detection, which is like to the deterministic detection-based method [7]

(See Fig 4-a). When interactions occur, the mixed proposal has more than one detections, these detections competed by equation (13). Therefore, this mixed proposal quite effectively models interactions and occlusions among targets (See Fig 4-b). For instance, when two similar appearance targets interacted on each other, the proposal for each target became a multiple peak distribution. Sampling particles in these proposals is like a blending step. On the other hand, although the two targets may share the same detections, the proposals of them are different to each other. There are two reasons: 1) the weights β_{kj} are computing in a joint space, for each target, the weights of detections are different. 2) The two targets have different information of dynamics. Therefore, for each target, particles will have high weights in its real state, and the posterior is approximate to its optimization. When two targets merge, we do not lose any target because a specific detection has high weight in respective proposal of each target.

Moreover, because each particle filter samples in a small space, we can obtain better approximation and significantly reduce computational cost than joint particle filter. In addition, we computed β_{kj} in each recursion, thus our method can automatically adapt to the different complex tracking situations.

4 Experiments and Results

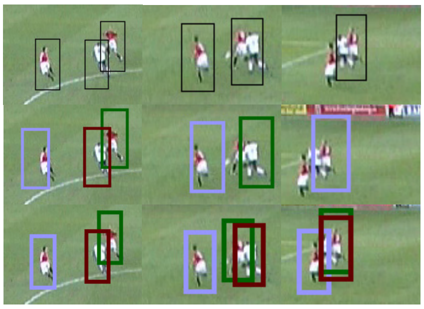
We evaluated our approach by tracking through sports video [13] and visual surveillance video [14] respectively. The test sequences consisted 2400 frames and 2106 frames recorded at a resolution of 640×480 and 320×240 pixels respectively. There were lots of complex interactions and occlusions in the selected sequences, and the appearing and disappearing of targets frequently took place.

Target states were define by $\mathbf{x}_t^i = (\mathbf{d}_t, \mathbf{d}_{t-1}, s_t, s_{t-1})$, where $\mathbf{d} = (X, Y)$ was the centre of rectangular box in the image coordinate system, s the scale factor. A constant velocity dynamic model was used, which could be best described by a second order autoregressive equation $\mathbf{x}_{k,t} = \mathbf{A}\mathbf{x}_{k,t-1} + \mathbf{B}\mathbf{x}_{k,t-2} + \mathbf{C}N(0, \mathbf{\Sigma})$, matrices $\mathbf{A}$, $\mathbf{B}$, $\mathbf{C}$ and $\mathbf{\Sigma}$ could be learned from a set of representative sequences where correct tracks had been obtained in our experiment. $N(0, \mathbf{\Sigma})$ was a Gaussian noise with zero mean and standard deviation of 1. λ_i in equation (4) was adjusted by 1000 detections in 300 frames. α in equation (11) was set to 0.6. In this section, we present our tracking results and make some comparisons with other methods.

4.1 Tracking Performance

In this subsection, we present our tracking results and make some comparisons with deterministic detection-based method [7]. The detected results of sports video were obtained by Adaboost detection [11]. For the surveillance video, we used the detections of [15] with the help of thermal images to get tracking results, and did not embed the prior model of detections.

Fig 5 shows the comparison between our method and [7] in the condition of interactions and occlusions. The first row is the detected results from 108 to 155, the second row is results of [7] and the third row is ours. In the frame 108, there were



Frame 108 Frame 126 Frame155

Figure 5: Comparison in the condition of interactions and occlusions. The first row is the detected results, the second row is results of deterministic detection-based method, and last row is our results. See text for more details.

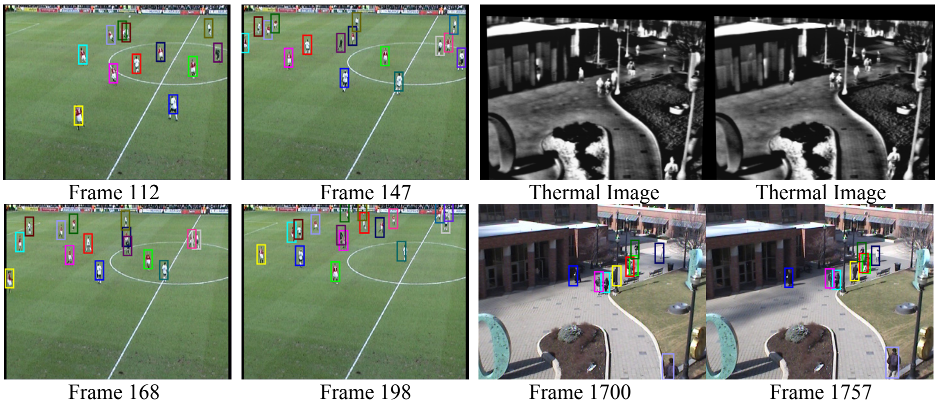


Figure 6: More tracking results of sports video and surveillance video. Please see the attached video for more details.

interactions among three players. After 18 frames, one player was lost and one player was assigned a wrong identification using [7]. 47 frames later, two players were lost and there were an incorrect tracking. Compared to the [7], our method maintained correct tracking in these challenging situations. In the frame 126, although we only got one detection for two players, the two mixed proposals of players shared this detection, high weight particles were sampling in this position. Because this detection had different weights in respective proposal and the different dynamic information of each target, we did not lose any target and maintain correct tracking. In the frame 155, although the proposal of the left red player shared one detection with other players, this detection had lower weight (computing in a joint space) than its dynamic information, it was not pulled by this detection.

Fig 6 shows the tracking results of sports video and surveillance video using our method.

4.2 Quantitative Evaluation

We made a statistical survey of 300 continued frames to evaluate tracking performance in the condition of interactions and made a comparison among four methods: PD-PF,

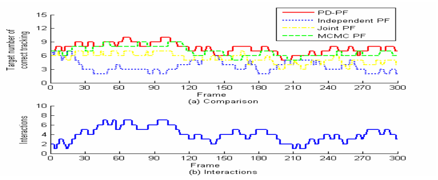


Figure 7: Tracking performance comparison in the condition of interactions. (a) The quantitative comparison among four methods. Tick marks (vertical line) show the number of correct tracking. (b) The number of interactions in these frames.

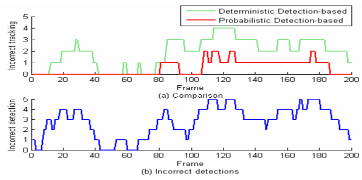


Figure 8: Influence of uncertain detections between two methods. (a) Comparison between two methods. Tick marks (vertical line) show the number of incorrect tracking. (b) The number of incorrect detections in 200 frames.

joint particle filter, independent particle filter and MCMC-particle filter. The person number of these selected frames is variable, and the maximum number in these frames is 12. We define that if two rectangles of persons have the overlapping region, one interaction would be counted. Fig 7 shows the quantitative evaluation of four methods. We can see that with the increasing number of interactions, independent particle filter cannot maintain correcting tracking, and our method have a better performance than joint particle filter and MCMC-particle filter.

4.3 Influence of Uncertain Detections

We made a statistical survey of 200 continued frames to evaluate the influence of uncertain detections between our method and deterministic detection-based method [7]. The evaluation was shown in Fig 8. Incorrect tracking includes false location, loss of target and false data association, and incorrect detections includes false detection, non-detection and false alarm. We can see that the uncertain detections have little influence on our method than the deterministic detection-based because of the prior of detection-reliability.

5 Conclusion

In this paper, a probabilistic detection-based particle filter has been presented. Experiments and evaluations show that our method not only obtains a better approximation of posterior than previous method in the condition of interactions and occlusions, but also minimize the influence of uncertain detections. However, this work also has some limitations: (1) With the increasing number of the interactions, the complexity of our method grows exponentially. (2) Some important parameters should be adjusted manually.

In the future, this work could be extended as the following. 1) In order to decrease computing complexity, a Markov chain Monte Carlo (MCMC) strategy for computing joint association probability should be considered. 2) A more suitable learning method to learn λ_i in equation (4) should be designed.

Acknowledgement: This work was supported in part by the NKBRPC (No.2006CB303100), NSFC Grant (No.60333010), NSFC Grant (No.60605001), the NHTRDP 863 Grant (No.2006AA01Z302) and (No.2007AA11Z225). We specially

thank Kenji Okuma and Yizheng Cai for providing their code on the web. We also thank Vinay Sharma and James W. Davis for providing us their detected results.

References

[1] P. Pμerez , C. Hue, J. Vermaak. Color-Based Probabilistic Tracking. *Proc. ECCV* pp. 661-675, 2002.

[2] J. Vermaak, A. Doucet, P. Pμerez. Maintaining multi-modality through mixture tracking. *Proc. IEEE ICCV*, pp.1110–1116, 2003.

[3] G. Gennari, G.D Hager. Probabilistic data association methods in visual tracking of groups. *Proc. IEEE CVPR*, pp.876-881, 2004.

[4] C. Rasmussen, G.D. Hager. Probabilistic data association methods for tracking complex visual objects. *IEEE Trans. PAMI*, pp.560-576, 2001.

[5] Z. Khan, T. Balch, F. Dellaert. Multi-target tracking with split and merged measurements. *Proc. IEEE CVPR*, pp.605-610, 2005.

[6] K. Okuma, A. Taleghani, N.D. Freitas, J.J. Little, D.G. Lowe. A boosted particle filter: Multitarget detection and tracking. *Proc. ECCV*, pp.28-39, 2004.

[7] Y. Cai, N.D. Freitas, J.J. Little. Robust visual tracking for multiple targets. *Proc. ECCV*, pp.125-135, 2006.

[8] Z. Khan, T. Balch, F. Dellaert. MCMC-based particle filtering for tracking a variable number of interacting targets. *IEEE Trans. PAMI*, pp.1805-1819, 2005.

[9] T. Zhao, R. Nevatia. Tracking multiple humans in complex situations. *IEEE Trans. PAMI*, pp.1208-1221, 2004.

[10] Y. Bar-Shalom, T.E. Fortmann. Tracking and Data Association, *New York: Academic Press*, 1988.

[11] P. Viola, M. Jones. Robust Real-Time Face Detection. *International Journal on Computer Vision*, pp.137-154, 2004.

[12] S. Cong, L. Hong, D. Wicker. Markov-chain Monte-Carlo approach for association probability evaluation. *IEE Proc. Control Theory Application*, pp.185-193, 2004

[13] "SCEPTRE Dataset" http://sceptre.king.ac.uk/sceptre/default.html

[14] "OTCBVS Benchmark Dataset 03: OSU Color Database"http://www.cse.ohio-state.edu/otcbvs-bench/

[15] J. Davis, V. Sharma. Fusion-Based Background-Subtraction using Contour Saliency. *Proc. IEEE CVPR*, pp.20-26, 2005.