

VISUAL TRACKING OF MULTIPLE INTERACTING OBJECTS THROUGH RAO-BLACKWELLIZED DATA ASSOCIATION PARTICLE FILTERING

Carlos R. del Blanco, Fernando Jaureguizar and Narciso García

Grupo de Tratamiento de Imágenes, Universidad Politécnica de Madrid, Madrid, 28040, Spain
 E-mails: {cda, fjn, narciso}@gti.ssr.upm.es

ABSTRACT

A multiple object visual tracking framework is presented, which is able to manage complex object interactions, missing detections and clutter. The main contribution is the ability to deal with complex situations in which the interacting objects can change their dynamics while they are occluded. This is achieved by explicitly estimating putative locations of the occluded objects. The tracking is modeled by a Rao-Blackwellized Data Association Particle Filter (RBDAPF), which has a tractable substructure that allows to analytically compute the object positions, while the object-measurement associations are approximated by Particle Filtering. Besides improving the accuracy, this filter decomposition reduces the computational cost, since the complexity with the number of objects becomes linear instead of exponential. The Particle Filter efficiently manages the measurements from visible and occluded objects, the clutter, and missing measurements to estimate the correct data associations that lead to a robust tracking. Experimental results on surveillance videos show that the proposed RBDAPF framework is able to track multiple interacting objects in complex situations.

Index Terms— Multi-object tracking, interactions, occlusions, Rao-Blackwellization, data association.

1. INTRODUCTION

Multiple object tracking is a fundamental task in many computer vision based systems such as surveillance, autonomous vehicle navigation, robotics, and behavior analysis. All these systems extract high level information from video sequences that is mainly based on the analysis of lower level data, such as trajectories that are obtained through multi-object trackers. While there exist reliable algorithms for the tracking of multiple non-interacting objects in constrained scenarios, the tracking of multiple interacting objects in uncontrolled scenarios is still a challenge.

The multiple object tracking task can be broken down into two subtasks: the detection of objects of interest, and the association between objects and detections along the time. The main challenges in the detection task arise from the fact that there can be missing detections due to occlusions or strong changes in the object appearance. There can be also additional detections generated by background structures, called clutter. In addition, the object detections are noisy, affecting the estimation of the object locations. Regarding the association task, the interaction between objects (especially when some of them experiment a partial or total occlusion) can yield erroneous associations that cause the interchange of object identities,

This work has been partially supported by the Ministerio de Ciencia e Innovación of the Spanish Government under project TEC2007-67764 (SmartVision).

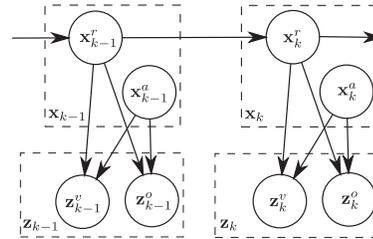


Fig. 1. Graph model representing the DBN.

and thus the corruption of their trajectories. On the other hand, the computational cost of the tracking inevitably grows exponentially with the number of objects.

Recursive Bayesian filtering is a popular framework that allows to probabilistically model the relationships that take place in the multi-object tracking. One solution is to use a Particle Filter to perform the tracking of all the objects, but without explicitly modeling the interactions [1]. However, its performance becomes lower as the number of objects increase, due to the sampling problems in large dimensional state spaces. In [2] a solution is presented that uses a mixture of Particle Filters to model the tracking of multiple people. In this case, each Particle Filter independently performs the tracking of one person. The both previous approaches do not account for interactions among objects, which can cause the exchanging of the object identities. An exclusion principle is presented in [3], which avoids that the same measure can be associated to two or more different objects. In addition, a Markov Chain Monte Carlo scheme is used to improve the sampling performance of the proposed Particle Filtering. A wider range of interactions are addressed in [4], which is able to deal with clutter and missing measurements. This is achieved by means of a Particle Filter framework that uses a Gibbs sampler to obtain the object-measurement associations. A similar strategy is followed in [5], but in this case a Joint Probabilistic Data Association filter is used. In order to improve the accuracy of the tracking and reduce the computational cost, the principle of Rao-Blackwellization has been used in the context of radar applications [6].

In this paper, a novel multiple object visual tracking framework is proposed, which is able to deal with complex object interactions, and has a computational cost that is linear with the number of objects. The main contribution is the explicit estimation of possible occluded objects, that allows to tackle interactions in which the objects change their dynamics while they are occluded. This event can not be managed for the aforementioned approaches, since they rely on the preservation of the object dynamics during the occlusion. The second major contribution is the adaptation of the Rao-

Blackwellized Data Association Particle Filter (RBDAPF) framework, conceived for radar applications [6], to real world visual tracking scenarios. The RBDAPF filter has a special substructure that allows to analytically compute the object positions, while the data associations between objects and measurements are approximated by a Particle Filter. This data association Particle Filter efficiently manages: the measurements from visible objects, the measurements from occluded objects, the measurements from clutter, and the missing measurements to generate the correct object trajectories. Experimental results on video surveillance sequences show that the proposed filtering framework is able to reliably track multiple interacting objects in complex real situations.

2. DYNAMIC BAYESIAN NETWORK FOR VISUAL TRACKING

The visual tracking task of multiple objects is modeled by a Dynamic Bayesian Network (DBN), which is used to recursively compute the joint probability of the object locations and the data association along the time, given a sequence of noisy measurements. This joint probability contains all the required information to yield an optimum estimate about the object trajectories in each time step.

Figure 1 shows a graphical model of the proposed DBN, where the nodes are random variables used to model the tracking, and the edges are probabilistic relationships among the random variables. All the information relative to the tracked objects in the time step k is grouped in a state vector \mathbf{x}_k , which is formed by the following hidden random variables

$$\mathbf{x}_k = \{\mathbf{x}_k^p, \mathbf{x}_k^v, \mathbf{x}_k^a\}, \quad (1)$$

where \mathbf{x}_k^p stores the 2D object positions over the image plane, \mathbf{x}_k^v contains the object velocities, and $\mathbf{x}_k^a = \{\mathbf{x}_k^{a(j)}\}_{j=1}^{N_m}$ encodes the data associations between the tracked objects and the N_m available measurements in the time step k . The j^{th} data association component relates the j^{th} measurement with an object $\mathbf{x}_k^{a(j)} = id_o$ or clutter $\mathbf{x}_k^{a(j)} = id_c$, where id_o is a unique identifier associated to each object, and id_c is an identifier for the clutter.

The noisy measurements are represented by the observed random variable $\mathbf{z}_k = \{\mathbf{z}_k^v, \mathbf{z}_k^o\}$, where \mathbf{z}_k^v contains the measurements of visible objects, and \mathbf{z}_k^o the ones of possible occluded objects. Each measurement $\mathbf{z}_k^{(j)} \in \mathbf{z}_k$, independently if its type, is a vector that contains the position of a putative object, $\mathbf{z}_k^{p(j)}$, a confidence value, $\mathbf{z}_k^{q(j)}$, that indicates the quality of the measurement based on the object appearance, and an identification value, $\mathbf{z}_k^{d(j)}$, that indicates what object detector has yielded the measurement. There are as many visible object detectors as tracked objects, each one focuses on detecting one specific object according to its visual appearance (in Section 5 the process of measurement acquisition is described). Also, there is another detector for occluded objects.

From a Bayesian perspective, the aim is to compute the posterior probability of the state vector $p(\mathbf{x}_k|\mathbf{z}_{1:k})$ using the prior information about the objects, and the sequence of available measurements until the current time step $\mathbf{z}_{1:k} = \{\mathbf{z}_1, \dots, \mathbf{z}_k\}$. Assuming that \mathbf{x}_k is a Markov process and the measurements $\mathbf{z}_{1:k}$ are conditionally independent given the state vector \mathbf{x}_k , the posterior probability can be recursively expressed by [7]

$$p(\mathbf{x}_k|\mathbf{z}_{1:k}) = \frac{p(\mathbf{z}_k|\mathbf{x}_k)p(\mathbf{x}_k|\mathbf{z}_{1:k-1})}{p(\mathbf{z}_k|\mathbf{z}_{1:k-1})} \quad (2)$$

$$p(\mathbf{x}_k|\mathbf{z}_{1:k-1}) = \int p(\mathbf{x}_k|\mathbf{x}_{k-1})p(\mathbf{x}_{k-1}|\mathbf{z}_{1:k-1})d\mathbf{x}_{k-1}. \quad (3)$$

Once the posterior probability has been computed, it is used to obtain an accurate estimation of the state vector $\tilde{\mathbf{x}}_k$ by means of the Minimum Mean Squared Error (MMSE) estimator. The estimated $\tilde{\mathbf{x}}_k$ contains the desired tracking information at the current time step.

However, the above posterior probability can not be analytically solved for the ongoing application due to the non-linearities and non-Gaussian processes involved in the multiple object tracking [7]. Therefore, it is necessary to resort to approximate inference techniques, such as Particle Filtering, to obtain a suboptimal solution.

3. RAO-BLACKWELLIZED DATA ASSOCIATION PARTICLE FILTER

Equation 2 assumes that there is no structure within the hidden variables that form the state vector. A more tractable expression can be obtained if the DBN has a tractable substructure, in such a way that some hidden variables of the state vector can be analytically marginalized out (i.e., they can be computed exactly), conditioned on the rest of variables, which are estimated using Particle Filtering. This technique is called Rao-Blackwellization, and it can be proven that the resulting expression for the posterior probability is more accurate [8]. Intuitively, it is expected that the variance be smaller because, on the one hand some variables are computed exactly, and on the other hand the variables that are estimated have a smaller dimension than the one of the whole state vector.

The proposed DBN has a tractable substructure that allows to marginalize out $\mathbf{x}_k^r = \{\mathbf{x}_k^p, \mathbf{x}_k^v\}$ conditioned on \mathbf{x}_k^a . This leads to the posterior probability expression

$$p(\mathbf{x}_k|\mathbf{z}_{1:k}) = p(\mathbf{x}_k^r, \mathbf{x}_k^a|\mathbf{z}_{1:k}) = p(\mathbf{x}_k^r|\mathbf{z}_{1:k}, \mathbf{x}_k^a)p(\mathbf{x}_k^a|\mathbf{z}_{1:k}), \quad (4)$$

where $p(\mathbf{x}_k^r|\mathbf{z}_{1:k}, \mathbf{x}_k^a)$ is assumed to be conditionally linear Gaussian, and therefore with an analytical expression. This assumption is based on twofold fact. First, the object dynamics is modeled as a linear system of constant velocity with additive Gaussian noise. And second, if it is known what measurement is related to each object, i.e the information given by \mathbf{x}_k^a , the linear likelihood model has only to deal with Gaussian perturbations due to the noise. In addition, the dynamics of each object $\mathbf{x}_k^{r(n)}$ becomes independent, simplifying the complexity of the tracking with regard to the number of objects from an exponential relationship to a linear one [6].

The analytical expression for computing each $p(\mathbf{x}_k^{r(n)}|\mathbf{z}_{1:k}, \mathbf{x}_k^a)$ is known as the Kalman Filter [7].

The other probability term of Equation 4, the data association posterior probability $p(\mathbf{x}_k^a|\mathbf{z}_{1:k})$, can be expressed as

$$p(\mathbf{x}_k^a|\mathbf{z}_{1:k}) = \frac{p(\mathbf{z}_k|\mathbf{z}_{1:k-1}, \mathbf{x}_k^a)p(\mathbf{x}_k^a|\mathbf{z}_{1:k-1})}{p(\mathbf{z}_k|\mathbf{z}_{1:k-1})}. \quad (5)$$

As the associations between objects and measurements are independent between consecutive time steps (due to the measurements are unordered), $p(\mathbf{x}_k^a|\mathbf{z}_{1:k})$ can be simplified as

$$p(\mathbf{x}_k^a|\mathbf{z}_{1:k}) = \frac{p(\mathbf{z}_k|\mathbf{z}_{1:k-1}, \mathbf{x}_k^a)p(\mathbf{x}_k^a)}{p(\mathbf{z}_k|\mathbf{z}_{1:k-1})}, \quad (6)$$

where $p(\mathbf{z}_k|\mathbf{z}_{1:k-1}, \mathbf{x}_k^a)$ is the data association likelihood, $p(\mathbf{x}_k^a)$ is the data association prior, and $p(\mathbf{z}_k|\mathbf{z}_{1:k-1})$ is simply a normalization constant.

The data association prior $p(\mathbf{x}_k^a)$ determines the possible associations between objects and measurements, for which end the following restrictions are imposed: first, each object can be associated only with one or none of the measurements. Second, each measurement

can be associated only to one object, although several measurements can be associated to the clutter. The mathematical expression of $p(\mathbf{x}_k^a)$ that fulfills the previous restrictions can be derived using a Markov Chain Monte Carlo technique as the one presented in [6].

The data association likelihood $p(\mathbf{z}_k | \mathbf{z}_{1:k-1}, \mathbf{x}_k^a)$ uses the measurements to evaluate each possible data association \mathbf{x}_k^a . It can be expressed as

$$p(\mathbf{z}_k | \mathbf{z}_{1:k-1}, \mathbf{x}_k^a) = \prod_{j=1}^{N_m} p(\mathbf{z}_k^{(j)} | \mathbf{z}_{1:k-1}, \mathbf{x}_k^{a(j)}), \quad (7)$$

where it has been considered that the measurements $\{\mathbf{z}_k^{(j)} = (\mathbf{z}_k^{p(j)}, \mathbf{z}_k^{q(j)}, \mathbf{z}_k^{d(j)})\}_{j=1}^{N_m}$ (see Section 2 for a description of each measurement component) are independent, and that each measurement only depends on its corresponding data association component. Each probability factor is computed as

$$p(\mathbf{z}_k^{(j)} | \mathbf{z}_{1:k-1}, \mathbf{x}_k^{a(j)}) = p(\mathbf{z}_k^{p(j)} | \mathbf{z}_{1:k-1}, \mathbf{x}_k^{a(j)}) p(\mathbf{z}_k^{q(j)} | \mathbf{z}_{1:k-1}, \mathbf{x}_k^{a(j)}), \quad (8)$$

where $p(\mathbf{z}_k^{p(j)} | \mathbf{z}_{1:k-1}, \mathbf{x}_k^{a(j)})$ is the data association likelihood based on the position information, given by

$$p(\mathbf{z}_k^{p(j)} | \mathbf{z}_{1:k-1}, \mathbf{x}_k^{a(j)}) = \begin{cases} c_d, & \text{if } \mathbf{x}_k^{a(j)} = id_c \\ \mathcal{N}(\mathbf{z}_k^{p(j)}; \tilde{\mathbf{z}}_k^{p(j)}, S_k), & \text{if } \mathbf{x}_k^{a(j)} = id_o, \end{cases} \quad (9)$$

where c_d is the clutter density, id_c is the clutter identifier, id_o is the object identifier stored in $\mathbf{z}_k^{d(j)}$, $\mathcal{N}(\mathbf{z}_k^{p(j)}; \tilde{\mathbf{z}}_k^{p(j)}, S_k)$ coincides with the likelihood of the Kalman Filter used to compute $p(\mathbf{x}_k^{r(n)} | \mathbf{z}_{1:k}, \mathbf{x}_k^a)$, where $\tilde{\mathbf{z}}_k^{p(j)}$ and S_k would be the predicted measurement and the covariance of the innovation term, respectively.

The other probability term in Equation 6 is the data association likelihood based on the object appearance (it is reminded that $\mathbf{z}_k^{q(j)}$ is a similarity value based on the object appearance), given by

$$p(\mathbf{z}_k^{q(j)} | \mathbf{z}_{1:k-1}, \mathbf{x}_k^{a(j)}) = \mathcal{N}(\mathbf{z}_k^{q(j)}; 0, \sigma_a^2). \quad (10)$$

where σ_a^2 is the expected object appearance variance.

In order to make the computation of $p(\mathbf{x}_k^a | \mathbf{z}_{1:k})$ feasible, a Particle Filtering based approach is used to approximate it.

4. PARTICLE FILTER APPROXIMATION OF THE DATA ASSOCIATION POSTERIOR

The data association posterior probability $p(\mathbf{x}_k^a | \mathbf{z}_{1:k})$ is approximated by a set of N_s weighted samples (or particles) as

$$p(\mathbf{x}_k^a | \mathbf{z}_{1:k}) \approx \sum_{i=1}^{N_s} w_k^{[i]} \delta(\mathbf{x}_k^a - \mathbf{x}_k^{a,[i]}), \quad (11)$$

where $\delta(x)$ is the Dirac delta function, $\{\mathbf{x}_k^{a,[i]}\}_{i=1}^{N_s}$ are the samples, and $\{w_k^{[i]}\}_{i=1}^{N_s}$ are the weights. The samples and weights are computed using the principle of importance sampling [9]. According to this, the samples are drawn from a proposal distribution $\mathbf{x}_k^{a,[i]} \sim q(\mathbf{x}_k^a | \mathbf{x}_{k-1}^a, \mathbf{z}_{1:k})$ that theoretically must have the same support than $p(\mathbf{x}_k^a | \mathbf{z}_{1:k})$. In practice, the more similar the proposal distribution is to the posterior, the greater the quality of the approximation made by

the Particle Filtering is. A popular choice is to use the data association prior as the proposal distribution: $q(\mathbf{x}_k^a | \mathbf{x}_{k-1}^a, \mathbf{z}_{1:k}) = p(\mathbf{x}_k^a)$ because it is simple to compute. In addition, it has proven to be a satisfactory approximation for the current tracking task.

The unnormalized weights are computed as

$$\tilde{w}_k^{[i]} \propto \frac{p(\mathbf{z}_k | \mathbf{z}_{1:k-1}, \mathbf{x}_k^{a,[i]}) p(\mathbf{x}_k^{a,[i]})}{q(\mathbf{x}_k^{a,[i]} | \mathbf{x}_{k-1}^a, \mathbf{z}_{1:k})} = p(\mathbf{z}_k | \mathbf{z}_{1:k-1}, \mathbf{x}_k^{a,[i]}), \quad (12)$$

and then they are normalized: $w_k^{[i]} = \left[\sum_{l=1}^{N_s} \tilde{w}_k^{[l]} \right]^{-1} \cdot \tilde{w}_k^{[i]}$.

The importance sampling has a significant problem, called degeneracy problem [7], which consists in all the weights except one have an insignificant value after a few iterations. To solve this problem, a resampling technique, called Sampling Importance Resampling (SIR), is used, which makes a random selection of the samples at each time step according to their weights. Thus, the samples with higher weights are selected more times, while the ones with insignificant weights are discarded.

5. MEASUREMENT ACQUISITION

The measurements $\mathbf{z}_k = \{\mathbf{z}_k^v, \mathbf{z}_k^o\}$ are acquired by two types of detectors. The first type is focused on obtaining measurements \mathbf{z}_k^v from visible objects. There exists one different detector per visible object, producing several measurements, although only one or none of them is generated by the object. The rest are false measurements, i.e. clutter due to similar background structures. The detectors are based on HSV color histograms that models the object appearance, which are used to compute similarity maps by means of the Bhattacharyya distance. Finally, the measurements are obtained by selecting the local maxima of the similarity maps. The stored information relative to each measurement is the position of the local maxima $\mathbf{z}_k^{p(j)}$, the Bhattacharyya distance $\mathbf{z}_k^{q(j)}$, and the detector identifier $\mathbf{z}_k^{d(j)}$, that relates one measurement to the detector that has generated it.

The second type of detectors is aimed to obtain measurements from occluded objects, and it is the same for all the objects. This detector uses the predicted object positions, given by the transition probability of the Kalman Filter $p(\mathbf{x}_k^p, \mathbf{x}_k^q | \mathbf{z}_{k-1}, \mathbf{x}_{k-1}^a)$, to compute probability maps of overlapped regions among the tracked objects (one map per object). The overlapped regions are candidates regions for inter-object occlusion. The discrete measurements are then obtained by sampling the overlapping probability maps, and are formed by the position of the sample $\mathbf{z}_k^{p(j)}$, a value representing the occlusion density $\mathbf{z}_k^{q(j)}$, and the overlapping map identifier $\mathbf{z}_k^{d(j)}$, that relates the measurement to the occluded object.

6. RESULTS

The proposed multi-object tracking algorithm has been tested in challenging situations showing different types of interactions and occlusions. Figure 2 illustrates a situation in which two people cross without changing their trajectories. The first row shows key frames of the people crossing. The second row shows the sampled posterior probability of the vector state (Equation 4) projected over the image plane, where the hypotheses relative to each object have been encoded with different colors. In addition, the envelop of the samples of each object has been drawn to improve the interpretation of the results. The third row shows the tracking results (bounding boxes and centers in different colors), in which each object has satisfactorily maintained its identity despite the crossing.

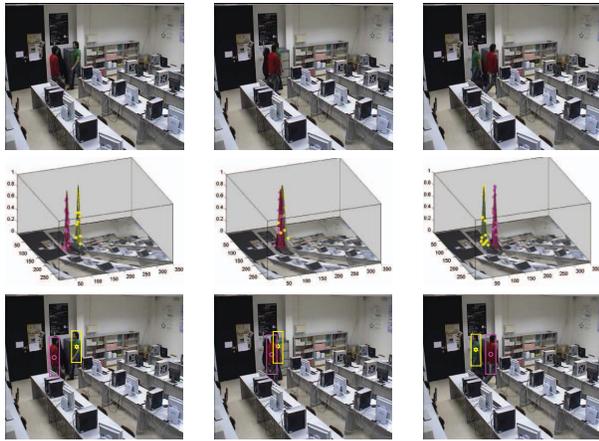


Fig. 2. People crossing situation.

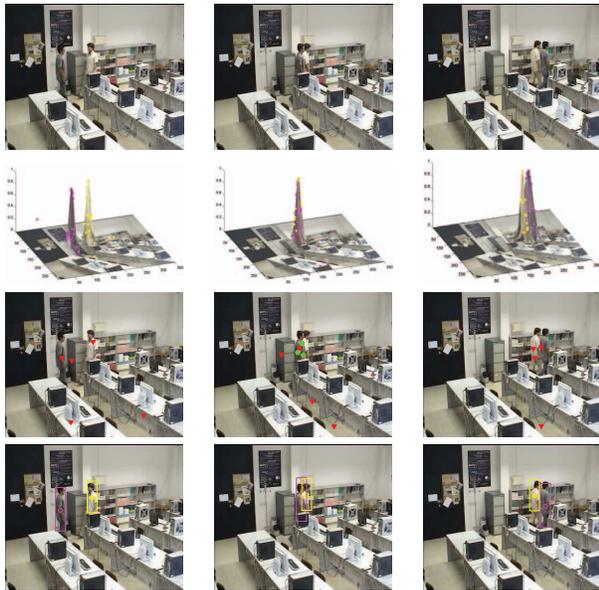


Fig. 3. Situation with a complex people interaction using measurements from visible and occluded objects.

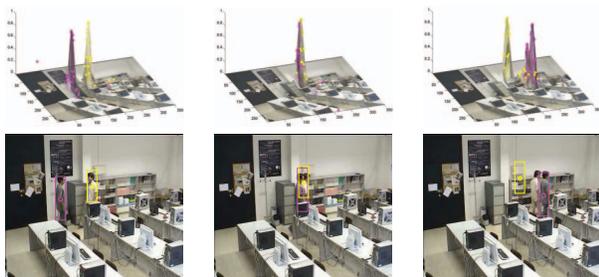


Fig. 4. Situation with a complex people interaction using measurements from only visible objects.

Figure 3 illustrates a more complex situation in which two people approach each other, and just when they are crossing, one of them changes its trajectory to follow the other. The first row contains three frames illustrating the interaction, the second row shows the sampled posterior probability, the third row shows the generated measurements from visible (red triangles) and occluded objects (green pentagons), and the last row shows the tracking results. Observe that the objects are correctly tracked due to the explicit estimation of the position of possible hidden objects. To demonstrate the efficiency of this approach, Figure 4 shows the results of processing the same sequence with the same algorithm, but without the computation of occluded object measures. As it can be observed, one object is lost since the algorithm expects that the object maintains its dynamics in absence of measurements, i.e. when there is an occlusion.

7. CONCLUSIONS

A novelty approach for multiple object tracking in visual imagery has been proposed, which is able to deal with complex object interactions. Especially, it succeeds in tracking objects that change their dynamics while they are occluded, situation that is not explicitly managed by the rest of multi-object tracking approaches. This is achieved by means of a Rao-Blackwellized Data Association Particle Filter, which has a special substructure that allows to analytically compute the object dynamics, while the data associations are approximated by a Particle Filter. In addition, this filter decomposition reduces the complexity of the algorithm with respect to the number of objects to a linear relationship. Experimental results on real surveillance videos with interacting objects have shown the efficiency and reliability of the presented tracking algorithm.

8. REFERENCES

- [1] M. Isard and J. MacCormick, "Bramble: a bayesian multiple-blob tracker," in *IEEE Proc. ICCV*, 2001, vol. 2, pp. 34–41.
- [2] K. Okuma, A. Taleghani, N. De Freitas, J. Little, and D. Lowe, "A boosted particle filter: Multitarget detection and tracking," in *Proc. ECCV*, 2004, pp. 28–39.
- [3] Z. Khan, T. Balch, and F. Dellaert, "Mcmc-based particle filtering for tracking a variable number of interacting targets," *IEEE Trans. PAMI*, vol. 27, pp. 1805–1918, 2005.
- [4] C. Hue, J. Le Cadre, and P. Perez, "Tracking multiple objects with particle filtering," *IEEE Trans. Aerosp. and Elect. Systems*, vol. 38, no. 3, pp. 791–812, 2002.
- [5] J. Vermaak, S. Godsill, and P. Perez, "Monte carlo filtering for multi target tracking and data association," *IEEE Trans. Aerosp. and Elect. Systems*, vol. 41, no. 1, pp. 309–332, 2005.
- [6] S. Särkkä, A. Vehtari, and J. Lampinen, "Rao-blackwellized particle filter for multiple target tracking," *Information Fusion*, vol. 8, no. 1, pp. 2–15, 2007.
- [7] M.S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-gaussian bayesian tracking," *IEEE Trans. Signal Processing*, vol. 50, no. 2, pp. 174–188, 2002.
- [8] A. Doucet, N.J. Gordon, and V. Krishnamurthy, "Particle filters for state estimation of jump markov linear systems," *IEEE Trans. Signal Processing*, vol. 49, no. 3, pp. 613–624, 2001.
- [9] 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.