

A Data Association Algorithm for Multiple Object Tracking in Video Sequences

M.H. Jaward, L. Mihaylova, N. Canagarajah and D. Bull

Abstract This paper presents a particle filtering algorithm for multiple object tracking. The proposed particle filter (PF) embeds a data association technique based on the joint probabilistic data association (JPDA) which handles the uncertainty of the measurement origin.

I. Introduction

Tracking a group of targets in video sequences has many surveillance applications. It has been used for security monitoring [1], [2], as well as for traffic flow measuring [3], accident detection on highways, and routine maintenance in nuclear facilities. The goal is to obtain a record of the trajectory of the moving objects such as humans or vehicles over space and time by processing the data. The moving target can undergo non-rigid deformations, rotations or occlusions.

During multiple target tracking, several challenging issues arise which do not exist in single target tracking. One of them is the management of multiple tracks caused by newly appearing targets and the disappearance of already existing targets. In earlier works, this is handled by using hybrid state estimation framework with joint tracking of all existing targets [4]. Joint tracking of targets avoids the necessity of maintaining several filters, but the identity of the individual targets might be lost, especially when targets are close to each other. A separate single tracking filter can be used for multiple target tracking, only when all targets are fairly well separated.

When the targets are close to each other, or their paths cross, it has been shown in aerospace applications [5], [6], [4], that the data association techniques can assist tracking without losing the identity of each target. However, some of the techniques developed for aerospace applications are not applicable to vision problems because of the particu-

larities of visual tracking based on different features such as colour and motion. In vision tracking there are no measurements in explicit form. It is computationally expensive to obtain external measurements in the form of blobs, but it is easier to obtain measurements from the image region specified by the state vector. For example, the clustering techniques proposed in [7] to evaluate the varying number of targets cannot be used as we do not have explicit external measurements.

In this paper, a sequential Monte Carlo version of a data association scheme is designed to track multiple targets and the track management is handled by existence probabilities calculated from the data association stage. This proposed scheme is simple and does not demand high computational resources. Tracking is based on multiple independent particle filters and the Probabilistic Data Association (PDA) algorithm which handles the uncertainty due to the measurement origin. The data association algorithm also helps to recover from full or partial occlusions. The proposed algorithm can estimate the number of active targets in the video sequence and can accordingly stop unwanted tracking filter(s).

The rest of the paper is organised as follows. Section 2 formulates the tracking problem and describes the motion and measurement models. Section 3 formulates the Joint Probability Data Association problem and presents the developed PF tracking algorithm based on Joint Probabilistic Data Association Algorithm (JPDA). Section 4 contains results demonstrating the PF-JPDA performance over natural video sequences. Section 5 summarises the results and discusses open issues for future research.

2. Problem Formulation

The aim is to sequentially track multiple objects described by the same specified colour histogram q^* , in a video sequence $\mathbf{Z}_k = \{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_k\}$, where \mathbf{z}_k denotes the image (vector of pixel values) at discrete time instant k . The state space approach

The authors are with the Department of Electrical and Electronic Engineering at the University of Bristol, Woodland Road, Bristol, BS8 1UB, United Kingdom, email: m.h.jaward@bristol.ac.uk, mila.mihaylova@ieee.org

requires to specify a motion model, i.e., the evolution of the state $p(\mathbf{x}_k|\mathbf{x}_{k-1})$, and a measurement model, linking the state and the current measurement, $p(\mathbf{z}_k|\mathbf{x}_k)$.

The Monte Carlo approach relies on a sample-based construction to represent the state pdf. Multiple particles (samples) of the state are generated, each one associated with a weight which characterises the quality of a specific particle. An estimate of the variable of interest is obtained by the weighted sum of particles. Two major stages can be distinguished: *prediction* and *update*. During prediction, each particle is modified according to the state model of the region of interest in the video frame, including the addition of random noise in order to simulate the effect of the noise on the state. In the update stage, each particle's weight is re-evaluated based on the new data. An inherent problem with particle filters is degeneracy, the case when only one particle has a significant weight. A *resampling* procedure helps to avoid degeneracy by eliminating particles with small weights and replicating the particles with larger weights.

The next two subsections describe the object motion model and the measurement likelihood function.

2.1. State Vector and Motion Model

The moving object is surrounded by a rectangular region. We describe the motion of the object by the random walk model,

$$\mathbf{x}_k = \mathbf{F}\mathbf{x}_{k-1} + \mathbf{v}_{k-1}, \quad (1)$$

where the state vector $\mathbf{x}_k = (x_k, y_k)^T$ comprises the coordinates of the centre of this image region. \mathbf{F} is the transition matrix ($\mathbf{F} = \mathbf{I}$ for our random walk model) and \mathbf{v}_{k-1} is the process noise, assumed to be white, zero-mean, Gaussian, with a covariance matrix \mathbf{Q} . Other variables can be added, such as width and height of the image region, velocities and scale change rate, depending on the application.

2.2. Colour Measurement Model

The work described in this paper is based on colour measurement cue. Other cues, e.g. texture, edges or motion and joint variations of any of them can be used. The choice of the specific cue depends on many factors such as the resolution of the video, background/ foreground environment and dynamic nature of the targets to be tracked. Following [8] and [9], we do not use the entire image \mathbf{z}_k as a

measurement, but rather we extract from the image the colour histogram q_k , computed inside the image region that is specified by the state vector \mathbf{x}_k . The centre is defined by (x_k, y_k) . Furthermore, we adopt the Gaussian density for the likelihood function of the measured colour histogram as follows:

$$p(q_k|\mathbf{x}_k) \propto \mathcal{N}(D_k; 0, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left\{-\frac{D_k^2}{2\sigma^2}\right\}, \quad (2)$$

where D_k is the distance between the reference histogram q^* of the objects to be tracked and the histogram q_k computed from the current frame \mathbf{z}_k in the region specified by the state vector \mathbf{x}_k . The standard deviation σ of the Gaussian density in (2) is a design parameter.

If the two histograms are calculated over U bins, the distance D_k between two histograms is derived in [8] from the Bhattacharya similarity coefficient:

$$D_k^2 = 1 - \sum_{u=0}^U \sqrt{q^*(u)q_k(u)}.$$

3. Joint Probability Data Association

The proposed here multiple target tracking technique is based on a sequential Monte Carlo filter and a Monte Carlo data association is avoiding the ambiguities caused by the different measurement origin. Data association is of crucial importance for multiple target tracking because of the necessity to relate each measurement to the correct object. Several methods have been proposed in the estimation and tracking literature [5], [6], [10] such as the nearest neighbour (NN) method. In general multi-target tracking deals with state estimation of an unknown number of targets. Some methods consider special cases with a constant or known number of targets. The observations are assumed to originate from different targets or from clutter. The clutter is a special type of the so-called false alarms, whose statistical properties are different from the targets. In some applications only one measurement is assumed available from each target, where in other applications several returns are available. This will reflect which data association method to use. In [5], this is referred to as the *nearest neighbour standard filter* and it relies only on the closest observation to any given state to perform the measurement update step. The method can also be implemented as a global optimisation, so that the total number of observations for tracking the statistical distance is minimised. Another multi-target tracking association method is the *Joint Probabilistic Data Association* (JPDA) which is an extension of the *Probability*

Data Association algorithm [5] to multiple targets. It estimates the states by a sum over all the association hypothesis weighted by the probabilities from the likelihood. The most general data association method, the multiple hypothesis tracking (MHT) technique, is computationally intensive, and calculates every possible update hypothesis [11]. In [12], the probabilistic MHT (PMHT) method is presented, and a maximum-likelihood method in combination with the expectation maximisation (EM) method. A comparative study of JPDAF and PMHT is made in [13]. In [14], the solution to the assignment problem for data association is proposed within the Bayesian framework by incorporating the association in the estimation equations. In [15], [16], this idea is suggested for particle filtering, when the problem of maintaining a track on a target in the presence of intermittent spurious objects. Samples are drawn from the overall target probability density. The so-called joint-filter in [17] is a solution to the joint data association and estimation problem for particle filters. The estimation is performed by a particle filter and a Gibbs sampler [18] is used for the data association.

The JPDA approach that we follow is an extension of the PDA developed in [5], [6] for single targets. In the JPDA, a known number of targets τ is assumed. The index $t = \{1, \dots, \tau\}$ designates one among the τ targets. The measurements at time step k are denoted as $\mathbf{z}_k = \{\mathbf{z}_k^j\}_{j=0}^{m_k}$, where an artificial measurement \mathbf{z}_0 is introduced to handle false alarms or clutter and the number of measurements is given by m_k . The measurement to target association probabilities are evaluated jointly across the targets. Let θ denotes the joint association event (the time index k is omitted for simplicity) and θ_t^j is the particular event which assigns the measurement j to the target t . By assuming that the estimation problem is Markovian and by applying the Bayes' theorem, the joint association probabilities are

$$P(\theta|\mathbf{Z}_k) = \frac{1}{c} p(\mathbf{z}_k|\theta, \mathbf{X}_k) P(\theta|\mathbf{X}_k), \quad (3)$$

where c is a normalisation constant and $\mathbf{X}_k = \{\mathbf{x}_1, \dots, \mathbf{x}_k\}$. Additionally, it is assumed that the measurements are detected independently of each other. The probability of the assignment θ conditioned on the sequence of the targets' states $P(\theta|\mathbf{X}_k)$ is approximated by

$$P(\theta|\mathbf{X}_k) = P_D^{\tau-n} (1 - P_D)^n P_{FA}^{m_k - (\tau - n)}, \quad (4)$$

where P_D denotes the probability of detection, n is

the number of \mathbf{z}_0 assignments and P_{FA} denotes the probability of false alarm.

The innovation between the j -th measurement \mathbf{z}_k^j and the predicted measurement $\hat{\mathbf{z}}_{t,k}^j$ of target t at time step k is $\nu_{t,k}^j = \mathbf{z}_k^j - \hat{\mathbf{z}}_{t,k}^j$ and $\mathbf{S}_{t,k}^j$ is the corresponding innovation covariance matrix. Then the normalised innovation $d_{t,k}^j$ is defined as

$$d_{t,k}^j = (\nu_{t,k}^j)^T (\mathbf{S}_{t,k}^j)^{-1} \nu_{t,k}^j. \quad (5)$$

By assuming that the measurement is of dimension M , it follows that the M -dimensional Gaussian association likelihood for the j -th measurement to the target t is

$$p(\nu_{t,k}^j) = \frac{1}{(2\pi)^{M/2} |\mathbf{S}_{t,k}^j|^{1/2}} \exp \left\{ -\frac{(d_{t,k}^j)^2}{2} \right\}, \quad (6)$$

where $|\mathbf{S}_{t,k}^j|$ is the determinant of $\mathbf{S}_{t,k}^j$.

Finally, the probability of an individual joint association event is given by

$$P(\theta|\mathbf{Z}_k) = P_D^{\tau-n} (1 - P_D)^n P_{FA}^{m_k - (\tau - n)} \prod_{\theta_t^j \in \theta} p(\nu_{t,k}^j). \quad (7)$$

As shown in [19], the number of associations increases exponentially with an increase in the number of measurements and targets, rendering the use of all hypothesis infeasible for even moderate values for these quantities. The number of possible associations λ , given that N_D of the τ targets have been detected, is

$$N_{\lambda_k}(N_D, \tau) = \frac{m_k! \tau!}{N_D! (m_k - N_D)! (\tau - N_D)!} \quad (8)$$

and so the total number of possible hypotheses (since N_D is not known) is

$$\sum_{N_D=0}^{\min(m_k, \tau)} N_{\lambda_k}(N_D, \tau). \quad (9)$$

Even for the case of three targets and three measurements, this amounts to 34 hypotheses. Therefore, in practice, it is common to apply gating techniques [5] to reduce the number of hypotheses and the computational cost respectively.

3.1. Monte Carlo JPDA

In this subsection, we describe the Monte Carlo JPDA used in our tracking algorithm. The Monte Carlo version of JPDA has been studied by several authors (see e.g. [20], [21]). A recent paper of Vermaak et al. [22] surveys most of the approaches developed earlier. In [22] marginal filtering distributions for each of the targets are

represented with Monte Carlo samples, instead of with Gaussian approximations, as it is in the standard JPDAF. The PF-JPDA algorithm implemented in our paper is given below. In the algorithm, $\mathbf{x}_{t,k}^{(i)}$ refers to the i th sample of the state vector corresponding to target t at time k .

The JPDA particle filter for multiple target tracking

1) Initialisation.

Set $k = 0$, generate N samples $\mathbf{x}_{t,0}^{(i)}$ for all targets $t = 1, \dots, \tau$ independently.

$\mathbf{x}_{t,0}^{(i)}$ is drawn from $p(\mathbf{x}_{t,0})$, with initial weights $w_{t,0}^{(i)} = 1/N$, for $i = 1, \dots, N$ particles and set $k = 1$.

2) For $i = 1, \dots, N$ predict new particles

$$\mathbf{x}_{t,k}^{(i*)} = \mathbf{F}\mathbf{x}_{t,k-1}^{(i)} + \mathbf{v}_{t,k-1}^{(i)}.$$

3) For each particles compute the weights for all measurements ($j = 0, \dots, m_k$) to target ($t = 1, \dots, \tau$) associations $w_{t,k}^{(i)} = \sum_{\theta} P(\theta|\mathbf{Z}_k)$ (Equation (7)) and normalise the weights for each target: $\tilde{w}_{t,k}^{(i)} = \frac{w_{t,k}^{(i)}}{\sum_{i=1}^N w_{t,k}^{(i)}}$.

4) For each target, generate a new set $\{\mathbf{x}_{t,k}^{(i)}\}_{i=1}^N$, by resampling with N times from $\{\mathbf{x}_{t,k}^{(i*)}\}_{i=1}^N$, where $P(\mathbf{x}_{t,k}^{(i)} = \mathbf{x}_{t,k}^{(i*)}, t) = \tilde{w}_{t,k}^{(i)}$.

5) Increase k and iterate to step 2.

3.2 Particularities of the JPDA Implementation

The algorithm assumes knowledge of the maximum number of targets, τ . It is based on the sequential Monte Carlo version of the JPDA algorithm [20], [21]. All objects to track are manually selected within rectangular regions. The reference histograms for each object are obtained from these selected regions with the help of colour cue. During the tracking process the number of objects is estimated. This is used to detect when an object disappears from the scene and to stop the corresponding tracking filter.

Let $P(H_l|\mathbf{Z}_k)$ ($l = 1, \dots, \tau$) denotes the posterior probability of the existence of l number of targets. In general, this probability depends on the ‘full’ hypothesis list which considers all possible hypotheses from frame 1 to frame k . However, it can be approximately estimated from the hypotheses assumed in the Monte Carlo JPDA. Using the total probability

theorem, the existence probability of l number of targets is given by

$$P(H_l|\mathbf{Z}_k) = \sum_{\theta \in \chi_l} P(\theta|\mathbf{Z}_k), \quad l = 1, \dots, \tau, \quad (10)$$

where χ_l is the event (hypothesis) that l number of targets exists. Joint existence probabilities of specific targets (such as the existence of target 1 with target 2 or target 1 with target 3) also can be estimated, in a similar way.

The probabilities in (10) give the number of targets active at time instant k . We also need to know which target(s) disappeared from the scene. This can be obtained from the posterior probabilities of existence of individual targets. Let $H_{t_1, t_2, \dots, t_\rho}$ be the hypothesis that t_1, t_2, \dots, t_ρ targets exist, with $\rho = \{1, 2, \dots, \tau\}$. Then the existence probability of this event is given by:

$$P(H_{t_1, t_2, \dots, t_\rho}|\mathbf{Z}_k) = \sum_{\theta \in \chi_{t_1, t_2, \dots, t_\rho}} P(\theta|\mathbf{Z}_k), \quad (11)$$

where $\chi_{t_1, t_2, \dots, t_\rho}$ denotes the events that targets $\{t_1, t_2, \dots, t_\rho\}$ exist. If a target (tracked by the algorithm) disappears from the scene, the changes in the number of objects is reflected in the probabilities $P(H_t|\mathbf{Z}_k)$. From the probabilities $P(H_{t_1, t_2, \dots, t_\rho}|\mathbf{Z}_k)$, we can obtain the identity of the target which just disappeared. The use of these probabilities in track management is further explained in the next section with an example. Accordingly, we can stop the corresponding filter and this reduces the computational complexity.

In many cases, the colour information of the object of interest varies in time. In such situations, adaptation of the initial reference histograms are required for continuous tracking similarly to [9]. The algorithm proposed in this paper can be used with or without adaptation. With adaptation, the reference models for some or all targets can be modified such as in [23]

4. Performance Investigation

The proposed algorithm is tested using real world image sequences. We show results with a video sequence obtained from www.visualsurveillance.org. The aim is to track two people as shown in figure 1(a). The maximum number of targets is two ($\tau = 2$). The algorithm is used with $N = 150$ samples for each filter. The PF is used with a system noise covariance of $\mathbf{Q} = \text{diag}\{5, 5\}$ and observation noise covariance of $\sigma^2 = 0.01$. The probability of detection is 0.9, and the probability of false alarm 0.1. The

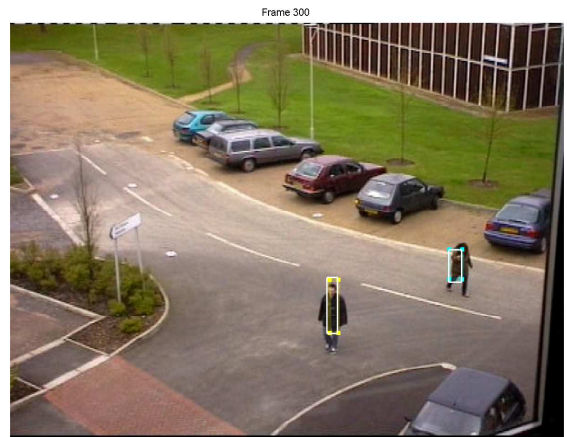
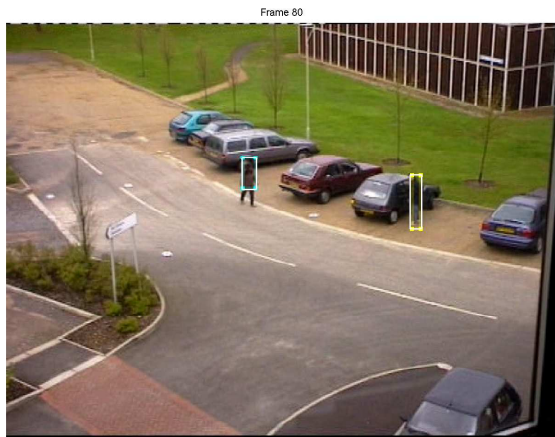
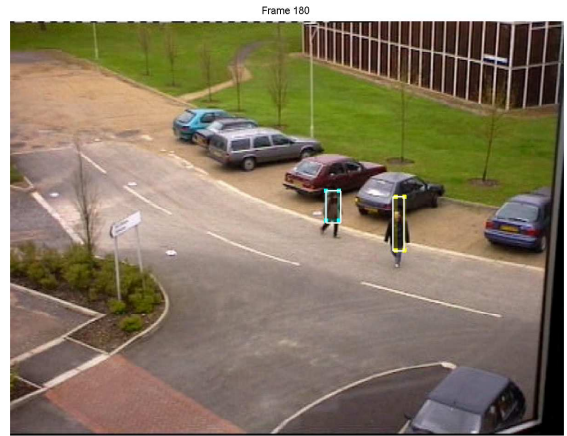
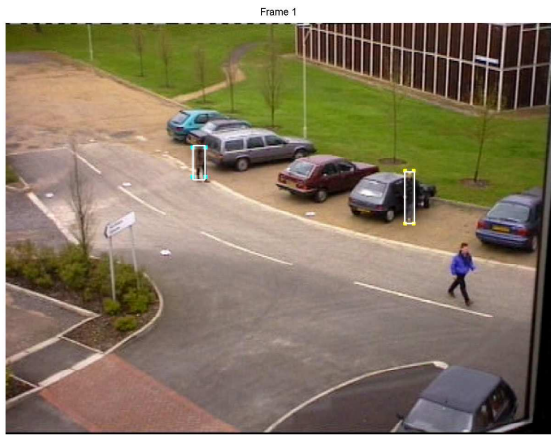


Fig. 1. Image frames of the video sequence: (a) Initial frame. Targets to be tracked are selected within a rectangle (b) Frame 80. Tracked players are marked with a rectangle.

Fig. 2. Image frames of the video sequence: (a) Frame 180- before occlusion (b) Frame 300- after occlusion. Tracked players are marked with a rectangle.

colour cue used $U = 128 \times 128 \times 128$ number of histogram bins.

The two targets are selected manually from the first frame. The person on the left is denoted as target #1 and selected second person on the right is denoted as target #2. A reference model adaptation is performed only for target #1 as it was found to be difficult to track without adaptation. The adaptation scheme proposed in [23] is used in this paper.

Figure 1 illustrates the tracking performance of the proposed algorithm. As shown in this figure 1(b), at frame 80 both filters are locked to their targets. Figure 2 shows the performance of proposed algorithm with a presence of partial occlusion. The figure in the top row shows that both filters are tracking their corresponding targets at frame 160. After the occlusion, as seen from the figure in the bottom row, the two filters are still successfully tracking the two objects. Variations of the number of targets tracked

is shown in Figure 3. As seen from this figure, before frame 360 the number of targets is two. When the target #1 disappears from the frame, the number of active targets becomes 1. This plot does not indicate which target disappeared and this information can be obtained from the existence probabilities as shown in Figure 4. Figure 4 presents the existence probability for targets #1 and #2. The probability for target #1 is very low after the frame 360 compared to that of target #2. This indicates that the disappeared target is the target #1. The probabilities of number of active targets change around frame 230. This is caused by the occlusion of two targets and at this period, the number of active targets becomes close to one.

5. Conclusions

This paper addresses the issue of data association of multiple targets. An algorithm for data association is designed for tracking objects in video sequences. Current investigations are

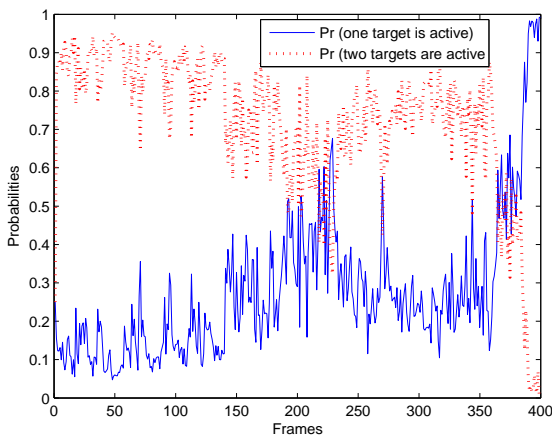


Fig. 3. Probabilities of the number of targets: one active object or two active targets.

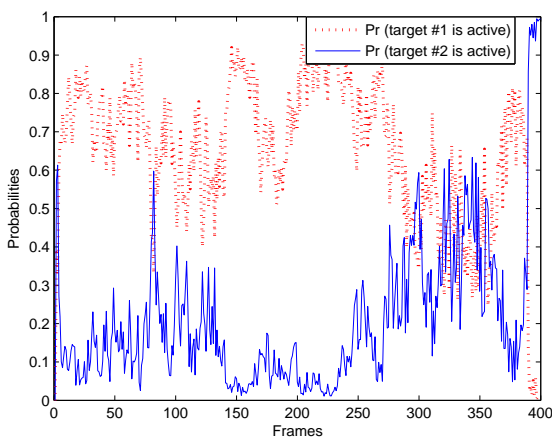


Fig. 4. Existence probabilities of individual targets.

concentrated on coping with full occlusions and using other features apart from colour.

Acknowledgements. The authors are grateful to the financial support by the UK MOD Data and Information Fusion Defence Technology Centre.

REFERENCES

- [1] E. Stringa and C.S. Regazzoni, "Real-time video-shot detection for scene surveillance applications," *IEEE Trans. Image Processing*, vol. 9, pp. 69–79, Jan. 2000.
- [2] I. Haritaoglu, D. Harwood, and L. Davis, " w^4 : who? when? where? what? a real time system for detecting and tracking people," in *Proc. IEEE Conf. Face Gesture Recognition*, 1998, pp. 222–227.
- [3] D. Koller, J. Weber, T. Huang, J. Malik, B. Rao G. Ogasawara, and S. Russell, "Toward robust automatic traffic scene analysis in real-time," in *Proc. Int. Conf. Pattern Recog.*, 1994, vol. 1, pp. 126–131.
- [4] W. Ng, J. Li, S. Godsill, and J. Vermaak, "A review of recent results in multiple target tracking," in *Proc. of the International Symp. on Signal Processing and Analysis*, 2005.
- [5] Y. Bar-Shalom and X.R. Li, *Estimation and Tracking: Principles, Techniques and Software*, Artech House, Boston, MA, 1993.
- [6] S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems*, Artech House, 1999.
- [7] W. Ng, J. Li, S. Godsill, and J. Vermaak, "Tracking variable number of targets using sequential monte carlo methods," in *Proc. of the IEEE Statistical Signal Processing Workshop*, 2005.
- [8] D. Comaniciu, V. Ramesh, and P. Meer, "Real-time tracking of non-rigid objects using mean shift," in *Proc. IEEE Conf. Comp. Vision Pattern Recog.*, 2000, pp. 142–149.
- [9] K. Nummiaro, E. Koller-Meir, and L. Van-Gool, "An adaptive color-based particle filter," *Image and Vision Computing*, vol. 21, pp. 99–110, 2003.
- [10] T. Kirubarajan and Y. Bar-Shalom, "Probabilistic data association techniques for target tracking in clutter," *Proc. of the IEEE*, vol. 92, no. 3, pp. 536–557, 2004.
- [11] D. Reid, "An algorithm for tracking multiple targets," *IEEE Trans. Automatic Control*, vol. 24, no. 6, pp. 84–90, 1979.
- [12] H. Gauvrit, J-P. Le Cadre, and C. Jauffret, "A formulation of multitarget tracking as an incomplete data problem," *IEEE Trans. Aerospace and Systems*, vol. 33, no. 4, pp. 1242–1257, Oct. 1997.
- [13] C. Rago, P. Willett, and R. Streit, "A comparison of the JPDAF and PMHT tracking algorithms," in *Proc. of the ICASSP*, 1995, vol. 5, pp. 3571–3574.
- [14] D. Avitzour, "A stochastic simulation bayesian approach to multitarget tracking," *IEE Proc. Radar, Sonar and Navig.*, vol. 142, pp. 41–44, 1995.
- [15] D.J. Salmond, D. Fisher, and N.J. Gordon, "Tracking on the presence of intermittent spurious objects and clutter," in *SPIE Conf. on Signal and Data Processing of Small Targets*, 1998.
- [16] N.J. Gordon, "A Hybrid Bootstrap Filter for Target Tracking in Clutter," *IEEE Trans. on Aerospace and Electronic Systems*, vol. 33, no. 1, pp. 353–358, 1997.
- [17] C. Hue, J-P. Le Cadre, and P. Perez, "Tracking multiple objects with particle filtering," *IEEE Trans. Aerospace and Systems*, vol. 38, no. 3, pp. 791–812, July 2002.
- [18] S. Gemen and D. Gemen, "Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, pp. 721–741, 1984.
- [19] B.Ristic, S. Arulampalam, and N.Gordon, *Beyond the Kalman Filter: Particle Filters for Tracking Applications*, Artech House Radar Library, 2004.
- [20] R. Karlsson and F. Gustafsson, "Monte Carlo data association for multiple target tracking," in *IEE Target tracking: Algorithms and applications*, Netherland, Oct 2001.
- [21] D. Schultz, W. Burgard, D. Fox, and A.B. Cremers, "Tracking multiple moving targets with a mobile robot using particle filters and statistical data association," in *IEEE International Conf. on Robotics and Automation*, 2001.
- [22] J. Vermaak, S.J. Godsill, and P. Perez, "Monte Carlo filtering for multi-target tracking and data association," *IEEE Trans. Aerospace and Systems*, vol. 41, no. 1, pp. 309–332, January 2005.
- [23] M.H. Jaward, N. Canagarajah, and D. Bull, "Adaptive tracking using combined colour and motion cues," *Under preparation for IEEE Trans. Image Processing*, 2006.