Hindawi Publishing Corporation
EURASIP Journal on Image and Video Processing
Volume 2011, Article ID 642532, 12 pages
doi:10.1155/2011/642532

Research Article

Combined Data Association and Evolving Particle Filter for Tracking of Multiple Articulated Objects

Harish Bhaskar[1] and Lyudmila Mihaylova[2]

[1] *Department of Computer Engineering, Khalifa University of Science Technology and Research, Sharjah Campus, UAE*
[2] *School of Computing and Communications, InfoLab21, Lancaster University, Lancaster LA1 4WA, UK*

Correspondence should be addressed to Harish Bhaskar, harishbhasky@gmail.com

Received 1 May 2010; Revised 19 November 2010; Accepted 14 January 2011

Academic Editor: Andreas Uhl

This paper proposes an approach for tracking multiple articulated targets using a combined data association and evolving population particle filter. A visual target is represented as a pictorial structure using a collection of parts together with a model of their geometry. Tracking multiple targets in video involves an iterative alternating scheme of selecting valid measurements belonging to a target from a clutter or other measurements that all fall within a validation gate. An algorithm with extended likelihood probabilistic data association and evolving groups of populations of particles representing a multiple-part distribution is designed. Variety in the particles is introduced using constrained genetic operators both in the sampling and resampling steps. We explore the effect of various model parameters on system performance and show that the proposed model achieves better accuracy than other widely used methods on standard datasets.

1. Introduction

Tracking articulated targets is a central problem in computer vision, with applications including robotics and surveillance. The problem of multiple articulated target tracking (MATT) deals with tracking a variable number of targets, each consisting of the same number of different constituent body parts, given noisy measurements at every instant of time from a dynamic scene, and simultaneously maintaining correct target identities irrespective of any visual perturbations [1–4]. However, these tasks are complicated by the nonrigid variation within the general class of objects that we wish to track (e.g., people, animals, etc.), appearance variations of targets, and the presence of occlusions. Different methods have been proposed in the literature to cope with these challenges, for example, [2, 5–7]. The pictorial structure approach proposed in [6] is an appealing approach for people modelling in view its simplicity and generality. A different, limb-based structure model is developed in [5], particularly suited for the detection and tracking of multiple people in crowded scenes.

The problem of MATT can be divided into the subproblems of estimation and data association. A popular approach to solving this estimation problem is to build linearised filters such as the extended Kalman filter (EKF) [8], under a Gaussian noise assumption. Consequently, sufficient statistics from such linearised filters are used for data association. However, with nonlinear models in the state equation and non-Gaussian noise assumption, such linearised models often lead to inaccurate solutions or even face divergence. The sequential Monte Carlo (SMC) methods, such as the particle filters (PFs) [9] have proven their potential for the estimation of nonlinear systems, with non-Gaussian noise assumption and multimodal distributions.

In this paper, we propose an approach combining evolving population particle filtering with extended likelihood data association for MATT applications. To account for the uncertainty in the origin of the measurement, the extended likelihood data association method [10] incorporates local attribute information of measurements weighted by probabilistic data association (PDA) for correctly identifying the measurement from the target as against the clutter.

On the other hand, the evolving population particle filter, provides iterative convergence of groups of particles through a specified kernel by introducing variety in the population using constrained genetic operators in both the sampling and resampling steps.

One of the main novelties of the method is that it conveniently integrates data association into evolving population particle filtering, thus allowing particles to regenerate both in sampling and resampling steps by simultaneously disregarding particle measurement that account for clutter within a specified validation gate. Our results suggest that the proposed integrated approach can considerably improve performance when compared individually to a Markov chain Monte Carlo (MCMC) combined data association technique or a generic particle filter (*non-MCMC filter*). Second, the geometrical constraints imposed by the picture structure representing the target are intrinsically modeled into the particle regeneration process through *constrained* genetic operations. Furthermore, some of the system parameters (such as the size of the validation gate) are learned from the data rather than specified by hand.

The remaining part of the paper is organised in the following way. Section 2 makes an overview of related works. Section 3 presents the proposed approach for multiple articulated object tracking. Results are given in Section 4. Finally, conclusions are summarised in Section 5.

2. Related Work

MATT is a highly challenging area of research within computer vision and tracking communities. The high degree of freedom of multiple articulated regions together with the interdependencies between them and with other targets requires efficient techniques able to cope effectively with the dynamic changes of the objects. In general, motion tracking of articulated objects in video consists of two distinct steps: *detection* and *tracking*. During the detection process, we aim to segment the human objects and their constituent body parts from the frames of the video sequences. In tracking, we are spatially locating these detected regions in time. A number of techniques have been proposed for the detection and tracking phases [2, 11–13]. In our paper, we assume a standard procedure applied for the detection step and propose a novel tracking methodology.

Tracking in recent years is often considered as a dynamic system estimation problem [14]. A number of different techniques have been proposed [15] in the past to estimate the variables of such dynamic systems. Some of the important methods include the Kalman filter [16], and unscented Kalman filter [17]. These methods assume that the posterior probability density of the system model is Gaussian. This assumption is more often restrictive and does not always suit different applications. In order to cope with the nonlinearity, extended techniques such as unscented Kalman filter [18], extended Kalman filter [19], approximation grid filter, and particle filters [17, 20] have been recommended. A particle filter approximates the posterior state probability density using a set of particles and propagating these particles over time with appropriate weighting coefficients often produces efficient tracking. Particle filters are robust to nonlinear, non-Gaussian systems with multi-modal distributions. However, even with a large population of particles, there may be no or little number of particles near the actual correct state. The second main drawback of particle filter methods is degeneracy. The problem of degeneracy refers to some particles having negligible weight as against the weight being concentrated on few others. Resampling techniques are employed to tackle degeneracy issues but sometimes when applied improperly can lead to sample impoverishment [21].

Population-based methods [22–24] are techniques that generate a collection of samples in parallel as against single independent or dependent samples. We can conveniently categorize these population-based methods into (a) MCMC type of methods and (b) methods based on importance sampling and resampling ideas. While MCMC methods are directed by theoretical convergence based on iterations, sampling/resampling techniques rely on processing a number of samples in parallel and sequential Monte. The population MCMC methods apply population moves that exchange variables between population members in order to generate the new target density. An example of a population move is the exchange move that swaps information between chains in a population. Similar types of moves are realised in the genetic algorithms, with the crossover, mutation, and exchange steps. In contrast, the sequential Monte Carlo methods [25] were constructed to sample from a sequence of related target distributions, using resampling techniques on the samples from previous target density. Commonly used resampling techniques include multinomial resampling [26], residual resampling technique [25], and stratified resampling [25].

One of the other main issues of MATT applications is the presence of multiple parts of the body and multiple targets that share similar feature characteristics, thus leading to uncertainty in the origin of measurements. Data association (DA) methods [27, 28] are used for correctly identifying the measurement that originated from the target from clutter of other multiple noisy measurements. It is assumed that the clutter is a model of false detections whose statistical properties are significantly different from those of the targets. Data association is of crucial importance to our problem because of the requirement to relate each measurement to the correct body part of the correct object of interest. There has been extensive studies in data association [10, 27, 29–36]. Most methods are restricted by assumptions on the number of targets, statistical properties of observations, and number of possible measurements.

In order to tackle the problem of MATT, it is important to combine estimation techniques and data association so that robust tracking is possible.

3. The Combined Data Association with Evolving Population Particle Filter for Multiple Articulated Object Tracking

In our proposed technique, we formulate MATT as an estimation problem combined with data association. Let us

begin by denoting the total number of targets in our scene to be $\mathfrak{T}$. We represent each target τ as a pictorial structure using a collection of n parts as in [6]. The objective of the proposed technique is to estimate the state vector $x_{k,\tau}$ of each target τ at time instant k by recursively updating the posterior distribution of each target $p(x_{k,\tau} \mid Z_k)$ based on a set $Z_k = \{z_1, z_2, \ldots, z_k\}$ of measurements at time k. The movement of each target is described by a general nonlinear state space model and estimated using the prediction step:

$$p(x_{k,\tau} \mid Z_{k-1}) = \int p(x_{k,\tau} \mid x_{k-1,\tau})\, p(x_{k-1,\tau} Z_{k-1})\, dx_{k-1,\tau}, \tag{1}$$

and update/filtering step

$$p(x_{k,\tau} \mid Z_k) = \int p(z_k \mid x_{k,\tau})\, p(x_{k,\tau} \mid Z_{k-1})\, dx_{k,\tau}, \tag{2}$$

where, $p(x_{k,\tau} \mid x_{k-1,\tau})$ is the dynamic model of state evolution and $p(z_k \mid x_{k,\tau})$ is the likelihood of any measurement z_k given the state $x_{k,\tau}$.

The proposed technique is meant to track multiple targets by taking into consideration the geometrical constraints of the various parts which represents the target before evaluating the marginal distribution. An evolving population Markov chain Monte Carlo (EPMCMC) filter is developed that encapsulates constrained evolution of the populations using specific genetic operators and robust association of measurements to targets incorporating local attribute information using the expected likelihood data association method [10]. The developed approach can be summarised as follows:

(1) initialisation: using pictorial structure type models as in [7], we localise all the $\mathfrak{T}$ targets along with the configuration of their n parts,

(2) for iterations $1, \ldots, T$,

  (a) evolve particles using *constrained* genetic operators as in EPMCMC filter,

  (b) the expected likelihood data association provides an efficient solution for data association of each object into the particle weights of the EPMCMC filter.

In the following sections, we describe the two steps in detail before outlining how they are integrated within the proposed framework.

3.1. EPMCMC Filter. The EPMCMC filter algorithm proceeds in three distinct steps. The first step involves initialising the populations of particles from their respective proposal distributions. Let $x_{k,\tau}^{(i)}$, $i = 1, \ldots, N$ represent the population of particles for target τ (each particle is a configuration vector containing the position and speed of the n body parts for the target) at time instant k. So, during initialisation: $x_{k=1,\tau}^{(i)} = \Gamma_{k=1,\tau}$, where $\Gamma_{k=1,\tau}$ is the distribution of particles around the initial localisation of different body parts of target τ. During initialisation, we also evaluate the initial weights of particles $w_{k=1,\tau}^{(i)}(x_{k,\tau}^{(i)})$ and normalise the weights to get $W_{k=1,\tau}^{(i)}$.

The EPMCMC filter iterates between the resampling and sampling steps. The resampling step is usually performed only when the effective sample size (ESS) is less than a predefined threshold $N_{\text{threshold}}$ and is measured as $(\sum_{i=1}^{N} (W_{k,\tau}^{(i)})^2)^{-1}$. During the resampling step, we perform local evolution of particles; that is, particles from the *same* population belonging to the *same* part of the target are combined iteratively using steps of crossover, mutation, or exchange. This approach introduces variety in particles by regenerating a unique and good population. Here, we also recalculate weights of particles in each population using a relevant likelihood metric and normalise them. An illustration of the genetic moves involved in the EPMCMC filter is presented in Figure 1.

In the sampling step, we perform global evolution of particles, that is, iteratively evolving particles from *different* populations belonging to *different* parts of the target using the steps of crossover, mutation, or exchange. However, the main difference from local evolution of particles is that in the global evolution we enforce geometrical constraints into the evolution, process. These geometrical constraints are derived from the pictorial structure model of the target and are based on the neighbourhood structure of every part of the target. When subjecting the particles to evolution using genetic operators, we allow particles of one part to be influenced by only the particles of its neighbours. For example, when performing the crossover operation, between two particles $x_{s,\tau}$ and $x_{q,\tau}$, we crossover chromosomes only based on the neighbourhood relationships that parts share with each other. That is, we restrict crossovers between the arms of one particle to the legs of the other and encourage crossovers between arms of one part together with the torso of the other, as these are neighbouring regions of the elastic pictorial structure model.

In addition, we evaluate the likelihood of each particle based on how well the image data supports the proposed hypothesized candidate parts. We map this support as the weighted summation of two subsequent terms: the first one indicates the goodness of fit of the part with the image data, and the second one measures the goodness of fit of pairs of candidate parts as connected in the pictorial structure model.

The proposed EPMCMC filter is given in Algorithm 1, the population MCMC move is presented in Algorithm 2, and the details for each step of the genetic algorithm (crossover, mutation, and exchange) are given in the next subsection.

3.1.1. Resample Moves. In the resample move step, equally weighted particles are chosen, and population MCMC is applied. We summarise the population MCMC algorithm as shown in Algorithm 2.

3.1.2. Crossver. In the proposed framework, the state vector is represented as a string of bits. The crossover point l_c is a random point on the string of bits of length l. The crossover operator cannot be applied to all parts of the state vector.

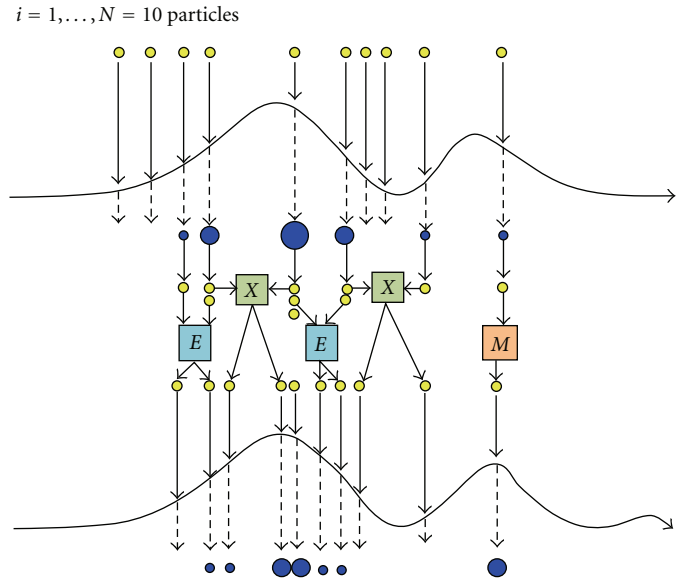


Figure 1: Illustration of the EP MCMC filter.

(1) Initialisation: at $k = 1$ generate initial samples $x_{k,\tau}^{(i)}$ and their respective weights $W_{k,\tau}^{(i)} = 1/N$.
(2) Sampling
  (i) For $k = 2, \ldots$.
  (ii) For each sample $i = 1, \ldots, N$.
    Draw $x_{k,\tau}^{(i)} \sim q_k$, where $x_{k,\tau}^{(i)}$ is the ith sample at kth time instant and q_k is a known transition prior at time instant k.
  (iii) Move the populations of samples by $x_{k,\tau}^{(i)} \sim G_k(x_{k-1,\tau}^{(i)})$, using the genetic transition kernel $G_k(\cdot)$ according to crossover, mutation or exchange steps as described below.
  (iv) Evaluate weights $w_{k,\tau}^{(i)}(x_{k,\tau}^{(i)})$ from the likelihood as presented in the next subsection and normalise the weights to obtain $W_{k,\tau}^{(i)} = w_{k,\tau}^{(i)} / \sum_{i=1}^{N} w_{k,\tau}^{(i)}$.
(3) Resampling
  (i) If $ESS \leq N_{\text{thr}_1}$, where the ESS is measured as $(\sum_{i=1}^{N} (W_{k,\tau}^{(i)})^2)^{-1}$ and N_{thr_1} is some threshold.
    (a) For t iterations
      (a1) Resample using one of the standard resampling techniques such as residual resampling.
      (a2) Perform population MCMC based moves to samples as mentioned in Section 3.1.1.
      (a3) Recompute the weights $w_{k,\tau}^{(i)}(x_{k,\tau}^{(i)})$ and normalise the weights to obtain $W_{k,\tau}^{(i)}$.
    (b) End
  (ii) Increment the time instant $k = k + 1$.
  (iii) Iterate steps (2). and (3).

Algorithm 1: The proposed evolving population Markov chain Monte Carlo filter.

(1) Initialisation
  (i) Select i^* resampled particles from each population $x_{k,\tau}^{(i^*)}$, where $i^* = 1, \ldots, N$.
(2) Iterate steps (2) and (3).
(3) (a) Mutation
    Perform Mutation as illustrated in Section 3.1.3.
  (b) CrossOver or Exchange Move Perform CrossOver or Exchange moves as illustrated in Section 3.1.2 or Section 3.1.4, respectively. Accept the move based on the Metropolis-Hastings rule, for example, from the probability $\min\{1, A\}$ as described in Section 3.1.4.

Algorithm 2: Population MCMC moves.

Some parts of the state vector may not undergo any changes, and thus for such components, the probability of crossover ρ_c is zero. This leaves the crossover being operated only on components that are expected to undergo random changes. The crossover operator functions on two distinguished offsprings (paired particles), for example, $x_{s,\tau}$, $x_{q,\tau}$. The algorithm for crossover is described below:

(i) draw a uniform random number u_c for every component $u_c \approx U(0,1)$, and if $u_c \leq \rho_c$, then perform crossover,

(ii) estimate a crossover point l_c based on a uniform random integer between 1 and the length of the component l,

(iii) generate two offsprings

$$\begin{aligned} x_{s,\tau} &= \left(x_{1s}, x_{2s}, \ldots, x_{(l_c-1)s}, x_{l_c q}, \ldots, x_{dq}\right), \\ x_{q,\tau} &= \left(x_{1q}, x_{2q}, \ldots, x_{(l_c-1)q}, x_{l_c s}, \ldots, x_{ds}\right), \end{aligned} \tag{3}$$

where ds and dq refers to the length of the samples $x_{s,\tau}$ and $x_{q,\tau}$, respectively.

For simplicity in the remaining derivation, we will omit the index τ. For crossover, the proposal distribution $q(\cdot)$ can be expressed as a product of the proposals for the two offsprings

$$q\left(x_k \mid x_{s,k-1}, x_{q,k-1}\right) \sim p\left(x_{s,k-1} \mid z_{k-1}\right) p\left(x_{q,k-1} \mid z_{k-1}\right). \tag{4}$$

Then, in the crossover operation, performed with the two offsprings $x_{s,k}$ and $x_{q,k}$, the particle weight can be expressed in the form

$$w_k^{cr,(i)} = \frac{p\left(z_k \mid x_{s,k}^{(i)}\right) p\left(z_k \mid x_{q,k}^{(i)}\right) p\left(x_{s,k}^{(i)} \mid x_{s,k-1}^{(i)}\right) p\left(x_{q,k}^{(i)} \mid x_{q,k-1}^{(i)}\right)}{q\left(x_{s,k-1}^{(i)} \mid z_{k-1}\right) q\left(x_{q,k-1}^{(i)} \mid z_{k-1}\right)}. \tag{5}$$

Then, the recursive weights can be written as

$$w_k^{cr,(i)} = w_{s,k-1}^{cr,(i)} w_{q,k-1}^{cr,(i)} \mathcal{L}_s\left(z_k, x_{s,k}^{(i)}\right) \mathcal{L}_q\left(z_k, x_{q,k}^{(i)}\right). \tag{6}$$

Here, $\mathcal{L}_s^{(i)}(z_k, x_{s,k}^{(i)}) = p(z_k \mid x_{s,k}^{(i)})$ is the likelihood function for the sth offspring, $\mathcal{L}_q^{(i)}(z_k, x_{q,k}^{(i)}) = p(z_k | x_{q,k}^{(i)})$ is the likelihood function of the qth offspring and $w_{s,k-1}^{cr,(i)}$ and $w_{q,k-1}^{cr,(i)}$ are the weights at $(k-1)$th time instant, for the sth and qth offspring, respectively.

Hence, in the case of the crossover, where there are paired particles with the same weight, we marginalise one of them and express the weights as a function of the proposal PDF of the other PDF.

3.1.3. Mutation. A probability of mutation ρ_m is initially defined for each component. Such a probability is chosen in order to make sure that components that need no stochastic fluctuations could be prohibited from undergoing mutation operation. For such components, the probability of mutation ρ_m is considered zero. The components are assumed as a vector string of binary units. According to the proposed mutation mechanism,

(i) draw a uniform random number u_m for every component $u_m \approx U(0,1)$, and if $u_m \leq \rho_m$, then perform mutation,

(ii) estimate a mutation point l_m based on a uniform random integer between 1 and length of the component l,

(iii) flip the mutation point l_m.

The weights are of the form

$$w_k^{\text{mutation},(i)} = \frac{p\left(x_k^{(i)} \mid x_{k-1}^{(i)}\right)}{Uq\left(x_k^{(i)} \mid x_{k-1}^{(i)}\right)}, \tag{7}$$

where, U is a uniform random number. During mutation, samples that undergo mutation are mutually independent. Therefore, the updated proposal distribution at time k is a factor of the proposal distribution at the previous iteration $k-1$.

3.1.4. Exchange. Consider two independent chains of samples, for example, $x_{s,k}$ and $x_{q,k}$. For their target distributions π_s and π_q, respectively, the swap of information between these two chains can be performed with a Metropolis-Hastings step. The swap occurs with probability $\min\{1, A\}$, where

$$A = \frac{\pi_s\left(x_{q,k}\right) \pi_q\left(x_{s,k}\right)}{\pi_s\left(x_{s,k}\right) \pi_q\left(x_{q,k}\right)}. \tag{8}$$

The genetic transition kernel combines the effectiveness of samples between various populations to create more efficient groups of samples.

3.2. Expected Likelihood Probabilistic Data Association (ELPDA). Since the state vector $x_{k,\tau} = \{x_{k,\tau,\eta}\}_{\eta=1}^{n}$ for target τ consists of the states for all body parts $\eta = 1, \ldots, n$, a data association problem needs to be resolved. In our work, we adopt the expected likelihood data association method from [10]. For the set of available measurements, we assume that one of the measurements originates from the target, and the rest are due to spurious clutter. In the case of tracking the pictorial structure of the human target, colour histograms are used for matching and the corresponding measurement equation is highly nonlinear. The data association problem is considered with respect to the whole pictorial structure (the whole graph) representing the target, for example, with respect to $x_{k,\tau}$.

We adapt the weights of particles $\widetilde{w}_{k,\tau}^{(i)}(x_{k-1:k}^{(i)})$ in the move step of the EPMCMC filter specified in Section 3.1.

Let m_k denote the number of available measurements at time step k. The measurements at time step k are denoted as

$z_k = \{z_k^j\}$, where $j = 0, \ldots, m_k$. If θ denotes any association event, then θ_k^j is a particular association event that assigns the jth measurement to target τ. According to [10], the conditional probability density function $p(\theta_k^j \mid Z_k)$ of the association event θ_k^j that the jth measurement within the validation gate is the measurement that originated from the target is given by

$$p\left(\theta_k^j \mid Z_k\right) = p\left(\theta_k^j \mid z_k, m_k, Z_{k-1}\right), \tag{9}$$

for the set of m_k measurements z_k that fall within the validation gate. We expand the conditional probability from above using the Bayesian rule to get

$$\begin{aligned} &p\left(\theta_k^j \mid z_k, m_k, Z_{k-1}\right) \\ &\quad \propto p\left(z_k \mid \theta_k^j, m_k, Z_{k-1}\right) \times p\left(\theta_k^j \mid m_k, Z_{k-1}\right). \end{aligned} \tag{10}$$

If the incorrect measurements have a uniform probability density function within the gating volume V_τ, and with the assumption of a normal measurement error for a correct measurement, we can rewrite

$$p\left(z_k \mid \theta_k^j, m_k, Z_{k-1}\right) = \begin{cases} {V_\tau}^{-(m_k-1)} P_G^{-1} \mathcal{N}\left(z_k^j; 0; S_k\right), \\ \forall\, j = 1, \ldots, m_k, \\ V_\tau^{-m_k}, \\ \forall\, j = 0, \end{cases} \tag{11}$$

where P_G is the probability of gating, S_k is the covariance matrix of the innovation vector z_k^j. Finally, it has been shown [10] that the probability $p(\theta_k^j \mid z_k, m_k, Z_k)$ of association events can be computed as

$$\begin{aligned} &p\left(\theta_k^j \mid m_k, Z_k\right) \\ &\quad = \begin{cases} \dfrac{1}{C} \times P_G^{(-1)} \\ \quad \times \dfrac{(P_D P_G)\mu_F(m_k - 1)}{m_k P_D P_G \mu_F(m_k - 1) + (1 - P_D P_G)\mu_F(m_k)} \\ \quad \times V_\tau^{-(m_k-1)} \times \mathcal{N}(z_k; 0; S_k), \\ \qquad\qquad\qquad \forall\, j = 1, \ldots, m_k, \\ \dfrac{(1 - P_D P_G)\mu_F(m_k)}{P_D P_G \mu_F(m_k - 1) + (1 - P_D P_G)\mu_F(m_k)}, \\ \qquad\qquad\qquad \forall\, j = 0, \end{cases} \end{aligned} \tag{12}$$

where P_D is the detection probability, μ_F is the probability mass function of the number of incorrect measurements, $P_D P_G$ refers to the probability that a target is detected and its measurements fall within the gate, and R_k is the measurement error covariance matrix.

Finally, we approximate the weights of the particles based on the computed expected likelihood using

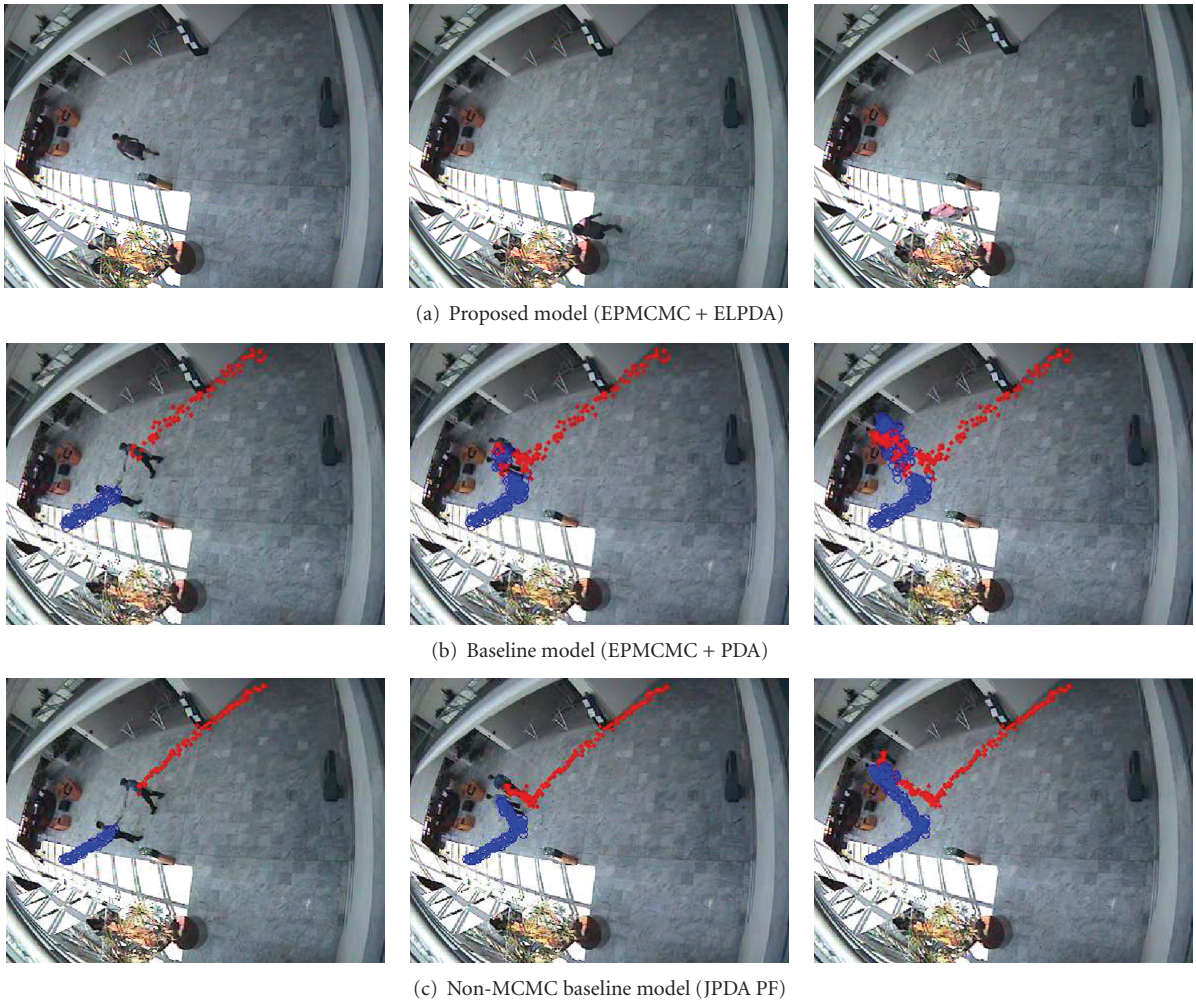
$$\begin{aligned} \widetilde{w}_{k,\tau}^{(i)} \propto \widetilde{w}_{k-1,\tau}^{(i)} &\times \frac{p\left(\theta_k^0 \mid m_k, z_{k-1}\right) V_\tau^{-(m_k)}}{q\left(x_{k,\tau}^{(i)} \mid x_{k-1,\tau}^{(i)}, z_k\right)} \\ &+ \frac{\sum_{j=1}^{m_k} p\left(\theta_k^j \mid m_k, z_{k-1}\right) V_\tau^{-(m_k-1)} P_G^{-1}}{q\left(x_{k,\tau}^{(i)} \mid x_{k-1,\tau}^{(i)}, z_k\right)} \\ &\times \frac{\mathcal{N}\left(z_k, z_k^j, R_k\right) p\left(x_k^{(i)} \mid x_{k-1,\tau}^{(i)}\right)}{q\left(x_{k,\tau}^{(i)} \mid x_{k-1,\tau}^{(i)}, z_k\right)}, \end{aligned} \tag{13}$$

where $i = 1, \ldots, N$.

4. Results

In this section, we perform systematic experiments evaluating (1) the accuracy of our proposed EPMCMC + ELPDA method with the EPMCMC + PDA algorithm and with a generic particle filter framework with a joint probabilistic data association (JPDA PF) proposed in [37, 38] and (2) the influence of the system parameters on the model including the geometric constraints, number of parts being tracked, the radius of the validation gate. We demonstrate our results on videos containing human targets, where the pictorial structure of the target is modeled as a graphical model of the parts of the body. The transition prior is assumed to be a constant velocity model [8] applied jointly with the pictorial structure for each target. The pictorial structure is represented as rectangles for each body part, and the state vector consists of the position and speed for the centre of each body part. To evaluate the performance of the models, we compute the *root mean square error distance* (RMSE) between the estimated center point of every part and its manually labeled counterpart. The results are presented in the form of a cumulative RMSE for all the targets in the video.

4.1. Multiple People Tracking. To track multiple people (together with their multiple body parts) in video, we use the CAVIAR [39] data set sequences. The CAVIAR dataset contains over 80 video sequences with one or more targets moving in real-time scenarios. We have chosen 11 videos from the data set at varying levels of complexity in terms of the motion characteristics, occlusion, and clutter (due to illumination changes). We have further divided the 11 selected sequences in 17 short clips with varying number of targets. Table 1 elaborates on the chosen sequences and the clips that have been extracted from these sequences along with the details of the frames, number of targets, the level of occlusion, and the presence of clutter. The tracked people have been manually annotated in these clips using a maximum 10-point model consisting of the centers of the head (1 point), torso (1 point), arms (2 points for each arm), and legs (2 points for each leg). We would like to particular highlight that the frames are of resolution



(a) Proposed model (EPMCMC + ELPDA)

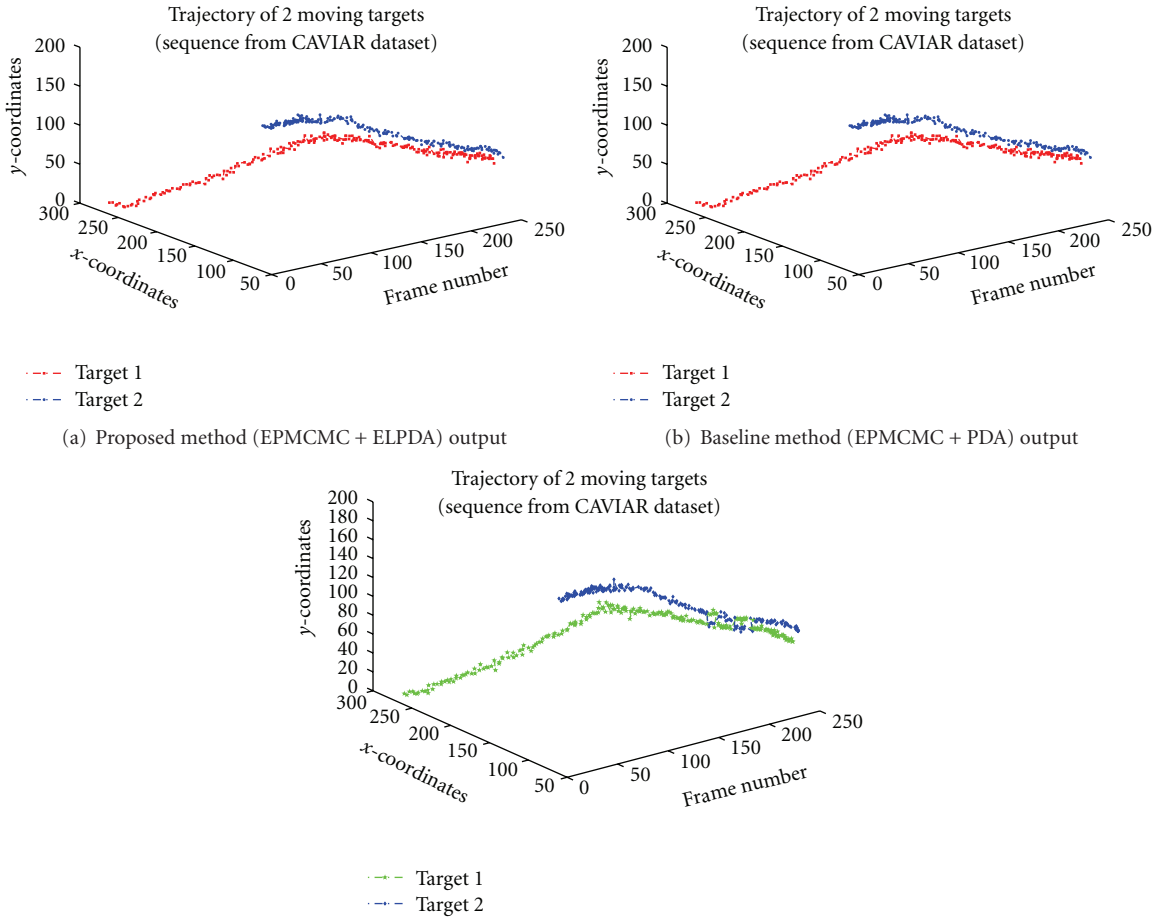
(b) Baseline model (EPMCMC + PDA)

(c) Non-MCMC baseline model (JPDA PF)

Figure 2: Sample video frames of the tracked body torso of two targets (trajectories presented in Figure 3).

384 × 288 captured at 25 frames per second. The scales of the human targets are fairly small in comparison with the resolution of the image, and thus in some of the images, a 10-point model gives a dense set of labeled ground-truth which is unusable; therefore, in such sequences, we use a 6-point model. We use the method of [7, 40] to initialise the multiple body parts composed multiple targets in first frame of each video using the pictorial structure model. For all our experiment described below unless mentioned otherwise, the number of Monte Carlo cycles is fixed to 100, the number of particles used is 500, and the size of the validation gate is 15. All our experiments are conducted on an Intel Duo Core processor with 3 GB RAM.

4.2. Comparison of Proposed with Related Techniques. One of the main novelties of the proposed EPMCMC + ELPDA method is its integration of expected likelihood into the weights of EPMCMC. The following experiments compares the accuracy of the proposed algorithm with the combined EPMCMC and PDA filter. We present the trajectories of two targets tracked by both techniques in Figure 3 and sample image frames from the corresponding tracked torso of the two targets in Figure 2. We also present some sample image frames of tracked torso's of multiple targets (four targets) from a different video sequence in Figure 4. The combined RMSE curves (in Figures 5 and 6) indicates that our technique has better accuracy than the baseline method. The mean combined RMSE values recorded for increasing Monte Carlo cycles are 2.5369 and 4.2667 for target 1 and 3.2547 and 4.3737 for target 2 using the proposed and baseline strategies, respectively. In terms of the computational demand, our algorithm takes 3419 msec on an average per image frame as against the combined EPMCMC and PDA methods that take 2169 msec (implemented in MATLAB). This is mainly due to the computation of the expected likelihood of all measurements that fall within the validation gate. We have also found that the cumulative error in the localisation of the target reduces by approximately 39% for an increased 1.2 seconds in the processing time. We presume that the tradeoff between the time taken for processing as against the accuracy of our technique acceptably good. A detailed analysis of the computation time per image frame across various different video clips is summarised in Table 1.



(a) Proposed method (EPMCMC + ELPDA) output

(b) Baseline method (EPMCMC + PDA) output

(c) Non-MCMC baseline method (JPDA PF) output

Figure 3: Trajectory of multiple targets (a) using proposed method (red/blue), (b) using baseline method (magenta/black) and (c) using Non-MCMC baseline method (green/blue), the JPDA PF proposed in [37, 38].

(a) (b) (c)

Figure 4: Sample video frames of the tracked torso of multiple targets.

4.3. Effect of Changing System Parameters. We study the effect of different changing system parameters and compare them between the proposed EPMCMC + ELPDA and the EPMCMC + PDA model. We do not compare our results to the generic particle filter framework, as these parameters are particularly relevant to the parts-based models. First, we examine the effect of increasing the number of targets being tracked on the performance of the models. Figure 7 suggests that increasing the number of targets from 1 to 2 decreases the accuracy of the proposed EPMCMC + ELPDA model, but increasing the targets beyond this does not significantly alter its accuracy.

Secondly, we investigate the effect of increasing the number of parts of each target that is used for tracking. In the lower levels of the model, we have a smaller collection of more salient parts representing the target followed by

Table 1: Tabular description of the chosen video clips (B2-Browse2, BWW1-BrowseWhileWhalking1, FC-Fight Chase, FR1-FightRunway1, LB-LeftBag, LBx-LeftBox, MC-MeetCrowd, MWS-MeetWalkSplit, MWT1-MeetWalkTogether1, RFF-RestFallenFloor, WBS1-Walk and ByShop1) and comparison of combined RMSE between EPMCMC + ELPDA model (Proposed), EPMCMC + PDA model (Baseline) and RMSE of *Torso alone* in generic particle filter framework (JPDA PF).

Video	Frames	Targets	Occlude	Clutter	Proposed	Baseline	JPDAPF	Time
B2	173	1	No	0	1.1821	2.0413	0.9648	3187
	198	1	Self	0	1.6669	2.3425	1.1578	3276
BWW1	361	1	Self	1	2.1638	2.6804	1.5673	3106
FC	129	1	No	0	0.5187	0.9835	0.2784	3016
	178	2	Yes	0	2.7861	4.0076	3.1279	3127
FR1	199	2	Yes	0	2.7883	3.9801	1.4531	3229
	201	3	Self	0	3.4412	7.6359	5.8763	3663
LB	401	3	Partial	1	3.6567	8.0015	5.1455	3841
	208	1	No	0	1.4009	2.2156	1.0151	2923
	146	1	Self	0	1.9768	3.8970	2.1412	3312
LBx	373	2	Partial	0	2.8990	5.7682	4.2349	3401
	370	1	Yes	1	2.5238	3.3532	2.0451	3198
MC	282	4	Partial	0	3.8982	6.4235	6.7347	3789
MWS	301	2	Partial	1	2.0679	4.2176	3.9567	3128
MWT1	323	2	Partial	0	2.5003	3.9134	2.2734	3215
RFF	388	1	Self	0	2.3027	4.5231	1.7235	3309
WBS1	876	5	Yes	0	3.5694	9.2722	7.0163	4789

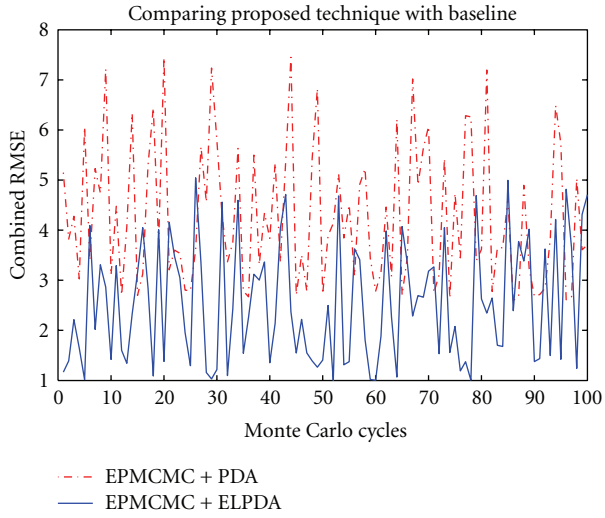


Figure 5: Combined RMSE versus the number of Monte Carlo cycles for target 1.

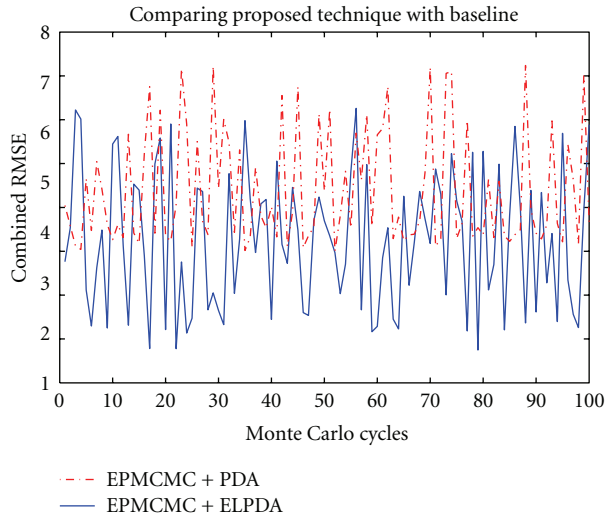


Figure 6: Combined RMSE versus the number of Monte Carlo cycles for target 2.

increasing number of lesser important parts. In the following experiment, we examine the impact of increasing the number of parts representing the target by comparing the model with level one (1 part: torso), level two (2 parts: head and torso), level three (6 parts: head, torso, two hands, and two legs), and level four (10 parts-head, torso, four for hands, and four for legs). The plot in Figure 8 suggests that increasing the representation of the target (using a larger subset of parts) increases accuracy.

In Figure 9, we present the results showing the effect of increasing the size of the validation gate against the accuracy of EPMCMC + ELPDA technique. The cumulative RMSE curves demonstrating the effect of increasing the size of the validation gate indicate that with increase gate size, accuracy increases, but beyond this, it does not improve the performance further.

Finally, in Figure 10, we demonstrate the effect of geometric constraints being enforced in the measurement of the likelihood against the accuracy of EPMCMC + ELPDA technique. The cumulative RMSE curves prove beyond doubt that with appropriate geometrical constraints on the pictorial structure model accuracy increases.

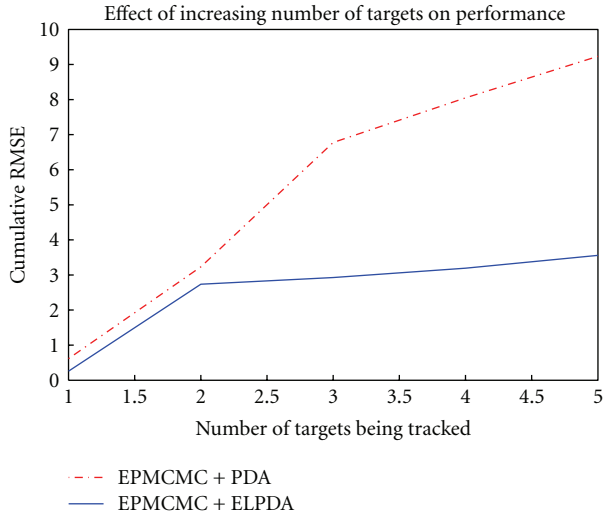


Figure 7: Cumulative RMSE versus the number of targets in the Video.

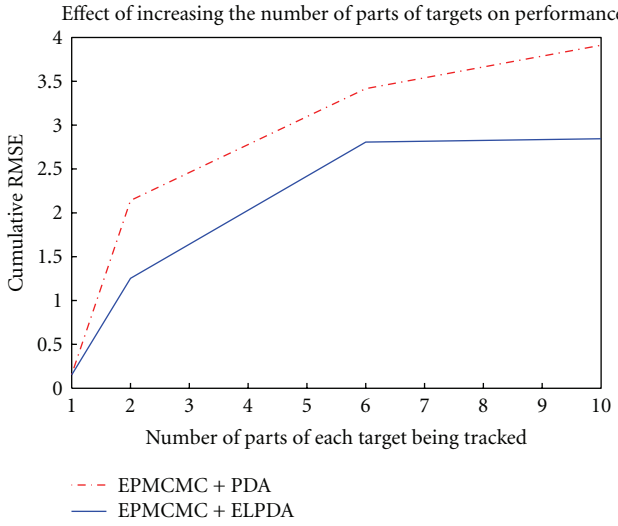


Figure 8: Cumulative RMSE versus the number of parts representing the target.

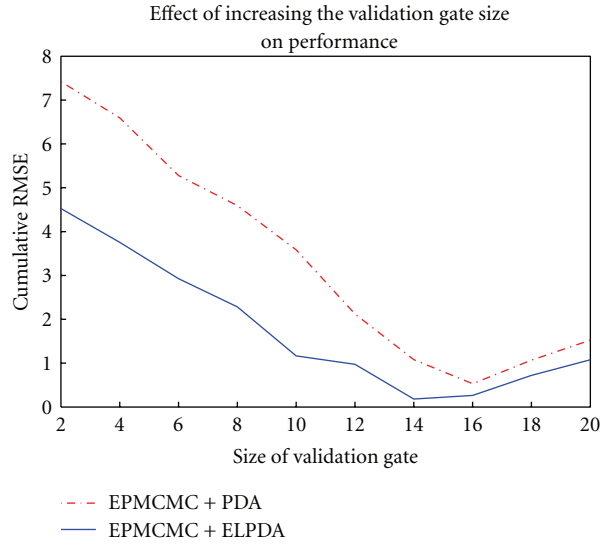


Figure 9: Cumulative RMSE versus the size of the validation gate.

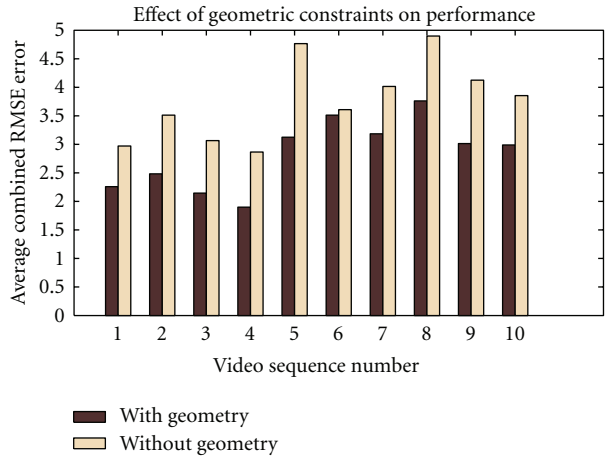
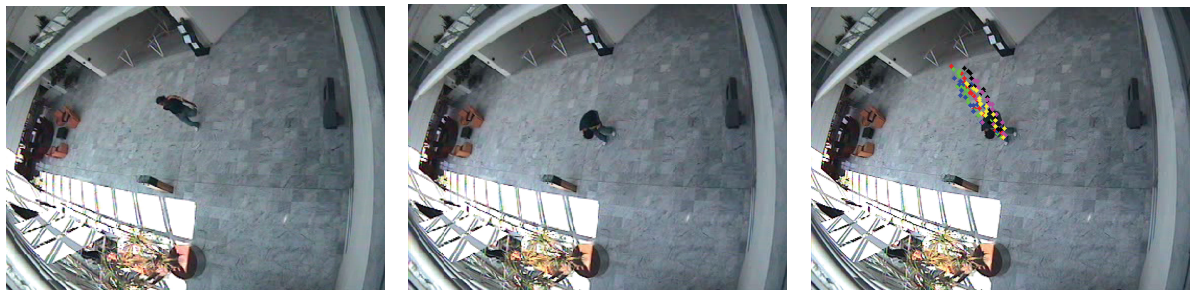


Figure 10: Effect of geometric constraints on combined RMSE for different video sequences (yellow without constraints/brown with constraints).

4.4. Failure Modes. In this subsection, we highlight some of common failure modes of the proposed framework. We attribute the missed detections of our proposed framework to three main conditions. First, when the object(s) of interest are subject to a high degree of self occlusion due to articulation of body parts, our framework fails to match the pictorial structure well to the image data and subsequently fail during tracking. The illustrations in Figure 11(a) show a scenario from our test set, where tracking multiple body parts accurately was highly difficult due to the aforementioned conditions. In such test sequences, we have resorted to models containing lower number of distinguishable body parts. In Figure 11(b), we highlight the second mode of failures that are mainly caused due to rapid changes in illumination conditions. Finally, we also describe conditions such as the image acquisition perspective and scale of the object(s) of interest as other causes for missed detections in Figure 11(c).

5. Conclusion

We have proposed an innovative method for combining extended likelihood data association with evolving population particle filtering for robust and accurate multiple target tracking. The evolving population filter introduces variety in the population of particles by combining them in both the sampling and resampling steps using constrained genetic operations. The extended likelihood data association filters those measurements that belong to the target from a clutter of other noisy measurements analysed within the validation gate. System parameters such



(a) Complex occlusion of body parts of a single target

(b) Object navigating in areas with rapid illumination changes

(c) Changing the image acquisition perspective and small object scale

Figure 11: Failure modes of the proposed EPMCMC + ELPDA model: (a) complex occlusion example; (b) rapid change in illumination conditions; (c) image acquisition perspectives.

the radius of validation gate are reestimated during each iteration, rather than fixed empirically, resulting in a model that outperforms similar recent methods on standard datasets.

Acknowledgments

The authors acknowledge the support of UK MOD Data and Information Fusion Defence Technology Centre under the Tracking Cluster Project no. DIFDTC/CSIPC1/02. The authors would also like to sincerely thank Professor Simon Godsill for his useful advices and discussions on sequential population Monte Carlo methods. They acknowledge the support from the (European Community's) Seventh Framework Programme (FP7/2007–2013) under Grant no. 238710 (Monte Carlo-based Innovative Management and Processing for an Unrivalled Leap in Sensor Exploitation).

References

[1] J. K. Aggarwal and Q. Cai, "Human motion analysis: a review," *Computer Vision and Image Understanding*, vol. 73, no. 3, pp. 428–440, 1999.

[2] D. A. Forsyth, O. Arikan, L. Ikemoto, J. O'Brien, and D. Ramanan, "Computational studies of human motion: part 1, tracking and motion synthesis," *Foundations and Trends in Computer Graphics and Vision*, vol. 1, no. 2-3, pp. 77–254, 2006.

[3] D. M. Gavrila, "The visual analysis of human movement: a survey," *Computer Vision and Image Understanding*, vol. 73, no. 1, pp. 82–98, 1999.

[4] T. B. Moeslund and E. Granum, "A survey of computer vision-based human motion capture," *Computer Vision and Image Understanding*, vol. 81, no. 3, pp. 231–268, 2001.

[5] M. Andriluka, S. Roth, and B. Schiele, "People-tracking-by-detection and people-detection-by-tracking," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '08)*, June 2008.

[6] P. F. Felzenszwalb and D. P. Huttenlocher, "Pictorial structures for object recognition," *International Journal of Computer Vision*, vol. 61, no. 1, pp. 55–79, 2005.

[7] D. Ramanan, D. A. Forsyth, and A. Zisserman, "Tracking people by learning their appearance," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 1, pp. 65–81, 2007.

[8] Y. Bar-Shalom and X. R. Li, *Estimation and Tracking: Principles,Techniques and Software*, Artech House, Norwood, Mass, USA, 1993.

[9] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter: ParticleFilter for Tracking Applications*, Artech House, London, UK, 2004.

[10] A. Marrs, S. Maskell, and Y. Bar-Shalom, "Expected likelihood for tracking in clutter with particle filters," in *Signal and data Processing of Small Targets*, vol. 4728 of *Proceedings of SPIE*, pp. 230–239, April 2002.

[11] R. Cucchiara, C. Grana, M. Piccardi, and A. Prati, "Detecting moving objects, ghosts, and shadows in video streams," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 10, pp. 1337–1342, 2003.

[12] D. Gerónimo, A. M. López, A. D. Sappa, and T. Graf, "Survey of pedestrian detection for advanced driver assistance systems," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 7, pp. 1239–1258, 2010.

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

[14] S. Zhou, R. Chellappa, and B. Moghaddam, "Adaptive visual tracking and recognition using particle filters," in *Proceedings of the IEEE International Conference on Multimedia & Expo (ICME '03)*, pp. 560–566, July 2003.

[15] T. Gandhi and M. M. Trivedi, "Pedestrian protection systems: issues, survey, and challenges," *IEEE Transactions on Intelligent Transportation Systems*, vol. 8, no. 3, pp. 413–430, 2007.

[16] S. Jung and K. Wohn, "Tracking and motion estimation of the articulated object: a hierarchical Kalman filter approach," *Real-Time Imaging*, vol. 3, no. 6, pp. 415–432, 1997.

[17] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter: Particle Filter for Tracking Applications*, vol. 2, Artech House, Norwood, Mass, USA, 2004.

[18] M. Briers, S. Maskell, and R. Wright, "A Rao-Blackwellised unscented Kalman filter," in *Proceedings of the 6th International Conference on Information Fusion*, pp. 55–61, ISIF, Queensland, Australia, 2003.

[19] D. Bizup and D. Brown, "The over-extended Kalman filter—use it!," in *Proceedings of the 6th International Conference on Information Fusion*, pp. 40–46, ISIF, Queensland, Australia, 2003.

[20] D. Schultz, D. Fox, and J. Hightower, "People tracking with anonimous and ID-sensors using Rao-Blackwellised particle filters," in *Proceedings of the International Conference on Artificial Intelligence (IJCAI '03)*, 2003.

[21] E. Wan and R. van der Merwe, "The unscented Kalman filter ," in *Kalman Filtering and Neural Networks*, S. Haykin, Ed., chapter 7, pp. 221–280, John Wiley & Sons, New York, NY, USA, 2001.

[22] O. Cappé, A. Guillin, J. M. Marin, and C. P. Robert, "Population monte carlo," *Journal of Computational and Graphical Statistics*, vol. 13, no. 4, pp. 907–929, 2004.

[23] Y. Iba and S. Coffa, "Population-based monte carlo algorithms," *Journal of Computational and Graphical Statistics*, vol. 13, no. 4, pp. 157–1933, 2000.

[24] A. Jasra, D. A. Stephens, and C. C. Holmes, "On population-based simulation for static inference," *Statistics and Computing*, vol. 17, no. 3, pp. 263–279, 2007.

[25] O. Cappe, R. Douc, and E. Moulines, "Comparison of resampling schemes for particle filtering," in *Proceedings of the 4th International Symposiumon Image and Signal Processing and Analysis (ISPA '05)*, Croatia, 2005.

[26] J. Hol, T. Shön, and F. Gustaffsson, "On resampling algorithms for particlefilters," in *Proceedings of the Nonlinear Statistical Signal Processing Workshop*, Cambridge, UK, September, 2006.

[27] S. Blackman and R. Popoli, *Design and Analysis of Modern Tracking Systems*, Artech House Radar Library, 1999.

[28] D. Salmond, D. Fisher, and N. Gordon, "Tracking and identification forclosely spaced objects in clutter," in *Proceedings of the European Control Conference*, IEEE, Brussels, Belgium, July 1997.

[29] T. Kirubarajan and Y. Bar-Shalom, "Probabilistic data association techniques for target tracking in clutter," *Proceedings of the IEEE*, vol. 92, no. 3, pp. 536–556, 2004.

[30] X. R. Li, "Engineer's guide to variable-structure multiple-model estimation for tracking," in *Multitarget-Multisensor Tracking: Applications and Advances*, Y. Bar-Shalom and W. D. Blair, Eds., vol. 3, chapter 3, pp. 499–567, Artech House, Norwood, Mass, USA, 2002.

[31] S. Maskell, M. Rollason, N. Gordon, and D. Salmond, "Efficient particle filtering for multiple target tracking with application to tracking in structured images," *Image and Vision Computing*, vol. 21, no. 10, pp. 931–939, 2003.

[32] Y. Bar-Shalom and W. Dale Blair, *Multitarget-Multisensor Tracking: Applications and Advances*, vol. 3, Artech House, Norwood, Mass, USA, 2000.

[33] M. Briers, S. Maskell, and M. Philpott, "Two-dimensional assignment with merged measurements using Lagrangian relaxation," in *Signal Processing of Small Targets*, Proceedings of SPIE, pp. 283–292, 2003.

[34] P. Horridge and S. Maskell, "Real-time tracking of hundreds of targets with efficient exact JPDAF implementation," in *Proceedings of International Conference on Information Fusion*, 2006.

[35] S. Maskell, M. Briers, and R. Wright, "Fast mutual exclusion," in *Signal Processing of Small Targets*, Proceedings of SPIE, 2004.

[36] L. Y. Pao, "Multisensor multitarget mixture reduction algorithms for tracking," *Journal of Guidance, Control, and Dynamics*, vol. 17, no. 6, pp. 1205–1211, 1994.

[37] M. H. Jaward, L. Mihaylova, N. Canagarajah, and D. Bull, "A data association algorithm for multiple object tracking in video sequences," in *Proceedings of the IEE Seminar on Target Tracking: Algorithms and Applications*, pp. 131–136, Birmingham, UK, 2006.

[38] M. H. Jaward, L. Mihaylova, N. Canagarajah, and D. Bull, "Multiple objectstracking using particle filters in video sequences," in *Proceedings of the IEEE Aerospace Conference*, Big Sky, Mont, USA, 2006.

[39] CAVIAR test case scenarios, 2005,http://homepages.inf.ed.ac.uk/rbf/.

[40] D. Ramanan and D. A. Forsyth, "Finding and tracking people from the bottom up," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CCVPR '03)*, pp. 467–474, June 2003.