Social Force Model-Based MCMC-OCSVM Particle PHD Filter for Multiple Human Tracking

Pengming Feng, Wenwu Wang, *Senior Member, IEEE*, Satnam Dlay, *Member, IEEE*, Syed Mohsen Naqvi, *Senior Member, IEEE*, and Jonathon Chambers, *Fellow, IEEE*

***Abstract*—Video-based multiple human tracking often involves several challenges, including target number variation, object occlusions, and noise corruption in sensor measurements. In this paper, we propose a novel method to address these challenges based on probability hypothesis density (PHD) filtering with a Markov chain Monte Carlo (MCMC) implementation. More specifically, a novel social force model (SFM) for describing the interaction between the targets is used to calculate the likelihood within the MCMC resampling step in the prediction step of the PHD filter, and a one class support vector machine (OCSVM) is then used in the update step to mitigate the noise in the measurements, where the SVM is trained with features from both color and oriented gradient histograms. The proposed method is evaluated and compared with state-of-the-art techniques using sequences from the CAVIAR, TUD, and PETS2009 datasets based on the mean Euclidean tracking error on each frame, the optimal subpattern assignment metric, and the multiple object tracking precision metric. The results show improved performance of the proposed method over the baseline algorithms, including the traditional particle PHD filtering method, the traditional SFM-based particle filtering method, multi-Bernoulli filtering, and an online-learning-based tracking method.**

***Index Terms*—Multiple human tracking, Markov chain Monte Carlo (MCMC) resampling, one class support vector machine (OCSVM), probable hypothesis density (PHD) filter, social force model.**

I. Introduction

VIDEO based multiple human tracking plays an important role in many applications such as surveillance, guidance, and homeland security, especially in enclosed environments such as an airport, campus or shopping mall. Tracking multiple human targets in the above situations presents several challenges including varying number of targets, object occlusion, and the adverse effect of environmental noise within measurements [1], [2]. Moreover, it is not always possible to associate measurements with particular targets which results in false alarms and missed detections [3]. In this work we attempt to address aspects of these challenges and focus on the problem of estimating the position of an unknown number of human targets, based on noisy observations, with the possible presence of missed detections and false alarms due to clutter.

Manuscript received June 23, 2016; revised October 24, 2016; accepted December 4, 2016. Date of publication December 9, 2016; date of current version March 15, 2017. This work was supported in part by the Engineering and Physical Sciences Research Council under Grant EP/K014307, and in part by the MOD University Defence Research Collaboration in Signal Processing. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Changsheng Xu.

P. Feng, S. Dlay, S. M. Naqvi, and J. Chambers are with the Communications, Sensors, Signal and Information Processing Group, School of Electrical and Electronic Engineering, Newcastle University, Newcastle upon Tyne NE1 7RU, U.K. (e-mail: P.Feng2@newcastle.ac.uk; Mohsen.Naqvi@newcastle.ac.uk; Satnam.Dlay@newcastle.ac.uk; Jonathon.Chambers@newcastle.ac.uk).

W. Wang is with the Center for Vision Speech and Signal Processing, Department of Electrical and Electronic Engineering, University of Surrey, Surrey GU2 7XH, U.K. (e-mail: w.wang@surrey.ac.uk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TMM.2016.2638206

A. Related Work

Most multiple human tracking approaches fall into one of three categories: achieving a more accurate dynamic model for prediction such as using an interaction model when predicting the position and velocity of a target [4], [5]; generating more stable recursive mathematical models such as unscented Kalman filters and MCMC particle filters [6]; and searching for more accurate measurement models [7], [8], for example, the tracking-learning-detection (TLD) [9] approach. However, the almost universally accepted mathematical framework used to describe multiple target tracking is that of filtering theory and, in particular, Bayesian filtering [10], where the posterior probability distribution is recursively predicted by propagating this distribution with the state model, which describes the motion of a target, and updates the state when a new observation becomes available [11]. The two such popular algorithms are the Kalman filter and particle filter. The Kalman filter is well-known to be optimal with the Gaussian linear model, while the particle filter can be employed to address the tracking problem with a non-Gaussian and non-linear model. These two methods can address the basic multiple target tracking problem when the number of targets is assumed to be known and fixed, however, they are not designed for dealing with the problem of a variable number of targets.

The random finite set (RFS) [12] probability hypothesis density (PHD) filter has been recently proposed to deal with the problem of tracking a varying number of targets. The advantage of the PHD filter is that it can estimate both the cardinality of targets and their states, and thus avoids the need for data association techniques as part of the multiple target framework [13], [14]. Moreover, it mitigates the computational complexity issue as often occurs in other multiple target tracking approaches such as multiple hypothesis tracking (MHT) [14] since it simply utilizes the first-order moment of the multi-target posterior rather than the posterior itself. The PHD filter will be used here and is the focus of this paper.

However, the tracking accuracy of the PHD filter is compromised due to the first order approximation of the RFS in both

the prediction and update steps, as well as noise in the measurements [15]. To address these limitations, we propose two new methods. First, the prediction of the states of the human targets is improved by considering the dynamics of the targets and the interactions between them via an interaction model [5]. This is different from conventional Bayesian filtering, where a fixed model is often used to predict the state of the target and then random noise is employed to generate new candidate states. Second, a background subtraction algorithm is used to generate the measurements for new born targets (detailed later in Section III-D.1), and then a classifier is employed to distinguish the clutter from the human targets based upon their different spatial characteristics, so that the measurement noise is mitigated effectively and the probability of false alarms and missed detections is thereby reduced.

In our work, the interaction information is exploited by a model constraining the motion between the targets. Within existing interaction models, the social force model (SFM) [4] is considered due to its ability in handling the interaction between human targets as well as their typical behaviour. Several researchers have used social force models to predict the states of humans based on their behaviour [4], [16]. Within the social force model, the behaviour of human targets is modelled via energy potentials which are adjusted by other targets and obstacles through repulsive forces [17]. We propose a product model suitable for likelihood calculation to characterise the social force interactions between the human targets based on multiple exponential-terms. This improves the MCMC resampling [18] of the PHD filter used in our work. To the best of our knowledge, no previous work has considered the interaction model in the prediction stage of MCMC-based PHD filtering algorithms. To compare the social force models for human tracking in [4], [16] with that proposed in this paper we consider their similarities, differences, advantages and disadvantages. In all these approaches, interaction forces between the targets are used to simulate the dynamic model of pedestrians; distance and angle between the targets, change of velocity and destination of individual targets are considered in the models; and the models are designed to use domain knowledge and thereby improve the performance of multiple human tracking. The major differences are that in [4] a sum function is used to combine the components of the social force model, but in [16] an exponential-term model is used and, as in (21) in Section II, a summation form is used to combine the model parameters. Our approach instead uses a product function as described by (27) and the influence of each model parameter is controlled by the variance terms in the exponential models. Moreover, a threshold is introduced to avoid calculating social forces when two targets are a large distance apart. The model in [4] has the advantage that it is simple, but the model in [16] is more flexible offering better tracking accuracy. Equation (21) also allows the influence from different model components to be more easily matched to different environments. Our model offers further improvement in accuracy in more complicated environments and these are demonstrated in Table IV in the simulation section. Finally, in terms of disadvantages, the approach in [4] has much less flexibility for use in different environments than our approach and that in [16].

B. Summary of Contributions

Our novel contributions include:

1) A new model is used to describe the social forces between targets to calculate likelihood values in the prediction stage of the particle PHD filter.
2) An MCMC resampling step is exploited to improve the prediction part of the particle PHD filter.
3) An OCSVM classifier which is trained by features from both color and oriented gradient histograms to mitigate measurement noise in background subtraction results, thereby further reducing the probability of false alarms and hence improving the performance of the PHD filter.

The remainder of the paper is structured as follows: In Section II, the background preliminaries of the traditional particle PHD filter and social force model are introduced, then the complete proposed multiple human tracking system framework is described in Section III, including the novel social force model, the MCMC resampling step and the OCSVM classifier; results and comparisons between the proposed approach and baseline methods are presented in Section IV. Finally, Section V provides a short conclusion and a discussion about the possible directions for future work.

II. Background and Preliminaries

A. Adapted Particle PHD Filter

In order to track a variable number of targets, the PHD filter is employed because of its relatively low computational complexity and good tracking performance. There are typically two forms of implementation of PHD filtering. One is based on numerical solutions [19] of the integrals in the prediction and measurement updating step of the PHD filter. The other is the Sequential Monte Carlo (SMC) approach based on the particle PHD filter. Here we use the SMC-PHD method because it performs well in the scenarios of non-Gaussian noise and nonlinear models; besides, it offers the flexibility to incorporate an OCSVM classifier to improve the update of the PHD filter in the presence of noisy measurements.

Assuming the set $\{\mathbf{x}_k^m\}_{m=1}^{m=M_k}$ includes the states of all the human targets, where $\mathbf{x}_k^m = [p_{k,x}^m, p_{k,y}^m, v_{k,x}^m, v_{k,y}^m, h_k^m, w_k^m]^T \in \mathbb{R}^6$ denotes the state of the mth target at discrete time k, including the 2D position $(p_{k,x}^m, p_{k,y}^m)$, velocity $(v_{k,x}^m, v_{k,y}^m)$, height and width of targets h_k^m, w_k^m; where $(\cdot)^T$ denotes the transpose operator and subscripts x, y are the horizontal and vertical coordinates of the target; M_k is the number of targets at time k. Denote the measurement set at time k as $\mathbf{Z}_k$, which includes $\mathbf{z}_k$ for each target. The basic principle of importance sampling in the particle filter is to represent a PDF $p(\mathbf{X}_k|\mathbf{Z}_k)$ by a set of random particles $\mathbf{x}_k^{m,i}$ having associated weights $w_k^{m,i}$, where $\mathbf{X}_k = \{\mathbf{x}_k^{m,i}, m = 1, ..., M_k, i = 1, ..., N\}$, which denotes all the particles utilized to describe the states of all human targets at time k, where N is the number of particles employed to describe the state of a target; in this paper, the notation with superscript $(\cdot)^{m,i}$ denotes a particle with index i employed to describe the state of the target with index m. Given a set of targets with states at time $k-1$, $\{\mathbf{x}_{k-1}^m\}_{m=1}^{m=M_{k-1}}$, the set of predicted particles and the associated weights from the state model

at time k is given by [10]

$$\left\{\tilde{\mathbf{x}}_k^{m,i}, w_{k-1}^{m,i}\right\}_{m=1,i=1}^{m=M_k,i=N}. \tag{1}$$

The particles are independently drawn from importance sampling density $q(\cdot)$ [10] to represent the prior distribution for the next step $p(\mathbf{X}_k|\mathbf{Z}_{k-1})$ [20], where "~" denotes the value resulting from the estimation step. Then the weights for the particles are calculated as [21]

$$w_k^{m,i} \propto w_{k-1}^{m,i} \frac{p(\mathbf{z}_k|\tilde{\mathbf{x}}_k^{m,i}) p(\tilde{\mathbf{x}}_k^{m,i}|\mathbf{x}_{k-1}^{m,i})}{q(\tilde{\mathbf{x}}_k^{m,i}|\mathbf{x}_{k-1}^{m,i}, \mathbf{z}_k)} \tag{2}$$

and by using the suboptimal choice of the particle sampling described in [22], the importance sampling density function $q(\cdot)$ is chosen as

$$q(\tilde{\mathbf{x}}_k^{m,i}|\mathbf{x}_{k-1}^{m,i}, \mathbf{z}_k) = p(\tilde{\mathbf{x}}_k^{m,i}|\mathbf{x}_{k-1}^{m,i}) \tag{3}$$

so the weights for each particle are calculated as [22]

$$w_k^{m,i} \propto w_{k-1}^{m,i} p(\mathbf{z}_k|\tilde{\mathbf{x}}_k^{m,i}) \tag{4}$$

thus the posterior distribution for a single target $p(\mathbf{x}_k^m|\mathbf{Z}_k)$ can be approximated as

$$p(\mathbf{x}_k^m|\mathbf{Z}_k) \approx \sum_{i=1}^{N} w_k^{m,i} \delta(\mathbf{x}_k^m - \tilde{\mathbf{x}}_k^{m,i}) \tag{5}$$

where $\delta(\cdot)$ denotes the Dirac delta measure and the posterior distribution for all targets $p(\mathbf{X}_k|\mathbf{Z}_k)$ can be calculated as $p(\mathbf{x}_k^m|\mathbf{Z}_k)_{m=1}^{M_k}$. In this way, the traditional particle filter is obtained, which can be utilized to implement the PHD filter.

To formulate the PHD filter, the RFS framework is employed. Assuming the particles for the PHD filter are independently drawn from the PDF $p(\mathbf{X}_{k-1}|\mathbf{Z}_{k-1})$, the resulting particles are employed to describe the states of M_k targets at time k, which are approximately distributed as $p(\mathbf{X}_k|\mathbf{Z}_k)$ [19]. In this case, the proposed filter is an approximation of the relationship between the prediction and updating step of the filter. Denoting $D(\cdot)$ as the PHD at a discrete time associated with the multi-target posterior density, the prediction and updating steps for the particle PHD filter can be described as follows:

1) Prediction: Particles $\tilde{\mathbf{x}}_k^{m,i}$ are drawn from the predicted particle set as (1) and fed into the prediction model of the particle PHD filter, which is described as [23]

$$D(\mathbf{X}_k|\mathbf{Z}_{k-1}) = \int \phi(\tilde{\mathbf{x}}_k^{m,i}|\mathbf{X}_k) D(\mathbf{X}_{k-1}|\mathbf{Z}_{k-1}) \delta\tilde{\mathbf{x}}_k^{m,i} + \Upsilon_k(\mathbf{X}_k) \tag{6}$$

where Υ_k is the intensity function of the new target birth RFS, $\phi(\tilde{\mathbf{x}}_k^{m,i})$ is the analogue of the state transition probability in the single target case which is calculated from

$$\phi(\tilde{\mathbf{x}}_k^{m,i}|\mathbf{X}_k) = e(\tilde{\mathbf{x}}_k^{m,i}|\mathbf{X}_k) + \beta(\tilde{\mathbf{x}}_k^{m,i}|\mathbf{X}_k) \tag{7}$$

in which $e(\cdot)$ is the probability that the target still exists at time k and $\beta(\cdot)$ is the intensity of the RFS for spawned targets. When exploiting the PHD filter with the particle filter, the PHD of states is represented by the weights of the particles, which include the survived particles and new-born particles.

In the traditional particle PHD filter, the particles employed to describe the new-born targets are selected randomly in the scene, however, in human tracking, the new-born targets can be obtained by employing a background subtraction step, as proposed in Section III-D.1 to be described later. In this case, assuming at time k, J_k new-born targets are obtained from the background subtraction, the initial weights assigned to the new born particles, which are employed to represent the new-born targets, are calculated as

$$\tilde{w}_k^{b,i} = \frac{1}{J_k \times N} \tag{8}$$

where $i = 1, ..., J_k \times N$ index the particles utilized to represent the new-born targets, then the weights are fed into (6). Since in a later step, the likelihood of all particles and the weights are calculated in the same way, for convenience, the new-born particles are added to the survived particles. With this method, we can obtain a particle set, which includes particles for both survived targets and new born targets

$$\{\tilde{\mathbf{x}}_k^i, \tilde{w}_{k|k-1}^i\}_{i=1}^{i=(M_{k-1}+J_k)\times N} \tag{9}$$

where i is the index of the i-th particle. The weights obtained from the prediction step are given as

$$\tilde{w}_{k|k-1}^i = \begin{cases} \phi(\tilde{\mathbf{x}}_k^i) w_{k-1}^i \; i = 1, ..., M_{k-1} \times N \\ \frac{\Upsilon_k}{J_k \times N} \; i = M_{k-1} \times N + 1, ..., (M_{k-1} + J_k) \times N. \end{cases} \tag{10}$$

In this way, the predicted PHD $D(\mathbf{X}_k|\mathbf{Z}_{k-1})$ at time k for target states $\mathbf{X}_k$ is obtained based on the weights of the particles.

2) Measurement Update: The update step of the PHD filter is defined as [23]:

$$D(\mathbf{X}_k|\mathbf{Z}_k) = D(\mathbf{X}_k|\mathbf{Z}_{k-1}) \times \left[p_M(\tilde{\mathbf{x}}_k^i) + \sum_{\forall \mathbf{z}_k \in \mathbf{Z}_k} \frac{\psi_{k,\mathbf{z}_k}(\tilde{\mathbf{x}}_k^i)}{\kappa_k + \langle \psi_{k,\mathbf{z}_k}(\tilde{\mathbf{x}}_k^i), D(\mathbf{X}_k|\mathbf{Z}_{k-1}) \rangle} \right] \tag{11}$$

where $p_M(\cdot)$ is the probability of missed detection, $\psi_{k,\mathbf{z}_k}(\tilde{\mathbf{x}}_k^i) = (1 - p_M(\tilde{\mathbf{x}}_k^i)) g_k(\mathbf{z}_k|\tilde{\mathbf{x}}_k^i)$, and $g_k(\mathbf{z}_k|\tilde{\mathbf{x}}_k^i)$ is the single-target likelihood defining the probability that a measurement $\mathbf{z}_k$ is generated by a target, κ_k is the clutter intensity, and $\langle f, g \rangle = \int f(x) g(x) dx$ [14].

In the particle PHD filter, the PHD $D(\cdot)$ is represented by the weights of particles. Once the new set of observations is available, we can substitute the approximation of $D(\mathbf{X}_k|\mathbf{Z}_{k-1})$ into (11) and the weights of each particle are updated based upon the receipt of the measurement $\mathbf{Z}_k$ as [23]

$$\tilde{w}_k^i = \left[p_M(\tilde{\mathbf{x}}_k^i) + \sum_{\forall \mathbf{z}_k \in \mathbf{Z}_k} \frac{\psi_{k,\mathbf{z}_k}(\tilde{\mathbf{x}}_k^i)}{\kappa_k + C_k(\mathbf{z}_k)} \right] \tilde{w}_{k|k-1}^i \tag{12}$$

where

$$C_k(\mathbf{z}_k) = \sum_{i=1}^{(M_{k-1}+J_k)\times N} \psi_{k,\mathbf{z}_k}(\tilde{\mathbf{x}}_k^i) \tilde{w}_{k|k-1}^i. \tag{13}$$

Algorithm 1: Adapted particle PHD filter.

Input: $\{\mathbf{x}_{k-1}^{m}\}_{m=1}^{m=M_{k-1}}$.
Output: $\{\mathbf{x}_{k}^{m}\}_{m=1}^{m=M_{k}}$ with M_k targets.
1: Generate (1) from $\{\mathbf{x}_{k-1}^{m}\}_{m=1}^{m=M_{k-1}}$ and feed into (6).
2: Select new-born particles as described in Section III-D.1.
3: Obtain (9) with weights as (10).
4: **for** $i = 1 : (M_{k-1} + J_k) \times N$ **do**
5: Calculate $g(\mathbf{z}_k|\tilde{\mathbf{x}}_k^i)$.
6: Update particle weights with (12).
7: **end for;**% Achieve particle set $\{\tilde{\mathbf{x}}_k^i, \tilde{w}_k^i\}_{i=1}^{i=(M_{k-1}+J_k)\times N}$ with updated weight.
8: Calculate M_k by (14) and (15).
9: Normalize $\tilde{w}_k^i$ with (16).
10: Initialize the cumulative probability $c_1 = 0$
11: Update $c_i = c_{i-1} + \tilde{w}_k^i$, $i = 2, ..., (M_{k-1} + J_k) \times N$.
12: Draw a starting point $\mu_1 \sim [0, (M_k \times N)^{-1}]$
13: **for** $j = 1, ..., M_k \times N$ **do**
14: $\mu_j = \mu_1 + (j-1)/(M_k \times N)$
15: **while** $\mu_j > c_i$ **do**
16: $i = i + 1$
17: **end while**
18: $\mathbf{x}_k^j = \tilde{\mathbf{x}}_k^i$
19: $w_k^j = N^{-1}$
20: **end for**
21: Clustering $\{\mathbf{x}_k^i, \frac{1}{N}\}_{i=1}^{i=M_k\times N}$, calculate (5) and output $\{\mathbf{x}_k^m\}_{m=1}^{m=M_k}$

Then the number of targets is calculated by the sum of all the weights for particles as follows [23]:

$$\tilde{M}_k = \sum_{i=1}^{(M_{k-1}+J_k)\times N} \tilde{w}_k^i \tag{14}$$

$$M_k = int(\tilde{M}_k) \tag{15}$$

where $int(\cdot)$ takes the integer nearest to $\tilde{M}_k$.

At each iteration k, $J_k \times N$ new particles are added to the old $M_{k-1} \times N$ particles for the new born targets. To limit the growth of the number of particles, and to avoid the problem of degeneracy, a resampling step is performed after the update step. Firstly, the weights for the particles are normalized as

$$\tilde{w}_k^i = \frac{\tilde{w}_k^i}{\tilde{M}_k}. \tag{16}$$

The algorithm for the adapted particle PHD filter with a resampling step at each time k is described as Algorithm 1 [24], where the input $\{\mathbf{x}_{k-1}^{m}\}_{m=1}^{m=M_{k-1}}$ represents the survived targets from the previous time $k-1$ and the output $\{\mathbf{x}_{k}^{m}\}_{m=1}^{m=M_{k}}$ denotes the tracking results in the form of the states of the targets.

The above method underpins the traditional particle PHD filter for multiple human tracking. However, using this method, the prediction of the states cannot be achieved accurately. We therefore exploit a social force model to improve the prediction of the states after step 3 and before step 4 of Algorithm 1 described in Section III-B, and an OCSVM classifier is utilized to improve the accuracy in step 5 as detailed in Section III-D.2.

B. Social Force Model for Multiple Human Tracking

Recently, modeling the behavior of pedestrians has been an important area of research mainly in evacuation dynamics and traffic analysis. Helbing *et al.* [25], [26] proposed the social force model for human tracking, where the human behavior, destination and velocity information are utilized to model the prediction for human targets.

Given a current set of target states $\{\mathbf{x}_{k}^{m}\}_{m=1}^{m=M_{k}}$ based on the position, velocity and walking behaviour of each target including its destination and avoiding collision with others [16], it is assumed in a social force model that every human target knows its current position and velocity, as well as its destination. In addition, it has social force with other targets if they are closer in distance than a pre-defined threshold. It is also often assumed that each target will predict the movement of other targets via a constant velocity model. Thus, the position information $\mathbf{p}_k^m = [p_{k,x}^m, p_{k,y}^m]^T$ and the velocity information $\mathbf{v}_k^m = [v_{k,x}^m, v_{k,y}^m]^T$, as the state of target $\mathbf{x}_k^m$ at time k, can be used to represent the social force between the targets. The social force model for target m is calculated between target m and all other targets. For example, the social force between targets m and n $(n \neq m)$ is calculated based upon the following parameters: the distance and angular displacement between m and n: $d_k^m(n)$ and $A_k^m(n)$; the change of velocity compared with target m: U_k^m and the cosine between velocity and destination path of target m: W_k^m [4].

The distance $d_k^m(n)$ can be calculated as [16]

$$d_k^m(n) = \|\mathbf{p}_k^m + t\mathbf{v}_k^m - \mathbf{p}_k^n - t\mathbf{v}_k^n\| \tag{17}$$

where $\|\cdot\|$ denotes the Euclidean norm and t is the time interval between frame $k-1$ and k. Since we assume each target intends to avoid collisions with other targets, the angular displacement between the velocity of the two targets is also considered as one of the important parameters for the social force model, which can be represented as factor $A_k^m(n)$ [4]

$$A_k^m(n) = 1 + \frac{(\mathbf{v}_k^m)^T \mathbf{v}_k^n}{\|\mathbf{v}_k^m\|\|\mathbf{v}_k^n\|}. \tag{18}$$

We also assume that each target m walks towards a destination $\mathbf{p}_o^m = [p_{o,x}^m, p_{o,y}^m]^T$, and in doing so tries to maintain a desired speed $\mathbf{u}^m = [u_x^m, u_y^m]^T$. These two components can be described as two energy functions U_k^m and W_k^m, which denote the change of velocity and cosine between the current velocity and destination path for target m respectively

$$U_k^m = \|(\mathbf{v}_k^m - \mathbf{u}^m)\| \tag{19}$$

$$W_k^m = \frac{(\mathbf{p}_o^m - \mathbf{p}_k^m)^T \mathbf{v}_k^m}{\|\mathbf{p}_o^m - \mathbf{p}_k^m\|\|\mathbf{v}_k^m\|} \tag{20}$$

where $\mathbf{v}_k^m$ denotes the velocity of target m at time k.

After calculating the above parameters, the overall social force for target m at time k can be written as [16]

$$S_k^m = \sum_{n\neq m} d_k^m(n) A_k^m(n) + \lambda_1 U_k^m + \lambda_2 W_k^m \tag{21}$$

where λ_1 and $\lambda_2 \in \mathbb{R}^+$ control the influence of the two regularizers. After the social force is obtained for each target, it can

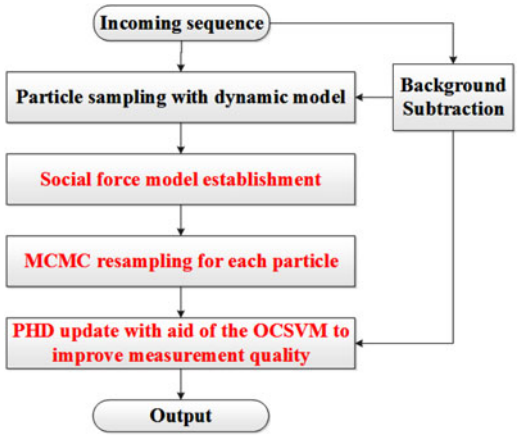

Fig. 1. Flowchart of the proposed system for multiple human tracking, where the red parts in the flowchart represent our contributions in the proposed system.

be incorporated into the prediction step of the particle filter. As mentioned earlier, the PHD filter can be easily influenced by the measurement noise, which may cause false alarms and missed detections; to address this problem, we use an OCSVM in the updating step to mitigate the noise effect. In the next section, we present details for our proposed particle PHD filter with a social force model-aided MCMC resampling step in prediction and an OCSVM in the updating step.

III. Social Force Model-Aided MCMC PHD Filter

A. Overview of the Proposed System

Fig. 1 gives an overview of our proposed algorithm. When an input frame is obtained from the video sequence, two main steps are performed based upon the fundamental Bayesian filtering framework: firstly, a social force model is established and an MCMC resampling step is performed which serves as the prediction part; secondly, background subtraction is employed in the PHD update step with an OCSVM classifier to obtain the states of the human targets as the resulting output.

B. Exponential-Term-Based Social Force Model

We exploit an exponential-term based energy function similar to that in [16] to describe the social force model for the likelihood calculation in the prediction stage of the MCMC-PHD filter. When a particle $\tilde{\mathbf{x}}_k^{m,i}$ is predicted to represent the state of target $m, \mathbf{x}_k^m$, at time k, its weight is predicted by the social force model representing interactions with other existing targets. Based upon (17), the distance between particle $\tilde{\mathbf{x}}_k^{m,i}$ from target m and the state $\mathbf{x}_k^n$ of target n can be used within an energy term

$$E_{k,d}^{m,i}(n) = e^{\frac{d_k^{m,i}(n)}{2\sigma_d^2}} \tag{22}$$

where σ_d controls the influence of the distance factor (denoted by subscript d) on the social force model. So the larger the distance between the predicted particle and the selected target, the higher the energy from the distance aspect, and $E_{k,d}^{m,i}(n)$ becomes minimum if the linear trajectories collide with each other. In this paper, obstacles in the scenes are also considered; the states of which are considered as targets with velocity $\mathbf{v}_k^m = [0,0]^T$ to calculate the social force model for each particle. Since the pedestrians will change their speed and angular velocity in order to avoid collision with others, by employing (18), the angular displacement factor for the social force model can be represented as

$$E_{k,\phi}^{m,i}(n) = (A_k^{m,i}(n))^{\beta} \tag{23}$$

where β controls the influence from the direction of the velocity and the subscript ϕ is used to represent angular displacement. Based on (22) and (23), the influence of multiple subjects can now be modeled as a weighted product. For example where particle $\tilde{\mathbf{x}}_k^{m,i}$ is assigned an energy with respect to each target n $(n \neq m)$ depending on its current distance and angular displacement ϕ [16] of the form

$$E_k^{m,i}(n) = E_{k,d}^{m,i}(n) E_{k,\phi}^{m,i}(n). \tag{24}$$

Two energy functions which denote the change of velocity and cosine between current velocity and destination path for particle $\mathbf{x}_k^{m,i}$ respectively can also be represented

$$E_{k,U}(m,i) = e^{-\frac{U_k^{m,i}}{2\sigma_v^2}} \tag{25}$$

$$E_{k,W}(m,i) = e^{-\frac{W_k^{m,i}}{2\sigma_D^2}} \tag{26}$$

where σ_v and σ_D control the influence of changing the velocity and destination on the social force of the target respectively.

To represent the state of target m, the overall interaction energy for particle $\mathbf{x}_k^{m,i}$ is predicted as

$$S_k^{m,i} = \prod_{n \neq m} E_k^{m,i}(n) E_{k,U}(m,i) E_{k,W}(m,i) \tag{27}$$

where the calculation of $S_k^{m,i}$ is different from those in [4] and [16] where a sum function instead of a product function was used.

The above equations can be used as the social force weight functions for establishing a posterior distribution within the prediction stage of the particle PHD filter. By calculating the social force from other targets, the estimated weight for prediction $s_k^{m,i}$ can be obtained by normalizing $S_k^{m,i}$. After calculating the social force, the particle set with their social force model results is obtained, which is used to derive the likelihood for particles in the MCMC resampling step described in Section III-C, to achieve more accurate prediction for particles.

C. Social Force Model-Based MCMC Resampling

In the traditional particle filter, an importance function is used in the sample selection step [27], [22], however, the MCMC based particle filter replaces the importance sampling step by building a Markov chain which exploits the posterior distribution [22], and thereby improves diversity among particles. In this paper, the MCMC resampling step is employed to improve the accuracy of the prior distribution, where the social force model is utilized to replace the likelihood function in the traditional MCMC particle filter [28].

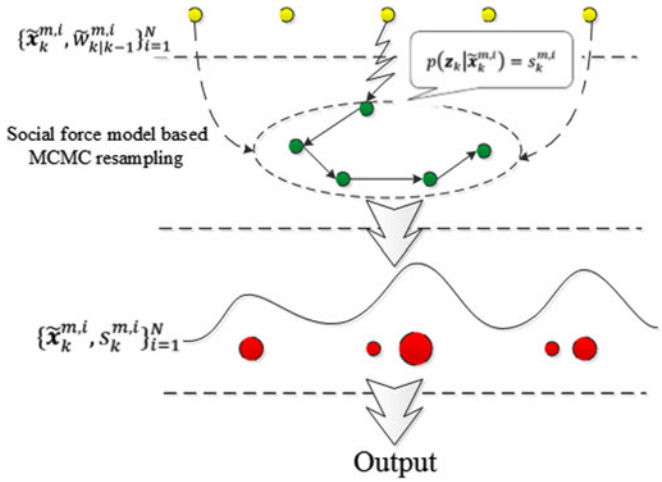

Fig. 2. Basic operation of the proposed social force model-aided MCMC based particle filter.

As described in [18], during the MCMC resampling, a particle $\tilde{\mathbf{x}}_k^{m,i}$ is propagated to a new state $\tilde{\mathbf{x}}_k^{m,i*}$ based on the following model:

$$\tilde{\mathbf{x}}_k^{m,i*} = \tilde{\mathbf{x}}_k^{m,i} + \mathbf{q} \tag{28}$$

where $\mathbf{q}$ denotes a zero-mean Gaussian noise vector. From the Metropolis-Hastings acceptance probability [22], we have the acceptance ratio calculated as

$$\alpha = \min\left\{1, \frac{p(\mathbf{z}_k|\tilde{\mathbf{x}}_k^{m,i*})p(\tilde{\mathbf{x}}_k^{m,i*}|\mathbf{x}_{k-1}^m)q(\tilde{\mathbf{x}}_k^{m,i}|\tilde{\mathbf{x}}_k^{m,i*})}{p(\mathbf{z}_k|\tilde{\mathbf{x}}_k^{m,i})p(\tilde{\mathbf{x}}_k^{m,i}|\mathbf{x}_{k-1}^m)q(\tilde{\mathbf{x}}_k^{m,i*}|\tilde{\mathbf{x}}_k^{m,i})}\right\}. \tag{29}$$

Since in our work, $q(\cdot|\tilde{\mathbf{x}}_k^{m,i})$ is symmetric in its arguments, that is

$$q(\tilde{\mathbf{x}}_k^{m,i*}|\tilde{\mathbf{x}}_k^{m,i}) = q(\tilde{\mathbf{x}}_k^{m,i}|\tilde{\mathbf{x}}_k^{m,i*}) \tag{30}$$

we can calculate the acceptance ratio α as

$$\alpha = \min\left\{1, \frac{p(\mathbf{z}_k|\tilde{\mathbf{x}}_k^{m,i*})p(\tilde{\mathbf{x}}_k^{m,i*}|\mathbf{x}_{k-1}^m)}{p(\mathbf{z}_k|\tilde{\mathbf{x}}_k^{m,i})p(\tilde{\mathbf{x}}_k^{m,i}|\mathbf{x}_{k-1}^m)}\right\}. \tag{31}$$

In this work, the likelihoods of particle state $p(\mathbf{z}_k|\cdot)$ are replaced by the results obtained from the social force model, thus

$$\alpha = \min\left\{1, \frac{s_k^{m,i*}p(\tilde{\mathbf{x}}_k^{m,i*}|\mathbf{x}_{k-1}^m)}{s_k^{m,i}p(\tilde{\mathbf{x}}_k^{m,i}|\mathbf{x}_{k-1}^m)}\right\}. \tag{32}$$

The state to be preserved is determined by drawing a point j from a uniform distribution. If $j < \alpha$ then the new state $\mathbf{x}_k^{m,i*}$ is retained, otherwise it is rejected. In this way, the social force model is fed into the MCMC resampling step for achieving more robust prediction.

Fig. 2 shows the steps of the social force model aided MCMC resampling step with an example target m.

As shown in Fig. 2, a particle $\tilde{\mathbf{x}}_k^{m,i}$ predicted from the state model is chosen as the initial value of the Markov chain. A new state $\tilde{\mathbf{x}}_k^{m,i*}$ is propagated as (28), then the acceptance ratio α [28] is calculated as in (32), where the likelihood for the particles $p(\mathbf{z}_k|\cdot)$ is obtained from the exponential term social force model s_k, described in Section III-B. After resampling the particles, which are obtained from the state model, a more robust prior distribution is achieved. The example pseudocode of this MCMC particle filter for target m is then summarised as Algorithm 2, where the inputs are the predicted particles for target m and the output is the posterior distribution from the prediction stage, and B denotes the number of burn-in period particles.

Algorithm 2: Social force model based MCMC resampling step (SFM-MCMC)

Input: Predicted particles for target m from state model $\{\tilde{\mathbf{x}}_k^{m,i}\}_{i=1}^{i=N}$

Output: Particles with predicted weights from the social force model aided MCMC resampling $\{\tilde{\mathbf{x}}_k^{m,i}, \tilde{w}_{k|k-1}^{m,i}\}_{i=1}^{i=N}$

```
Initialize the Markov chain by the predicted particles from the state transition function using the states of target at k − 1.
for i = 1:N + B do
    Propagate x̃_k^{m,i*} from x̃_k^{m,i} with (28) .
    Calculate s_k^{m,i} and s_k^{m,i*} for x̃_k^{m,i} and x̃_k^{m,i*} with (27).
    Compute α with (32).
    Draw a point j from a uniform distribution.
    if j < α then
        retain the new state: x̃_k^{m,i} = x̃_k^{m,i*}.
    else reject the new state.
    end if
end for
Discard the first B particles of the iterations.
```

We note that the acceptance of a proposed state depends on the likelihood found by calculating the social force from the other targets. The predicted weight for particles $\tilde{w}_{k|k-1}^i$ can be replaced by the results obtained from the social force model

$$\tilde{w}_{k|k-1}^i = s_k^i. \tag{33}$$

In this case, after resampling all the predicted particles, we can obtain a set of predicted particles with estimated weights from the social force model $\{\tilde{\mathbf{x}}_k^i, \tilde{w}_{k|k-1}^i\}_{i=1}^{i=(M_k+J_k)\times N}$. In the next section, we will present the measurement we use for updating.

D. Robust Measurement Model

Besides the state model, another important step is the measurement model for particle updating, in this work, two main steps are employed to obtain a robust measurement model: background subtraction and a one class support vector machine.

1) Background Subtraction: In multiple human tracking, the measurement and the new-born targets are difficult to select from the video sequence. In the prediction part, instead of sampling particles for a new-born target, we employed a background subtraction step to facilitate the sampling step. The background subtraction results can also be utilized as the random measurement set for the proposed particle PHD filter [24]. In this paper, we

used the codebook method [29], [30] for background subtraction which is robust to capture structural background motion over a long period of time under limited memory. In this method, samples at each pixel are clustered into the set of codewords based on a color distortion metric together with brightness bounds. Not all pixels are represented with the same number of codewords. The background is encoded on a pixel-by-pixel basis. Background/foreground detection involves testing the difference of the current image from the background model with respect to color and brightness differences. If an incoming pixel satisfies two conditions, it is classified as background: first, the color distortion to a codeword is less than the detection threshold; second, its brightness lies within the brightness range of that codeword. Otherwise, it is classified as foreground [30]. Some background subtraction results are shown in Section IV-C.1. The results can be used to select the new-born targets and build up an RFS for the measurement set [31]. The center of each block $\mathbf{c}_k = [c_{k,x}, c_{k,y}]^T$ which contains the localization information, can be employed as one part of the measurement [24], so the likelihood for each particle based upon the foreground position $g_b(\mathbf{c}_k|\tilde{\mathbf{x}}_k^i)$ can be calculated as

$$g_b(\mathbf{c}_k|\tilde{\mathbf{x}}_k^i) = e^{-\frac{(\mathbf{p}_k^i - \mathbf{c}_k)^T(\mathbf{p}_k^i - \mathbf{c}_k)}{\sigma_R^2}} \tag{34}$$

which shows the distance between the state of the particles and the foreground information, where $\mathbf{p}_k^i = [p_{k,x}^i, p_{k,y}^i]^T$ denotes the position of the targets taken from the particle $\tilde{\mathbf{x}}_k^i$ and σ_R is the standard deviation of the measurement model in the Bayesian filtering model.

However, the raw background subtraction results generally contain many artifacts, which include small 'salt and pepper' terms and large noise patches caused by the problem of poor illumination and similar color between the foreground and background. The noise patches may be regarded as a new born target in the prediction step of the PHD filter and cause the occurrence of false alarms [32]. To address this issue, in this paper, we propose to use an OCSVM classifier [33] to distinguish the human targets from noise as described next.

2) One Class Support Vector Machine: The basic idea is that given a data set drawn from an underlying probability distribution p, the OCSVM estimates a function f to describe its 'support region' (where a sample of p most likely comes from), where the corresponding values of the function f are larger than a particular threshold value [34].

To design the classifier, based on a training dataset, the following quadratic optimisation problem needs to be solved:

$$\min_{\mathbf{w},\varsigma,\rho} \ \frac{1}{2}\|\mathbf{w}\|^2 + \frac{1}{\nu L}\sum_{i=1}^{L}\varsigma_i - \rho$$
$$\text{subject to } (\mathbf{w}^T\Phi(\tilde{\mathbf{x}}_k^i)) \geq \rho - \varsigma_i, \ \ \xi_i \geq 0 \tag{35}$$

where $\mathbf{w}$ is the normal vector, $\nu \in (0, 1]$, ρ is from the Lagrangian model of the SVM, which is set to be zero in this work and the nonzero slack variables $\varsigma = [\varsigma_1, ..., \varsigma_L]$ are introduced to allow for the possibility of outliers (the data points which are not drawn from the supporting region) and $\Phi(\cdot)$ is a nonlinear kernel function which maps the original data into a different space for better separation. For a test particle $\tilde{\mathbf{x}}_k^i$, the decision function for estimating whether it comes from the determined distribution is

$$f(\tilde{\mathbf{x}}_k^i) = (\mathbf{w}^T\Phi(\tilde{\mathbf{x}}_k^i)) - \rho. \tag{36}$$

In the application of multiple human tracking, the features from both color and oriented gradient [35] of multiple human regions are employed for training the OCSVM classifier, which can be used to estimate the likelihood function value for each particle. Given a particle $\tilde{\mathbf{x}}_k^i$ at time instance k, the features from both the color and oriented gradient histogram are extracted based upon the position, width and height information of $\tilde{\mathbf{x}}_k^i$ and the corresponding likelihood function, $\vartheta_k(\tilde{\mathbf{x}}_k^i)$ can be estimated as

$$\vartheta_k(\tilde{\mathbf{x}}_k^i) = e^{(\varpi \cdot f(\tilde{\mathbf{x}}_k^i))} \tag{37}$$

where ϖ is a constant we set for calculating the weights for the particles, thereby controlling the influence of the sub-likelihood from the OCSVM. Its value is chosen empirically in our work. In this way, the likelihood for each particle is obtained and these weights can then be taken as the input to the updating step of the PHD filter.

E. Particle PHD Updating and Resampling

After obtaining the particles from the MCMC resampling step, we can achieve a particle set with their estimated weights described by their social force as described in Section III-C and the likelihood $g(\mathbf{z}_k|\tilde{\mathbf{x}}_k^i)$ is calculated based upon the results from both background subtraction and the OCSVM

$$g(\mathbf{z}_k|\tilde{\mathbf{x}}_k^i) = \vartheta_k(\mathbf{x}_k^i)g_b(\mathbf{c}_k|\tilde{\mathbf{x}}_k^i). \tag{38}$$

By feeding (38) into (12), the weights for particles are updated. The number of human targets and the particles are resampled as in Algrorithm 1 described in Section II.

F. Summary of the Proposed System

A summary of the proposed system is given in Algorithm 3, which we refer to as SFM-MCMC-OCSVM-PHD.

IV. Simulation Experiments

In this section, simulations are provided to examine the performance of our system and to compare with results from other recent methods.

A. Dataset Selection and Parameter Setup

In order to evaluate the performance of the proposed system for multiple human tracking, particularly to handle the situation of varying number of targets, close interactions and occlusions, we firstly chose sequences from three different publicly available video datasets: one from the PETS2009 dataset [36] where 3–6 human targets are walking in an outdoor campus environment, one sequence from the CAVIAR dataset [37] where 1–5 human targets are walking in a shopping mall environment and

Algorithm 3: Social force model-aided MCMC-OCSVM particle PHD filter (SFM-MCMC-OCSVM-PHD)

Input: Video sequence with ℓ frames.
Output: $\{\mathbf{x}_k^m\}_{m=1}^{M_k}$ and M_k.

1: OCSVM classifier training.
2: Initialize targets states in the first frame $\{\mathbf{x}_1^m\}_{m=1}^{m=M_0}$.
3: **for** $k = 2{:}\ell$ **do**
4: Background subtraction to extract the measurement set $\mathbf{Z}_k$ for targets and the estimated positions of the new born targets.
5: Predict particles for both survived targets and new born targets separately as described in Algorithm 1.
6: Calculate social force s_k^i for each particle.
7: SFM-MCMC resampling with Algorithm 2.
8: Calculate $g(\mathbf{z}_k|\tilde{\mathbf{x}}_k^i)$ by (34), (37) and (38).
9: Update the PHD weights with (12).
10: Calculate M_k by (14).
11: Resampling of particles with the method described in Algorithm 1.
12: Output tracking results at time k, $\{\mathbf{x}_k^m\}_{m=1}^{M_k}$ and M_k.
13: **end for**

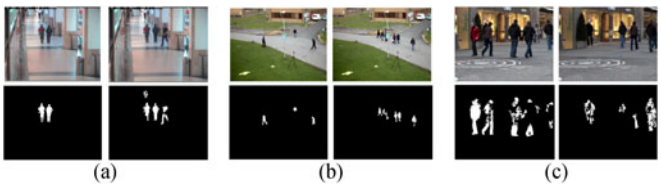

Fig. 3. Selected frames and examples for background subtraction results from the three selected sequences from three different datasets, i.e., (a) is from the "EnterExitCrossingPaths1cor" sequence from the CAVIAR dataset, (b) is from the "PETS09_ View001_ S2_ L1" sequence from the PETS2009 dataset, and (c) is from "TUD_ Stadtmitte" sequence from the TUD dataset, where we can find the human target boundaries are extracted successfully, but there is still much environmental noise which may cause missed detections and false alarms. In order to mitigate such noise, an OCSVM classifier will be employed based upon the features from both color and oriented gradient histograms of human targets.

one from the TUD dataset [38] where 5–7 human targets are walking in an outdoor-shopping mall environment. In order to make more reliable evaluation, 17 more sequences from the CAVIAR dataset are also employed. All sequences are recorded at a resolution of 320×240 pixels at 25 frames/sec and each sequence contains around 200 frames, including human targets appearing, disappearing and occlusion in the scenario, selected example frames are given in Fig. 3. 100 particles are employed for each target and for the MCMC step, 20 burn-in particles are used for each target in the MCMC resampling step. The set up of the remaining parameters is discussed in the following sections. The dynamic and measurement models which were used to predict and update the particles are described as

$$\mathbf{x}_k = \mathbf{F}\mathbf{x}_{k-1} + \omega_k \tag{39}$$

$$\mathbf{z}_k = \mathbf{H}\mathbf{x}_k + \mathbf{v}_k \tag{40}$$

TABLE I
BACKGROUND SUBTRACTION PARAMETERS FOR EACH SEQUENCE

	α_b	β_b	ε_b
CAVIAR dataset	0.5	2	30
PETS 2009	0.7	1.5	20
TUD dataset	0.7	1.7	10

where the state and measurement transformation matrices $\mathbf{F}$ and $\mathbf{H}$ are given as

$$\mathbf{F} = \begin{pmatrix} 1 & 0 & \Delta t & 0 & 0 & 0 \\ 0 & 1 & 0 & \Delta t & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \quad \mathbf{H} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{pmatrix}^T \tag{41}$$

where Δt is the time interval between frame k and $k+1$ which is set as 1 in the simulations, the zero-mean noise vector ω_k for prediction in the state model has covariance structure $cov\{\omega_k\} = Diag\{25, 25, 16, 16, 4, 4\}$ and for $\mathbf{v}_k$ $cov\{\mathbf{v}_k\} = Diag\{25, 25\}$. The missed detection probability $p_M = 0.01$, the survival probability $e = 0.99$, the new born intensity $\Upsilon = 0.1$ and clutter intensity $\kappa = 0.01$. The parameters for background subtraction, exponential-term based social force model and OCSVM classifier are selected empirically, which are shown as follows:

1) Parameters for Background Subtraction: For the background subtraction method described in Section III-D.1, the parameters that need to be set include the shadow bound α_b, the highlight bound β_b and the color detection threshold ε_b which for each sequence are given as Table I, which were found empirically to yield best performance. For other parameters, we used the default values as those set in [30], for example, the color sampling bandwidth is set to be 20 for all these three datasets. Due to space limitation, the set-up for the remaining parameters is omitted but can be found from [30].

2) Parameters for Exponential-Term-Based Social Force Model: The exponential-term based social force model introduced in Section III-B has many parameters, such as σ_d, β, σ_v, and σ_D, which control the influence from distance, angular displacement, change of velocity and destination respectively. In this paper, the parameters are selected based on pilot tests, we use a sequence from PETS2009 to perform simulations with different values of the above four parameters and use the mean Euclidean error for each target position as the evaluation measure to select the best parameter set, which is shown in Table II. From the experiment and comparison, we found when the parameters are chosen as: $\sigma_d = 4$, $\beta = 4$, $\sigma_v = 4$, and $\sigma_D = 2$, the exponential-term based social force model performs the best, therefore these settings are adopted for simulations in subsequent sections.

3) Parameters for OCSVM-Classifier: In our work, to obtain the OCSVM classifier, the training dataset $\mathbf{S} = [\mathbf{s}_1, ..., \mathbf{s}_L]$ is employed, where each training data $\mathbf{s}$ is a 593×1 vector,

TABLE II
EXPERIMENTAL VALUES FOR PARAMETERS USED IN SOCIAL FORCE MODEL

σ_d	β	σ_v	σ_D	Mean of Euclidean error
1	1	1	1	5.15
1	1	1	2	5.11
1	1	1	4	5.33
1	1	1	8	5.12
1	2	1	2	5.15
1	4	1	2	5.14
1	8	1	2	5.33
1	4	2	2	5.53
1	4	4	2	5.21
1	4	8	2	5.41
2	4	4	2	5.12
4	4	4	2	5.05
8	4	4	2	5.22
16	4	4	2	5.09
32	4	4	2	5.18

TABLE III
PARAMETERS VALUES USED IN OCSVM CLASSIFIER AND SYSTEM EVALUATION

ϖ value	1	25	50	75	100
OSPA value	4.49	4.77	4.86	4.25	4.77

containing human features extracted from the training frame, including 512 parameters from the color histogram and 81 from oriented gradient histogram. The OCSVM classifier is trained by 82 sets of features extracted from different human targets. The influence of the OCSVM is controlled by parameter ϖ in (37), which is also chosen based on experiments. The OSPA results with respect to the different values of ϖ are shown as Table III. From the comparison, $\varpi = 75$ is found to perform the best, hence is employed in the later simulations.

B. Performance Metrics

Several measures are employed to examine the performance of the proposed particle PHD filter and compare the results from the related algorithms, including the Euclidean error in each frame, the optimal subpattern assignment (OSPA) [40], [41], and the multiple object tracking precision (MOTP) [42]. For a sequence with ℓ frames, assuming at time k, the tracking system gives the tracking results $\mathbf{O}_k = \{\mathbf{o}_k^1, ..., \mathbf{o}_k^n\}$ with n targets while $\mathbf{Y} = \{\mathbf{y}_k^1, ..., \mathbf{y}_k^m\}$ is the ground truth information with m targets. These measures are defined below. In addition, the computational complexity has also been considered in our evaluations.

1) Mean and Standard Deviation of Euclidean Error on Each Frame: The localization error for each target in terms of mean and variance can be used as a performance metric to evaluate the accuracy and stability of our proposed tracking system. The mean of Eucidean errors (MEE) at frame number k is denoted by

$$MEE_k = \frac{1}{n}\sum_{i=1}^{n} \|\mathbf{o}_k^i - \mathbf{y}_k^i\| \tag{42}$$

and its standard deviation (SD) is given by

$$SD_k = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\|\mathbf{o}_k^i - \mathbf{y}_k^i\| - MEE_k)^2}. \tag{43}$$

2) OSPA: In multiple human tracking, the accuracy not only depends on the error between the estimated position and the ground truth information of the targets in the scenario, but also the missed detections and false alarms. Dominic *et al.* proposed a metric to evaluate the tracking system by error from both distance and the number of targets [39] which is used by Ristic *et al.* [40] for evaluating multiple human tracking algorithms. As described in [39], given the set of tracking results $\mathbf{O}_k$ and the ground truth information $\mathbf{Y}_k$, the distance between $\mathbf{O}_k$ and $\mathbf{Y}_k^m$, $d_k^c(\mathbf{o}_k^n, \mathbf{y}_k^m) := \min(c, d(\mathbf{o}_k^n, \mathbf{y}_k^m))$ with cut off at $c > 0$ and $1 \le p \le \infty$, is calculated as [40]

$$d_{k,p}^c(\mathbf{O}_k, \mathbf{Y}_k) := \left(\frac{1}{n}\left(\min_{\pi \in \prod_n}\sum_{i=1}^{m} d^c(\mathbf{o}_k^i, \mathbf{y}_k^{\pi(i)})^p + c^p(n-m)\right)\right)^{\frac{1}{p}} \tag{44}$$

for $m \le n$, and $d_{k,p}^c(\mathbf{O}, \mathbf{Y}) = d_p^c(\mathbf{Y}, \mathbf{O})$ for $m > n$. The function d_p^c is named as the OSPA metric of order p with cut-off c. In this paper, we use $c = 20$ and $p = 2$ in our evaluations. Based on the OSPA metric, a new evaluation measure for multiple target tracking has been recently proposed, named optimal subpattern assignment for multiple target tracking (OSPAMT) [41], however, in this paper OSPA and the following MOTP measure, which is also employed in [43], are sufficient for comparative evaluation.

3) MOTP: The MOTP [42] is the total error in the estimated position for matched object-hypotheses pairs over all frames, averaged by the total number of matches made. It shows the ability of the tracker to estimate the precise object positions, which can be calculated as

$$\text{MOTP}_k = \frac{\sum_{i=1,k=1}^{i=n,k=\ell} error_k^i}{\sum_{k=1}^{\ell} c_k} \tag{45}$$

where $error_k^i$ denotes the Euclidean error for target i at time k and c_k is the total number of matched targets at time k.

C. Evaluation of Tracking Results

In this section, our proposed exponential-term based SFM-MCMC is compared with the traditional SFM proposed in [4] and the S-SFM proposed by Pellegrini *et al.* in [16]. The proposed SFM-MCMC-OVSCM-PHD filter is compared with the traditional particle PHD filter in [24]. First, the comparison between the particle PHD filter and SFM-MCMC-PHD filter is made, followed by the comparison between the SFM-MCMC-PHD and SFM-MCMCOCSVM-PHD filters.

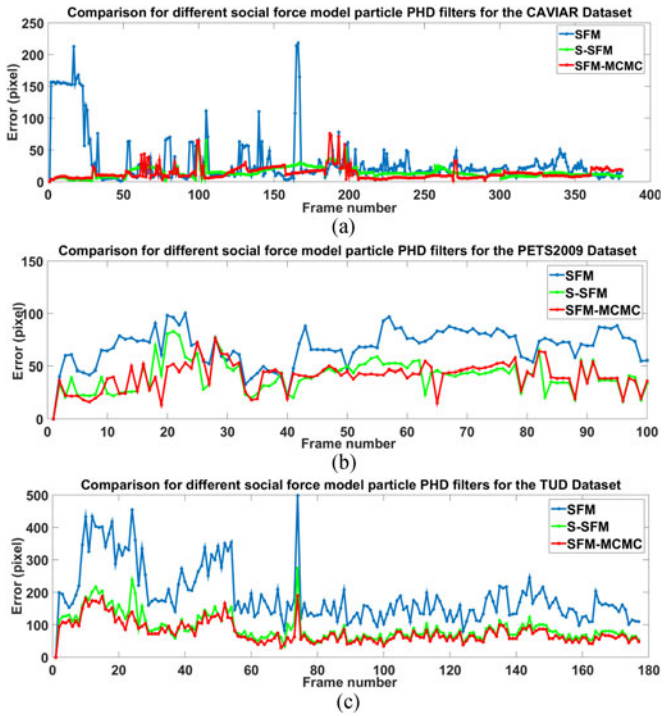

Fig. 4. Comparison in terms of Euclidean tracking error of the three social force models when employed by the particle PHD filter for multiple human tracking. (a) is the comparison for the "EnterExitCrossingPaths1cor" sequence from the CAVIAR dataset, (b) is for the "PETS09_ View001_ S2_ L1" sequence from the PETS2009 dataset, and (c) is for the "TUD_ Stadtmitte" sequence from the TUD datasett. The blue line denotes the traditional SFM [4], the green line denotes S-SFM [16] and the red line denotes the SFM-MCMC algorithm proposed in our paper.

1) Background Subtraction Results: In Fig. 3 we show some selected frames and results from the background subtraction for three datasets we employed, from which, we can find most of the targets appearing in the scenario, however, there is still much noise from the environment, which may cause false alarms and hence influence the performance. What's more, sometimes it may fail to detect the targets because of occlusion and the poor lightning. In this case, the OCSVM classifier is employed to aid calculation for each particle, in this way, the noise is mitigated, in the later section, the improvement made by employing OCSVM will be shown.

2) Social Force Model Results: By employing the parameters in Table II, and three example sequences selected from different datasets, our proposed social force model is first compared with the traditional SFM [4] and the S-SFM [16]. Fig. 4 shows the comparison of Eulidean tracking error in each frame between the above methods for the three sequences. The MEE over all the frames and their SD are also compared in Table IV. From Table IV and Fig. 4 we can see that the proposed social force model consistently attains better performance for the three sequences in terms of both the MEE and SD, as compared with the two baseline social force models. The improvement of the proposed social force model comes from the exponential-term model employed to describe the parameters such as the distance, angle, change of velocity, and the destination used in the model, with their influence controlled by the variance terms in the exponential-term model. In addition, in the proposed social force model, we have employed a threshold to control the modelling of the social forces between two targets, by excluding those that are far apart from each other in terms of distance (i.e. greater than the pre-defined threshold). This essentially avoids the influence from unnecessary targets, hence improving the tracking accuracy when more targets are present in the environment. After the SFM-MCMC resampling, the predicted weights for particles are updated. Fig. 5 shows an example distribution of predicted particle weights, for frame 11 of the ' PETS09_ View001_ S2_ L1′ sequence from the PETS2009 dataset.

From the figure we can find that the particles with higher social force are given higher weights than others, the redundant peaks in the figure are because of the noise patches which will be mitigated in the updating step. Compared with the traditional particle PHD filter, where the particles are given the same weights in the prediction stage, the weights in our SFM-MCMC-PHD filter are determined based on the SFM which leads to more accurate prediction.

3) OSPA Evaluation: In order to evaluate our proposed system in terms of both localization and cardinality, the OSPA metric has also been employed. In this paper, we use $c = 20$ and $p = 2$. Comparisons for the three example sequences as in the previous experiment are shown in Fig. 6, where the black line denotes the OSPA value from the traditional PHD filter, the blue line corresponds to our proposed SFM-MCMC-PHD particle PHD filter and the red line denotes our proposed SFM-MCMC-OCSVM-PHD algorithm. To perform more reliable evaluation, the average OSPA values for all the 20 sequences based upon different methods have been obtained and are shown in Table V. From the above comparison, we can observe that the improvement of the proposed system comes from both the exponential-term based SFM-MCMC resampling step and the OCSVM likelihood calculation step. The background subtraction is an integrated component as in [25] which is used to determine the measurement foreground pixels. Without background subtraction, none of the methods under study would operate, so the individual improvement from background subtraction is not provided in this paper. By using the OCSVM, the average OSPA value for the 20 sequences is further reduced by 4.71 pixels since the OCSVM can distinguish the measurement of the human targets from the noise in the environment. To examine the difference in OSPA results between the traditional particle PHD filter and our proposed SFM-MCMC-OCSVM-PHD filter, one-way ANOVA based F-test [44] is performed. We obtained $F = 8.74$, p-value $= 0.0051$ and the degree of freedom (1,42), where the F value is the ratio of the between-group variability to the within-group variability and the p-value is the probability of a more extreme result than the value we actually achieved when the null hypothesis is true. Using the degree of freedom value and significant value 0.05, the critical value F_{crit} is found to be 4.07 from the F-distribution table given in [44]. According to the test, the results are accepted as statistically significant if $F \geq F_{crit}$ and the p-value is less than the significant value. From the test results, we can confirm the difference

TABLE IV
COMPARISON OF THE MEAN OF THE EUCLIDEAN TRACKING ERRORS OVER THE FRAMES AND THEIR STANDARD DEVIATION BY THREE SOCIAL FORCE MODELS FOR THE "ENTEREXITCROSSINGPATHS1COR" SEQUENCE FROM THE CAVIAR DATASET, THE "PETS09_ VIEW001_ S2_ L1" SEQUENCE FROM THE PETS2009 DATASET, AND THE "TUD_ STADTMITTE" SEQUENCE FROM THE TUD DATASET

	CAVIAR			PETS2009			TUD		
	SFM [4]	S-SFM [16]	SFM-MCMC	SFM [4]	S-SFM [16]	SFM-MCMC	SFM [4]	S-SFM [16]	SFM-MCMC
MEE (pixel)	31.81	14.28	13.22	68.25	40.76	39.41	188.70	89.60	77.0
SD (pixel)	39.55	9.99	8.26	17.20	14.95	13.29	85.12	41.06	33.64

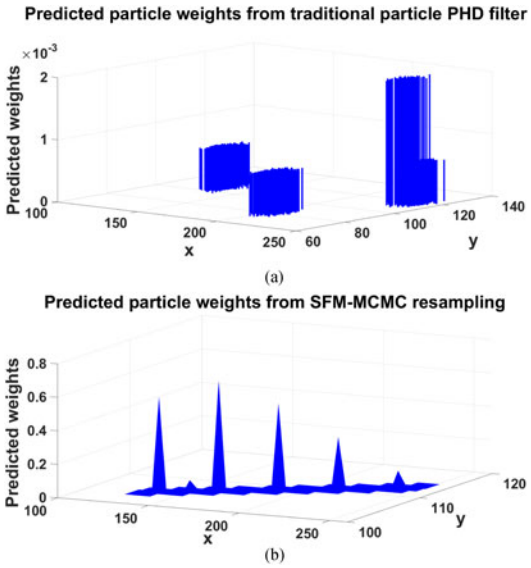

Fig. 5. Comparison between the two distributions of the predicted particle weights, where (a) is from the traditional particle PHD filter and (b) is from the SFM-MCMC-PHD filter; for frame 11 of the "PETS09_ View001_ S2_ L1" sequence from the PETS2009 dataset, where the ground truth position of the targets are (142,102), (233,115), and (200,95).

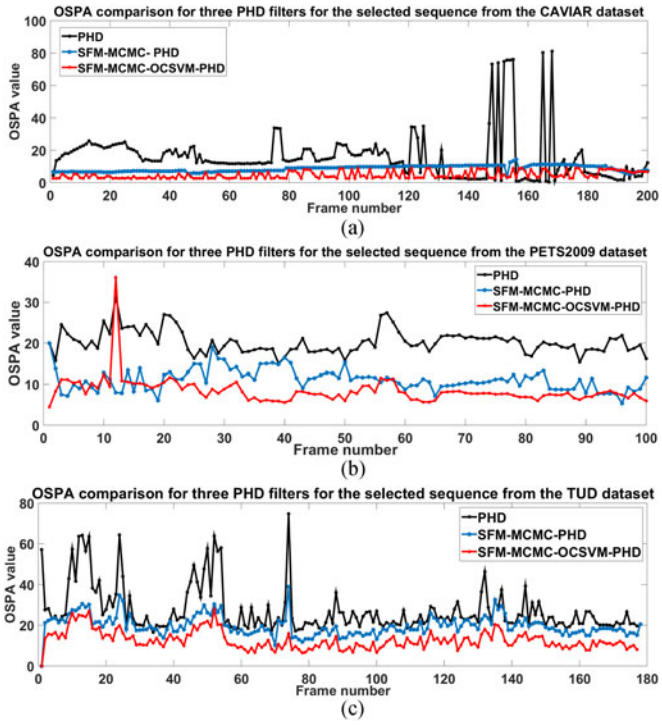

Fig. 6. Performance evaluation with OSPA measure for our proposed social force model aided MCMC particle PHD filter and the traditional particle PHD filter for multiple human tracking. The performance is examined with the "EnterExitCrossingPaths1cor" sequence from the CAVIAR dataset, the "PETS09_ View001_ S2_ L1" sequence from the PETS2009 dataset, and the "TUD_ Stadtmitte" sequence from the TUD dataset. The black line in the figure denotes the OSPA value from the traditional PHD particle PHD filter; the blue line shows the result by adding a social force model aided MCMC resampling step in the prediction stage of the particle PHD filter, and the red line denotes the OSPA value from our proposed SFM-MCMC-OCSVM-PHD filter.

in OSPA results between our proposed SFM-MCMC-OCSVM-PHD and traditional PHD filter is statistically significant.

4) MOTP Evaluation: Besides the Euclidean error in each frame, standard deviation and OSPA are used for evaluation, MOTP is also employed to evaluate the proposed tracking system. The MOTP results for the three selected sequences from different datasets are shown in Table VI. From the MOTP comparison, it can be observed clearly that our proposed method can greatly improve the tracking accuracy over the traditional PHD filter. For the CAVIAR dataset, the MOTP value is reduced by 6.14 pixels by employing the SFM-MCMC-PHD filter and then further reduced by 2.71 by employing the SFM-MCMC-OCSVM-PHD filter. For the PETS2009 dataset, the MOTP value is reduced by 1.67 and 2.92 pixels respectively. For the TUD dataset, the MOTP value is reduced by 2.56 and a further 1.79 pixels respectively. The average MOTP value over all the 20 sequences is reduced from 10.70 to 8.63 pixels by employing the SFM-MCMC-PHD filter and a further 6.31 pixels by SFM-MCMC-OCSVM-PHD filter. The reduction of MOTP is mainly due to the utilization of the social force model aided MCMC step for resampling in the prediction stage, so that a more accurate posterior distribution is achieved. Moreover, the OCSVM classifier in the updating step helps to mitigate the measurement noise from the environment so that the problems of missed detections and false alarms are mitigated. By performing a one way ANOVA based F test for our proposed SFM-MCMC-OCSVM-PHD filter and the traditional particle PHD filter, we obtained $F = 6.86$, p-value $= 0.0131$ and the degree of freedom (1,34). By setting the significant value to be 0.05, the critical value F_{crit} is found to be 4.13 from the F-distribution table given in [44]. From the test results, we can observe the difference in MOTP results between our proposed SFM-MCMC-OCSVM-PHD and the traditional PHD filter is statistically significant.

TABLE V
COMPARISON OF OSPA OVER 20 SEQUENCES FOR PROPOSED PHD FILTERS

	PHD [24]	SFM-MCMC-PHD	OCSVM-PHD	SFM-MCMC-OCSVM-PHD
OSPA (pixel)	21.93	13.54	12.42	8.83
Improvement	–	38.25%	43.36%	59.73%

TABLE VI
MOTP COMPARISON FOR THREE SEQUENCES USING THE TRADITIONAL AND PROPOSED PARTICLE PHD FILTERS

	CAVIAR	PETS2009	TUD
PHD [24]	12.49	8.23	16.42
SFM-MCMC-PHD	6.35	6.56	13.86
OCSVM-PHD	5.73	6.15	14.21
SFM-MCMC-OCSVM-PHD	3.64	4.84	12.07

TABLE VII
COMPARISON OF OSPA RELATED TO THE NUMBER OF PARTICLES

Number of particles	50	100	500	1000
Run-time/frame	1.44s	1.84s	4.85s	5.10s
OSPA (pixel)	22.52	21.32	21.06	21.03

5) Computational Complexity: In this section, the computational complexity is also examined through the run-time. Since we are using the particle PHD filter in this system, the number of particles plays an important role in affecting the computational complexity. In order to select the most suitable number of particles, as an example, the OSPA measure and run-time as a function of the number of particles are calculated by employing a sequence from the PETS2009 dataset, the results of which are shown in Table VII. It can be observed that the increase in particle number has a bigger impact on the computational cost as compared with that on the OSPA results. Similar results have been observed for other sequences. In our simulations, the number of particles is selected empirically based on these experiments. We found that the number of particles chosen as 100 tends to provide a good compromise between run time and tracking performance.

The computational complexity of the proposed tracking system has also been considered. Compared with RFS, the particle PHD filter has a smaller computational cost since only the first moment of the posterior is employed instead of the posterior itself. However, the main growth of the time complexity is from the background subtraction and the OCSVM part of the proposed tracking system. If the times needed for determining the brightness and color conditions are denoted as T_B and T_C respectively, and the update time is T_U, the total processing time for a single image pixel can be expressed as

$$T = N_B T_B + N_C (T_B + T_C) + N_U (T_B + T_C + T_U) \quad (46)$$

where N_B is the number of codewords rejected after testing the brightness condition, N_C is the number of codewords rejected after testing both the brightness and color conditions and N_U is 1 if a matching codeword is found and 0 otherwise. Furthermore, the computational complexity of the OCSVM classifier is $O(m^3)$ where m is the number of the training patterns. As compared with the traditional PHD filter [24], the time complexity of the proposed algorithm becomes higher due to the introduction of the SFM and OCSVM steps. The average run-time (calculated using 20 video sequences from the three different datasets) is shown in Table VIII. This run-time comparison is made by implementing the algorithms with MATLAB (version R2015a) with a 3.4 GHz I5 processor.

TABLE VIII
RUN-TIME COMPARISONS FOR PROPOSED PHD FILTERS

	PHD [24]	SFM-MCMC-PHD	SFM-MCMC-OCSVM-PHD
Run-time/frame	1.39s	1.43s	1.95s

TABLE IX
OSPA COMPARISON OF THREE RECENT METHODS AND THE PROPOSED METHOD OVER 20 SEQUENCES

	PHD [24]	Method in [46]	Method in [45]	Proposed method
OSPA (pixels)	21.93	15.39	12.95	8.83

From this table, we can find that the run time for the traditional particle PHD filter is 1.39 s/frame, and the overhead for the social force model aided MCMC resampling step is 0.04 s/frame and for the one class SVM is 0.52 s/frame, the run-time increases 2.9% by employing the social force based MCMC resampling step, and grows 36% by employing the OCSVM. However, the tracking accuracy has been improved by 38.25% and 34.9% by employing them respectively. From the comparison, we find that the increase of time complexity is mainly due to the use of the OCSVM classifier when calculating the features from the color and oriented gradient histograms for each particle.

6) Comparison With State-of-the-art Methods: We also compared our work with two recent multiple human tracking methods proposed in [45] and [46]. In [45], online learning of non-linear motion patterns and robust appearance models are used for multiple target tracking and in [46] a background subtraction based multi-Bernoulli filtering method is proposed for visual tracking. The mean of the OSPA measure is employed for evaluation, which is shown in Table IX. All the three methods are evaluated on the 20 sequences from the CAVIAR, PETS2009 and TUD datasets.

From the comparison in the table we observe that the multi-Bernoulli filter proposed in [46] performs better than the particle PHD filter proposed in [24]. However, in [46], a kernel based

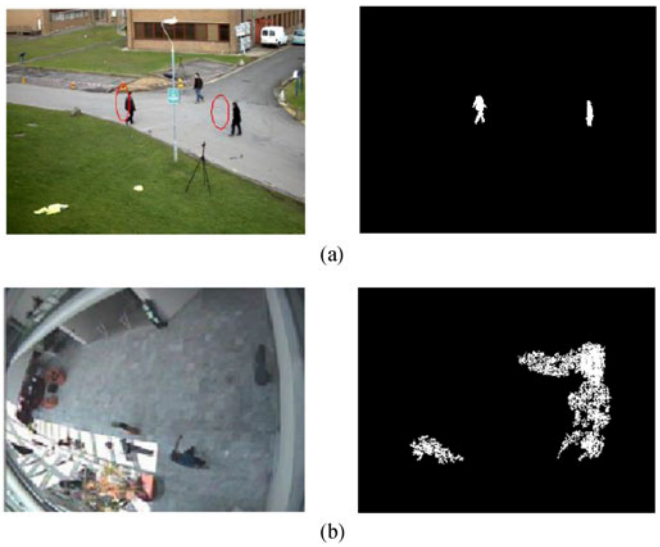

Fig. 7. Selected examples of failure cases of the proposed tracking system. In (a) the tracking results are influenced by the failure of detection from background subtraction, and in (b) the brightness of the environment causes the failure of the background subtraction.

background subtraction method was employed instead of the codebook method which we employed in our system. As such, the quality of the measurement is generally worse than that in our proposed system. Moreover, the social force model based MCMC resampling step provides more accurate predictions of target states. The appearance modelling based method in [45], however, does not address the challenge of varying number of targets hence it generates an OSPA value that is higher than our proposed method.

7) Examples of Tracking Failures: Nevertheless, the tracking results of the proposed system can be degraded by the following factors: increase in the number of targets and variations in lighting and color of the targets which will influence the results from background subtraction. For example, for the sequence Browse1 from the CAVIAR dataset, the lighting in the environment is varying, and there is a large amount of noise in the results obtained from the background subtraction, hence it fails to provide accurate foreground measurement, which also leads to possible failure of the tracking system. What is more, when human targets are very crowded in the visual scene, the social force model can be influenced by false alarms.

Fig. 7 shows selected failure cases of the proposed tracking system. We would like to note that these tracking scenarios pose common challenges to many existing methods such as [47]–[50] as well as the baseline methods [24], [46]. The codebook method for background subtraction is adopted in our study as it is the preference in other works in this field [17], [30] but there may be other approaches to further improve the detection results and hence the quality of the measurement set. However, a detailed comparison of such techniques is outside of the scope of this work.

The proposed method gives advantages over the baselines in certain scenarios especially with group target movements, varying number of targets, and large amount of environmental noise. However, to address the aforementioned common challenges, more powerful techniques in appearance modeling [51]–[53], background subtraction and occlusion handling [2] are required which will be the focus of our future research.

V. Conclusion

In this paper, we have presented a novel method based upon the particle PHD filter for multiple human tracking, where in the prediction step a novel social force model was used in an MCMC chain for resampling the particles; the posterior from the MCMC resampling was used as the posterior distribution in the particle updating step; in the particle likelihood calculation step, an OCSVM classifier was used to mitigate the adverse impact of measurement noise. The simulations showed improvement of our proposed method in terms of both localization and cardinality. In our future work, we will explore alternative classifiers, e.g. an online one-class classifier, for dealing with the situation where only a limited length of video sequence is available. In addition, sparsity in either the observed feature space or the parameter space of the tracking model could be exploited to further reduce the computational complexity of the proposed method.

Acknowledgment

The authors would like to thank the Associate Editor and the anonymous reviewers for their valuable input to improving the paper.

References

[1] C.-T. Chu, J. Hwang, H. Pai, and K. M. Lan, "Tracking human under occlusion based on adaptive multiple kernels with projected gradients," *IEEE Trans. Multimedia*, vol. 15, no. 7, pp. 1602–1615, Nov. 2013.

[2] T. Zhang, K. Jia, C. Xu, Y. Ma, and N. Ahuja, "Partial occlusion handling for visual tracking via robust part matching," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2014, pp. 1258–1265.

[3] X. Li *et al.*, "Visual tracking with spatio-temporal dempster-shafer information fusion," *IEEE Trans. Image Process.*, vol. 22, no. 8, pp. 3028–3040, Aug. 2013.

[4] M. Luber *et al.*, "People tracking with human motion predictions from social forces," in *Proc. IEEE Int. Conf. Robot. Autom.*, May 2010, pp. 464–469.

[5] X. Li *et al.* "A survey of appearance models in visual object tracking," *ACM Trans. Intell. Syst. Technol.*, vol. 4, no. 4, pp. 1–58, 2013.

[6] S. Liu, T. Zhang, X. Cao, and C. Xu, "Structural correlation filter for robust visual tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2016, pp. 4312–4320.

[7] E. Polat and M. Ozden, "A nonparametric adaptive tracking algorithm based on multiple feature distributions," *IEEE Trans. Multimedia*, vol. 8, no. 6, pp. 1156–1163, Dec. 2006.

[8] X. Li, A. Dick, C. Shen, A. Hengel, and H. Wang, "Incremental learning of 3D-DCT compact representations for robust visual tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 4, pp. 863–881, Apr. 2013.

[9] Z. Kalal, K. Mikolajczyk, and J. Matas, "Tracking-learning-detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 7, pp. 1409–1422, Jul. 2012.

[10] 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 Process.*, vol. 50, no. 2, pp. 2454–2467, Feb. 2002.

[11] M. Barnard *et al.*, " Robust multi-speacker tracking via dictrionry learning and identity modelling," *IEEE Trans. Multimedia*, vol. 16, no. 3, pp. 864–880, Apr. 2014.

[12] E. Maggio, M. Taj, and A. Cavallaro, "Efficient multitarget visual tracking using random finite sets," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 8, pp. 1016–1027, Aug. 2008.

[13] K. Panta, B. Vo, and S. Singh, "Novel data association schemes for the probability hypothesis density filter," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 43, no. 2, pp. 556–570, Apr. 2007.
[14] E. Maggio and A. Cavallaro, *Video Tracking*. Hoboken, NJ, USA: Wiley, 2011.
[15] V. Kilic, M. Barnard, W. Wang, and J. Kittler, "Audio assisted robust visual tracking with adaptive particle filtering," *IEEE Trans. Multimedia*, vol. 17, no. 2, pp. 186–200, Feb. 2015.
[16] S. Pellegrini, A. Ess, K. Schindler, and L. Gool, "You'll never walk alone: Modeling social behavior for multi-target tracking," in *Proc. IEEE 12th Int. Conf. Comput. Vis.*, Sep.-Oct. 2009, pp. 261–268.
[17] A. Ur-Rehman, S. M. Naqvi, L. S. Mihaylova, and J. Chambers, "Multi-target tracking and occlusion handling with learned variational bayesian clusters and a social force model," *IEEE Trans. Signal Process.*, vol. 64, no. 5, pp. 1320–1335, Mar. 2015.
[18] W. Gilks and C. Berzuini, "Following a moving target-monte carlo inference for dynamic Bayesian models," *J. Roy. Statist. Soc. Ser. B, Statist. Methodology*, 2001, pp. 127–146, 2001.
[19] S. Li, H. Wang, and T. Chai, "A t-distribution based particle filter for target tracking," in *Proc. Amer. Control Conf.*, 2006, pp. 1–6.
[20] X. Li, C. Shen, A. Dick, Z. Zhang, and Y. Zhuang, "Online metric-weighted linear representations for robust visual tracking," *IEEE Trans. Pattern Anal. Mach. Intelli.*, vol. 38, no. 5, pp. 931–950, May 2016.
[21] P. Cui, L. Sun, F. Wang, and S. Yang, "Contexual mixture tracking," *IEEE Trans. Multimedia*, vol. 11, no. 2, pp. 333–341, Feb. 2009.
[22] B. Ristic, S. Arulampalam, and N. Gordon, *Beyond the Kalman Filter: Particle Filters for Tracking Applications*. Norwood, MA, USA: Artech House, 2004.
[23] D. E. Clark and J. Bell, "Convergence results for the particle PHD filter," *IEEE Trans. Signal Process.*, vol. 54, no. 7, pp. 2652–2661, Jul. 2006.
[24] Y. Wang, J. Wu, A. A. Kassim, and W. Huang, "Tracking a variable number of human groups in video using probability hypothesis density," in *Proc. IEEE 18th Int. Conf. Pattern Recog.*, Aug. 2006, vol. 3, pp. 1127–1130.
[25] D. Helbing, I. Farkas, and T. Vicsek, "Simulating dynamical features of escape panic," *Nature*, vol. 407, no. 6803, pp. 487–490, 2000.
[26] D. Helbing, I. Farkas, P. Molnar, and T. Vicsek, "Simulation of pedestrian crowds in normal and evacuation situations," *Pedestrian Evacuation Dyn.*, vol. 21, no. 2, pp. 21–58, 2002.
[27] Z. Khan, T. Tucker, and F. Dellaert, "MCMC-based particle filtering for tracking a variable number of interacting targets," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 11, pp. 1805–1819, Nov. 2005.
[28] Z. Zhao and M. Kumar, "An MCMC-based particle filter for multiple target tracking," in *Proc. 15th Int. Conf. Inf. Fusion*, 2012, pp. 1676–1682.
[29] P. Feng, M. Yu, S. Naqvi, and J. Chambers, "Deep learning for posture analysis in fall detection," in *Proc. 19th IEEE Int. Conf. Digital Signal Process.*, Aug. 2014, pp. 12–17.
[30] K. Kim, T. H. Chalidabhongse, D. Harwood, and L. Davis, "Real-time foreground-background segmentation using codebook model," *Real-Time Imag.*, vol. 11, no. 3, pp. 172–185, 2005.
[31] P. Feng, W. Wang, S. M. Naqvi, S. Dlay, and J. A. Chambers, "Social force model aided robust particle PHD filter for multiple human tracking," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Mar. 2016, pp. 1–5.
[32] P. Feng, W. Wang, S. M. Naqvi, and J. A. Chambers, "Adaptive retrodiction particle PHD filter for multiple human tracking," *IEEE Signal Process. Lett.*, vol. 23, no. 11, pp. 1592–1596, Nov. 2016.
[33] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural-Comput.*, vol. 13, no. 7, pp. 1443–1471, 2001.
[34] P. Feng, M. Yu, S. M. Naqvi, W. Wang, and J. A. Chambers, "A robust student's-t distribution PHD filter with OCSVM updating for multiple human tracking," in *Proc. Eur. Signal Process. Conf.*, 2015, pp. 1–5.
[35] S. Zhang, X. Yu, Y. Sui, S. Zhao, and L. Zhang, " Object tracking with multi-view support vector machines," *IEEE Trans. Multimedia*, vol. 17, no. 3, pp. 265–278, Mar. 2015.
[36] I. Goldberg and M. J. Atallah, "Privacy enhancing technologies," in *Proc. 9th Int. Symp.*, 2009. [Online]. Available: http://dx.doi.org/10.1007/978-3-642-03168-7
[37] R. Fisher, "Caviar case scenarios," 2003, accessed on 2013. [Online]. Available: http://groups.inf.ed.ac.uk/vision/CAVIAR/CAVIARDATA1/
[38] M. Andriluka, S. Roth, and B. Schiele, "Monocular 3D pose estimation and tracking by detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2010, pp. 623–630.
[39] D. Schuhmacher, B. T. Vo, and B. Vo, "A consistent metric for performance evaluation of multi-object filters," *IEEE Trans. Signal Process.*, vol. 56, no. 8, pp. 3447–3457, Aug. 2008.
[40] B. Ristic, B.-N. Vo, and D. Clark, "Performance evaluation of multi-target tracking using the OSPA metric," in *Proc. 13th IEEE Int. Conf. Inf. Fusion*, Jul. 2010, pp. 1–7.
[41] T. Vu and R. Evans, "A new performance metric for multiple target tracking based on optimal subpattern assignment," in *Proc. 17th IEEE Int. Conf. Inf. Fusion*, Jul. 2014, pp. 1–8.
[42] K. Bernardin and R. Stiefelhagen, "Evaluating multiple object tracking performance: The clear MOT metrics," *J. Image Video Process.*, vol. 2008, pp. 1–10, 2008.
[43] H. B. Shitrit, J. Berclaz, F. Fleuret, and P. Fua, "Multi-commodity network flow for tracking multiple people," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 8, pp. 1614–1627, Aug. 2014.
[44] G. A. Bluman, *Elementary Statistics*. New York, NY, USA: McGraw Hill, 2013.
[45] B. Yang and R. Nevatia, "Multi-target tracking by online learning of non-linear motion patterns and robust appearance models," in *Proc. IEEE Comput. Vis. Pattern Recog.*, Jun. 2012, pp. 1918–1928.
[46] R. Hoseinnezhad, B. N. Vo, and B. T. Vo, "Visual tracking in background subtracted image sequences via multi-Bernoulli filtering," *IEEE Trans. Signal Process.*, vol. 61, no. 2, pp. 392–397, Jan. 2013.
[47] E. Maggio, E. Piccardo, C. Regazzoni, and A. Cavallaro, "Particle PHD filtering for multi-target visual tracking," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Apr. 2007, vol. 1, pp. 1101–1104.
[48] L. Wang, N. Yung, and L. Xu, "Multiple human tracking by iterative data association and detection step," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 5, pp. 1886–1899, Oct. 2014.
[49] C. Huang, Y. Li, and R. Nevatia, "Multiple target tracking by learning-based hierarchical association of detection responses," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 4, pp. 898–910, Apr. 2013.
[50] L. Wen *et al.*, "UA-DETRAC: A new benchmark and protocol for multi-object detection and tracking," *CoRR,* 2016.
[51] T. Zhang, S. Liu, N. Ahuja, M. Yang, and B. Ghanem, "Robust visual tracking via consistent low-rank sparse learning," *Int. J. Comput. Vis.*, vol. 111, no. 2, pp. 171–190, 2015.
[52] T. Zhang, B. Ghanem, S. Liu, and N. Ahuja, "Robust visual tracking via structured multi-task sparse learning," *Int. J. Comput. Vis.*, vol. 101, no. 2, pp. 367–383, 2013.
[53] T. Zhang *et al.*, "Structural sparse tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Jun. 2015, pp. 150–158.

Pengming Feng was born in Jilin, China. He received the B.Sc. degree in automatic control from the Beijing University of Chemical Technology, Beijing, China, in 2012, the M.Sc. degree in digital communication systems from Loughborough University, Loughborough, U.K., in 2013, and the Ph.D. degree in intelligent signal processing from Newcastle University, Newcastle upon Tyne, U.K., in 2016.

During his Ph.D. program, he worked as part of the University Defence Research Collaboration, sponsored by the U.K. Defence Science and Technology Laboratory and Engineering and Physical Science Research Council, on the project entitled "Signal Processing in Networked Battlespace." His research interests include multiple target tracking, machine learning, and sparse representation.

Wenwu Wang (M'02–SM'11) was born in Anhui, China. He received the B.Sc. degree in automatic control in 1997, the M.E. degree in control science and control engineering in 2000, and the Ph.D. degree in navigation guidance and control in 2002, all from the Harbin Engineering University, Harbin, China.

In May 2002, he then joined Kings College, London, U.K., as a Postdoctoral Research Associate and transferred to Cardiff University, Cardiff, U.K., in January 2004, where he worked in the area of blind signal processing. In May 2005, he joined the Tao Group Ltd. (now Antix Labs Ltd.), Reading, U.K., as a DSP Engineer working on algorithm design and implementation for real-time and embedded audio and visual systems. In September 2006, he joined Creative Labs, Ltd., Egham, U.K., as an R&D Engineer, working on 3D spatial audio for mobile devices. Since May 2007, he has been with the Centre for Vision Speech and Signal Processing, University of Surrey, Guildford, U.K., where he is currently a Reader in *Signal Processing*, and a Co-Director of the Machine Audition Laboratory. During spring 2008, he was a Visiting Scholar with the Perception and Neurodynamics Laboratory and the Center for Cognitive Science, Ohio State University, Columbus, OH, USA. He has authored or coauthored more than 150 publications, including two books, *Machine Audition: Principles, Algorithms and Systems* (IGI Global, 2010) and *Blind Source Separation: Advances in Theory, Algorithms and Applications* (Springer, 2014). His current research interests include blind signal processing, sparse signal processing, audio–visual signal processing, machine learning and perception, machine audition (listening), and statistical anomaly detection.

Dr. Wang is a Member of the Ministry of Defence University Defence Research Collaboration (UDRC) in *Signal Processing* (since 2009), a Member of the BBC Audio Research Partnership (since 2011), an Associate Member of Surrey Centre for Cyber Security (since 2014), and a Member of the MRC/EPSRC Microphone Network (since 2015). He is currently an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING. He is also Publication Co-Chair of ICASSP 2019 (to be held in Brighton, U.K.). He was a Tutorial Speaker at ICASSP 2013, UDRC Summer Schools 2014, 2015, and 2016, and SpaRTan/MacSeNet Spring School 2016.

Satnam Dlay (M'15) received the B.Sc. (Hons.) degree in electrical and electronic engineering and the Ph.D. degree in VLSI design from Newcastle University, Newcastle upon Tyne, U.K.

In 1984, he was a Postdoctoral Research Associate with Newcastle University and helped to establish an Integrated Circuit Design Centre, funded by the EPSRC. In 1984, he was a Lecturer with the Department of Electronic Systems Engineering, University of Essex, Essex, U.K. In 1986, he rejoined the Newcastle University as a Lecturer with the School of Electrical, Electronic and Computer Engineering, where he was promoted to Senior Lecturer in 2001. In recognition of his major achievements, he has been appointed to a Personal Chair in Signal Processing Analysis. He has authored more than 250 research papers ranging from biometrics and security, biomedical signal processing, and implementation of signal processing architectures.

Prof. Dlay is a College Member of the Engineering and Physical Science Research Council (EPSRC). He serves on many editorial boards and has played an active role in numerous international conferences in terms of serving on technical and advisory committees as well as organizing special sessions. He was a recipient of a scholarship from the EPSRC and the Charles Hertzmann Award.

Syed Mohsen Naqvi (S'07–M'10–SM'15) received the Ph.D. degree in signal processing from Loughborough University, Loughborough, U.K., in 2009.

He was a Postdoctoral Research Associate with the EPSRC U.K.-funded projects and REF Lecturer from 2009 to 2015. Prior to his postgraduate studies in Cardiff and Loughborough Universities, U.K., he served the National Engineering and Scientific Commission, Islamabad, Pakistan, from January 2002 to September 2005. He is a Lecturer of signal and information processing with the School of Electrical and Electronic Engineering, Newcastle University, Newcastle upon Tyne, U.K. He has more than 70 publications with the main focus of his research being on multimodal (audio–video) signal and information processing. He has successfully (co)-supervised 15 Ph.D. and M.Sc. students and is (co)-supervising 10 Ph.D. and M.Sc. students. He organized special sessions on multi-target tracking in FUSION 2013 and 2014, delivered seminars, and was a speaker at UDRC Summer School 2015 and 2016. His research interests include multimodal processing for human behaviour analysis, multitarget tracking, and source separation, all for machine learning.

Dr. Naqvi is Fellow of the Higher Education Academy, a Member of the University Defence Research Collaboration (UDRC) in Signal Processing, and a Member of the International Society of Information Fusion.

Jonathon Chambers (S'85–M'85–SM'98–F'11) received the Ph.D. and D.Sc. degrees in signal processing from the Imperial College of Science, Technology, and Medicine, Imperial College London, London, U.K., in 1990 and 2014, respectively.

From 1991 to 1994, he was a Research Scientist with the Schlumberger Cambridge Research Centre, Cambridge, U.K. In 1994, he returned to Imperial College London as a Lecturer in signal processing and was promoted to a Reader (Associate Professor) in 1998. From 2001 to 2004, he was the Director of the Centre for Digital Signal Processing and a Professor in signal processing with the Division of Engineering, King's College London, London, U.K. From 2004 to 2007, he was a Cardiff Professorial Research Fellow with the School of Engineering, Cardiff University, Cardiff, U.K. Between 2007 and 2014, he led the Advanced Signal Processing Group, School of Electronic, Electrical, and Systems Engineering, Loughborough University, Loughborough, U.K., where he is currently a Visiting Professor. In 2015, he joined the School of Electrical and Electronic Engineering, Newcastle University, Newcastle upon Tyne, U.K., where he is a Professor of signal and information processing and the Head of the Communications, Sensors, Signal and Information Processing Group. He is also a Guest Professor with Harbin Engineering University, Harbin, China. He is a coauthor of the books *Recurrent Neural Networks for Prediction: Learning Algorithms, Architectures and Stability* (Wiley, 2001) and *EEG Signal Processing* (Wiley, 2007). He has advised more than 70 researchers through to Ph.D. graduation, and authored or coauthored more than 500 conference and journal articles, many of which are in IEEE journals. His research interests include adaptive signal processing and machine learning and their applications.

Dr. Chambers is a Fellow of the Royal Academy of Engineering, U.K. He was the Technical Program Chair of the 15th International Conference on Digital Signal Processing and the 2009 IEEE Workshop on Statistical Signal Processing, both held in Cardiff, U.K., and a Technical Program Co-Chair for the 36th IEEE International Conference on Acoustics, Speech, and Signal Processing, Prague, Czech Republic. He has served on the IEEE Signal Processing Theory and Methods Technical Committee for six years, and the IEEE Signal Processing Society Awards and Conference Boards for three years. He is currently a Member of the IEEE Signal Processing Society Jack Kilby Medal Committee. He has also served as an Associate and Senior Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING for three terms. He was the recipient of the first QinetiQ Visiting Fellowship in 2007, for his outstanding contributions to adaptive signal processing and his contributions to QinetiQ, as a result of his successful industrial collaboration with the international defence systems company QinetiQ.