Article

Tracking Multiple Video Targets with an Improved GM-PHD Tracker

Xiaolong Zhou [1,2], Hui Yu [3], Honghai Liu [2,*] and Youfu Li [4]

[1] College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China; zxl@zjut.edu.cn
[2] School of Computing, University of Portsmouth, Portsmouth PO1 3HE, UK
[3] School of Creative Technologies, University of Portsmouth, Portsmouth PO1 2DJ, UK; hui.yu@port.ac.uk
[4] Department of Mechanical and Biomedical Engineering, City University of Hong Kong, Hong Kong, China; meyfli@cityu.edu.hk
* Correspondence: honghai.liu@port.ac.uk; Tel./Fax: +44-23-9284-5541

Abstract: Tracking multiple moving targets from a video plays an important role in many vision-based robotic applications. In this paper, we propose an improved Gaussian mixture probability hypothesis density (GM-PHD) tracker with weight penalization to effectively and accurately track multiple moving targets from a video. First, an entropy-based birth intensity estimation method is incorporated to eliminate the false positives caused by noisy video data. Then, a weight-penalized method with multi-feature fusion is proposed to accurately track the targets in close movement. For targets without occlusion, a weight matrix that contains all updated weights between the predicted target states and the measurements is constructed, and a simple, but effective method based on total weight and predicted target state is proposed to search the ambiguous weights in the weight matrix. The ambiguous weights are then penalized according to the fused target features that include spatial-colour appearance, histogram of oriented gradient and target area and further re-normalized to form a new weight matrix. With this new weight matrix, the tracker can correctly track the targets in close movement without occlusion. For targets with occlusion, a robust game-theoretical method is used. Finally, the experiments conducted on various video scenarios validate the effectiveness of the proposed penalization method and show the superior performance of our tracker over the state of the art.

Keywords: robot vision; video targets tracking; probability hypothesis density; weight penalization; multi-feature fusion

1. Introduction

Tracking targets in video is an ever-increasing field of research with a wide spectrum of applications in vision-based robotic intelligence, including robot navigation, intelligent surveillance, human behaviour understanding, human-robot interactions, and so on. Despite many excellent research works [1–5] having been explored, an effective and accurate solution to the problem remains challenging.

Recently, the random finite set approach for target tracking [6–17] has attracted considerable attention. The probability hypothesis density (PHD) filter [6] uses the first-order statistical moment of the multi-target posterior density, providing a computationally-tractable alternative to data association. However, it is generally intractable due to the "curse of dimensionality" in numerical integration. The Gaussian mixture PHD filter (GM-PHD) [7] does not suffer from this problem, because its posterior intensity function can be propagated analytically in time.

Although the GM-PHD filter originates from radar tracking [7–10], recently, it has been widely explored for visual tracking [11–17]. For simplicity, the GM-PHD filter-based tracker is called the GM-PHD tracker in this paper. For example, Pham *et al.* [11] used the GM-PHD tracker to track multiple objects from colour images. They showed that the PHD was proportional to the approximated density from colour likelihood. They also used this GM-PHD tracker to track 3D locations of heads of people using multiple cameras [12]. Wu and Hu [13] combined the modified detection method with the PHD filter to build a multi-target visual tracking framework. They first generated observations by detecting the foreground objects and then estimated the target state using a GM-PHD filter. Furthermore, Wu *et al.* [14] proposed an auction algorithm to calculate target trajectories automatically. Zhou *et al.* [15] incorporated entropy distribution into the GM-PHD filter to automatically and efficiently estimate the birth intensity and, finally, robustly tracked the newborn video targets. Furthermore, they used game theory to handle the mutual occlusion problem and proposed an integrated system to robustly track the multiple video targets [16]. Pollard *et al.* [17] used a homographic transformation to compensate the camera motion and to combine geometric and intensity-based criteria for object detection and combined the GMC-PHD filter to track the targets from an aerial video.

Despite significant progress of the GM-PHD tracker, robust and reliable tracking of multiple targets in video is still far from being solved, especially in noisy video data and tracking targets in close movement.

To eliminate the noisy data in the video, the tracker should have the ability of accurately determining the birth intensity of the newborn targets in the GM-PHD filter. Conventionally, the birth intensity must cover the whole state space [18] when no prior localization information on the newborn targets was available. Such a requirement entails a high computational cost and can easily be interfered by clutters. To remedy this, Maggio *et al.* [19] assumed that the birth of a target occurred in a limited space around the measurements. They drew the newborn particles from the centre of the measurement set. However, the proposed method could easily be interfered by clutters and the measurements originating from the survival targets. Recently, Zhou *et al.* [15] proposed an effective method based on entropy distribution to automatically and correctly estimate the birth intensity. They first initialized the birth intensity using the previously-obtained target states and measurements and then updated it using the currently-obtained measurements. The entropy distribution was incorporated to remove those noises that were irrelevant to the measurements, and the coverage rate was computed to further eliminate the noises.

Generally, each measurement is assumed to correspond to one target and *vice versa* in multi-target tracking. This so-called one-to-one assumption expresses that a target can only be associated with one measurement. However, in the GM-PHD tracker, this one-to-one assumption is violated whenever multiple measurements are close to one target. In other words, the efficiency of the GM-PHD tracker may degrade when targets come near each other. To remedy this, Yazdian-Dehkordi *et al.* proposed a competitive GM-PHD (CGM-PHD) tracker [20] and a penalized GM-PHD (PGM-PHD) tracker [21] to refine the weights of the close moving targets in the update step in the GM-PHD filter. However, they did not provide continuous trajectories for the targets. By considering this point, Wang *et al.* [22] proposed a collaborative penalized GM-PHD (CPGM-PHD) tracker, in which they utilized the track label of each Gaussian component to collaboratively penalize the weights of those close moving targets with the same identity. However, the aforementioned trackers are merely suitable for point target tracking, which may fail in video target tracking. Compared to the simple point representations of the target state and the measurement in point target tracking, the representations are more complicated in video target tracking. Both the location and the size of video targets are considered for modelling the target state and the measurement. As video targets move closely, the aforementioned trackers (GM-PHD tracker, CGM-PHD tracker, PGM-PHD tracker and CPGM-PHD tracker) may track multiple targets with the same identity (shown as in Figure 1a) or with switched identities (shown as in Figure 1b).

As targets move close enough, mutual occlusion may occur. As a result, the measurements originating from targets within the occlusion region will be merged into one measurement. Without an occlusion handling method, the tracker may fail to track them. Because occlusion handling is not the main contribution of this paper, we incorporate our previous reported game-theoretical method [23] into the tracker to solve the mutual occlusion problem. In this paper, we propose an improved GM-PHD tracker to robustly track targets in a video, especially to track targets in close movement. The pipeline of the proposed tracker is shown in Figure 2, and the main contributions are listed as follows.

(1) An improved GM-PHD tracker with multi-feature fusion-based weight penalization is proposed to effectively track targets in a video, especially to track the targets in close movement.

(2) A weight matrix of all updated weights is constructed, and an effective ambiguous weights determination method is proposed. The conventional trackers (the CGM-PHD, PGM-PHD and CPGM-PHD trackers) only consider the total weight for ambiguous weights determination, which is not applicable for Case 2. In contrast, we utilize the total weight and predicted target states to effectively determine the ambiguous weights for Case 1 and Case 2, respectively. In this paper, Case 1 is the case that one target is associated with multiple measurements; while Case 2 is the case that one target is associated with one incorrect measurement. More details of Case 1 and Case 2 are stated in Section 2.3.

(3) Multiple features that include spatial-colour appearance, histogram of oriented gradient and target area are fused and incorporated into the tracker to penalize the ambiguous weights. By doing so, the weights of the mismatched targets can be greatly reduced, and thus, the tracking accuracy is improved.

(a)

(b)
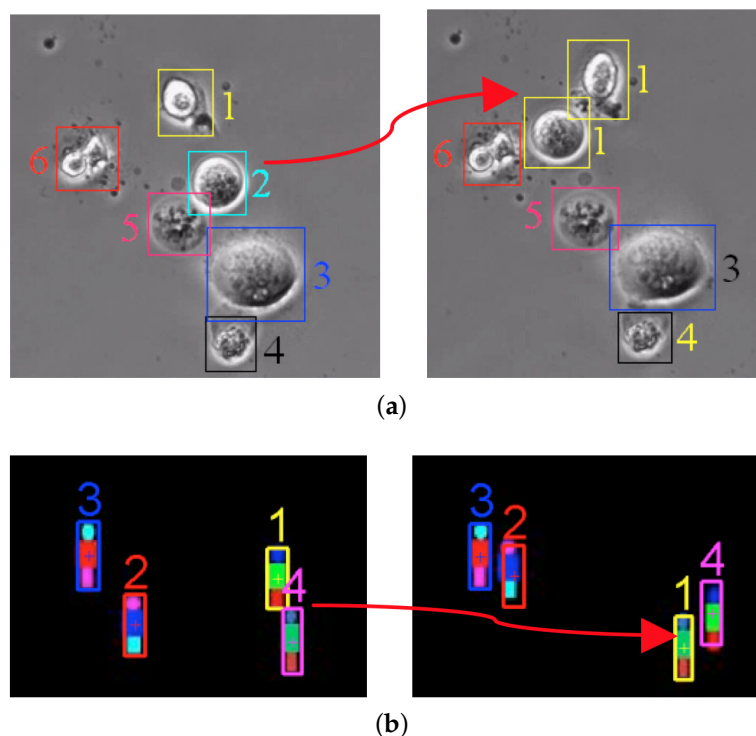
Figure 1. Tracking targets in close movement with the conventional Gaussian mixture probability hypothesis density (GM-PHD) tracker. (**a**) Mistracking two cells (Cells 1 and 2 in the left image) with the same identity (Cell 1 in the right image); (**b**) mistracking two targets (Targets 1 and 4 in the left image) with switched identities (Targets 4 and 1 in the right image).
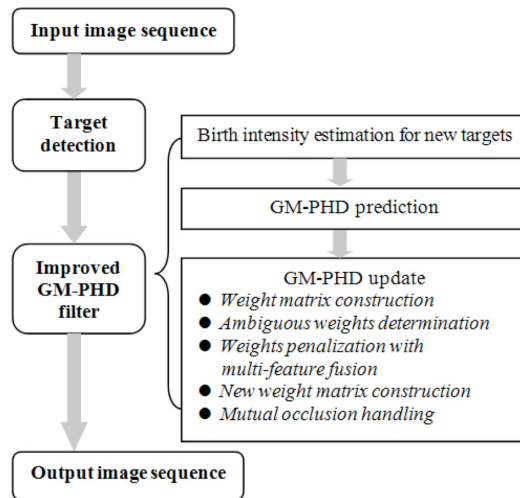
Figure 2. Pipeline of the proposed tracker.

The rest of this paper is organized as follows. Section 2 simply introduces the GM-PHD filter and its drawbacks. Section 3 presents the proposed tracker in detail. Some experimental results are discussed in Section 4, followed by concluding remarks in Section 5.

2. Problem Formulation

2.1. Target State and Measurement Representation

For an input image sequence, the kinematic state of a target i at time t is denoted by $\mathbf{x}_t^i = \{\mathbf{l}_t^i, \mathbf{v}_t^i, \mathbf{s}_t^i\}$, where $\mathbf{l}_t^i = \{l_{x,t}^i, l_{y,t}^i\}$, $\mathbf{v}_t^i = \{v_{x,t}^i, v_{y,t}^i\}$ and $\mathbf{s}_t^i = \{w_t^i, h_t^i\}$ are the location, velocity and bounding box size of the target, respectively; $i = 1, \cdots, N_t$, and N_t denotes the number of targets at time t. The measurement originating from a target j at time t is denoted by $\mathbf{z}_t^j = \{\mathbf{l}_{z,t}^j, \mathbf{s}_{z,t}^j\}$, where $j = 1, \cdots, N_{m,t}$, and $N_{m,t}$ denotes the number of measurements at time t. The target state set and measurement set at time t are denoted by $\mathbf{X}_t = \{\mathbf{x}_t^1, \cdots, \mathbf{x}_t^{N_t}\}$ and $\mathbf{Z}_t = \{\mathbf{z}_t^1, \cdots, \mathbf{z}_t^{N_{m,t}}\}$, respectively. The measurements are obtained by object detection, and any object detection method can be used in our tracker. To show the robust performance of the proposed tracker, a simple background subtraction algorithm [15] is utilized to obtain the measurements.

2.2. The GM-PHD Filter

The GM-PHD filter was first proposed by Vo and Ma [7] in 2006. It is a closed-form solution to the PHD filter recursion, whose posterior intensity function is estimated by a sum of weighted Gaussian components that can be propagated analytically in time. More details of the GM-PHD filter are in the literature [7]. Generally, the GM-PHD filter can be implemented in the prediction and update steps.

Step 1: Prediction. Suppose that PHD $D_{t-1}(\mathbf{x}_{t-1})$ at time $t-1$ has the form $D_{t-1}(\mathbf{x}_{t-1}) = \sum_{i=1}^{J_{t-1}} \omega_{t-1}^{(i)} \mathbf{N}(\mathbf{x}_{t-1}; \mathbf{m}_{t-1}^{(i)}, \mathbf{P}_{t-1}^{(i)})$, then the predicted PHD $D_{t|t-1}(\mathbf{x}_t)$ is given by:

$$D_{t|t-1}(\mathbf{x}_t) = \gamma_t(\mathbf{x}_t) + p_{sv} \sum_{i=1}^{J_{t-1}} \omega_{t-1}^{(i)} \mathbf{N}(\mathbf{x}_t; \mathbf{m}_{sv,t|t-1}^{(i)}, \mathbf{P}_{sv,t|t-1}^{(i)}) \tag{1}$$

where $\mathbf{m}_{sv,t|t-1}^{(i)} = \mathbf{F}_{t-1}\mathbf{m}_{t-1}^{(i)}$ and $\mathbf{P}_{sv,t|t-1}^{(i)} = \mathbf{Q}_{t-1} + \mathbf{F}_{t-1}\mathbf{P}_{t-1}^{(i)}\mathbf{F}_{t-1}^T$. $\gamma_t(\mathbf{x}_t)$ and p_{sv} denote the probabilities of newborn targets and survival targets, respectively. $\mathbf{N}(\cdot; \mathbf{m}, \mathbf{P})$ denotes a Gaussian component with the mean $\mathbf{m}$ and covariance $\mathbf{P}$. $\mathbf{F}_{t-1}$ is the motion transition matrix.

Step 2: Update. The predicted PHD can be expressed as a Gaussian mixture $D_{t|t-1}(\mathbf{x}_t) = \sum_{i=1}^{J_{t|t-1}} \omega_{t|t-1}^{(i)} \mathbf{N}(\mathbf{x}_t; \mathbf{m}_{t|t-1}^{(i)}, \mathbf{P}_{t|t-1}^{(i)})$, then the posterior PHD $D_t(\mathbf{x}_t)$ at time t is given by:

$$D_t(\mathbf{x}_t) = (1 - p_d)D_{t|t-1}(\mathbf{x}_t) + \sum_{\mathbf{z}_t \in \mathbf{Z}_t} D_{g,t}(\mathbf{x}_t; \mathbf{z}_t) \tag{2}$$

$$D_{g,t}(\mathbf{x}_t; \mathbf{z}_t) = \sum_{i=1}^{J_{t|t-1}} \omega_{g,t}^{(i)}(\mathbf{z}_t) \mathbf{N}(\mathbf{x}_t; \mathbf{m}_{g,t}^{(i)}(\mathbf{z}_t), \mathbf{P}_{g,t}^{(i)}(\mathbf{z}_t)) \tag{3}$$

$$\omega_{g,t}^{(i)}(\mathbf{z}_t) = \frac{p_d \omega_{t|t-1}^{(i)} \mathbf{N}(\mathbf{z}_t; \mathbf{m}_{h,t}^{(i)}, \mathbf{P}_{h,t}^{(i)})}{\lambda_t c_t(\mathbf{z}_t) + p_d \sum\limits_{i=1}^{J_{t|t-1}} \omega_{t|t-1}^{(i)} \mathbf{N}(\mathbf{z}_t; \mathbf{m}_{h,t}^{(i)}, \mathbf{P}_{h,t}^{(i)})} \tag{4}$$

where $\mathbf{m}_{g,t}^{(i)}(\mathbf{z}_t) = \mathbf{m}_{t|t-1}^{(i)} + K(\mathbf{z}_t - \mathbf{H}_t \mathbf{m}_{t|t-1}^{(i)})$, $K = \mathbf{P}_{t|t-1}^{(i)} \mathbf{H}_t^T (\mathbf{H}_t \mathbf{P}_{t|t-1}^{(i)} \mathbf{H}_t^T + \mathbf{R}_t)^{-1}$, $\mathbf{P}_{g,t}^{(i)}(\mathbf{z}_t) = (\mathbf{I} - K\mathbf{H}_t)\mathbf{P}_{t|t-1}^{(i)}$, $\mathbf{m}_{h,t}^{(i)} = \mathbf{H}_t \mathbf{m}_{t|t-1}^{(i)}$, $\mathbf{P}_{h,t}^{(i)} = \mathbf{H}_t \mathbf{P}_{t|t-1}^{(i)} \mathbf{H}_t^T + \mathbf{R}_t$. p_d is the detection probability. λ_t and $c_t(\mathbf{z}_t)$ are the average rate and probability density of the spatial distribution of Poisson distributed clutters, respectively. $\mathbf{H}_t$ and $\mathbf{R}_t$ are the measurement matrix and the covariance matrix of the measurement noise, respectively.

To predict the newborn targets, we need to find the peak (the mean of Gaussian) of intensity $\gamma_t(\mathbf{x}_t)$, *i.e.*, the position where the targets are most probable to appear. To automatically and accurately estimate the birth intensity, our previous work [15] is utilized in this paper. Furthermore, we employ the pruning and merging algorithms [7] to prune the irrelevant components and to merge the same intensity components into one component. The peaks of the intensity are the points of the highest local concentration of the expected number N_t of targets. Finally, we can estimate the target states with N_t ordered mean with the largest weights.

2.3. Drawbacks of the GM-PHD Filter

The GM-PHD filter recursively propagates the first-order moment associated with the multi-target posterior density to avoid the complicated data association problem and, consequently, can be efficiently used in multiple video targets' tracking. However, as targets come near each other, multiple measurements may associate with one target or incorrect targets. Normally, each predicted state $\mathbf{x}_{t|t-1}^{(i)}$ of target i is associated with only one measurement $\mathbf{z}_t^j$ originating from target i, which means that the weight of the i-th predicted target updated by the jt-h measurement should be far greater than those weights updated by other measurements. However, in the real-world scenarios, two possible cases could violate this one-to-one association. As a result, the GM-PHD filter may track multiple targets with the same identity or with switched identities. Figure 3 is a pictorial example of the aforementioned two cases when tracking two targets in close movement.

Case 1: One predicted target (shown as target $\mathbf{x}_{t|t-1}^1$ in Figure 3a) may be associated with more than one measurement (shown as the measurements $\mathbf{z}_t^1$ and $\mathbf{z}_t^2$ in Figure 3a). In such a case, there are at least two updated weights for the same target (shown as $\bar{\omega}_t^{(1,1)}$ and $\bar{\omega}_t^{(1,2)}$ in Figure 3a), whose values are far greater than other updated weights. $\bar{\omega}_t^{(i,j)}$ is the normalized weight of target i updated by measurement j. For simplicity, indices i and j are used to represent the i-th predicted target state $\mathbf{x}_{t|t-1}^i$ and the j-th measurement $\mathbf{z}_t^j$, respectively. As a result, the GM-PHD tracker tracks Targets 1 and 2 with the same Identity 1 (shown as the right image in Figure 3a).

Case 2: one predicted target may be associated with another measurement that is not originated from this target. As shown in Figure 3b, measurement $\mathbf{z}_t^1$ should theoretically be associated with Target 1, while measurement $\mathbf{z}_t^2$ should be associated with Target 2. However, $\bar{\omega}_t^{(1,2)}$ is actually greater than $\bar{\omega}_t^{(1,1)}$, while $\bar{\omega}_t^{(2,1)}$ is greater than $\bar{\omega}_t^{(2,2)}$. As a result, the GM-PHD tracker tracks Targets 1 and 2 with switched Identities 2 and 1, respectively (shown as the right image in Figure 3b).
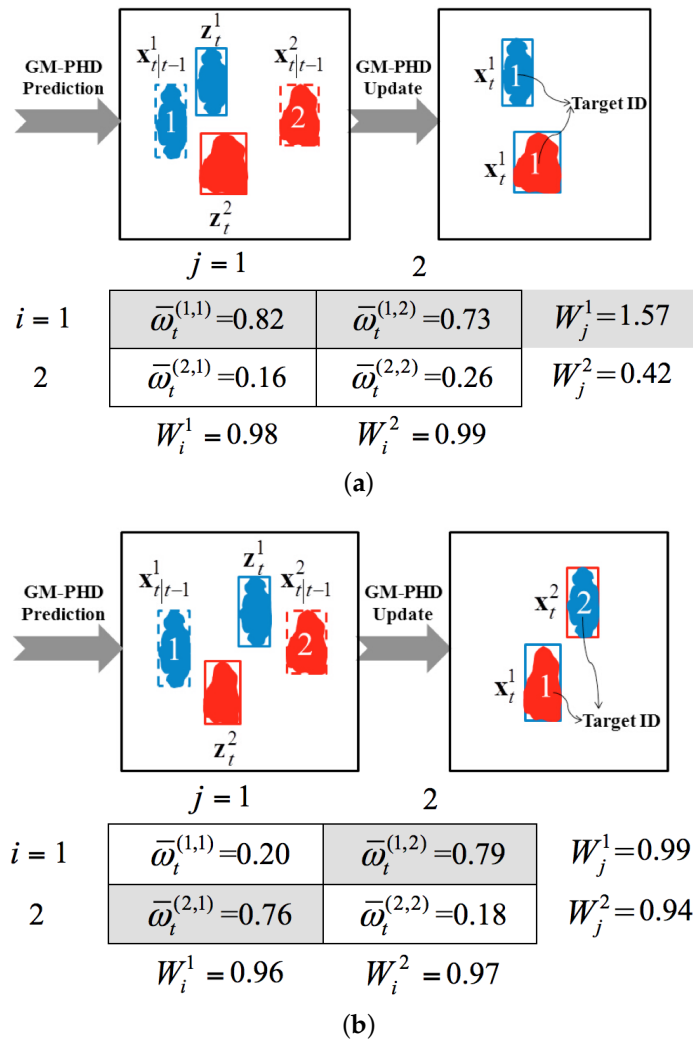
Figure 3. A pictorial example of tracking two targets in close movement. (**a**) Case 1: two targets with the same identity; (**b**) Case 2: two targets with switched identities.

To improve the aforementioned drawbacks, an improved GM-PHD tracker with weight penalization is proposed.

3. Improved GM-PHD Tracker with Weight Penalization

The way of improving the drawbacks is to penalize the weights of those targets that move closely. First, a weight matrix that consists of all updated weights is constructed. Then, an ambiguous weight is defined, and the corresponding methods for searching ambiguous weights are proposed. Finally, multiple features are fused and incorporated into the tracker to penalize the ambiguous weights.

3.1. Weight Matrix Construction

Figure 4 is a symbolic representation of updated weights. For better clarification, the matrix that includes the weights of all targets updated by all measurements is called the weight matrix (shown as in Figure 4). In the weight matrix, the *i*-th row represents the weights of the *i*-th predicted target updated by all measurements, while the *j*-th column represents the weights of all predicted targets updated by the *j*-th measurement. $W_j^i = \sum_{j=1}^{N_{m,t}} \bar{\omega}_t^{(i,j)}$ in the figure is the total weight of the *i*-th row,

while $W_i^j = \sum_{i=1}^{J_{t|t-1}} \bar{\omega}_t^{(i,j)}$ is the total weight of the *j*-th column. $N_{m,t}$ and $J_{t|t-1}$ are the numbers of measurements and predicted target states, respectively.

$$\bar{\omega}_t^{(i,j)} = \frac{p_d \omega_{t|t-1}^{(i)} \mathbf{N}(\mathbf{z}_t^j; \mathbf{m}_{h,t}^{(i)}, \mathbf{P}_{h,t}^{(i)})}{\lambda_t c_t(\mathbf{z}_t^j) + p_d \sum_{i=1}^{J_{t|t-1}} \omega_{t|t-1}^{(i)} \mathbf{N}(\mathbf{z}_t^j; \mathbf{m}_{h,t}^{(i)}, \mathbf{P}_{h,t}^{(i)})} \tag{5}$$

	$j = 1$	2	$\cdots$	$N_{m,t}$	
$i = 1$	$\bar{\omega}_t^{(1,1)}$	$\bar{\omega}_t^{(1,2)}$	$\cdots$	$\bar{\omega}_t^{(1,N_{m,t})}$	W_j^1
2	$\bar{\omega}_t^{(2,1)}$	$\bar{\omega}_t^{(2,2)}$	$\cdots$	$\bar{\omega}_t^{(2,N_{m,t})}$	W_j^2
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$J_{t|t-1}$	$\bar{\omega}_t^{(J_{t|t-1},1)}$	$\bar{\omega}_t^{(J_{t|t-1},2)}$	$\cdots$	$\bar{\omega}_t^{(J_{t|t-1},N_{m,t})}$	$W_j^{J_{t|t-1}}$
	$W_i^1 \approx 1$	$W_i^2 \approx 1$	$\cdots$	$W_i^{N_{m,t}} \approx 1$	

Figure 4. Weight matrix: a symbolic representation of updated weights.

3.2. Ambiguous Weight Determination

As stated in Section 2.2, the peaks of the updated GM-PHD are the points of the highest local concentration of the expected number N_t of targets. However, an incorrect estimate of the multi-target state may be obtained when targets move in a close space (as explained in the cases listed in Section 2.3). To remedy this, the incorrect weights should be penalized. In this paper, the weights of those close moving targets are defined as the ambiguous weights. Before penalization, the weight matrix should be analysed first to determine the ambiguous weights. In the CGM-PHD tracker [20], PGM-PHD tracker [21] and CPGM-PHD tracker [22], the weight of target *i* is determined as an ambiguous weight once the total weight W_j^i of the *i*-th row is greater than one. However, this method is not applicable to Case 2 (as stated in Section 2.3) since the total weight W_j^i may be less than one when targets approach each other. To remedy this, both the total weight W_j^i and the predicted target states are utilized to determine the ambiguous weights of Case 1 and Case 2, respectively.

(1) Ambiguous weights' determination for Case 1

Normally, as targets are all correctly associated, the total weights W_j^i should be approximate to one according to Equation (5). However, when targets move closely and simultaneously, multiple measurements are closer to one target *i* compared to the other targets; Gaussians in the *i*-th row in the weight matrix related to these measurements may have large enough weights. As a result, the total weights W_j^i may be greater than one. In other words, for a given weight matrix, if the total weight W_j^i of the *i*-th row satisfies the following condition:

$$W_j^i > 1 \tag{6}$$

this weight matrix is determined as an ambiguous weight matrix. The ambiguous weight matrix shows the possibility that one or more ambiguous weights may be involved in this matrix. To further determine the ambiguous weights, the expected targets' number and weight index in the matrix are used.

First, the expected number N_t of targets is calculated according to the method proposed in Section 2.2.

Then, the first N_t largest weights in the ambiguous weight matrix are selected as the ambiguous candidates.

Finally, if more than one candidate is in the same row in the matrix, these candidates are determined as the ambiguous weights. Otherwise, no ambiguous weights are involved. In other words, if more than one candidate has the same row index i, the corresponding weights $\bar{\omega}_t^{(i,j)}$ and $\bar{\omega}_t^{(i,j')}$ are determined as the ambiguous weights. $j' \neq j$ and $j' \in \{1, 2, \cdots, N_{m,t}\}$. The related measurements j and j' are determined as the ambiguous measurements, which are prone to be associated with the same target i. Consequently, the ambiguous weights $\bar{\omega}_t^{(i,j)}$ and $\bar{\omega}_t^{(i,j')}$ should be penalized. For example, the weights $\bar{\omega}_t^{(1,1)}$ and $\bar{\omega}_t^{(1,2)}$ in Figure 3a can be determined as the ambiguous weights according to the proposed method.

(2) Ambiguous weights' determination for Case 2

To determine the ambiguous weights for Case 2, those targets that move closely should be determined first. Targets i and i' are regarded as two close moving targets when:

$$\left\| \mathbf{l}_{t|t-1}^i - \mathbf{l}_{t|t-1}^{i'} \right\| < \left\| \mathbf{s}_{t|t-1}^i \right\| + \left\| \mathbf{s}_{t|t-1}^{i'} \right\| \tag{7}$$

where $\mathbf{l}_{t|t-1}^i$ (or $\mathbf{l}_{t|t-1}^{i'}$) and $\mathbf{s}_{t|t-1}^i$ (or $\mathbf{s}_{t|t-1}^{i'}$) are the location and size of the predicted state $\mathbf{x}_{t|t-1}^i$ (or $\mathbf{x}_{t|t-1}^{i'}$) of the target i (or i'), respectively. $\|\cdot\|$ represents the Euclidean norm.

Then, the ambiguous weights of Case 2 can be determined according to the measurements originating from the close moving targets. For two close moving targets i and i', if more than one measurement satisfies the following condition, these weights $\bar{\omega}_t^{(i,j)}$ can be regarded as the ambiguous weights.

$$\left\| \mathbf{l}_{z,t}^j - \mathbf{l}_{t|t-1}^i \right\| < \left\| \mathbf{l}_{t|t-1}^i - \mathbf{l}_{t|t-1}^{i'} \right\| \tag{8}$$

where $\mathbf{l}_{z,t}^j$ is the location of the j-th measurement $\mathbf{z}_t^j$.

After the ambiguous weights between the measurement j and the target i have been determined, multiple features that include the spatial-colour appearance, histogram of oriented gradient and target area are fused to penalize these ambiguous weights.

3.3. Multi-Feature Fusion

(1) Spatial-colour appearance

A colour histogram of a target is a representation of the distribution of colours inside this target's region in an image. Colour histogram-based appearances [24–27] are effective and efficient at capturing the distribution characteristics of visual features inside the target regions for visual tracking. In this section, a spatial constraint colour histogram appearance model (so-called spatial-colour appearance model) is presented.

The appearance of a target i is modelled as a Gaussian mixture $q_i = q_i(\omega_k^i, \mu_k^i, \Sigma_k^i)$, representing the colour distribution of a target's pixels [24]. $k = 1, \cdots, K$, and K is the number of Gaussian components. The measure of the similarity $P_s(i,j)$ between the measurement j and the target i is then defined by:

$$P_s(i,j) = \exp \left\{ \frac{1}{N_j} \sum_{\Omega_j} \log \left\{ \sum_{k=1}^K \omega_k^i \mathbf{N}(c_{\mathbf{l}^j}; \mu_k^i, \Sigma_k^i) \right\} \right\} \tag{9}$$

$$\mathbf{N}\left(c; \mu, \sum\right) = \frac{\exp\left\{ -\frac{1}{2}(c-\mu)' \Sigma^{-1}(c-\mu) \right\}}{\sqrt{2\pi |\Sigma|}} \tag{10}$$

where $c_{\mathbf{l}^j} = (r_{\mathbf{l}^j}, g_{\mathbf{l}^j}, I_{\mathbf{l}^j})$ is the colour of the pixel located in $\mathbf{l}^j$ within the support region Ω_j of the measurement j. N_j is the number of foreground pixels in Ω_j. $g_{\mathbf{l}^j} = G_{\mathbf{l}^j}/(R_{\mathbf{l}^j} + G_{\mathbf{l}^j} + B_{\mathbf{l}^j})$, $r_{\mathbf{l}^j} = R_{\mathbf{l}^j}/(R_{\mathbf{l}^j} + G_{\mathbf{l}^j} + B_{\mathbf{l}^j})$ and $I_{\mathbf{l}^j} = (R_{\mathbf{l}^j} + G_{\mathbf{l}^j} + B_{\mathbf{l}^j})/3$. Figure 5 is a schematic diagram of the colour distribution of the foreground pixels within a measurement's region.
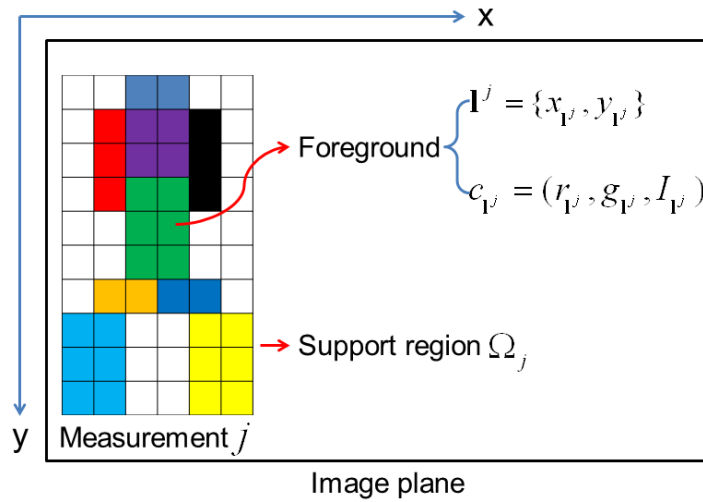
Figure 5. A schematic diagram of the colour distributions of the foreground pixels and the support region of the measurement *j*.

However, the aforementioned appearance model may fail when targets have similar colour distributions. To remedy this, a Gaussian spatial constraint [26] is incorporated, and the measure of the similarity is improved by:

$$P_s(i,j) = \exp \left\{ \frac{1}{N_j} \sum_{\Omega_j} \log \left\{ \mathbf{N}(\mathbf{l}_j; \mathbf{l}_t^i, \textstyle\sum_t^i) \sum_{k=1}^{K} \omega_k^i \mathbf{N}(c_{\mathbf{l}^j}; \mu_k^i, \textstyle\sum_k^i) \right\} \right\} \tag{11}$$

where $\mathbf{N}(\mathbf{l}_j; \mathbf{l}_t^i, \sum_t^i)$ is the Gaussian spatial constraint of the locations of the foreground pixels, and $\sum_t^i = [(w_t^i/2)^2, 0; 0, (h_t^i/2)^2]$. $\mathbf{l}_t^i = \{l_{x,t}^i, l_{y,t}^i\}$ and $\{w_t^i, h_t^i\}$ are the location and size of bounding box of the target *i* at time *t*, respectively.

(2) Histogram of oriented gradient [28]

The gradient $G(x,y)$ and orientation $O(x,y)$ of each pixel in the target region is calculated by:

$$G(x,y) = \sqrt{[I(x+1,y) - I(x-1,y)]^2 + [I(x,y+1) - I(x,y-1)]^2} \tag{12}$$

$$O(x,y) = \arctan \left\{ \frac{I(x,y+1) - I(x,y-1)}{I(x+1,y) - I(x-1,y)} \right\} \tag{13}$$

where $I(x,y)$ is the location of pixel in the image *I*.

The weighted oriented gradient histogram $q_h^i(u)$ of target *i* is formed by dividing the orientation into 36 bins (10° each step).

$$q_h^i(u) = C \sum_{r=1}^{n_i} k \left(\left\| (\mathbf{l}_r^i - \mathbf{l}_0^i)/h \right\|^2 \right) G(\mathbf{l}_r^i) \delta[b(\mathbf{l}_r^i) - u] \tag{14}$$

where $u = 1, 2, \cdots, 36$, $C = 1/\sum_{i=1}^{n_i} k(\|\mathbf{l}_r^i\|^2)$ is a normalization function, n_i is the number of pixels in target *i*'s region, $k(\cdot)$ is an isotropic kernel profile, $\mathbf{l}_r^i$ is the location of pixel *r*, *h* is the bandwidth, δ is the Kronecker delta function and $b(\mathbf{l}_r^i)$ associates the pixel *r* with the histogram bin.

The gradient of oriented histogram likelihood between the measurement *j* and the target *i* is defined by:

$$P_h(i,j) = \frac{1}{\sqrt{2\pi}\sigma_h} \exp \left\{ \frac{-d_h^2[q_h^i(u), q_h^j(u)]}{2\sigma_h} \right\} \tag{15}$$

$$d_h^2[q_h^i(u), q_h^j(u)] = \sqrt{1 - \rho[q_h^i(u), q_h^j(u)]} \tag{16}$$

$$\rho[q_h^i(u), q_h^j(u)] = \sum_{u=1}^{36} \sqrt{q_h^i(u) \cdot q_h^j(u)} \tag{17}$$

where σ_h is the Gaussian variance, which is set as 0.3 in our experiments.

(3) Target area

The degree of change between the areas of the target i and measurement j is defined by:

$$P_a(i,j) = \frac{\min\{S_i, S_j\}}{\max\{S_i, S_j\}} \tag{18}$$

where S_i and S_j represent the areas of target i and measurement j, respectively. It is reasonable to state that the larger the $P_a(i,j)$ is, the more possible it is that the measurement j is generated from the target i, because the size of the same target changes slightly between two consecutive frames.

(4) Multi-feature fusion

In this paper, the aforementioned features are fused to robustly penalize the ambiguous weight between the measurement j and the target i.

$$P_f(i,j) = (P_s(i,j) + P_h(i,j) + P_a(i,j))/3 \tag{19}$$

Obviously, the larger the $P_f(i,j)$ is, the more possibility there is that the measurement j is generated from the target i. In fact, if a measurement j is truly generated from a target i, the $P_f(i,j)$ should approximately be one.

3.4. Weight Penalization

The ambiguous weight $\bar{\omega}_t^{(i,j)}$ can be penalized according to the multi-feature fusion.

$$\bar{\omega}_t^{(i,j)} = \bar{\omega}_t^{(i,j)} \cdot P_f(i,j) \tag{20}$$

After all of the ambiguous weights have been penalized, all of the weights in the j-th column in the weight matrix should be further normalized by:

$$\bar{\omega}_t^{(i,j)} = \bar{\omega}_t^{(i,j)} / W_i^j \tag{21}$$

where $i = 1, \cdots, J_{t|t-1}$.

4. Experimental Evaluation

Our tracker can be employed for various scenarios, such as person tracking for human behaviour surveillance and analysis, car tracking for traffic surveillance, human hand and object tracking for human-object interactions, cell tracking for biomedical application, and so on.

In this section, we first evaluate our weight penalization method on several kinds of scenarios that include synthetic image sequences, outdoor human surveillance and cell moving surveillance scenarios by comparing to the state-of-the-art weight penalization methods. We then qualitatively test the proposed tracker on three more challenging scenarios and quantitatively compare it to several state-of-the-art trackers.

To quantitatively evaluate the tracking performance, the CLEAR MOTmetrics [23] is used. This returns a precision score MOTP (multi-object tracking precision) and an accuracy score MOTA

(multi-object tracking accuracy) that is composed of a miss rate (MR), a false positive rate (FPR) and a mismatch rate (MMR).

$$\text{MOTP} = \frac{\sum_{i,t}[S(gb_t^i \cap tb_t^i)/S(gb_t^i \cup tb_t^i)]}{\sum_t c_t} \tag{22}$$

$$\text{MOTA} = 1 - \frac{\sum_t(m_t + fp_t + mme_t)}{\sum_t g_t} \tag{23}$$

where $S(\cdot)$ represents the area. gb_t^i is the ground truth box, and tb_t^i is the associated tracked box of the target i for time t. c_t is the number of matched targets for time t. m_t, fp_t, mme_t and g_t are the numbers of misses, false positives, mismatches and ground truths, respectively, for time t.

4.1. Experimental Parameter Setup

Parameters of the tracker involved in the experiments are set as follows. Similarly as set in our previous work [15], we have the state transition model as $\mathbf{F}_t = [\mathbf{I}_2, T\mathbf{I}_2, \mathbf{0}_2; \mathbf{0}_2, \mathbf{I}_2, \mathbf{0}_2; \mathbf{0}_2, \mathbf{0}_2, \mathbf{I}_2]$ and $\mathbf{Q}_t = \delta_v^2[T^4\mathbf{I}_2/4, T^3\mathbf{I}_2/2, \mathbf{0}_2; T^3\mathbf{I}_2/2, T^2\mathbf{I}_2, \mathbf{0}_2; \mathbf{0}_2, \mathbf{0}_2, T^2\mathbf{I}_2]$, where $\mathbf{0}_n$ and $\mathbf{I}_n$ are the $n \times n$ zero and identity matrices. $T = 1$ frame is the interval between two consecutive time steps. $\delta_v = 3$ is the standard deviation of the state noise. We also set the measurement model as $\mathbf{H}_t = [\mathbf{I}_2, \mathbf{0}_2, \mathbf{0}_2; \mathbf{0}_2, \mathbf{0}_2, \mathbf{I}_2]$ and $\mathbf{R}_t = \delta_w^2\mathbf{I}_4$, where $\delta_w = 2$ is the standard deviation of the measurement noise. The values of residual parameters involved in our tracker are set as: $p_d = 0.99$, $p_{sv} = 0.95$, $\lambda_t = 0.01$, $c_t(\mathbf{z}_t) = \text{(image area)}^{-1}$ and $\sigma_h = 0.3$.

4.2. Evaluation of the Proposed Weight Penalization Method

We evaluate the proposed weight penalization method on three scenarios, including a synthetic image sequence, an outdoor human surveillance scenario and a cell moving surveillance scenario. Moreover, to demonstrate the effectiveness of the proposed method, it is also compared to the conventional GM-PHD tracker [7] and the CPGM-PHD tracker [22].

(1) Qualitative analysis

Tracking on a synthetic image sequence: A synthetic image sequence is used to validate the effectiveness of the proposed weight penalization method. Figures 6 and 7 show the tracking results and the corresponding weight matrices obtained by the trackers, respectively. At $t = 48$, all of the trackers can successfully track all of the targets (shown as in Figure 6a). At $t = 49$, Targets 1 and 4 approach very close, as well as Targets 2 and 3. Without any weight penalization method, the conventional GM-PHD tracker tracks Target 2 with the wrong Identity 3, while switching the identities for Targets 1 and 4 (shown as in Figure 6b). According to the method proposed in the CPGM-PHD tracker, two ambiguous weights for Case 1 are determined and rearranged (shown as in Figure 7b), and the corresponding targets are tracked with correct identities (shown as Targets 2 and 3 in Figure 6c). However, the CPGM-PHD tracker cannot correctly track the targets with switched identities for Case 2 (shown as Targets 1 and 4 in Figure 6c). On the contrary, our tracker determines the ambiguous weights for both Case 1 and Case 2 and penalizes the ambiguous weights by fusing the multiple target features. By doing so, four ambiguous weights are determined to be rearranged (shown as in Figure 7c), and all of the targets are tracked with correct identities (shown as in Figure 6d). Figure 8 shows the trajectories of the tracked targets. The results demonstrate that the trajectories obtained by our tracker are closer to the ground truth.

To show the effectiveness of multi-feature fusion, we also perform our tracker with a single feature, such as the target area feature and the colour appearance feature. The tracking results are shown as in Figure 6e,f. Since the target areas of two closely moving targets (Targets 2 and 3) are almost the same, the difference between the measurements originating from them is negligible. If we only use target area to penalize the weights, the tracker should perform just like a conventional GM-PHD tracker (shown as Figure 6e). Although the areas of Targets 2 and 3 are almost the same,

their appearances are totally different. Therefore, penalizing the weights with the colour appearance feature can correctly track these two targets. However, many similarities occur in Targets 1 and 4, which results in a mismatched result.

Tracking on an outdoor human surveillance scenario: An outdoor human surveillance sequence is used to further evaluate the proposed weight penalization method. Figure 9 shows the tracking results by the GM-PHD tracker, the CPGM-PHD tracker and our tracker, respectively. Without the weight penalization method, the conventional GM-PHD tracker tracks the close moving targets with the same identities at $t = 89$ (shown as two Target 1s in Figure 9a). Both the CPGM-PHD and proposed trackers can successfully track the closely moving targets at $t = 89$ (shown as in Figure 9b,c). However, both of the GM-PHD and CPGM-PHD trackers track the merged measurement as one single target, as mutual occlusion occurs in targets at $t = 90$ (shown as Target 1 in Figure 9d,e). On the contrary, our tracker can correctly track the targets in mutual occlusion (shown as Targets 1 and 5 in Figure 9f) by incorporating the mutual occlusion handling method. Figure 10 demonstrates the trajectories of the tracked targets. The GM-PHD tracker and CPGM-PHD tracker cannot correctly track the targets in consecutive time steps and, thus, results in many mismatches. By contrast, the results obtained by our tracker are closer to the ground truth and, thus, largely reduce the mismatches.

Tracking on a cell moving surveillance scenario: A cell moving surveillance sequence captured from the phase contrast microscopy video is tested to evaluate the robustness of the proposed tracker. The high density of the cell population makes cells move in a relatively close space. Without weight penalization, the GM-PHD tracker may track the closely moving cells with the same identity. As shown in the left image in Figure 11b, two cells are assigned with the same ID 27. On the contrary, both trackers (the CPGM-PHD tracker and the proposed tracker) with the weight penalization method can successfully track the cells with the correct identities (shown as the middle and right images in Figure 11b). Nevertheless, our tracker can achieve more exact cell states. Moreover, our tracker can successfully track the mitosis cell as a newborn cell (shown as in Figure 11c) because of the incorporation of an effective birth intensity estimation method.
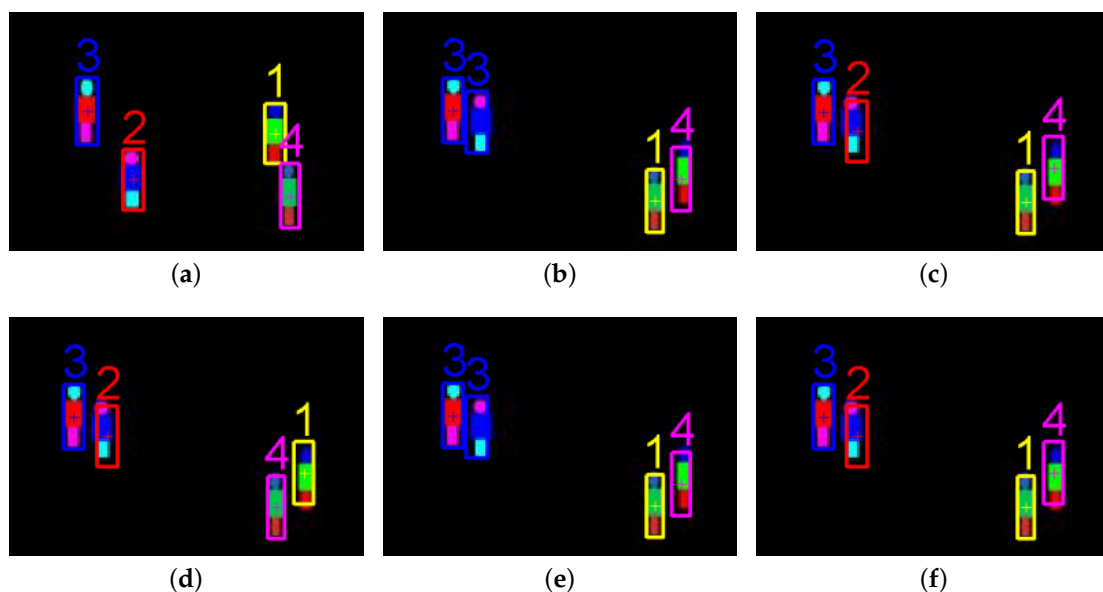
Figure 6. Tracking results on a synthetic image sequence. (**a**) Tracked targets at $t = 48$ by all of the trackers; (**b**) tracked targets at $t = 49$ by the GM-PHD tracker; (**c**) tracked targets at $t = 49$ by the collaborative penalized GM (CPGM)-PHD tracker; (**d**) tracked targets at $t = 49$ by our tracker with multi-feature fusion; (**e**) tracked targets at $t = 49$ by our tracker with the target area feature; (**f**) tracked targets at $t = 49$ by our tracker with the colour appearance feature.

|	j=1	2	3	4
i=1	0.1874	5.30×10^{-21}	2.65×10^{-4}	0.7314
2	5.25×10^{-29}	0.1117	1.13×10^{-14}	4.38×10^{-19}
3	3.87×10^{-33}	0.8883	0.9997	1.84×10^{-39}
4	0.8126	6.46×10^{-24}	4.37×10^{-34}	0.2684

(**a**)

	j=1	2	3	4
i=1	0. 0528	1.83×10^{-24}	3.86×10^{-22}	0.9436
2	2.36×10^{-29}	0.9995	4.71×10^{-16}	4.26×10^{-18}
3	6.84×10^{-32}	0.0005	0.9999	5.95×10^{-35}
4	0.9472	6.26×10^{-26}	8.22×10^{-33}	0.0564

(**b**)

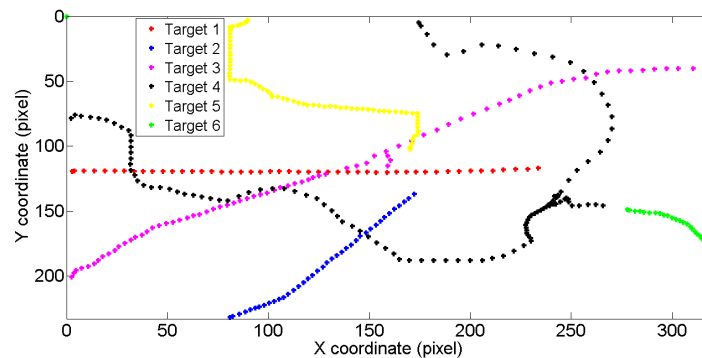|	j=1	2	3	4
i=1	0.9535	4.74×10^{-20}	1.94×10^{-4}	0.0042
2	3.43×10^{-30}	0.9972	8.24×10^{-15}	1.62×10^{-18}
3	4.92×10^{-26}	0.0028	0.9998	6.84×10^{-39}
4	0.0465	5.78×10^{-23}	3.19×10^{-34}	0.9958

(**c**)

Figure 7. Updated weight matrices at *t* = 49 on a synthetic image sequence. (**a**) For the GM-PHD tracker; (**b**) for the CPGM-PHD tracker; (**c**) for our tracker.
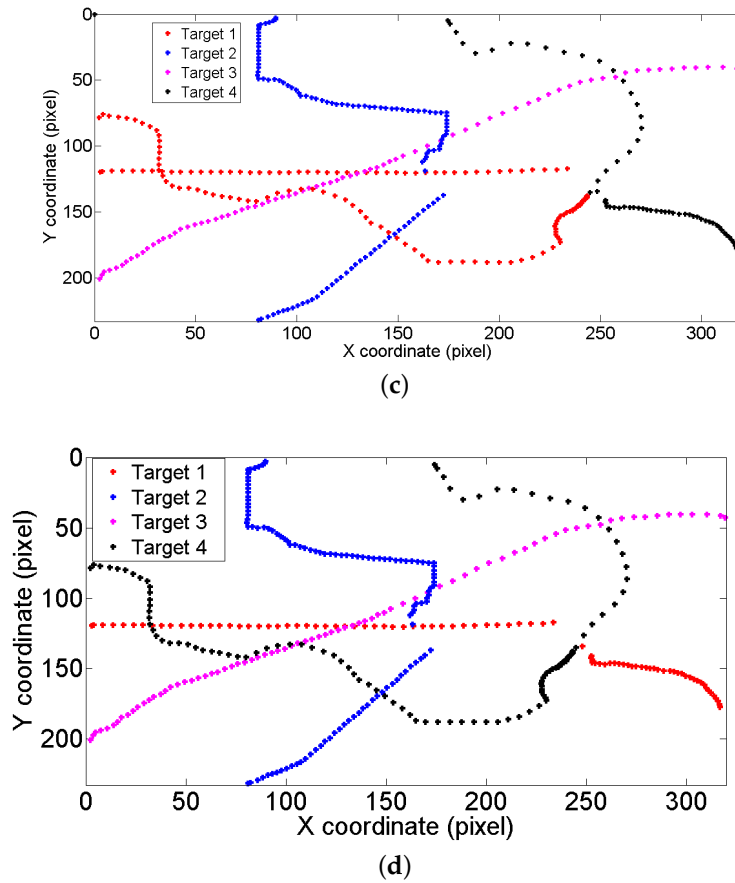

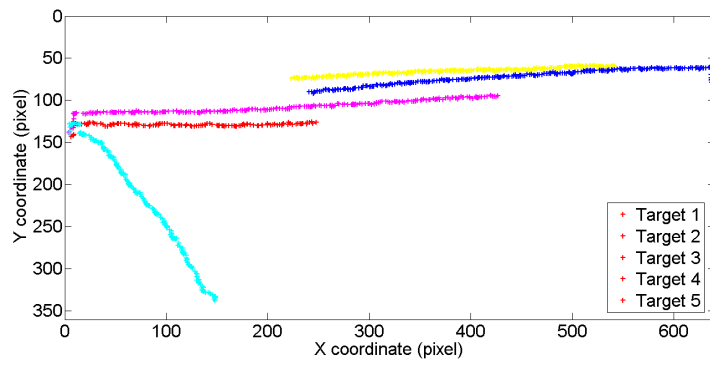
(**a**)

(**b**)

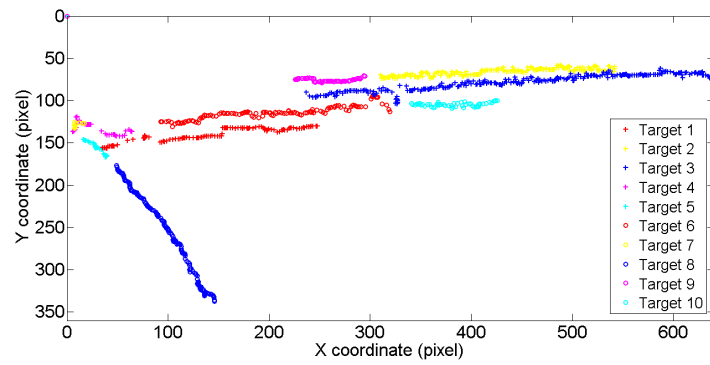**Figure 8.** *Cont.*

(**c**)

(**d**)
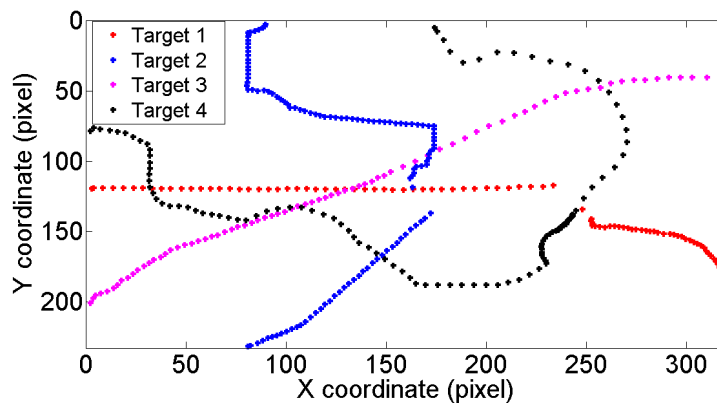
Figure 8. Tracking trajectories on a synthetic image sequence. (**a**) Ground truth; (**b**) for the GM-PHD tracker; (**c**) for the CPGM-PHD tracker; (**d**) for our tracker.
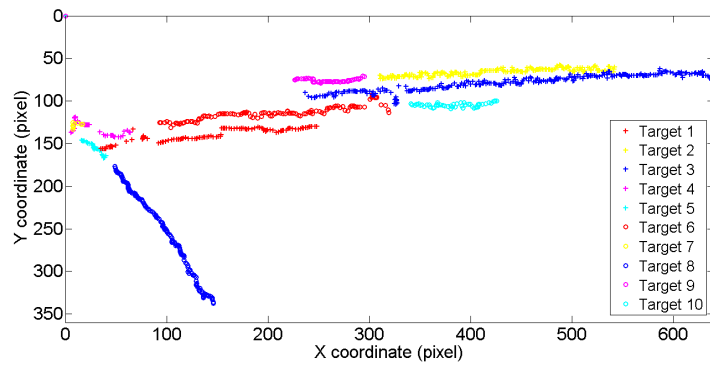

(**e**)　　　　　　　　　(**f**)　　　　　　　　　(**g**)

(**h**)　　　　　　　　　(**i**)　　　　　　　　　(**j**)

Figure 9. Tracking results on an outdoor human surveillance scenario. (**a**,**d**) Tracked targets by the GM-PHD tracker; (**b**,**e**) tracked targets by the CPGM-PHD tracker; (**c**,**f**) tracked targets by our tracker.
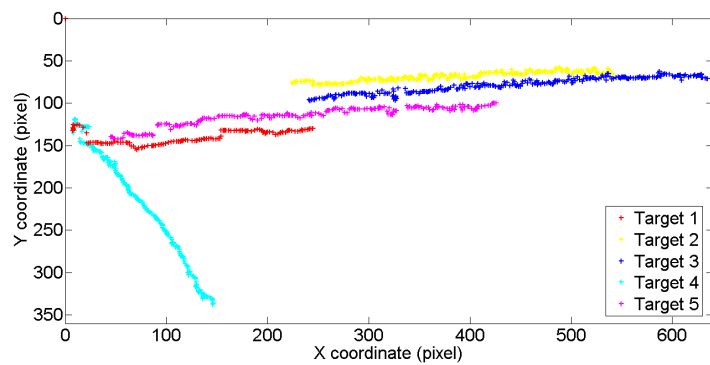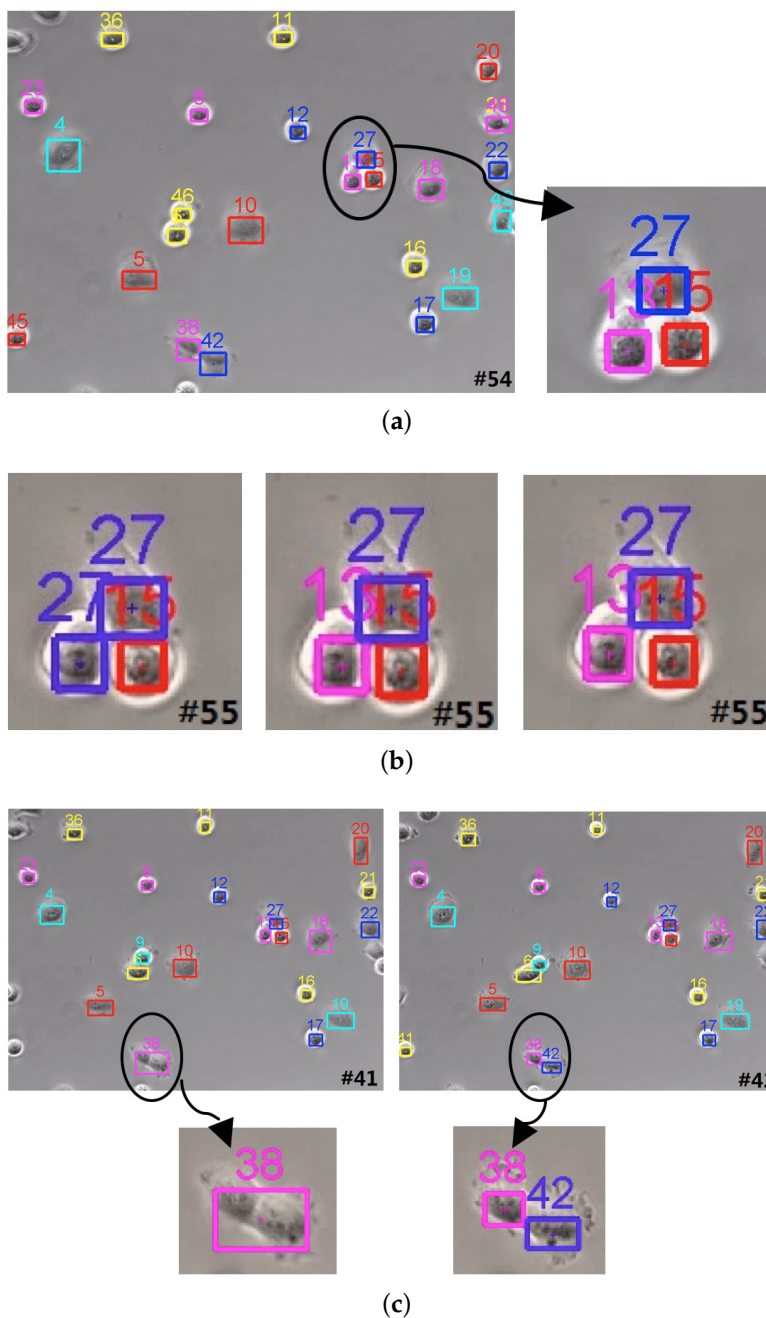
Figure 10. Tracking trajectories on an outdoor human surveillance scenario. (**a**) Ground truth; (**b**) for the GM-PHD tracker; (**c**) for the CPGM-PHD tracker; (**d**) for our tracker.

Figure 11. Tracking results on a cell moving surveillance scenario. (**a**) Left: tracked targets at *t* = 54; right: local tracked targets; (**b**) from left to right: tracked targets at *t* = 55 by the GM-PHD tracker, by the CPGM-PHD tracker and by our tracker, respectively; (**c**) from left to right: tracked targets by our tracker at *t* = 41 and *t* = 42, respectively.

(2) Quantitative analysis

We quantitatively evaluate the tracking performance according to the CLEAR MOT metrics. Table 1 shows the corresponding tracking performance comparison of the GM-PHD tracker, CPGM-PHD tracker and our tracker tested on the above-mentioned surveillance scenarios. The results show that tracking with our tracker can achieve better scores, both in MOTP and MOTA, on the tested sequences. Moreover, to show the effectiveness of the proposed weight penalization method for tracking the closely moving targets, the mismatch rate is also compared (shown as in Table 2).

By determining the ambiguous weights for two cases and incorporating multiple target features to penalize the ambiguous weights, our tracker can reduce the mismatch rate and, thus, improve the tracking accuracy.

Table 1. Tracking performance comparison of the GM-PHD tracker, CPGM-PHD tracker and our tracker. MOTA, multi-object tracking accuracy; MOTP, multi-object tracking precision.

Tracker	Performance	Synthetic Images	Outdoor Human Surveillance	Cells Moving
GM-PHD	MOTA	0.8586	0.6265	0.5128
tracker [7]	MOTP	0.9266	0.8567	0.4283
CPGM-PHD	MOTA	0.9863	0.7038	0.6842
tracker [22]	MOTP	0.9536	0.8724	0.5581
Our	MOTA	1	0.9348	0.7218
tracker	MOTP	0.9675	0.9273	0.6065

Table 2. Mismatch rate (MMR: %) comparison of the GM-PHD tracker, CPGM-PHD tracker and our tracker.

Tracker	Synthetic Images	Outdoor Human Surveillance	Cells Moving
GM-PHD tracker	12.88	8.52	18.75
CPGM-PHD tracker	1.37	2.76	7.13
Our tracker	0	0.94	2.68

4.3. Evaluation of the Proposed Tracker

We first qualitatively evaluate our tracker on three more challenging surveillance scenarios, including interactive person tracking for public surveillance, person and luggages tracking for subway station surveillance of PETS2006 [29] and crowd person tracking for campus surveillance of PETS2009 [30], and then quantitatively compare our tracker with the state-of-the-art trackers according to the CLEAR MOT metrics. Moreover, the computational cost of our tracker on tested surveillance scenarios is also presented and discussed.

(1) Qualitative and quantitative analysis

Figure 12 shows tracking results of our tracker tested on the above-mentioned three challenging surveillance scenarios. In Figure 12a, three persons move closely and frequently interact with each other. At $t = 939$, Person 1 and Person 2 get close, and occlusion occurs at $t = 945$. At $t = 959$, three persons get close, and a long-term occlusion occurs among them. Although persons get close and interact frequently, our tracker can successfully track all three persons with correct identities with time due to the effective weight penalization and occlusion handling method. In Figure 12b, three persons get close at $t = 775$, and serious occlusion occurs with almost the same appearance at $t = 782$. With our multi-feature fusion scheme, as well as occlusion handling method, the persons involved are accurately tracked. It is noted that when a person is moving with luggage in his or her hand, both the person and luggage are tracked as one single target (shown as Targets 12 and 16 in Figure 12b). However, when they are separated, they are tracked as two targets with different identities (shown as Targets 21 and 27 in Figure 12b). In Figure 12c, a large number of persons and many interactions are involved. For example, at $t = 160$ and $t = 165$, Persons 3, 9 and 12, as well as Persons 4 and 5 are walking together. Similarly, at $t = 293$, Persons 3 and 23, as well as Persons 5 and 18 get close, and occlusions occur. Even though, our tracker still can effectively track those persons with correct identities.

To show the superior performance of our tracker, it is quantitatively compared to the state-of-the-art trackers according to the CLEAR MOT metrics. We compare the MOTA and MOTP

scores of our tracker with the scores reported in [31–33] on the subway station surveillance scenario of PETS2006, and the scores reported in [34–36] on the campus surveillance scenario of PETS2009, respectively. The results in Tables 3 and 4 show that our tracker achieves a better MOTP score on tracking precision and a comparable MOTA score on tracking accuracy. The reason for the lower MOTA score than the score reported in [35] is that we implement object detection using a simple background subtraction method. This simple method tends to generate a large number of noise in variable environment. Although our tracker can eliminate a large number of noise, some noises may still be tracked as the targets. To further achieve a high MOTA score, a more robust object detection method should be incorporated.

Table 3. Tracking performance comparison on the subway station surveillance scenario of PETS2006.

	GM-PHD Tracker [7]	Tracker in [31]	Tracker in [32]	Tracker in [33]	Our Tracker
MOTA	0.3440	0.9875	0.9221	0.9656	0.8861
MOTP	0.4286	0.5816	0.4980	0.5687	0.6346

Table 4. Tracking performance comparison on the campus surveillance scenario of PETS2009.

	GM-PHD Tracker [7]	Tracker in [34]	Tracker in [35]	Tracker in [36]	Our Tracker
MOTA	0.4617	0.7591	0.8932	0.7977	0.8826
MOTP	0.4976	0.5382	0.5643	0.5634	0.6055

(a)

(b)

(c)
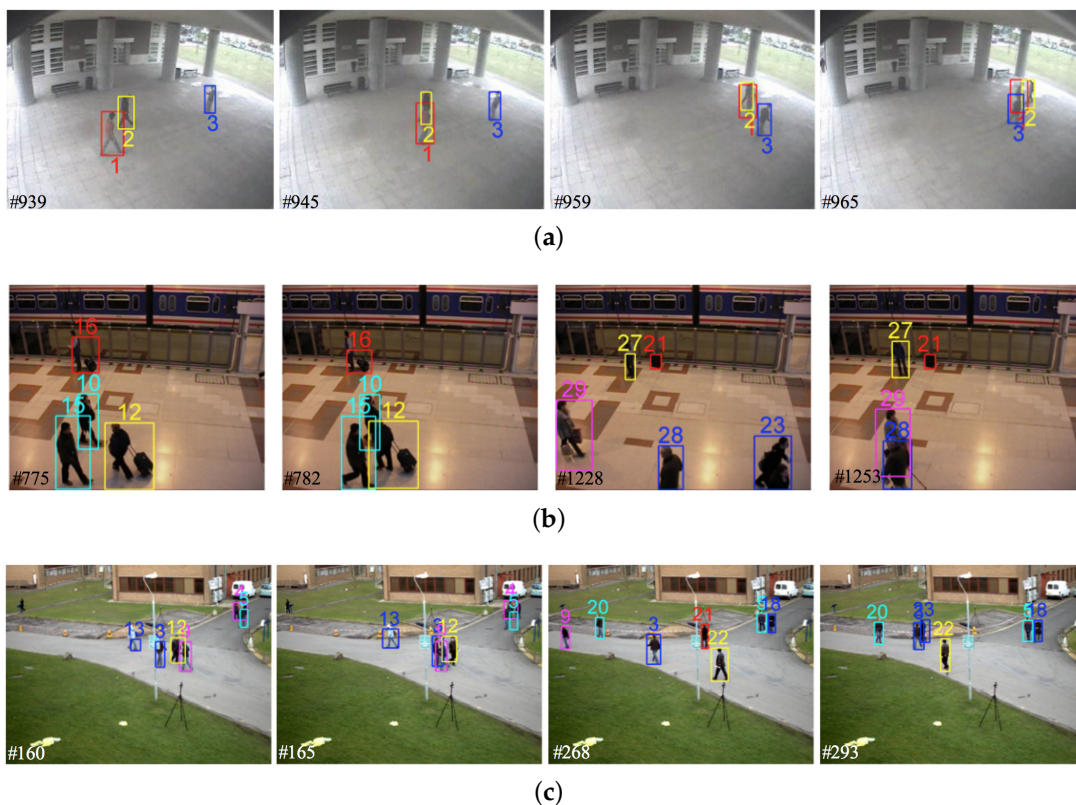
Figure 12. Tracking results of our tracker on three challenging surveillance scenarios. (**a**) Tracking interactive persons for public surveillance; (**b**) tracking persons and luggage for subway station surveillance; (**c**) tracking crowd persons for campus surveillance.

(2) Computational cost

The proposed tracker is implemented in MATLAB using a computer with Inter(R) Core(TM) i7-4600U CPU 2.10 GHz and 4 GB of memory. Without any code optimization, the average runtimes of the tested surveillance videos are shown as in Figure 13. The majority of the runtimes are consumed in game theory-based mutual occlusion handling, because it is a pixel-wise iteration process. In addition, tracking a larger number of targets also increases the computational burden. In cell moving surveillance, subway station surveillance and campus surveillance scenarios, a large number of targets, as well as many occlusions are involved, which cost more computational time and slow down the processing speed.
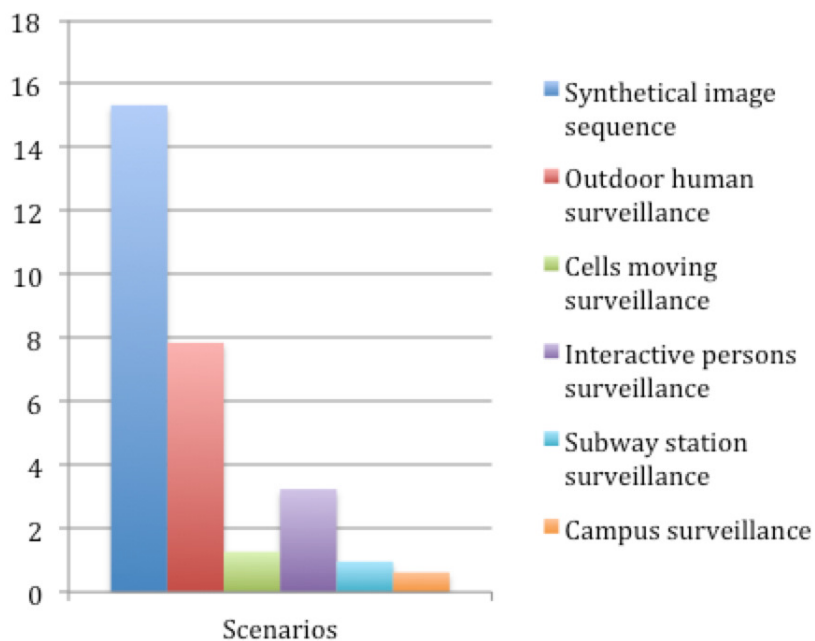

Figure 13. Processing time (unit: fps) of our tracker on the tested scenarios.

5. Conclusions

We have developed a robust GM-PHD tracker to track targets in close movement in video. We incorporated an entropy-based birth intensity estimation method to effectively eliminate the false positives caused by noises. Particularly, we presented a weight penalization method to accurately track the targets in close movement.

The majority of the leading methods in the state of the art only considered ambiguous weight penalization for Case 1. Besides, only the total weight was used for ambiguous weight determination. However, both Case 1 and Case 2 could cause incorrect tracking. In this paper, we constructed a weight matrix and used both the total weight and target state to determine the ambiguous weights for both cases in the matrix. We then fused multiple target features, including the spatial-colour appearance, histogram of oriented gradient and target area, to penalize the ambiguous weights. By doing so, those weights between the target and the irrelevant measurements can be greatly penalized and, thus, lead to an improved tracking accuracy with a low mismatch rate. Moreover, fusing multiple features took advantage of single feature merit and leveraged the corresponding weights.

We experimentally validated our tracker on a variety of scenarios and qualitatively and quantitatively compared our tracker to the conventional GM-PHD tracker, as well as the state-of-the-art trackers. The results demonstrated that our tracker achieved an improvement in precision and accuracy.

However, the processing speed of our tracker was not fast enough, which limited the real-time application. To remedy this, employing a more efficient occlusion handling method will be helpful and will be explored in our future works.

Author Contributions: Xiaolong Zhou wrote the paper and conceived of the experiment. Hui Yu contributed the experimental data and to the analysis. Honghai Liu and Youfu Li supervised the overall work and reviewed the paper.

Conflicts of Interest: The authors hereby declare no conflict of interest.

References

1. Zhou, S.; Fei, F.; Zhang, G.; Liu, Y.; Li, W.J. Hand-Writing Motion Tracking with Vision-Inertial Sensor Fusion: Calibration and Error Correction. *Sensors* **2014**, *14*, 15641–15657.

2. Choi, Y.J.; Kim, Y.G. A Target Model Construction Algorithm for Robust Real-Time Mean-Shift Tracking. *Sensors* **2014**, *14*, 20736–20752.

3. Bai, T.; Li, Y.; Zhou, X. Learning local appearances with sparse representation for robust and fast visual tracking. *IEEE Trans. Cybern.* **2015**, *45*, 663–675.

4. Kwon, J.; Lee, K.M. Tracking by sampling and integrating multiple trackers. *IEEE Trans. Pattern Anal. Mach. Intell.* **2014**, *36*, 1428–1441.

5. Park, C.; Woehl, T.; Evans, J.; Browning, N. Minimum cost multi-way data association for optimizing multitarget tracking of interacting objects. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *37*, 611–624.

6. Mahler, R. Multitarget bayes filtering via first-order multitarget moments. *IEEE Trans. Aerosp. Electron. Syst.* **2003**, *39*, 1152–1178.

7. Vo, B.N.; Ma, W. The Gaussian mixture probability hypothesis density filter. *IEEE Trans. Signal Process.* **2006**, *54*, 4091–4104.

8. Yoon, J.H.; Kim, D.Y.; Bae, S.H.; Shin, V. Joint initialization and tracking of multiple moving objects using doppler information. *IEEE Trans. Signal Process.* **2011**, *59*, 3447–3452.

9. Panta, K.; Clark, D.E.; Vo, B.N. Data association and track management for the Gaussian mixture probability hypothesis density filter. *IEEE Trans. Aerosp. Electron. Syst.* **2009**, *45*, 1003–1016.

10. Panta, K.; Clark, D.E.; Vo, B.N. Extended target tracking using a Gaussian-mixture PHD filter. *IEEE Trans. Aerosp. Electron. Syst.* **2012**, *48*, 3268–3286.

11. Pham, N.T.; Huang, W.; Ong, S.H. Tracking multiple objects using probability hypothesis density filter and color measurements. In Proceedings of the 2007 IEEE International Conference on Multimedia and Expo, Beijing, China, 2–5 July 2007; pp. 1511–1514.

12. Pham, N.T.; Huang, W.; Ong, S.H. Probability hypothesis density approach for multi-camera multi-object tracking. In Proceedings of the 8th Asian Conference on Computer Vision, Tokyo, Japan, 18–22 November 2007; pp. 875–884.

13. Wu, J.; Hu, S. Probability hypothesis density filter based multi-target visual tracking. In Proceedings of the 2010 29th Chinese Control Conference (CCC), Beijing, China, 29–31 July 2010; pp. 2905–2909.

14. Wu, J.; Hu, S.; Wang, Y. Probability-hypothesis-density filter for multitarget visual tracking with trajectory recognition. *Opt. Eng.* **2010**, *49*, 97011–97019.

15. Zhou, X.; Li, Y.; He, B. Entropy distribution and coverage rate-based birth intensity estimation in GM-PHD filter for multi-target visual tracking. *Signal Process.* **2014**, *94*, 650–660.

16. Zhou, X.; Li, Y.; He, B.; Bai, T. GM-PHD-based multi-target visual tracking using entropy distribution and game theory. *IEEE Trans. Ind. Inform.* **2014**, *10*, 1064–1076.

17. Pollard, E.; Plyer, A.; Pannetier, B.; Champagnat, F.; Besnerais, G.L. GM-PHD filters for multi-object tracking in uncalibrated aerial videos. In Proceedings of the 12th International Conference on Information Fusion (FUSION'09), Seattle, WA, USA, 6–9 July 2009; pp. 1171–1178.

18. Clark, R.D.; Vo, B.N. Improved SMC implementation of the PHD filter. In Proceedings of the 2010 13th Conference on Information Fusion (FUSION), Edinburgh, UK, 26–29 July 2010; pp. 1–8.

19. Maggio, I.E.; Taj, M.; Cavallaro, A. Efficient multi-target visual tracking using Random Finite Sets. *IEEE Trans. Circuits Syst. Video Technol.* **2008**, *18*, 1016–1027.

20. Yazdian-Dehkordi, M.; Azimifar, Z.; Masnadi-Shirazi, M. Competitive Gaussian mixture probability hypothesis density filter for multiple target tracking in the presence of ambiguity and occlusion. *IET Radar Sonar Navig.* **2012**, *6*, 251–262.

21. Yazdian-Dehkordi, M.; Azimifar, Z.; Masnadi-Shirazi, M. Penalized Gaussian mixture probability hypothesis density filter for multiple target tracking. *Signal Process.* **2012**, *92*, 1230–1242.

22. Wang, Y.; Meng, H.; Liu, Y.; Wang, X. Collaborative penalized Gaussian mixture PHD tracker for close target tracking. *Signal Process.* **2014**, *102*, 1–15.

23. Zhou, X.; Li, Y.; He, B. Game-theoretical occlusion handling for multi-target visual tracking. *Pattern Recognit.* **2013**, *46*, 2670–2684.

24. Wang, Y.; Wu, J.; Kassim, A.; Huang, W. Occlusion reasoning for tracking multiple people. *IEEE Trans. Circuits Syst. Video Technol.* **2009**, *19*, 114–121.

25. Zhou, X.; Li, Y.; He, B. Multi-target visual tracking with game theory-based mutual occlusion handling. In Proceedings of the 2013 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), Tokyo, Japan, 3–7 November 2013; pp. 4201–4206.

26. Zhang, X.; Hu, W.; Luo, G.; Manbank, S. Kernel-bayesian framework for object tracking. In Proceedings of the 8th Asian Conference on Computer Vision, Tokyo, Japan, 18–22 November 2007; pp. 821–831.

27. Wu, J.; Hu, S.; Wang, Y. Adaptive multifeature visual tracking in a probability-hypothesis-density filtering framework. *Signal Process.* **2013**, *93*, 2915–2926.

28. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110.

29. PETS2006. Available online: http://www.cvg.reading.ac.uk/PETS2006/data.html (accessed on 1 December 2015).

30. PETS2009. Available online: http://www.cvg.reading.ac.uk/PETS2009/a.html (accessed on 1 December 2015).

31. Zulkifley, M.A.; Moran, B. Robust hierarchical multiple hypothesis tracker for multiple-object tracking. *Expert Syst. Appl.* **2012**, *39*, 12319–12331.

32. Joo, S.W.; Chellappa, R. A multiple-hypothesis approach for multiobject visual tracking. *IEEE Trans. Image Process.* **2007**, *16*, 2849–2854.

33. Torabi, A.; Bilodeau, G.A. A multiple hypothesis tracking method with fragmentation handling. In Proceedings of the 2009 Canadian Conference on Computer and Robot Vision (CRV'09), Kelowna, BC, Canada, 25–27 May 2009; pp. 8–15.

34. Yang, J.; Shi, Z.; Vela, P.; Teizer, J. Probabilistic multiple people tracking through complex situations. In Proceedings of the 9th International Symposium on Privacy Enhancing Technologies (PETS'09), Seattle, WA, USA, 5–7 August 2009; pp. 79–86.

35. Andriyenko, A.; Schindler, K.; Roth, S. Discrete-continuous optimization for multi-target tracking. In Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Providence, RI, USA, 16–21 June 2012; pp. 1926–1933.

36. Park, C.; Woehl, T.; Evans, J.; Browning, N. Online multiperson tracking-by-detection from a single, uncalibrated camera. *IEEE Trans. Pattern Anal. Mach. Intell.* **2011**, *33*, 1820–1833.