GM-PHD-Based Multi-Target Visual Tracking Using Entropy Distribution and Game Theory

Xiaolong Zhou, Youfu Li, *Senior Member*, *IEEE*, Bingwei He, and Tianxiang Bai

***Abstract*—Tracking multiple moving targets in a video is a challenge because of several factors, including noisy video data, varying number of targets, and mutual occlusion problems. The Gaussian mixture probability hypothesis density (GM-PHD) filter, which aims to recursively propagate the intensity associated with the multi-target posterior density, can overcome the difficulty caused by the data association. This paper develops a multi-target visual tracking system that combines the GM-PHD filter with object detection. First, a new birth intensity estimation algorithm based on entropy distribution and coverage rate is proposed to automatically and accurately track the newborn targets in a noisy video. Then, a robust game-theoretical mutual occlusion handling algorithm with an improved spatial color appearance model is proposed to effectively track the targets in mutual occlusion. The spatial color appearance model is improved by incorporating interferences of other targets within the occlusion region. Finally, the experiments conducted on publicly available videos demonstrate the good performance of the proposed visual tracking system.**

***Index Terms*—Birth intensity estimation, Gaussian mixture probability hypothesis density (GM-PHD) filter, multi-target visual tracking (MTVT), mutual occlusion handling.**

I. Introduction

MULTI-TARGET VISUAL TRACKING (MTVT) is used to locate and identify multiple moving targets at each image frame in a video sequence. An MTVT is crucial in intelligent video surveillance systems and in activity analysis or high-level event understanding in many industrial applications [1]–[5]. The problem of MTVT extends the single-target visual tracking to a situation where the number of moving targets is unknown and varies with time. Recently, many researchers have successfully explored the Gaussian mixture probability hypothesis density (GM-PHD) filter [6]–[8] in a multi-target tracking in video. Compared with traditional association-based techniques, the GM-PHD filter effectively overcomes the difficulty caused by the data association. In this paper, we develop a system that combines the GM-PHD filter with object detection to track multiple moving targets in a video. However, noisy video data, varying number of targets, and mutual occlusion problems make this development a challenge.

To track the varying number of targets in a noisy video, the proposed system must track the newborn targets accurately as they enter the scene. In other words, an important issue in the GM-PHD filter is automatically and accurately determining the birth intensity of the newborn targets. Conventionally, the birth intensity must cover the whole state space [9] when no prior localization information on the newborn targets is available. Such requirement entails high computational cost and can easily be interfered by clutters. To narrow the search space, Wang *et al.* [6] manually preset the means of Gaussian in the birth intensity according to the scene information, such as edges or shop entrances. However, presetting the birth intensity initially requires knowledge of the scene information, which involves human interactions. To automatically estimate the birth intensity, Maggio *et al.* [10] assume that the birth of a target occurs in a limited volume around the measurements. They draw the newborn particles from a mixture of Gaussians centered at the components of the measurements set. However, the proposed method could easily be interfered by clutters and the measurements originating from the survival targets. To eliminate the negative effect of the survival targets, Wang *et al.* [11] classify the measurements into two parts, namely, the measurements originating from the newborn targets and those originating from the survival targets. However, the measurements originating from the newborn targets may contain some noises. In such a case, directly determining the birth intensity by the measurements originating from the newborn targets will result in many false positives.

In addition, mutual occlusion may occur in the interacting targets as they move close together. Once occlusion occurs, the measurements originating from these targets within the occlusion region will be merged into one measurement. Without an occlusion handling algorithm, the system may fail to track the targets in mutual occlusion. Currently, extensive methods, such as multiple camera fusing methods [12], [13], Monte Carlo-based probabilistic methods [14], [15], and appearance model-based deterministic methods [16]–[19], have been presented to solve the mutual occlusion problems. The problem of tracking multiple interacting targets in mutual occlusion is still far from being completely solved, thereby remaining an open issue. Compared with the two other classes of occlusion handling

Manuscript received August 01, 2012; revised July 28, 2013, and October 12, 2013; accepted November 20, 2013. Date of publication December 05, 2013; date of current version May 02, 2014. This work was supported in part by Research Grants Council of Hong Kong (CityU 118311), in part by the City University of Hong Kong (7008176), and in part by the National Natural Science Foundation of China (61273286 and 51175087). Paper No. TII-13-0369.

X. Zhou is with the College of Computer Science and Technology, Zhejiang University of Technology, Hangzhou 310023, China (e-mail: mexlz@hotmail.com).

Y. Li is with the Department of Mechanical and Biomedical Engineering, City University of Hong Kong, Kowloon 852, Hong Kong (e-mail: meyfli@cityu.edu.hk).

B. He is with the School of Mechanical Engineering and Automation, Fuzhou University, Fuzhou 350108, China (e-mail: mebwhe@fzu.edu.cn).

T. Bai is with the Department of Research and Development, Advanced Semiconductor Materials (ASM) Pacific Technology Ltd., Kwai Chung 852, Hong Kong (e-mail: tianxiangbai@gmail.com).

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/TII.2013.2294156

methods, tracking with the appearance model-based deterministic methods offers several advantages, including generality, flexibility, computational efficiency, and large amount of information [16]. For example, Vezzani *et al.* [16] use an appearance-driven tracking model to overcome large- and long-lasting occlusions. They generate two different images to represent the target model: the appearance image and a probability mask. The appearance image contains the red, green, and blue (RGB) colors of each point of the target, and the corresponding probability mask reports the reliability of these colors. Based on this target model, the authors classify the invisible regions into dynamic occlusions, scene occlusions, and apparent occlusions. Xing *et al.* [17] build a dedicated observation model that maintains three discriminative cues, namely, appearance, size, and motion. The target appearance is modeled as the color histogram in hue, saturation, and value color space in discriminative region of the target. The mutual occlusion problem is then handled by a two-way Bayesian inference method. However, the aforementioned appearance models cannot deal with interacting targets having similar color distributions and are thus expected to fail in tracking. To remedy this problem, Papadourakis and Argyros [18] model the target by using an ellipse and a Gaussian mixture model (GMM). The ellipse accounts for the position and spatial distribution of an object, and a GMM represents the color distribution of the object. The occlusion-handling method proposed is based on both the spatial and appearance components of a target's model. Similarly, Hu *et al.* [19] model the human body as a vertical ellipse and use the spatial color mixture of the Gaussian appearance model [20] to model the spatial layout of the colors of a person. The occlusion is deduced using the current states of the interacting targets and handled using the proposed appearance model. However, the aforementioned appearance models do not consider mutual interferences among the interacting targets. Hence, the tracking precision may be greatly affected as mutual occlusion occurs.

In this paper, we attempt to solve the aforementioned problems. We propose an entropy distribution-based algorithm [21] to automatically and accurately estimate the birth intensity. We also propose a game theory-based algorithm to robustly handle the mutual occlusion problem. Entropy, the term that usually refers to the Shannon entropy [22], is a measure of the uncertainty in a random variable. Game theory, which was first proposed by Nash [23], is the study of multi-person decision making. Nash stated that in noncooperative games, sets of optimal strategies [called Nash equilibrium (NE)] are used by the players in a game such that no player can benefit by unilaterally changing his or her strategy if the strategies of the other players remain unchanged. Game theory has been successfully explored in visual tracking [24]–[27]. For example, Yang *et al.* [24] formulate the game-theoretical multi-target tracking for kernel-based tracker. They propose a kernel-based interference model and construct a game to bridge the joint motion estimation with the NE of the game. Inspired by the work of [24], a robust game-theoretical occlusion-handling algorithm based on the improved appearance model is proposed. The main contributions of this paper are as follows.

1) A new birth intensity estimation algorithm is proposed. The birth intensity is first initialized using the previously obtained target states and measurements, and then updated based on the entropy distribution and coverage rate using the currently obtained measurements. By doing so, the noises within the initialized birth intensity will be greatly eliminated.
2) An improved spatial color appearance with interferences by other targets within the occlusion region is modeled. Compared with the conventional color histogram-based appearance model, the proposed model is more robust even when targets in occlusion have similar color distributions.
3) A robust game-theoretical mutual occlusion-handling algorithm is proposed. Unlike in other conventional occlusion-handling algorithms, a noncooperative game is constructed to bridge the joint measurements estimation and the NE of the game.

The rest of this paper is organized as follows. Section II presents the backgrounds on the probability hypothesis density (PHD) filter and the GM-PHD filter. Section III first introduces the measurements classification and birth intensity initialization simply, and then describes the entropy distribution-based and coverage rate-based birth intensity update in detail. Section IV first introduces a simple two-step occlusion reasoning algorithm, and then presents a game-theoretical algorithm to solve the mutual occlusion problem. Some experimental results on publicly available videos are discussed in Section V, and followed by concluding remarks in Section VI.

II. Problem Formulation

For an input image frame of a video sequence at time t, a target region is approximated with a rectangle. The kinematic state of a target i at time t is denoted by $\mathbf{x}_t^i = \{\mathbf{l}_t^i, \mathbf{v}_t^i, \mathbf{s}_t^i\}$. $\mathbf{l}_t^i = \{l_{x,t}^i, l_{y,t}^i\}$, $\mathbf{v}_t^i = \{v_{x,t}^i, v_{y,t}^i\}$, and $\mathbf{s}_t^i = \{w_t^i, h_t^i\}$ are the location, velocity, and size of the target, respectively. $i = 1, \ldots, N_t$, where N_t is the number of targets at time t. Similarly, the model of a measurement j at time t is denoted by $\mathbf{z}_t^j = \{\mathbf{l}_{z,t}^j, \mathbf{s}_{z,t}^j\}$. $j = 1, \ldots, N_{m,t}$, where $N_{m,t}$ is the number of measurements at time t. The target states set and measurements set at time t are denoted by $\mathbf{X}_t = \{\mathbf{x}_t^1, \ldots, \mathbf{x}_t^{N_t}\}$ and $\mathbf{Z}_t = \{\mathbf{z}_t^1, \ldots, \mathbf{z}_t^{N_{m,t}}\}$, respectively. In this paper, an MTVT problem is formulated as the multi-target GM-PHD filtering.

A. PHD Filter

By definition [28], the PHD $D_t(\mathbf{x}_t)$ is the density whose integral on any region S of the state space is the expected number of target N_t contained in S. $\mathbf{x}_t$ is the element of $\mathbf{X}_t$. In general, one cycle of the PHD filter has two steps: prediction and update.

1) Prediction: Suppose that the PHD $D_{t-1}(\mathbf{x}_{t-1})$ at time $t-1$ is known, the predicted PHD is given by

$$D_{t|t-1}(\mathbf{x}_t) = \gamma_t(\mathbf{x}_t) + \int [p_{\mathrm{sv},t}(\mathbf{x}_{t-1}) f_{t|t-1}(\mathbf{x}_t|\mathbf{x}_{t-1}) + p_{\mathrm{sp},t}(\mathbf{x}_t)] \times D_{t-1}(\mathbf{x}_{t-1}) d\mathbf{x}_{t-1} \quad (1)$$

where $f_{t|t-1}(\mathbf{x}_t|\mathbf{x}_{t-1})$ denotes the single-target Markov transition density. $\gamma_t(\mathbf{x}_t)$, $p_{\mathrm{sv},t}(\mathbf{x}_{t-1})$, and $p_{\mathrm{sp},t}(\mathbf{x}_t)$ denote the probabilities of newborn targets, survival targets, and spawned targets, respectively.

2) Update: The predicted PHD is updated with the measurements $\mathbf{Z}_t$ obtained at time t. The number of clutters is assumed to be Poisson distributed with the average rate of λ_t, and the probability density of the spatial distribution of clutters is $c_t(\mathbf{z}_t)$. $\mathbf{z}_t$ is the element of $\mathbf{Z}_t$. Let the detection probability be $p_{d,t}(\mathbf{x}_t)$. Then, the updated PHD is given by

$$D_t(\mathbf{x}_t) = [1 - p_{d,t}(\mathbf{x}_t)]D_{t|t-1}(\mathbf{x}_t) + \sum_{\mathbf{z}_t \in \mathbf{Z}_t} \frac{p_{d,t}(\mathbf{x}_t) g_t(\mathbf{z}_t|\mathbf{x}_t) D_{t|t-1}(\mathbf{x}_t)}{\lambda_t c_t(\mathbf{z}_t) + \int p_{d,t}(\mathbf{x}_t) g_t(\mathbf{z}_t|\mathbf{x}_t) D_{t|t-1}(\mathbf{x}_t) d\mathbf{x}_t} \quad (2)$$

where $g_t(\mathbf{z}_t|\mathbf{x}_t)$ denotes the single-target likelihood.

B. GM-PHD Filter

The GM-PHD filter is a closed solution to the PHD filter. To implement it, certain assumptions are needed: 1) each target follows a linear dynamical model where the process and observation noises are Gaussian: $f_{t|t-1}(\mathbf{x}_t|\mathbf{x}_{t-1}) = N(\mathbf{x}_t; \boldsymbol{F}_t\mathbf{x}_{t-1}, \boldsymbol{Q}_t)$ and $g_t(\mathbf{z}_t|\mathbf{x}_t) = N(\mathbf{z}_t; \boldsymbol{H}_t\mathbf{x}_t, \boldsymbol{R}_t)$.$N(\cdot; \mathbf{m}, \mathbf{P})$ denotes a Gaussian component with the mean $\mathbf{m}$ and the covariance $\mathbf{P}$. $\boldsymbol{F}_t$ and $\boldsymbol{H}_t$ are the transition and the measurement matrices, respectively. $\boldsymbol{Q}_t$ and $\boldsymbol{R}_t$ are the covariance matrices of the process noise and the measurement noise, respectively; 2) the survival and detection probabilities are independent of the target state: $p_{\mathrm{sv},t}(\mathbf{x}_{t-1}) = p_{\mathrm{sv}}$ and $p_{d,t}(\mathbf{x}_t) = p_d$; and 3) the birth intensity can be represented by $\gamma_t(\mathbf{x}_t) = \sum_{i=1}^{J_{\gamma,t}} \omega_{\gamma,t}^{(i)} \mathrm{N}(\mathbf{x}_t; \mathbf{m}_{\gamma,t}^{(i)}, \mathbf{P}_{\gamma,t}^{(i)})$, where $J_{\gamma,t}$, $\omega_{\gamma,t}^{(i)}$, $\mathbf{m}_{\gamma,t}^{(i)}$, and $\mathbf{P}_{\gamma,t}^{(i)}$ are the Gaussian mixture parameters [29].

According to [29], the GM-PHD filter is implemented as follows.

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

$$D_{t|t-1}(\mathbf{x}_t) = \gamma_t(\mathbf{x}_t) + p_{\mathrm{sv}} \times \sum_{i=1}^{J_{t-1}} \omega_{t-1}^{(i)} \mathrm{N}(\mathbf{x}_t; \mathbf{m}_{\mathrm{sv},t|t-1}^{(i)}, \mathbf{P}_{\mathrm{sv},t|t-1}^{(i)}). \quad (3)$$

Update: The $D_{t|t-1}(\mathbf{x}_t)$ can be expressed as a Gaussian mixture of the form $D_{t|t-1}(\mathbf{x}_t) = \sum_{i=1}^{J_{t|t-1}} \omega_{t|t-1}^{(i)} N(\mathbf{x}_t; \mathbf{m}_{t|t-1}^{(i)}, \mathbf{P}_{t|t-1}^{(i)})$. Then, the posterior intensity is given by

$$D_t(\mathbf{x}_t) = (1 - p_d) D_{t|t-1}(\mathbf{x}_t) + \sum_{\mathbf{z}_t \in \mathbf{Z}_t} D_{g,t}(\mathbf{x}_t; \mathbf{z}_t) \quad (4)$$

$$D_{g,t}(\mathbf{x}_t; \mathbf{z}_t) = \sum_{i=1}^{J_{t|t-1}} \omega_{g,t}^{(i)}(\mathbf{z}_t) N(\mathbf{x}_t; \mathbf{m}_{g,t}^{(i)}(\mathbf{z}_t), \mathbf{P}_{g,t}^{(i)}(\mathbf{z}_t)) \quad (5)$$

$$\omega_{g,t}^{(i)}(r_t, \mathbf{z}_t) = \frac{p_d \omega_{t|t-1}^{(i)} N(\mathbf{z}_t; \mathbf{m}_{h,t}^{(i)}, \mathbf{P}_{h,t}^{(i)})}{\lambda_t c_t(\mathbf{z}_t) + p_d \sum_{i=1}^{J_{t|t-1}} \omega_{t|t-1}^{(i)} N(\mathbf{z}_t; \mathbf{m}_{h,t}^{(i)}, \mathbf{P}_{h,t}^{(i)})} \quad (6)$$

where $\mathbf{m}_{\mathrm{sv},t|t-1}^{(i)} = \boldsymbol{F}_t \mathbf{m}_{t-1}^{(i)}$, $\mathbf{P}_{\mathrm{sv},t|t-1}^{(i)} = \boldsymbol{Q}_t + \boldsymbol{F}_t \mathbf{P}_{t-1}^{(i)} \boldsymbol{F}_t^T$, $\mathbf{m}_{g,t}^{(i)}(\mathbf{z}_t) = \mathbf{m}_{t|t-1}^{(i)} + K(\mathbf{z}_t - \boldsymbol{H}_t \mathbf{m}_{t|t-1}^{(i)})$, $\mathbf{P}_{g,t}^{(i)}(\mathbf{z}_t) = (\mathbf{I} - K\boldsymbol{H}_t)\mathbf{P}_{t|t-1}^{(i)}$, $K = \mathbf{P}_{t|t-1}^{(i)} \boldsymbol{H}_t^T \bullet (\boldsymbol{H}_t \mathbf{P}_{t|t-1}^{(i)} \boldsymbol{H}_t^T + \boldsymbol{R}_t)^{-1}$, $\mathbf{m}_{h,t}^{(i)} = \boldsymbol{H}_t \mathbf{m}_{t|t-1}^{(i)}$, and $\mathbf{P}_{h,t}^{(i)} = \boldsymbol{R}_t + \boldsymbol{H}_t \mathbf{P}_{t|t-1}^{(i)} \boldsymbol{H}_t^T$.

The spawned targets in the prediction step of the PHD filter (1) usually come from the requirements of military applications for radar tracking, e.g., an airplane sends a missile [11]. For simplicity, we assume that all targets in our tracking scenario consist of survival targets and newborn targets. The prediction and update steps discussed above indicate that the number of components of the predicted and posterior intensities increases with time. To solve this problem, we use the pruning and merging algorithms proposed by Vo and Ma [29] to prune the components that are irrelevant to the target intensity and to merge the components that share the same intensity peak into one component. The peaks of the intensity are the points of the highest local concentration of the expected number N_t of the targets. The estimate of the multi-target states is the set of N_t ordered of the mean with the largest weights.

As shown in (3), the birth intensity $\gamma_t(\mathbf{x}_t)$ needs to be accurately estimated before the prediction step. As shown in (4), the predicted PHD is updated by the measurements. Once the mutual occlusion occurs, the measurements originating from the targets within the occlusion region will be merged into one measurement. The merging will affect the update results of the filter and ultimately the tracking performance. This paper focuses on solving the aforementioned problems.

III. Birth Intensity Estimation

A new birth intensity estimation algorithm based on the entropy distribution and coverage rate is proposed. Fig. 1 shows an illustration of the proposed birth intensity estimation process in one cycle of the GM-PHD filter. The measurements $\mathbf{Z}_t$ are obtained by object detection and are classified into two parts: the birth measurements $\mathbf{Z}_{b,t}$ and the survival measurements $\mathbf{Z}_{s,t}$. The birth intensity is first initialized using the previously obtained target states $\mathbf{X}_{t-1}$ and measurements $\mathbf{Z}_{t-1}$. The initialized birth intensity $\gamma_{\mathrm{ini},t}(\mathbf{x}_t)$ is then updated using the birth measurements $\mathbf{Z}_{b,t}$.

A. Object Detection

The measurements are obtained by object detection. Any object detection method can be incorporated into our tracking system. To show the robustness of the proposed algorithm for tracking targets in a noisy video, a simple background subtraction algorithm for object detection is utilized. The static background image is assumed to be already known. First, each pixel in the background image is modeled as red, green, and blue channels. Then, the difference between the current image and the background image for each channel is calculated; the pixel is labeled as a foreground if the difference of one channel is larger than the threshold τ_1. Finally, the morphological operator is employed to eliminate the isolated noises, and the eight-connected component labeling algorithm is used to connect the

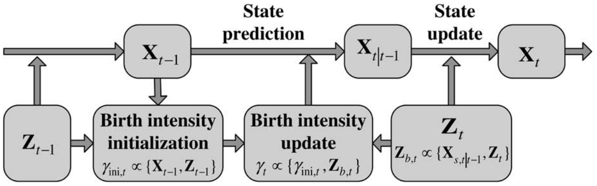
State prediction
State update
$\mathbf{X}_{t-1}$
$\mathbf{X}_{t|t-1}$
$\mathbf{X}_t$
$\mathbf{Z}_{t-1}$
Birth intensity initialization
$\gamma_{\mathrm{ini},t} \propto \{\mathbf{X}_{t-1}, \mathbf{Z}_{t-1}\}$
Birth intensity update
$\gamma_t \propto \{\gamma_{\mathrm{ini},t}, \mathbf{Z}_{b,t}\}$
$\mathbf{Z}_t$
$\mathbf{Z}_{b,t} \propto \{\mathbf{X}_{s,t|t-1}, \mathbf{Z}_t\}$

Fig. 1. Birth intensity estimation process in one cycle of the GM-PHD filter.

detected foreground pixels to a set of regions. Each connected region is enclosed by a rectangle. The state (location and size) of one rectangle represents one measurement.

Although the morphological operator can remove some isolated noises of small sizes, noises of big sizes caused by an unstable environment may still exist in the measurements. Furthermore, the measurements can be affected by choosing different values of $\tau_1 (0<\tau_1<255)$. The smaller the τ_1 is, the larger the number of noises is while the more the foreground pixels of true targets are. In experiments, we choose $\tau_1 = 20$ to ensure all true targets are detected regardless of the number of noises.

B. Measurements Classification

The measurements obtained may be generated by the survival targets, newborn targets, and noises. To eliminate interferences by those measurements generated by the survival targets, we classify the measurements into two parts: the birth measurements $\mathbf{Z}_{b,t}$ originating from the newborn targets and the survival measurements $\mathbf{Z}_{s,t}$ originating from the survival targets. The jth measurement $\mathbf{z}_t^j$ is regarded as the survival measurement $\mathbf{z}_{s,t}^j$, if it satisfies

$$\mathbf{z}_{s,t}^j = \left\{ \mathbf{z}_t^j \middle| \left\| \mathbf{l}_{z,t}^j - \mathbf{l}_{s,t|t-1}^i \right\| < \mathbf{v}_{\max}^i \bullet T \right\} \tag{7}$$

where $\mathbf{l}_{z,t}^j \in \mathbf{z}_t^j$ and $\mathbf{l}_{s,t|t-1}^i \in \mathbf{x}_{s,t|t-1}^i$, $\mathbf{x}_{s,t|t-1}^i = \boldsymbol{H}_t \boldsymbol{F}_t \mathbf{x}_{t-1}^i$ is the predicted state of $\mathbf{x}_{t-1}^i$, $\mathbf{v}_{\max}^i = \max\{\|\mathbf{v}_1^i\|, \ldots, \|\mathbf{v}_{t-1}^i\|\}$ is the maximum velocity of a target i up to time $t-1$ ($t > 0$, t is an integer), $i = 1, \ldots, N_{t-1}$, $j = 1, \ldots, N_{m,t}$, $T = 1$ frame is the interval between two consecutive time steps, and $\| \bullet \|$ is the Euclidean norm (hereinafter the same). The residual measurements are the birth measurements

$$\mathbf{Z}_{b,t} = \mathbf{Z}_t - \mathbf{Z}_{s,t}. \tag{8}$$

C. Birth Intensity Initialization

Based on the target states, $\mathbf{X}_{t-1} = \{\mathbf{x}_{t-1}^i\}_{i=1}^{N_{t-1}} = \{\mathbf{l}_{t-1}^i, \mathbf{v}_{t-1}^i, \mathbf{s}_{t-1}^i\}_{i=1}^{N_{t-1}}$ and the measurements $\mathbf{Z}_{t-1} = \{\mathbf{z}_{t-1}^j\}_{j=1}^{N_{m,t-1}} = \{\mathbf{l}_{z,t-1}^j, \mathbf{s}_{z,t-1}^j\}_{j=1}^{N_{m,t-1}}$, the measurements $\mathbf{Z}_{\mathrm{cnew},t-1}$ originating from the candidate newborn targets are obtained by

$$\mathbf{z}_{\mathrm{tra},t-1}^j = \left\{ \mathbf{z}_{t-1}^j \middle| \left\| \mathbf{l}_{z,t-1}^j - \mathbf{l}_{t-1}^i \right\| < \frac{1}{2} \|\mathbf{s}_{t-1}^i\| \right\} \tag{9}$$

$$\mathbf{Z}_{\mathrm{cnew},t-1} = \mathbf{Z}_{t-1} - \mathbf{Z}_{\mathrm{tra},t-1} \tag{10}$$

where $\mathbf{Z}_{\mathrm{tra},t-1} = \{\mathbf{z}_{\mathrm{tra},t-1}^j\}_{j=1}^{N_{\mathrm{tra},t-1}}$ is the measurement originating from the tracked targets at time $t-1$. $N_{\mathrm{tra},t-1}$ is the number of measurements in $\mathbf{Z}_{\mathrm{tra},t-1}$. At the initial time step ($t=0$), all measurements obtained are regarded as $\mathbf{Z}_{\mathrm{cnew},t-1}$, because no target is tracked at first. The birth intensity for the next time step is then initialized by a Gaussian mixture

$$\begin{aligned} \gamma_{\mathrm{ini},t}(\mathbf{x}_t) &= p(\mathbf{x}_t|\boldsymbol{\theta}) = \sum_{m=1}^{M} \pi_m p(\mathbf{x}_t|\boldsymbol{\theta}_m) \\ &= \sum_{m=1}^{M} \pi_m N(\mathbf{x}_t; \mathbf{u}_m, \boldsymbol{\varphi}_m) \end{aligned} \tag{11}$$

where $\boldsymbol{\theta}_m = \{\boldsymbol{\mu}_m, \boldsymbol{\varphi}_m\}$ is a parameter set of the mth Gaussian component that contains the mean $\boldsymbol{\mu}_m = \boldsymbol{H}_{t-1}^{-1} \mathbf{z}_{\mathrm{cnew},t-1}^m$ and the covariance $\boldsymbol{\varphi}_m = \boldsymbol{H}_{t-1}^{-1} \boldsymbol{R}_{t-1} (\boldsymbol{H}_{t-1}^{-1})^T$. $\boldsymbol{\theta} = \{\pi_1, \ldots, \pi_M, \boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_M\}$ is a parameter set of M components. $M = N_{m,t-1} - N_{\mathrm{tra},t-1}$ is the number of measurements in $\mathbf{Z}_{\mathrm{cnew},t-1}$. π_m is the weight of the mth Gaussian component and satisfies the following condition

$$\sum_{m=1}^{M} \pi_m = 1. \tag{12}$$

D. Entropy Distribution-Based Birth Intensity Update

After initializing the birth intensity, the next step is to update the parameter set $\boldsymbol{\theta}$ according to $\mathbf{Z}_{b,t}$. Due to the noisy video data, the initialized birth intensity may contain some noises. To eliminate such noises, the Shannon entropy [22] is used to model the prior distribution of the parameter set $\boldsymbol{\theta}$. For a random variable $\mathbf{X}$ with n outcomes $\{x_1, x_2, \ldots, x_n\}$, the Shannon entropy $H(\mathbf{X})$ is defined as

$$H(\mathbf{X}) = -\sum_{i=1}^{n} p(x_i) \log p(x_i) \tag{13}$$

where $p(x_i)$ is the probability mass function of x_i. We select the negative exponent with the Shannon entropy dependent only on the mixture weight π_m as the prior distribution of $\boldsymbol{\theta}$

$$p(\boldsymbol{\theta}) = \exp\{-H(\pi_m)\} \tag{14}$$

where $\boldsymbol{H}(\pi_m) = -\sum_{m=1}^{M} \pi_m \log \pi_m$ is the entropy measure. Then, we use the birth measurements $\mathbf{Z}_{b,t} = \{\mathbf{z}_{b,t}^i\}_{i=1}^{N_{\mathrm{bm},t}}$ to update $\boldsymbol{\theta}$, where $N_{\mathrm{bm},t}$ is the number of birth measurements. By doing so, the weights of the components within the initialized birth intensity those are irrelevant to the birth measurements will rapidly become small. The corresponding components should be removed once their weights become negative.

Given the $\mathbf{Z}_{b,t}$, its log-likelihood can be given by

$$\log p(\mathbf{Z}_{b,t}|\boldsymbol{\theta}) = \sum_{i=1}^{N_{\mathrm{bm},t}} \log \sum_{m=1}^{M} \pi_m g(\mathbf{z}_{b,t}^i|\boldsymbol{\theta}_m) \tag{15}$$

where $g(\mathbf{z}_{b,t}^i|\boldsymbol{\theta}_m)$ represents the single-target likelihood in the mth Gaussian component. The parameter set $\boldsymbol{\theta}$ can be estimated by the criterion of the maximum a posterior (MAP)

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \{\log p(\mathbf{Z}_{b,t}|\boldsymbol{\theta}) + \log p(\boldsymbol{\theta})\}. \tag{16}$$

To estimate π_m, we set the derivative of the log-posterior with respect to π_m to zero under the constraint from (12)

$$\frac{\partial}{\partial \pi_m} \left(\log p(\mathbf{Z}_{b,t}|\boldsymbol{\theta}) + \log p(\boldsymbol{\theta}) + \lambda \left(\sum_{m=1}^{M} \pi_m - 1 \right) = 0 \tag{17}$$

where λ is a Lagrange multiplier. Substituting (14) and (15) into (17) yields

$$\sum\nolimits_{i=1}^{N_{\mathrm{bm},t}} W_m(\mathbf{z}_{b,t}^i)/\pi_m + \log \pi_m + \lambda + 1 = 0 \tag{18}$$

$$W_m(\mathbf{z}_{b,t}^i) = \frac{\pi_m g(\mathbf{z}_{b,t}^i|\boldsymbol{\theta}_m)}{\sum_{m=1}^{M} \pi_m g(\mathbf{z}_{b,t}^i|\boldsymbol{\theta}_m)} \tag{19}$$

where $W_m(\mathbf{z}_{b,t}^i)$ reflects how much the $\mathbf{z}_{b,t}^i$ belongs to the mth Gaussian component. Multiplying both sides of (18) by π_m and summing over m using the constraint from (12), the following is obtained

$$\lambda = -N_{\mathrm{bm},t} - 1 - \sum\nolimits_{m=1}^{M} \pi_m \log \pi_m. \tag{20}$$

Given $\{\pi_m\}_{m=1}^M$, λ is calculated by (20). Substituting it into (18) and multiplying both sides by π_m yields

$$N_m + \pi_m \log \pi_m + \pi_m(\lambda + 1) = 0 \tag{21}$$

where $N_m = \sum_{i=1}^{N_{\mathrm{bm},t}} W_m(\mathbf{z}_{b,t}^i)$.

Given λ, the goal is to calculate π_m from (21). As $0<\pi_m<1$, we use the Taylor expansion to expand $\log \pi_m$ at $\pi_m = 1$ and select the first order to approximate $\log \pi_m$

$$\log \pi_m \approx \pi_m - 1. \tag{22}$$

Substituting (22) into (21), π_m is obtained

$$\pi_m = -\lambda/2 \pm \sqrt{\lambda^2/4 - N_m}, \quad (0<\pi_m<1). \tag{23}$$

Similarly, the MAP estimations of $\boldsymbol{\mu}_m$ and $\boldsymbol{\varphi}_m$ are obtained

$$\boldsymbol{\mu}_m = \frac{1}{N_m} \sum_{i=1}^{N_{\mathrm{bm},t}} W_m(\mathbf{z}_{b,t}^i) \cdot \boldsymbol{H}_t^{-1} \mathbf{z}_{b,t}^i \tag{24}$$

$$\boldsymbol{\varphi}_m = \frac{1}{N_m} \sum_{i=1}^{N_{\mathrm{bm},t}} W_m(\mathbf{z}_{b,t}^i) \cdot (\boldsymbol{H}_t^{-1} \mathbf{z}_{b,t}^i - \boldsymbol{\mu}_m) \cdot (\boldsymbol{H}_t^{-1} \mathbf{z}_{b,t}^i - \boldsymbol{\mu}_m)^T. \tag{25}$$

After each iteration step, the components within the birth intensity whose weights are negative are removed from the mixing components set. The weights of the remaining components are then normalized for the next iteration step. The iteration terminates when the difference rate of the log-posterior is smaller than the preset threshold $\tau_2 (0<\tau_2<1,\ \tau_2 = 0.5$ in our experiments); the updated parameter set $\boldsymbol{\theta}$ of the birth intensity is then obtained.

E. Coverage Rate-Based Birth Intensity Update

Fig. 2 is a pictorial example that shows the probable overlapping between the birth intensity component $\boldsymbol{\theta}_m$ and the birth measurement $\mathbf{z}_{b,t}^i$ ($m = 6$ and $i = 4$ in this figure), where $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ are the components for the newborn targets and $\boldsymbol{\theta}_3 - \boldsymbol{\theta}_6$ are the components for the noises.

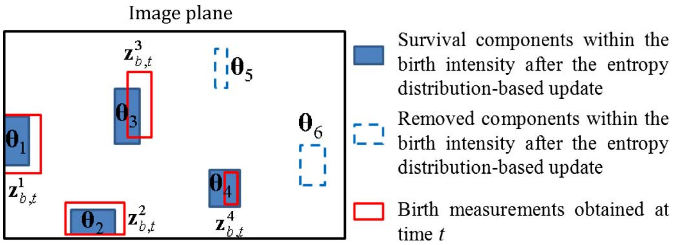

Fig. 2. Pictorial example of probable overlapping between the birth intensity components and the birth measurements.

After the entropy distribution-based birth intensity update step, the Gaussian components within the initialized birth intensity that are irrelevant to the birth measurements are removed (shown as $\boldsymbol{\theta}_5$ and $\boldsymbol{\theta}_6$ in Fig. 2). However, some noises may still exist in the birth intensity (shown as $\boldsymbol{\theta}_3$ and $\boldsymbol{\theta}_4$ in Fig. 2). To further eliminate these noises, a coverage rate-based method is proposed according to the fact that newborn targets enter the scene gradually (shown as $\boldsymbol{\theta}_1$ and $\boldsymbol{\theta}_2$ in Fig. 2), whereas noises appear irregularly. The coverage rate consists of two parts: the intersection rate and the area rate. For each survival Gaussian component $\boldsymbol{\theta}_m$ within the birth intensity, we define the intersection rate $I_{r,t}^m$ and the area rate $A_{r,t}^m$ as

$$I_{r,t}^m = S(\boldsymbol{\theta}_m \cap \mathbf{z}_{b,t}^i)/S(\boldsymbol{\theta}_m) \tag{26}$$

$$A_{r,t}^m = S(\mathbf{z}_{b,t}^i)/S(\boldsymbol{\theta}_m) \tag{27}$$

where $S(\bullet)$ is a function to compute the area. $\boldsymbol{\theta}_m \cap \mathbf{z}_{b,t}^i$ is the intersection part of the mth Gaussian component and the ith birth measurement. Here, the ith birth measurement closest to the mth Gaussian component is selected. As shown in Fig. 2, the component is probably a newborn target when $I_{r,t}^m$ approximates to 1 and, at the same time, $A_{r,t}^m$ is larger than 1. Otherwise, the weight of this component should be greatly reduced. To distinguish the newborn targets from the noises, the weight π_m is updated by

$$\pi_m = \pi_m \cdot \left(1 - \exp\left(\frac{-(I_{r,t}^m)^2}{2\sigma^2}\right)\right) \cdot \left(1 - \exp\left(\frac{-(A_{r,t}^m - 1)^2}{2\sigma^2}\right)\right) \tag{28}$$

where σ is the standard deviation that is preset as $\sigma = 0.2$ to control the width of the distribution. Once the weight is below the given threshold τ_3 ($0<\tau_3<1$, $\tau_3 = 0.1$ in the experiments), the corresponding component is removed. Once all the components are updated, the weights are normalized and the birth intensity is finally obtained.

IV. Mutual Occlusion Handling

To track the targets in mutual occlusion, a robust occlusion-handling algorithm based on the game theory is proposed. First, the mutual occlusion region is determined by a two-step occlusion reasoning algorithm. Then, the spatial color appearance model is improved by incorporating the interferences of other targets within the occlusion region. Finally, a noncooperative game is constructed to obtain the optimal locations

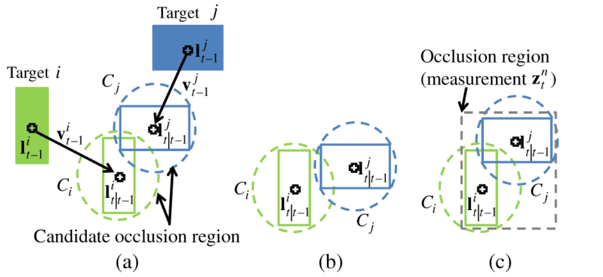

Fig. 3. Illustration of occlusion reasoning. (a) Occlusion prediction. (b) No occlusion occurs. (c) Occlusion occurs and occlusion region is determined.

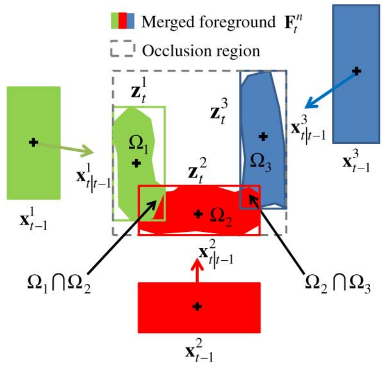

Fig. 4. Pictorial example of mutual occlusion analysis.

of the measurements originating from the targets within the occlusion region.

A. Occlusion Reasoning

Fig. 3 shows an illustration of occlusion reasoning that includes occlusion prediction and occlusion determination.

1) Occlusion Prediction: In Fig. 3(a), C_i (or C_j) is a circle at center $\mathbf{l}^i_{t|t-1}$ (or $\mathbf{l}^j_{t|t-1}$) with radius $\|\mathbf{s}^i_{t|t-1}\|$ (or $\|\mathbf{s}^j_{t|t-1}\|$). $\mathbf{l}^i_{t|t-1}$ (or $\mathbf{l}^j_{t|t-1}$) and $\mathbf{s}^i_{t|t-1}$ (or $\mathbf{s}^j_{t|t-1}$) are the location and size of the predicted state $\mathbf{x}^i_{t|t-1}$ (or $\mathbf{x}^j_{t|t-1}$) of the target i (or j), respectively. The candidate occlusion region is predicted only when $C_i \cap C_j \neq \emptyset$ ($i \neq j$), i.e.,

$$\left\|\mathbf{l}^i_{t|t-1} - \mathbf{l}^j_{t|t-1}\right\| < \left\|\mathbf{s}^i_{t|t-1}\right\| + \left\|\mathbf{s}^j_{t|t-1}\right\|. \tag{29}$$

Otherwise, no occlusion occurs.

2) Occlusion Determination: Two possible situations are possible in the candidate occlusion region: no occlusion and occlusion [shown in Fig. 3(b) and (c)]. As the overlap between the targets in occlusion always increases gradually, the size of the first detected merged measurement is always larger than the size of the corresponding single target. To further determine the occlusion region, the measurements (detections) obtained at current time t is incorporated. If a measurement $\mathbf{z}^n_t = \{\mathbf{l}^n_{z,t}, \mathbf{s}^n_{z,t}\}$ $(n = 1, \ldots, N_{m,t})$ within the candidate occlusion region satisfies (30), this measurement is regarded as an occlusion region

$$\left\|\mathbf{s}^n_{z,t}\right\| > \varepsilon \bullet \max\left\{\left\|\mathbf{s}^i_{t|t-1}\right\|, \left\|\mathbf{s}^j_{t|t-1}\right\|\right\} \tag{30}$$

where ε is a scale factor. The size of the detected target may slightly be changed between the consecutive frames because of the changes in the target's pose or because of the depth of view. Compared with the size of the target before mutual occlusion, the size of the target after mutual occlusion is largely changed because it is merged with other targets. Consequently, we set $\varepsilon = 1.2$ to determine the occlusion region correctly.

B. Occlusion Analysis

As mutual occlusion occurs, the occlusion region $\mathbf{z}^n_t$ that contains the merged foreground $\mathbf{F}^n_t$ can be determined by occlusion reasoning. Then, the identities (IDs) and number N^n_o of the targets involved in this occlusion region can be determined. Fig. 4 shows a pictorial example of the mutual occlusion analysis. At time $t-1$, three targets are isolated with the states $\mathbf{x}^1_{t-1}$, $\mathbf{x}^2_{t-1}$, and $\mathbf{x}^3_{t-1}$, respectively. At time t, mutual occlusion occurs and the measurements $\mathbf{z}^1_t$, $\mathbf{z}^2_t$, and $\mathbf{z}^3_t$ originating from these three targets are merged into one measurement $\mathbf{z}^n_t$. In this figure, Ω_i represents the support region within the bounding box of the target i ($i = 1, 2, 3$).

To track the targets in the occlusion region, the measurement $\mathbf{z}^n_t$ should be correctly segmented. In other words, the goal is to obtain the optimal individual measurements $(\mathbf{z}^{1*}_t, \ldots, \mathbf{z}^{N^n_o*}_t)$. Given $\mathbf{F}^n_t$, the optimal solution $(\mathbf{z}^{1*}_t, \ldots, \mathbf{z}^{N^n_o*}_t)$ is to maximize the similarity probability between $(\mathbf{z}^1_t, \ldots, \mathbf{z}^{N^n_o}_t)$ and $\mathbf{F}^n_t$

$$\begin{aligned}\left(\mathbf{z}^{1*}_t, \ldots, \mathbf{z}^{N^n_o*}_t\right) &= \underset{\{\mathbf{z}^i_t\}^{N^n_o}_{i=1}}{\arg\max}\, P\left(\mathbf{z}^1_t, \ldots, \mathbf{z}^{N^n_o}_t \middle| \mathbf{F}^n_t\right)\\ &= \underset{\{\mathbf{z}^i_t\}^{N^n_o}_{i=1}}{\arg\max}\, P\left(\mathbf{z}^i_t, \mathbf{z}^{-i}_t \middle| \mathbf{F}^n_t\right)\\ &= \underset{\{\mathbf{z}^i_t\}^{N^n_o}_{i=1}}{\arg\max}\, P\left(\mathbf{z}^i_t \middle| \mathbf{F}^n_t, \mathbf{z}^{-i}_t\right) P\left(\mathbf{z}^{-i}_t \middle| \mathbf{F}^n_t\right)\end{aligned} \tag{31}$$

where $\mathbf{z}^{-i}_t = \{\mathbf{z}^j_t\}^{N^n_o}_{j=1, j\neq i}$.

To obtain the optimal solution of (31), an improved spatial color appearance with interferences of other targets within the occlusion region is modeled to measure the similarity probability. In addition, a robust game-theoretical algorithm is proposed to bridge the optimal solution of (31) and the constructed game.

C. Improved Appearance Model With Target Interferences

The appearance of a target i is modeled as a GMM $q^i = q^i(\omega^i_k, \mu^i_k, \Sigma^i_k)$, representing the color distribution of the target pixels, where $(\omega^i_k, \mu^i_k, \Sigma^i_k)$ represents the weight, mean, and covariance matrix of the kth Gaussian component of the mixture, respectively, $k = 1, \ldots, K$, and K is the number of Gaussian components. The measure of the similarity $P_s(p^i, q^i)$ between the candidate p^i (for a target i after occlusion) and the model q^i (for a target i before occlusion) is defined as the probability that p^i's colors are drawn from

q^i with a spatial constraint [30]. Equation (32) is shown at the bottom of the page, where $N(c;\mu,\Sigma)=[2\pi|\Sigma|]^{-1/2}\exp\{-\frac{1}{2}(c-\mu)'\Sigma^{-1}(c-\mu)\}$, $\Sigma_t^i=[(w_t^i/2)^2,0;0,(h_t^i/2)^2]$, $c_{\mathbf{l}^i}=(r_{\mathbf{l}^i},g_{\mathbf{l}^i},I_{\mathbf{l}^i})$ is the color of the pixel located in $\mathbf{l}^i$ within the target region Ω_i of p^i, $g_{\mathbf{l}^i}=G_{\mathbf{l}^i}/(R_{\mathbf{l}^i}+G_{\mathbf{l}^i}+B_{\mathbf{l}^i})$, $r_{\mathbf{l}^i}=R_{\mathbf{l}^i}/(R_{\mathbf{l}^i}+G_{\mathbf{l}^i}+B_{\mathbf{l}^i})$, $I_{\mathbf{l}^i}=(R_{\mathbf{l}^i}+G_{\mathbf{l}^i}+B_{\mathbf{l}^i})/3$, N_i is the number of foreground pixels in Ω_i, and $N(\mathbf{l}^i;\mathbf{l}_t^i,\Sigma_t^i)$ is the spatial constraint of the foreground pixels.

As occlusion occurs, the interferences of other targets within the occlusion region also need to be considered. The measure of the similarity is then improved as shown in equation (33) at the bottom of the page, where p^j is the appearance of the other targets within the occlusion region. $\Psi_1=\Omega_i-\Omega_i\cap\Omega_j$, $\Psi_2=\Omega_i\cap\Omega_j$, $\sigma_1=\sum_{k=1}^{K}\omega_k^i N(c_{\mathbf{l}^i};\mu_k^i,\Sigma_k^i)$, $\sigma_2=\sum_{j=1}^{N_o^n}\sum_{k=1}^{K}\omega_k^j N(c_{\mathbf{l}^i};\mu_k^j,\Sigma_k^j)$, and $N(\mathbf{l}_t^i)=N(\mathbf{l}^i;\mathbf{l}_t^i,\Sigma_t^i)$. Note that if no occlusion or overlap occurs between the targets, (33) degenerates to (32).

D. Game-Theoretical Mutual Occlusion Handling

To obtain the optimal solution of (31), an algorithm based on game theory is proposed. In game theory, a noncooperative game is one in which players make decisions independently. As mutual occlusion occurs, the individual measurements involved in the occlusion region compete to independently maximize the similarity probability between the measurements and the foreground. Therefore, constructing a noncooperative game to bridge the joint measurements estimation with NE of the game is reasonable. We construct an n-person, nonzero-sum, noncooperative game and assume that the size of the target remains constant during the occlusion. With this assumption, the estimation of the measurements $(\mathbf{z}_t^1,\ldots,\mathbf{z}_t^{N_o^n})$ is simplified to the estimation of the locations $(\mathbf{l}_{z,t}^1,\ldots,\mathbf{l}_{z,t}^{N_o^n})$ of the measurements.

Normally, a game consists of three components: the player, the strategy of the player, and the corresponding utility. In the constructed game, these components are defined as follows:

Player: The individual measurement $\mathbf{z}_t^i$ originating from the target $i\in\{1,\ldots,N_o^n\}$ within the occlusion region.

Strategy: Motion of the player, i.e., the location $\mathbf{l}_{z,t}^i$ of the player $\mathbf{l}_{z,t}^i=\{l_{x,t}^i,l_{y,t}^i\}\in\mathbb{R}^2$.

Utility: $U^i(\mathbf{l}_{z,t}^i,\mathbf{l}_{z,t}^{-i})=P(\mathbf{z}_t^i|\mathbf{F}_t^n,\mathbf{z}_t^{-i})P(\mathbf{z}_t^{-i}|\mathbf{F}_t^n)$, where $\mathbf{l}_{z,t}^{-i}=\{\mathbf{l}_{z,t}^j\}_{j=1,j\neq i}^{N_o^n}$.

To find an NE of the game, the *best response* should be defined first.

Definition 1 [31]: The *best response* of a player i to the profile of strategies $\mathbf{l}_{z,t}^{-i}$ is the strategy of that player such that

$$U^i(\mathbf{l}_{z,t}^i,\mathbf{l}_{z,t}^{-i})\geq U^i(\mathbf{l}_{z,t}^{i'},\mathbf{l}_{z,t}^{-i})\quad\forall\mathbf{l}_{z,t}^{i'}\in\mathbb{R}^2. \tag{34}$$

Hence, an NE of the game is a strategy profile for which the strategy of every player is a *best response* to the strategies of other players.

Definition 2 [31]: $(\mathbf{l}_{z,t}^{1*},\ldots,\mathbf{l}_{z,t}^{N_o^n*})$ is an NE for the game with utility $\{U^i(\mathbf{l}_{z,t}^i,\mathbf{l}_{z,t}^{-i})\}_{i=1,\ldots,N_o^n}$ if the strategy of every player is a *best response* to the strategies of other players.

$$U^i(\mathbf{l}_{z,t}^{i*},\mathbf{l}_{z,t}^{-i*})\geq U^i(\mathbf{l}_{z,t}^i,\mathbf{l}_{z,t}^{-i*}),\quad\text{for every player } i. \tag{35}$$

Given the $\mathbf{l}_{z,t}^{-i*}$, the goal is to determine the *best response* of the player i, i.e.,

$$\begin{aligned}\max U^i(\mathbf{l}_{z,t}^i,\mathbf{l}_{z,t}^{-i*})&=\max P(\mathbf{z}_t^i|\mathbf{F}_t^n,\mathbf{z}_t^{-i*})P(\mathbf{z}_t^{-i*}|\mathbf{F}_t^n)\\&\propto\max P(\mathbf{z}_t^i|\mathbf{F}_t^n,\mathbf{z}_t^{-i*})\end{aligned} \tag{36}$$

where $P(\mathbf{z}_t^i|\mathbf{F}_t^n,\mathbf{z}_t^{-i*})$ is the similarity probability between $\mathbf{z}_t^i$ and $\mathbf{F}_t^n$ with interferences of other fixed measurements $\mathbf{z}_t^{-i*}$. Maximizing the $P(\mathbf{z}_t^i|\mathbf{F}_t^n,\mathbf{z}_t^{-i*})$ is equal to maximizing the measure of similarity $P_s(p^i,q^i|p^j)$, where p^i, q^i, and p^j are the color appearances of the measurement $\mathbf{z}_t^i$ originating from the target i, of the corresponding target model before occlusion, and of the other measurement j ($j\neq i$) within the occlusion region, respectively. To obtain the *best response* $\mathbf{l}_{z,t}^{i*}$, the derivative of $P_s(p^i,q^i|p^j)$ with respect to $\mathbf{l}_{z,t}^i$ is set to zero

$$\sum_{\Psi_1}\mathbf{l}^i-N_i'\bullet\mathbf{l}_{z,t}^i+\sum_{\Psi_2}\frac{\sigma_1}{\sigma_2}\bullet\mathbf{l}^i-\mathbf{l}_{z,t}^i\bullet\sum_{\Psi_2}\frac{\sigma_1}{\sigma_2}=0 \tag{37}$$

where N_i' is the number of foreground pixels in the support region Ψ_1. $\mathbf{l}_{z,t}^i$ is calculated by (37) and is regarded as the *best response* $\mathbf{l}_{z,t}^{i*}$ of the player i

$$\mathbf{l}_{z,t}^{i*}=\left(\sum_{\Psi_1}\mathbf{l}^i+\sum_{\Psi_2}\frac{\sigma_1}{\sigma_2}\bullet\mathbf{l}^i\right)\Bigg/\left(N_i'+\sum_{\Psi_2}\frac{\sigma_1}{\sigma_2}\right). \tag{38}$$

The location $\mathbf{l}_{z,t}^i$ of the player i is initialized by the corresponding predicted target's location $\mathbf{l}_{t|t-1}^i$, $i\in\{1,\ldots,N_o^n\}$. Given the initialized $\mathbf{l}_{z,t}^{-i*}$, the *best response* $\mathbf{l}_{z,t}^{i*}$ of the player i can be calculated by (38). $\mathbf{l}_{z,t}^{i*}$ can be iteratively updated until the process reaches an equilibrium. The

$$P_s(p^i,q^i)=\exp\left\{\frac{1}{N_i}\sum_{\Omega_i}\log\left\{N(\mathbf{l}^i;\mathbf{l}_t^i,\Sigma_t^i)\sum_{k=1}^{K}\omega_k^i N(c_{\mathbf{l}^i};\mu_k^i,\Sigma_k^i)\right\}\right\} \tag{32}$$

$$P_s(p^i,q^i|p^j)=\exp\left\{\frac{1}{N_o^n}\left[\sum_{\Psi_1}\log(N(\mathbf{l}_t^i)\bullet\sigma_1)+\sum_{\Psi_2}\frac{\sigma_1}{\sigma_2}\log(N(\mathbf{l}_t^i)\bullet\sigma_1)\right]\right\} \tag{33}$$

equilibrium is obtained when the maximum component of the difference vector Δl satisfies (39). Δl is the difference of the *best response* sets between the consecutive iteration cycles

$$\max(\Delta l) < T_{\text{NE}} \tag{39}$$

where $\Delta l = |\{\mathbf{l}_{z,t}^{1*},\ldots,\mathbf{l}_{z,t}^{N_o^n *}\}_{\text{iteration}_j} - \{\mathbf{l}_{z,t}^{1*},\ldots,\mathbf{l}_{z,t}^{N_o^n *}\}_{\text{iteration}}$ $j-1|(j=1,2,\ldots)\{\mathbf{l}_{z,t}^{1*},\ldots,\mathbf{l}_{z,t}^{N_o^n *}\}_{\text{iteration}_0}$ is the initialized locations' set. T_{NE} is the given threshold. The smaller the T_{NE} is, the more is the iteration time needed and the more precise the results are. In the experiments, we set $T_{\text{NE}}=1$ pixel to achieve a tradeoff between the efficiency and precision. When the iteration terminates at iteration cycle j, the *best response* set $\{\mathbf{l}_{z,t}^{1*},\ldots,\mathbf{l}_{z,t}^{N_o^n *}\}_{\text{iteration}_j}$ is determined as the NE of the game. This NE is regarded as the optimal segmentations of the measurement. The measurements $(\mathbf{z}_t^{1*},\ldots,\mathbf{z}_t^{N_o^n *})$ with the $\{\mathbf{l}_{z,t}^{1*},\ldots,\mathbf{l}_{z,t}^{N_o^n *}\}_{\text{iteration}_j}$ are then incorporated into the filter to update the states of the targets within the occlusion region.

V. Experimental Results and Discussions

The proposed visual tracking system is tested on the publicly available videos. In particular, the contributions of the proposed birth intensity estimation algorithm and the game-theoretical occlusion-handling algorithm are assessed.

The state transition model is a constant velocity model [8] with $\boldsymbol{F}_t = [\boldsymbol{I}_2, T \bullet \boldsymbol{I}_2, \boldsymbol{0}_2; \boldsymbol{0}_2, \boldsymbol{I}_2, \boldsymbol{0}_2; \boldsymbol{0}_2, \boldsymbol{0}_2, \boldsymbol{I}_2]$ and $\boldsymbol{Q}_t = \sigma_v^2[T^4 \bullet \boldsymbol{I}_2/4, T^3 \bullet \boldsymbol{I}_2/2, \boldsymbol{0}_2; T^3 \bullet \boldsymbol{I}_2/2, T^2 \bullet \boldsymbol{I}_2, \boldsymbol{0}_2; \boldsymbol{0}_2, \boldsymbol{0}_2, T^2 \bullet \boldsymbol{I}_2]$, where $\boldsymbol{0}_n$ and $\boldsymbol{I}_n$ are the $n \times n$ zero and identity matrices, respectively. $T = 1$ frame is the interval between two consecutive time steps. $\sigma_v = 3$ is the standard deviation of the state noise. The measurements follow the measurement likelihood with $\boldsymbol{H}_t = [\boldsymbol{I}_2, \boldsymbol{0}_2, \boldsymbol{0}_2; \boldsymbol{0}_2, \boldsymbol{0}_2, \boldsymbol{I}_2]$ and $\boldsymbol{R}_t = \sigma_w^2 \boldsymbol{I}_4$, where $\sigma_w = 2$ is the standard deviation of the measurement noise. We set the values of the parameters used in the GM-PHD filter as follows: detection probability $p_d = 0.99$, survival probability $p_{\text{sv}} = 0.95$, average distributed rate of clutters $\lambda_t = 0.01$, and spatial distribution of clutters $c_t(\mathbf{z}_t) = (\text{image area})^{-1}$.

A. Experimental Results

We evaluated the proposed tracking system on the following video datasets: 700 frames from "PETS 2000,"[1] 630 frames from "BEHAVE,"[2] 200 frames from "ViSOR#1,"[3] 415 frames from "ViSOR#2,"[3] 951 frames from "PETS2006,"[4] 922 frames from "CAVIAR,"[5] and 795 frames from "PETS2009."[6] All videos are captured from a single static camera. The challenging issues involved are listed in Table I.

TABLE I
Challenging Issues Involved in the Videos

Datasets	Challenging issues involved in the videos
PETS2000	Three newborn targets with a few noises No occlusion
BEHAVE	Three newborn targets with a large number of noises No occlusion
ViSOR#1	Six newborn targets with a few noises No occlusion
ViSOR#2	Two newborn targets with no noises Two interacting targets with similar color distributions Partial and total occlusion Targets merge once and split once
PETS2006	Thirteen newborn targets with a few noises Three interacting targets with similar color distributions Partial and total occlusion Targets frequently merge and split
CAVIAR	Three newborn targets with a large number of noises Two interacting targets with similar color distributions Partial occlusion Targets frequently merge and split
PETS2009	A large number of interacting targets Partial and total occlusion Targets frequently merge and split

1) Qualitative Analysis: To demonstrate the performance of the *improved tracking system (ITS)* in handling the listed challenging issues, it is compared with the *standard GM-PHD filter-based tracking system (STS)* whose birth intensity is estimated using the state-of-the-art algorithm proposed by Wang *et al.* [11].

a) Noises Elimination: Fig. 5 shows the detection results and the tracking results of the *STS* and the *ITS* for the first three video sequences (no occlusion involved) listed in Table I. The numbers involved are the target IDs, which are managed using the algorithm proposed by Vo and Ma [29]. Due to the environmental uncertainty, such as illumination changes or waving of trees in the wind, some noises exist in the detection results (shown in the first row in Fig. 5). Tracking with the *STS* is easily interfered by the noises, thereby resulting in many false positives (shown in the second row in Fig. 5). In contrast, the proposed birth intensity estimation algorithm greatly eliminates the noises. Tracking with the *ITS* achieves less false positives compared with the *STS* (shown in the third row in Fig. 5).

b) Occlusion Handling: Figs. 6–9 show the detection results and the tracking results of the *STS* and the *ITS* for the last four video sequences (occlusion involved) listed in Table I.

In Fig. 6 ("ViSOR#2"), two targets with similar color distributions are involved without interferences of noises. Both the *STS* and the *ITS* can track the targets accurately as they enter the scene. However, as targets move close together, they are detected as one merged measurement (shown as $t = 166$ in Fig. 6). Without occlusion handling, the *STS* fails to track target 1 but tracks the merged measurement as target 2 from $t = 166$. As the two targets split, target 1 is retracked as newborn target 3 from $t = 245$. In contrast, the *ITS* incorporates the game-theoretical occlusion-handling algorithm based on the proposed appearance model. Although the targets have similar color distributions, the *ITS* can successfully track them for the entire occlusion period even when under total occlusion (shown as $t = 191$ in Fig. 6).

[1] Available: ftp://ftp.pets.rdg.ac.uk/pub/PETS2000.

[2] Available: http://groups.inf.ed.ac.uk/vision/BEHAVEDATA/.

[3] Available: http://imagelab.ing.unimore.it/visor/video_categories.asp.

[4] Available: http://www.cvg.rdg.ac.uk/PETS2006/data.html.

[5] Available: http://homepages.inf.ed.ac.uk/rbf/CAVIARDATA1/.

[6] Available: http://www.cvg.rdg.ac.uk/PETS2009/a.html.

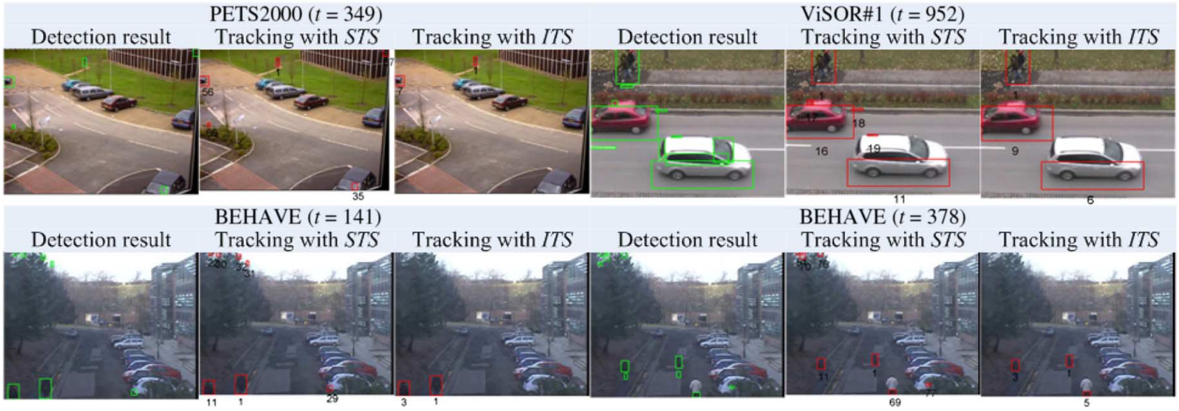

Fig. 5. Tracking results of the datasets "PETS2000," "ViSOR#1," and "BEHAVE."

In Fig. 7 ("PETS2006"), the targets frequently merge and split. The *STS* loses the targets (shown as $t = 508$ and $t = 777$ in Fig. 7) as mutual occlusion occurs (targets merge) and then retracks the merged measurement as a newborn target (shown as $t = 778$ in Fig. 7). As the two targets split, they are also tracked as the newborn targets. In contrast, the *ITS* performs robustly regardless of the merging or splitting of the targets. In particular, the *ITS* can handle three targets with similar color distributions merging together from $t = 777$ to $t = 802$ (shown as $t = 777$, $t = 778$, and $t = 788$ in Fig. 8). However, as the targets enter the scene as a group, the *ITS* tracks the merged measurement as one newborn target (shown as $t = 1122$ in Fig. 7) because no prior information about these individual targets is available. In such a case, the merged measurement cannot be determined as the occlusion region by the proposed occlusion reasoning algorithm. Instead, the merged measurement is tracked as one single target.

In Figs. 8 ("CAVIAR") and 9 ("PETS2009"), the *STS* fails to track the targets as the mutual occlusion occurs, whereas the *ITS* can successfully track the targets in occlusion. Particularly, as several occlusions simultaneously occur in different target groups (shown as $t = 723$ and $t = 728$ in Fig. 9), the *ITS* still can robustly track the targets in each occlusion region.

2) Quantitative Analysis: The CLEAR MOT metrics [32] are used to evaluate the tracking performance. The metrics return a multi-object tracking precision (MOTP) score and a multi-object tracking accuracy (MOTA) score. The MOTA score is composed of the miss rate (MR), the false positive rate (FPR), and the mismatch rate (MMR)

$$\mathrm{MOTP} = \frac{\sum_{i,t}\left[S(\mathrm{gb}_t^i \cap \mathrm{tb}_t^i)/S(\mathrm{gb}_t^i \cup \mathrm{tb}_t^i)\right]}{\sum_t c_t} \tag{40}$$

$$\mathrm{MOTA} = 1 - \frac{\sum_t (m_t + \mathrm{fp}_t + \mathrm{mme}_t)}{\sum_t g_t} \tag{41}$$

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

We compare the *ITS* with the *STS* and other state-of-the-art tracking systems according to the CLEAR MOT metrics.

a) Comparison With the STS: Without robust birth intensity estimation and occlusion-handling algorithms, the *STS* produces either a large FPR, particularly when a large number of noises are involved, or a large MR when mutual occlusion frequently occurs (shown in Table II). Take, e.g., the dataset "BEHAVE" that contains a small number of targets (e.g., two targets at $t = 141$ in Fig. 5) with a relatively large number of noises (e.g., five noises at $t = 141$ in Fig. 5) at some time steps. Consequently, the number of false positives fp_t is larger than the number of ground truth g_t when tracking by the *STS*. According to (41), the FPR $\left(\sum_t \mathrm{fp}_t / \sum_t g_t\right)$ is larger than 1 and ultimately makes the MOTA score negative (shown in Table II). As mutual occlusion occurs, the *STS* may lose the targets or track the merged measurement as one target. This scenario results in a large MR (shown in the last four datasets in Table II). In contrast, the *ITS* can accurately estimate the birth intensity and robustly handle the mutual occlusion problem. The results in Table II show that the *ITS* outperforms the *STS* both in terms of MOTP and MOTA scores.

b) Comparison With the State-of-the-Art Tracking Systems: We also compare the *ITS* with the state-of-the-art systems reported by Joo and Chellappa [33], Torabi and Bilodeau [34], and Zulkifley and Moran [35] for the dataset "PETS2006" (shown in Table III) and by Andriyenko *et al.* [36], Breitenstein *et al.* [37], and Yang *et al.* [38] for the dataset "PETS2009" (shown in Table IV). The results in Table III show that the *ITS* achieves a good MOTP score and a low MOTA score. The results in Table IV show that the results of the *ITS* outperforms the previously published results by Breitenstein *et al.* [37] and Yang *et al.* [38] in terms of both precision and accuracy. Compared with the previously published results by Andriyenko *et al.* [36], the *ITS* achieves a better MOTP score but gets a lower MOTA score. The reason for the low MOTA score is the use of a simple background subtraction method for object detection. This approach tends to generate a large number of noises in variable environment. Although our system can eliminate a large number of noises, some noises may still be tracked as the targets. For example, target 17 at $t = 846$ in Fig. 8 is a false positive. This can be improved by using a highly robust object detection method.

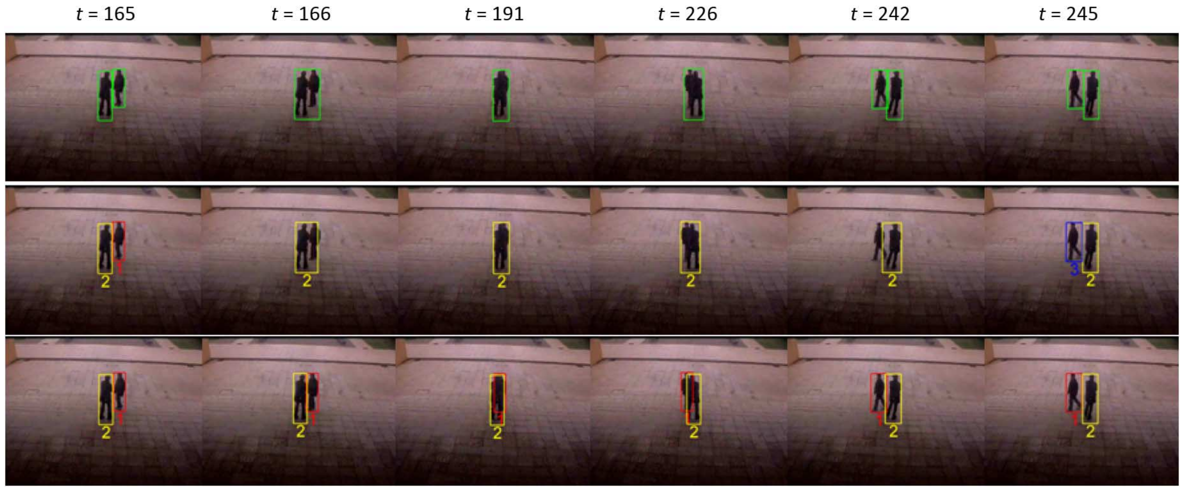

Fig. 6. Tracking results of the dataset “ViSOR#2.” First row: detection results. Second row: tracking results with *STS*. Third row: tracking results with *ITS*.

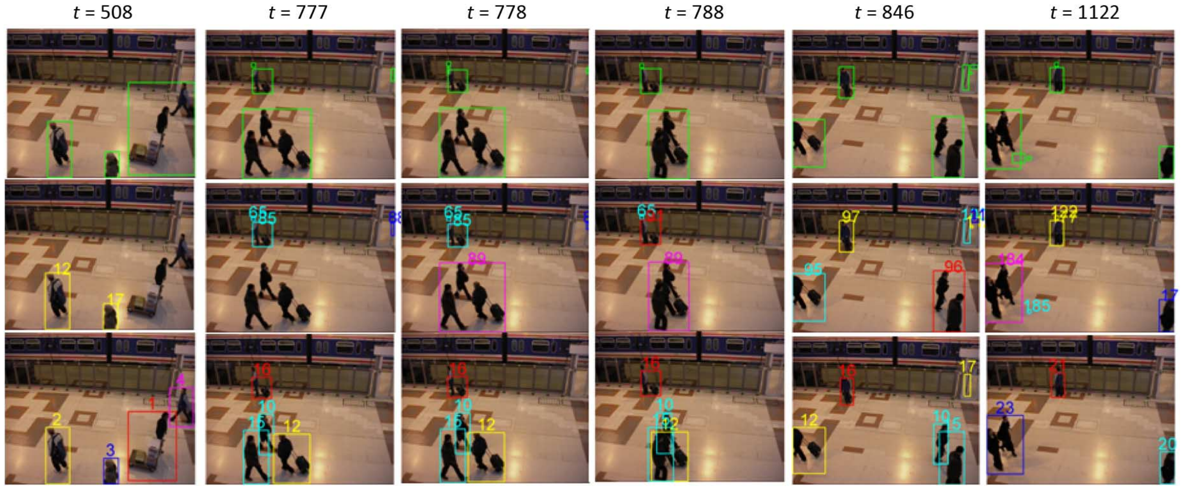

Fig. 7. Tracking results of the dataset “PETS2006.” First row: detection results. Second row: tracking results with *STS*. Third row: tracking results with *ITS*.

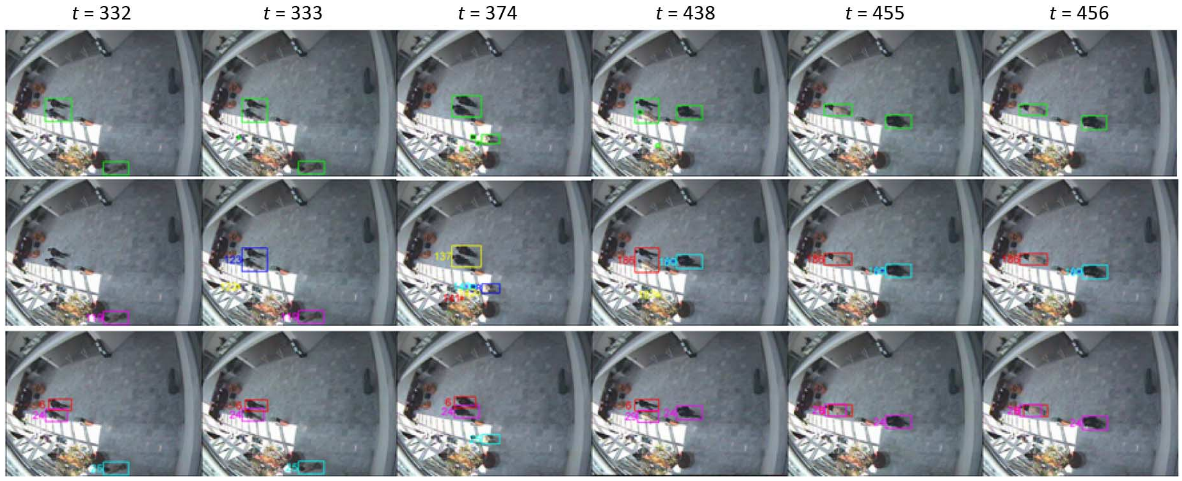

Fig. 8. Tracking results of the dataset “CAVIAR.” First row: detection results. Second row: tracking results with *STS*. Third row: tracking results with *ITS*.

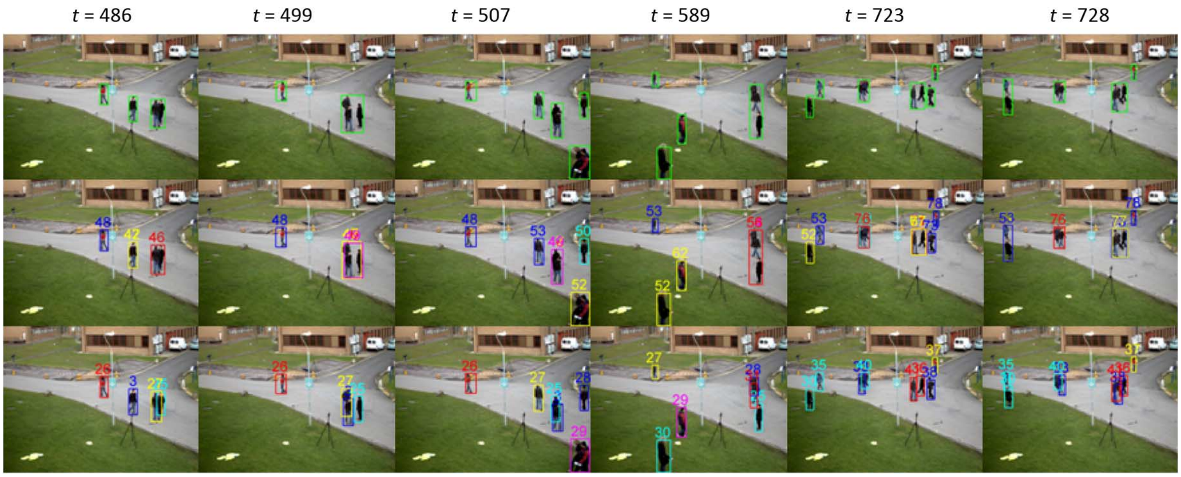

Fig. 9. Tracking results of the dataset "PETS2009." First row: detection results. Second row: tracking results with *STS*. Third row: tracking results with *ITS*.

TABLE II
TRACKING PERFORMANCE COMPARISON BETWEEN THE *ITS* AND THE *STS*

Dataset	System	MOTP (%)	MOTA (%)	FPR (%)	MR (%)	MMR (%)
PETS2000	*ITS*	77.86	99.32	0.20	0.48	0
	STS	76.25	66.67	33.04	0.29	0
BEHAVE	*ITS*	64.14	86.75	5.02	8.23	0
	STS	62.96	–131.43	231.12	0.31	0
ViSOR#1	*ITS*	70.82	95.26	2.93	1.81	0
	STS	69.93	90.52	6.09	3.39	0
ViSOR#2	*ITS*	85.46	99.36	0.13	0.38	0.13
	STS	67.92	89.53	0.13	10.08	0.26
PETS2006	*ITS*	62.92	86.16	6.43	7.12	0.29
	STS	42.86	34.4	49.21	14.65	1.74
CAVIAR	*ITS*	80.64	78.65	19.56	0.96	0.83
	STS	65.78	33.93	52.73	10.13	3.21
PETS2009	*ITS*	58.47	87.21	0.11	11.45	1.23
	STS	49.76	46.17	0.23	19.94	6.12

TABLE III
TRACKING PERFORMANCE COMPARISON BETWEEN THE *ITS* AND THE STATE-OF-THE-ART TRACKING SYSTEMS ON THE DATASET "PETS2006"

System (%)	Our *ITS*	Joo and Chellappa [33]	Torabi and Bilodeau [34]	Zuikifley and Moran [35]
MOTP	62.92	49.8	56.87	58.16
MOTA	86.16	92.21	96.56	98.75

TABLE IV
TRACKING PERFORMANCE COMPARISON BETWEEN THE *ITS* AND THE STATE-OF-THE-ART TRACKING SYSTEMS ON THE DATASET "PETS2009"

System (%)	Our *ITS*	Andriyenko et al. [36]	Breitenstein et al. [37]	Yang et al. [38]
MOTP	58.47	56.4	56.3	53.8
MOTA	87.21	89.3	79.7	75.9

B. Discussions

Although all the aforementioned experiments validate the capacity of our tracking system to handle the challenging issues listed in Table I, other issues need to be discussed further.

1) Tracking Newborn Group Targets: To invoke the proposed occlusion-handling algorithm, the prior information about the targets before merging is necessary. However, as targets enter the scene as a group (shown as target 23 at $t = 1122$ in Fig. 7 and as target 29 at $t = 507$ in Fig. 9), the occlusion-handling algorithm cannot be invoked. In such a case, the targets are only tracked as one group of newborn targets. The effective object detection methods should thus be incorporated to accurately detect the targets as they first appear in the scene.

2) Processing Speed: The proposed system is implemented in MATLAB using a computer with Inter Core 2 Duo 2.20 GHz and 2 GB of memory. Without any code optimization, the average runtimes for the first three datasets (no occlusion involved) listed in Table I are about 2–10 frames per second (fps), whereas those for the last four datasets (occlusion involved) are about 0.4–1.2 fps. More than 95% of the runtimes are consumed in two parts. One part is for determining the NE of the proposed mutual occlusion-handling algorithm because it is a pixel-wise iteration process. The other part is for target ID management. The target ID is managed according to the algorithm proposed by Vo and Ma [29], in which each tracked target is labeled with an independent ID. The result is a large computational cost once a large number of noises are tracked as targets. To remedy the aforementioned drawbacks, employing highly efficient appearance model and ID management method will be helpful and will be explored in our future works.

VI. CONCLUSION

In this paper, an MTVT system that combined the GM-PHD filter with object detection was developed to track multiple 2-D moving targets in a video. Specifically, two key issues involving the GM-PHD filter were investigated and remedied.

Due to the environmental uncertainty, the video data may contain some noises. To eliminate the interferences by noises, an improved measurement dependent birth intensity estimation algorithm was proposed. Unlike in existing measurement dependent birth intensity estimation algorithms, the entropy distribution and coverage rate were incorporated in the proposed algorithm. By incorporating the entropy distribution, the noises within the birth intensity that were irrelevant to the birth measurements were removed. To further eliminate the noises within the birth intensity, the coverage rate between each survival birth intensity component and the corresponding birth measurement was computed and used to update the weights of the component. By doing so, the components could be removed once their weights were less than the given threshold.

As mutual occlusion occurs, the measurements originating from the targets within the occlusion region will be merged into one measurement. The GM-PHD filter may associate this measurement with one of the targets in occlusion region while losing the other targets or it may lose all the targets in occlusion region. To solve this challenging problem, a game-theoretical mutual occlusion-handling algorithm was proposed. To improve the robustness of mutual occlusion handling, an improved spatial color appearance with interferences by other targets within the occlusion region was modeled. Compared with the conventional color histogram-based appearance model, the improved model was more robust even when the targets involved in the occlusion region had similar color distributions. Based on this improved appearance model, an n-person, nonzero-sum, and noncooperative game was constructed. The individual measurements originating from the individual targets within the occlusion region were regarded as the players in the constructed game competing for maximum utilities using certain strategies. The NE of the game was selected as the optimal estimations of the locations of the players. By doing so, the targets in mutual occlusion could be successfully tracked.

The experiments on publicly available video sequences were conducted to evaluate the proposed tracking system. Compared with the *STS* and with the state-of-the-art tracking systems, our system showed great improvements in precision and accuracy.

In the future, we will investigate the algorithms for tracking newborn group targets and for improving tracking efficiency.

References

[1] D. I. Kosmopoulos, A. S. Voulodimos, and A. D. Doulamis, "A system for multicamera task recognition and summarization for structured environments," *IEEE Trans. Ind. Informat.*, vol. 9, no. 1, pp. 161–171, Feb. 2013.

[2] D. Kosmopoulos, N. Doulamis, and A. Voulodimos, "Bayesian filter based behavior recognition in workflows allowing for user feedback," *Comput. Vis. Image Understanding*, vol. 116, no. 3, pp. 422–434, 2012.

[3] C. Tran and M. M. Trivedi, "3-D posture and gesture recognition for interactivity in smart spaces," *IEEE Trans. Ind. Informat.*, vol. 8, no. 1, pp. 178–187, Feb. 2012.

[4] G. Wang, L. Tao, H. Di, X. Ye, and Y. Shi, "A scalable distributed architecture for intelligent vision system," *IEEE Trans. Ind. Informat.*, vol. 8, no. 1, pp. 91–99, Feb. 2012.

[5] T. Zhang, S. Liu, C. Xu, and H. Lu, "Mining semantic context information for intelligent video surveillance of traffic scenes," *IEEE Trans. Ind. Informat.*, vol. 9, no. 1, pp. 149–160, Feb. 2013.

[6] Y. Wang, J. Wu, W. Huang, and A. Kassim, "Gaussian mixture probability hypothesis density for visual people tracking," in *Proc. 10th Int. Conf. Inf. Fusion*, Quebec, Canada, 2007, pp. 1–6.

[7] E. Pollard, A. Plyer, B. Pannetier, F. Champagnat, and G. L. Besnerais, "GM-PHD filters for multi-object tracking in uncalibrated aerial videos," in *Proc. 12th Int. Conf. Inf. Fusion*, Seattle, WA, USA, 2009, pp. 1171–1178.

[8] J. Wu, S. Hu, and Y. Wang, "Probability-hypothesis-density filter for multitarget visual tracking with trajectory recognition," *Opt. Eng.*, vol. 49, no. 12, pp. 12970-11–12970-19, Dec. 2010.

[9] B. Ristic, D. Clark, and B.-N. Vo, "Improved SMC implementation of the PHD filter," in *Proc. 13th Int. Conf. Inf. Fusion*, Edinburgh, U.K., 2010, pp. 1–8.

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

[11] Y. Wang, J. Wu, A. Kassim, and W. Huang, "Data-driven probability hypothesis density filter for visual tracking," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 18, no. 8, pp. 1085–1095, Aug. 2008.

[12] Z. Wu, N. I. Hristov, T. L. Hedrick, T. H. Kunz, and M. Betke, "Tracking a large number of objects from multiple views," in *Proc. IEEE 12th Int. Conf. Comput. Vis.,* 2009, pp. 1546–1553.

[13] S. M. Khan and M. Shah, "Tracking multiple occluding people by localizing on multiple scene planes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 3, pp. 505–519, Mar. 2009.

[14] T. Zhao, R. Nevatia, and B. Wu, "Segmentation and tracking of multiple humans in crowded environments," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 7, pp. 1198–1211, Jul. 2008.

[15] Z. Khan, T. Balch, and F. Dellaert, "MCMC data association and sparse factorization updating for real time multitarget tracking with merged and multiple measurements," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 1960–1972, Dec. 2006.

[16] R. Vezzani, C. Grana, and R. Cucchiara, "Probabilistic people tracking with appearance models and occlusion classification: The AD-HOC system," *Pattern Recognit. Lett.*, vol. 32, pp. 867–877, Apr. 2011.

[17] J. Xing, H. Ai, L. Liu, and S. Lao, "Multiple player tracking in sports video: A dual-mode two-way Bayesian inference approach with progressive observation modeling," *IEEE Trans. Image Process.*, vol. 20, no. 6, pp. 1652–1667, Jun. 2011.

[18] V. Papadourakis and A. Argyros, "Multiple objects tracking in the presence of long-term occlusions," *Comput. Vis. Image Understanding*, vol. 114, no. 7, pp. 835–846, Jul. 2010.

[19] W. Hu, X. Zhou, M. Hu, and S. Manbank, "Occlusion reasoning for tracking multiple people," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 19, no. 1, pp. 114–121, Jan. 2009.

[20] H. Wang, D. Suter, K. Schindler, and C. Shen, "Adaptive object tracking based on an effective appearance filter," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 9, pp. 1661–1667, Sep. 2007.

[21] X. Zhou, Y. F. Li, B. He, T. Bai, and Y. Tang, "Birth intensity online estimation in GM-PHD filter for multi-target visual tracking," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst.*, Vila Moura, Algarve, Portugal, 2012, pp. 3893–3898.

[22] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 379–423, Jul. 1948.

[23] J. Nash, "Two-person cooperative games," *Econometrica*, vol. 21, no. 1, pp. 128–140, Jan. 1953.

[24] M. Yang, T. Yu, and Y. Wu, "Game-theoretical multiple target tracking," in *Proc. IEEE 11th Int. Conf. Comput. Vis.*, Rio de Janeiro, Brazil, 2007, pp. 1–8.

[25] D. Gu, "A game theory approach to target tracking in sensor networks," *IEEE Trans. Syst. Man Cybern. B: Cybern.*, vol. 41, no. 1, pp. 2–13, Feb. 2011.

[26] C. Soto, B. Song, and A. K. Roy-Chowdhury, "Distributed multi-target tracking in a self-configuring camera network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Miami, FL, USA, 2009, pp. 1486–1493.

[27] H.-Y. Shi, W.-L. Wang, N.-M. Kwok, and S.-Y. Chen, "Game-theory for wireless sensor networks: A survey," *Sensors*, vol. 12, no. 7, pp. 9055–9097, Jul. 2012.

[28] R. Mahler, "Multitarget Bayes filtering via first-order multitarget moments," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 39, no. 4, pp. 1152–1178, Oct. 2003.

[29] B.-N. Vo and W. K. Ma, "The Gaussian mixture probability hypothesis density filter," *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4091–4104, Nov. 2006.

[30] X. Zhang, W. Hu, G. Luo, and S. Manbank, "Kernel-Bayesian framework for object tracking," in *Proc. 8th Asian Conf. Comput. Vis.*, Tokyo, Japan, 2007, pp. 821–831.

[31] E. N. Barron, *Game Theory: An Introduction*, New York, NY, USA: Wiley, 2008.

[32] K. Bernardin and R. Stiefelhagen, "Evaluating multiple object tracking performance: The CLEAR MOT Metrics," *EURASIP J. Image Video Process.*, vol. 2008, pp. 1–10, Feb. 2008.

[33] S. W. Joo and R. Chellappa, "A multiple-hypothesis approach for multiobject visual tracking," *IEEE Trans. Image Process.*, vol. 16, no. 11, pp. 2849–2854, Nov. 2007.

[34] A. Torabi and G. A. Bilodeau, "A multiple hypothesis tracking method with fragmentation handling," in *Proc. Can. Conf. Comput. Robot Vis.*, Kelowna, British Columbia, Canada, 2009, pp. 8–15.

[35] M. A. Zulkifley and B. Moran, "Robust hierarchical multiple hypothesis tracker for multiple-object tracking," *Expert Syst. Appl.*, vol. 39, no. 16, pp. 12319–12331, Nov. 2012.

[36] A. Andriyenko, K. Schindler, and S. Roth, "Discrete-continuous optimization for multi-target tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recog.*, Providence, RI, USA, 2012, pp. 1926–1933.

[37] M. D. Breitenstein, F. Reichlin, B. Leibe, E. Koller-Meier, and L. V. Gool, "Online multiperson tracking-by-detection from a single, uncalibrated camera," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 9, pp. 1820–1833, Sep. 2011.

[38] J. Yang, Z. Shi, P. Vela, and J. Teizer, "Probabilistic multiple people tracking through complex situations," in *Proc. IEEE Workshop Perform. Eval. Track. Surv.*, Miami, FL, USA, 2009, pp. 79–86.

Xiaolong Zhou, photograph and biography not available at the time of publication.

Youfu Li (S'91-M'92-SM'01), photograph and biography not available at the time of publication.

Bingwei He, photograph and biography not available at the time of publication.

Tianxiang Bai, photograph and biography not available at the time of publication.