Enhanced GM-PHD Filter Using CNN-Based Weight Penalization for Multi-Target Tracking

Zeyu Fu, Syed Mohsen Naqvi, Jonathon A. Chambers

***Abstract*—In this paper, an enhanced Gaussian mixture-probability hypothesis density filter (GM-PHD) using convolutional neural network (CNN) based weight penalization is proposed to track multiple targets in video. Existing GM-PHD filter based tracking methods are not always able to accurately track the targets when they are in close proximity, especially with noisy detection responses or in a crowded environments. To address this issue, a measurement classification step which combines a confidence score with a gating technique is presented to discard the false measurements and initialise new-born targets. High level human features extracted from a pre-trained CNN are utilized to penalize the ambiguous weights in the weight matrix. In addition, we integrate an improved track management scheme with occlusion handling to form the tracks of confirmed targets and maintain the track continuity. Experimental results on two publicly available benchmark video sequences validate the efficacy of our proposed method in video-based multi-target tracking.**

***Index Terms*—Multi-target tracking, tracking by detection, GM-PHD filter, weight penalization**

I. Introduction

Video-based multi-target tracking has been an emerging technique in the last decade, since it is crucial in many applications such as intelligent video surveillance, behavior analysis, assistive technology and human-computer interactions [1], [2]. Many researchers have been seeking higher-level tracking systems to locate a number of targets, retrieve their trajectories, and recognise their identities from video sequences. However, there still exist many challenging problems caused by more realistic environments such as the presence of noise, occlusions, background clutter, and illumination changes. To address these challenges, traditional approaches have involved explicit association between measurements and targets in multi-target tracking such as multiple hypotheses tracking (MHT) [3] and the joint probabilistic data association filter (JPDAF) [4]. Recently, tracking-by-detection with data association driven by the recent advancements in object detection has become the leading paradigm for multi-target tracking in video [5]–[9]. Nevertheless, these methods are not capable of effectively accounting for target birth and death, and suffer from being computationally expensive in data association.

Different from data association-based tracking approaches, the random finite set (RFS) based PHD filter [10] which originated from radar tracking has been successfully explored in video-based multi-target tracking [1], [11]–[14], since this online state estimation technique has the ability to deal with varying number of targets, reduce missed detections, and mitigate spatial noise. Two implementations of approximating the PHD filter function have been made by a Gaussian mixture as in the GM-PHD filter [15] or the sequential Monte Carlo (SMC) method via a set of weighted random particles known as the SMC-PHD filter [16]. In fact, the GM-PHD filter has been extended to handle nonlinearity in dynamical and measurement models, as well as being able to avoid the computational complexity in terms of using a clustering method in the SMC-PHD filters, as a result, it can be more efficiently applied in real world applications [17]. However, the performance of conventional GM-PHD filter based tracking methods may degrade significantly when tracking closely spaced targets, especially with noisy detections or in a crowded environment. To address this issue, a collaborative penalized GM-PHD filter was proposed by Wang et al. [18], which employed track identities to perform weight refinement by collaboratively penalizing the weights of targets with the same identities. But this method may no longer be applicable to achieve effective performance with noisy observations in a real world tracking scenario. Zhou et al. [13] explored the fusion of multiple features to penalize the ambiguous weights, so as to improve the tracking accuracy. However, both trackers would track the merged measurements within the occlusion area as one single target without an occlusion handling method.

In this paper, we propose an enhanced GM-PHD filter using CNN-based weight penalization for video-based multi-target tracking, as outlined by the block diagram in Fig. 1. The proposed system exploits the confidence score of detection results and incorporates our previous work [19] namely a gating technique to apply measurement classification. Recently, deep convolutional neural networks (CNNs) have outperformed heuristic, hand-crafted features in terms of appearance modelling [20]. Therefore, a pre-trained CNN model [21] is integrated to improve the weight penalization, following by improved track management with occlusion handling to achieve better tracking performance.

II. The Proposed Tracking System

For a video-based multi-target tracking system, the state of a target m is represented by a six dimensional vector $\mathbf{x}_k^m = [p_{x,k}^m, p_{y,k}^m, v_{x,k}^m, v_{y,k}^m, w_k^m, h_k^m]^T$ and contains the actual 2D image location, velocity and the bounding box size of the target respectively, where $m = 1, \ldots, M_k$, and M_k

This work was supported by the Engineering and Physical Sciences Research Council (EPSRC) Grant number EP/K014307 and the MOD University Defence Research Collaboration in Signal Processing.

Z. Fu, S. M. Naqvi and J. A. Chambers are with the Intelligent Sensing and Communications Research Group, School of Engineering, Newcastle University, United Kingdom, E-mails:{z.fu2, mohsen.naqvi, jonathon.chambers}@newcastle.ac.uk.

978-1-5386-1663-5/17/$31.00 ©2017 IEEE

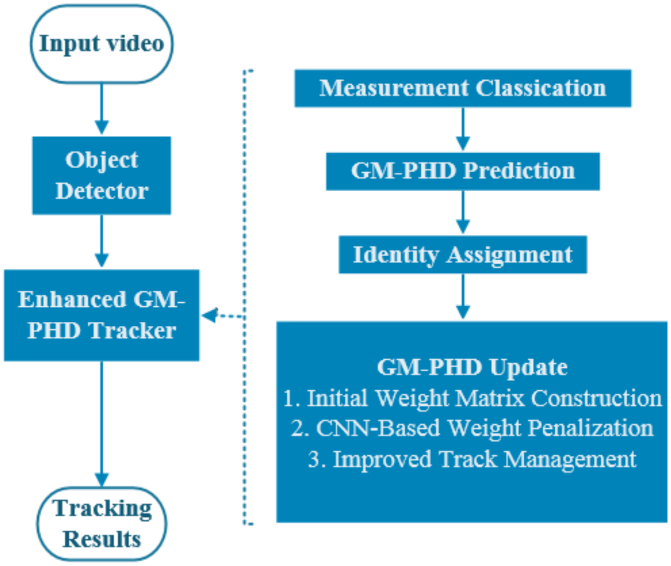

Fig. 1. Block diagram of the proposed tracking system for video-based multi-target tracking

denotes the number of tracked targets at time k. The observed measurement vector $\mathbf{z}_k^n = [\bar{p}_{x,k}^n, \bar{p}_{y,k}^n, \bar{w}_k^n, \bar{h}_k^n]^T$, typically contains the n-th target location and size information, where $n = 1, \ldots, N_k$, and N_k is the number of measurements at time k . Based upon the random finite set (RFS) framework, a multiple target state and a multiple target measurement at time k can be represented by two finite sets: $\mathbf{X}_k = \{\mathbf{x}_k^1, ..., \mathbf{x}_k^{M_k}\}$ and $\mathbf{Z}_k = \{\mathbf{z}_k^1, ..., \mathbf{z}_k^{N_k}\}$.

A. The Gaussian Mixture PHD Filter

The PHD filter proposed by Mahler [10] is a natural extension of the single-target Bayesian framework to multi-targets; representing the multi-target states and multi-target measurements, as well as recursively propagating the first-order moment of the multi-target posterior $p_k(\mathbf{X}_k|\mathbf{Z}_{1:k})$, referred to as the intensity function $\nu_k(\mathbf{x}|\mathbf{Z}_{1:k})$ abbreviated by $\nu_k(\mathbf{x})$. The GM-PHD filter proposed by Vo and Ma [15] introduces a closed-form solution to the PHD recursion. The posterior PHD intensity function can be represented by a sum of weighted Gaussian components that are propagated analytically in time [17]. Given a posterior intensity ν_{k-1} in a Gaussian mixture form at time $k-1$, then

$$\nu_{k-1}(\mathbf{x}) = \sum_{j=1}^{J_{k-1}} w_{k-1}^j \mathcal{N}(\mathbf{x}; \mathbf{m}_{k-1}^j, \mathbf{P}_{k-1}^j) \tag{1}$$

where J_{k-1} denotes the number of Gaussian components at time $k-1$, w_{k-1}^j is the corresponding weight of the j-th Gaussian component, and $\mathcal{N}(\cdot; \mathbf{m}, \mathbf{P})$ illustrates Gaussian components with mean $\mathbf{m}$ and covariance $\mathbf{P}$. Since the PHD filter requires an additional algorithm to provide target identity information, we employ the method in [17] to assign a unique label I_{k-1}^j as a hidden identity to individual Gaussian components to form the identity set $\mathcal{I}_{k-1} = \{I_{k-1}^1, ..., I_{k-1}^{J_{k-1}}\}$, allowing these labels to propagate through time without affecting the GM-PHD recursion. The GM-PHD prediction can also be represented by a Gaussian mixture at time k [15],

$$\nu_{k|k-1}(\mathbf{x}) = \nu_{k|k-1}^s(\mathbf{x}) + \gamma_k(\mathbf{x}) \tag{2}$$

$$\gamma_k(\mathbf{x}) = \sum_{j=1}^{J_{\gamma,k}} w_{\gamma,k}^j \mathcal{N}(\mathbf{x}; \mathbf{m}_{\gamma,k}^j, \mathbf{P}_{\gamma,k}^j) \tag{3}$$

$$\nu_{k|k-1}^s(\mathbf{x}) = e_{k|k-1} \sum_{j=1}^{J_{k-1}} w_{k-1}^j \mathcal{N}(\mathbf{x}; \mathbf{F}\mathbf{m}_{k-1}^j, \mathbf{Q} + \mathbf{F}\mathbf{P}_{k-1}^j \mathbf{F}^T) \tag{4}$$

where $\nu_{k|k-1}^s(\mathbf{x})$ denotes the predicted intensity of survival targets and $\gamma_k(\mathbf{x})$ is the predicted intensity of new-born targets with new identities $\mathcal{I}_{\gamma,k} = \{I_{\gamma,k}^1, ..., I_{\gamma,k}^{J_{\gamma,k}}\}$, $\mathbf{F}$ is the state transition matrix and $\mathbf{Q}$ is the process noise covariance. The spawned targets are treated as new-born targets in this paper. The predicted intensity of the GM-PHD filter can be modelled as,

$$\nu_{k|k-1}(\mathbf{x}) = \sum_{j=1}^{J_{k|k-1}} w_{k|k-1}^j \mathcal{N}(\mathbf{x}; \mathbf{m}_{k|k-1}^j, \mathbf{P}_{k|k-1}^j) \tag{5}$$

and meanwhile its identity set is given as, $\mathcal{I}_{k|k-1} = \mathcal{I}_{k-1} \cup \mathcal{I}_{\gamma,k}$. Once the new set of observations is available, the GM-PHD update at time k can be given as [15],

$$\nu_k(\mathbf{x}) = p_M \nu_{k|k-1}(\mathbf{x}) + \sum_{\mathbf{z} \in \mathbf{Z}_k} \sum_{j=1}^{J_{k|k-1}} w_k^j(\mathbf{z}) \mathcal{N}(\mathbf{x}; \mathbf{m}_{k|k}^j(\mathbf{z}), \mathbf{P}_{k|k}^j) \tag{6}$$

where

$$w_k^j(\mathbf{z}) = \frac{(1-p_M) w_{k|k-1}^j q_k^j(\mathbf{z})}{\kappa_k(\mathbf{z}) + (1-p_M)\sum_{i=1}^{J_{k|k-1}} w_{k|k-1}^i q_k^i(\mathbf{z})} \tag{7}$$

$$q_k^j(\mathbf{z}) = \mathcal{N}(\mathbf{z}; \mathbf{H}\mathbf{m}_{k|k-1}^j, \mathbf{R} + \mathbf{H}\mathbf{P}_{k|k-1}^j \mathbf{H}^T) \tag{8}$$

where $\mathbf{H}$ is the observation matrix, $\mathbf{R}$ is the observation noise covariance, p_M denotes the missing detection probability, and κ_k denotes the clutter density. Each predicted Gaussian component gives rise to $(1 + |\mathbf{Z}_k|)$ updated components assigned with the same identity label, i.e., $I_k^j = I_{k|k-1}^j$. More details of the GM-PHD filter framework are available in [15] and [17].

B. Measurement Classification

The observable measurements can be obtained by an object detector at each time step k. Due to the imperfections in the object detector, there is much potential uncertainty in the original detection results, which increases the inefficiency of the PHD update and birth prediction. In order to build a robust measurement model, these noisy measurements must be classified as three subsets: survival measurement set, birth measurement set and background clutter. Firstly, we use the detection confidence score $c_k \in [0, 1]$ associated with each detection to categorize the spurious measurement set $\mathbf{\Gamma}_k = \{\mathbf{z}_{k,f} : c_k < c_{th}\}$ that will be discarded, where c_{th} is the confidence threshold, and its complementary set $\mathbf{Z}_{k,r} = \mathbf{Z}_k \setminus \mathbf{\Gamma}_k$ which will be retained as a real measurement set. This score can be either provided from the detector or computed via the intersection-over-union (IOU) [22]. As the number of new-born targets is unknown, and any initialization

$$
\begin{array}{c c}
 & \begin{array}{cccc} n=1 & 2 & \cdots & N_k \end{array} \\
\begin{array}{c} j=1 \\ 2 \\ \vdots \\ J_{k|k-1} \end{array} &
\left[\begin{array}{cccc}
w_k^{(1,1)} & w_k^{(1,2)} & \cdots & w_k^{(1,N_k)} \\
w_k^{(2,1)} & w_k^{(2,2)} & \cdots & w_k^{(2,N_k)} \\
\vdots & \vdots & \vdots & \vdots \\
w_k^{(J_{k|k-1},1)} & w_k^{(J_{k|k-1},2)} & \cdots & w_k^{(J_{k|k-1},N_k)}
\end{array}\right]
\end{array}
$$

Fig. 2. A symbolic representation of updated weights [13]

or prior information is unavailable for predicting the new-born targets, we adopt the approach in [19], where an adaptive gating method is used to further extract the birth measurement set $\mathbf{Z}_{k,b}$ from the real measurement set $\mathbf{Z}_{k,r} = \mathbf{Z}_{k,b} \cup \mathbf{Z}_{k,s}$, where $\mathbf{Z}_{k,s}$ is the survival measurement set. Therefore, each measurement $\mathbf{z}_{k,b}$ in $\mathbf{Z}_{k,b}$ will be initialised as a new target trajectory with a new identity, and it will be eventually infused with the birth prediction in (3).

C. CNN-Based Weight Penalization

Ideally, multi-target tracking generally follows a one-to-one matching scheme where one target can be only associated with one unique measurement [13]. While this correspondence in the existing GM-PHD based methods can be lost in the closely spaced target tracking scenario. Suppose there are two targets moving near each other, one of which is matched with multiple effective measurements while the other suffers from this special type of miss-detection [18]. In this case, the method in [13] exploited two possible failure cases, including multiple targets moving with the same identities or with switched identities since some predicted targets are possible to be associated with the measurements not originally owned by them. However, there is another scenario which must be addressed, which is that multiple targets will be tracked as a single target, if they are associated with only one measurement (merged measurements) within the occlusion region.

Therefore, we propose to incorporate a pre-trained CNN [21] for appearance modelling to enhance the robustness of weight penalization. Firstly, a weight matrix $\mathbf{W}_k \in \mathbb{R}^{J_{k|k-1} \times N_k}$ as shown in Fig. 2 can be generated by the updated weights from (7), where the j-th row denotes the weights updated by all valid measurements obtained from Section II-B, and $J_{k|k-1}$ and N_k are the number of predicted targets and measurements respectively. The high-level features achieved by a pre-trained CNN from the target region are used to penalize the updated weights in the weight matrix. This CNN model is built with two convolutional layers followed by a max pooling layer and six residual layers, see [21] for details. We employ the Bhattacharyya distance to calculate the following similarity score in terms of feature space between the j-th predicted target and n-th measurement at time k,

$$\theta_k(j,n) = \frac{1}{\sqrt{2\pi\sigma_\theta^2}} \exp\left(- \frac{\{S_k(j,n)\}^2}{2\sigma_\theta^2} \right) \tag{9}$$

where

$$S_k(j,n) = \sqrt{1 - (\mathbf{f}_k^j)^T \mathbf{d}_k^n} \tag{10}$$

and $\mathbf{f}_k^j$ and $\mathbf{d}_k^n$ are the feature vectors of the j-th predicted target and the n-th measurement respectively, and σ_θ^2 denotes the variance of the similarity score. Therefore, all of the updated weights in the weight matrix can be refined as,

$$w_k^{(j,n)} = w_k^{(j,n)} \times \theta_k(j,n). \tag{11}$$

D. Improved Track Management

Based on the penalization step, we integrate an improved track management scheme with occlusion handling to enable the tracker to correctly extract the confirmed tracks, and discard false tracks that are least reliable. Firstly, we adapt the method in [17] to select targets with the maximum weights as a collection of possible tracks. The index of maximum weight can be found from the j-th row of the penalized weight matrix, where $j = 1, ..., J_{k|k-1}$,

$$\tilde{n} = \underset{n=1:N_k}{\arg\max}(w_k^{(j,n)}). \tag{12}$$

Algorithm 1: Improved Track Management ($k > 1$)

```
Input  : Penalized weight Matrix W_k ∈ ℝ^{J_{k|k-1} × N_k}
Output: Tracking Results X_k.
Initialization: W_con = ∅ and W_ten = ∅
for j = 1 : J_{k|k-1} do
    Compute the index of maximum weight ñ with (12);
    if w_k^{(j,ñ)} ≥ w_th then
        Compute the index set of confirmed tracks:
          W_con = W_con ∪ {j, ñ};
        Merged Target Segmentation:
        Search for the ambiguous weights in W_k with
          the same value of ñ, and select the targets with
          smaller weights among them as covered targets,
          which will remain unchanged during the update
          step.
    end
    else
        Compute the index set of tentative tracks:
          W_ten = W_ten ∪ {j, ñ}
        Delete the tracks using W_ten after missing T_miss
          frames;
        Append tentative tracks which are not deleted to
          the confirmed tracks.
    end
    Obtain X_k from the confirmed tracks.
end
```

In this way, targets can be confirmed with $w_k^{(j,\tilde{n})} \geq w_{th}$ and labelled with the same identity as that in prediction, where w_{th} denotes the threshold of a target confirmation. In contrast, the rest of the targets which fail to reach w_{th} are tentatively eliminated after a certain value of T_{miss} frames. Nevertheless, when the occlusion occurs to confirmed targets, their maximum weights are likely to share the same column

(a)

(b)

Fig. 3. Qualitative tracking results of our proposed tracking system on (a) PETS2009-S2L1, (b) TUD-Stadtmitte datasets.

index $\tilde{n}$ in $\mathbf{W}_k$, resulting in targets with merged measurements being tracked as a single target within the occlusion region. To remedy this, we set up a detailed algorithm as summarized in Algorithm 1 in order to segment the merged targets and better manage the target tracks.

III. Experiments

In this section, we validate the proposed tracking system on two commonly used and challenging benchmark datasets: PETS2009-View001-S2L1 (PETS2009) [23], TUD-Stadtmitte [24]. For both datasets, we adopt the annotated ground truths and detections originally provided by the website.[1] The following parameters are utilized to implement the tracker, including the missed detection probability $p_M = 0.05$, the survival probability $e = 0.99$, the clutter intensity $\kappa = 0.0001$, the variance for the similarity score σ_θ^2 is empirically set to 25, and the thresholds of confidence score and target confirmation used are $c_{th} = 0.3$ and $w_{th} = 0.5$ respectively.

We employ the standard CLEAR MOT [25] which is currently the most widely accepted evaluating tool in multi-target tracking to examine the performance of our proposed system. This performance measure mainly entails two metrics, Multiple Object Tracking Precision (MOTP) as a precision score, is designed to measure the average position errors in 2D image plane between estimated tracking results and ground truth in percentage, and Multiple Object Tracking Accuracy (MOTA) as an accuracy score, is comprised of false negatives ratio (FNR), false positives ratio (FPR) and the number of identity switches (IDS). Evaluation measures with (↑) indicate that higher is better, and with (↓) denote lower is better.

[1] http://www.milanton.de/

TABLE I
Quantitative comparison between proposed method and different approaches on PETS2009 dataset. The best results are shown in bold, the second best are underlined

Method	MOTP (↑)	MOTA (↑)	IDS (↓)	FPR (↓)	FNR (↓)
Breitenstein [5]	56.0%	79.7%	-	-	-
GAC [6]	58.3%	**81.4%**	**19**	-	-
Gomez [26]	**75.0%**	51.1%	27	**3.7%**	45.2%
Yoon [27]	57.4%	66.6%	34	15.1%	18.0%
GSDL [19]	61.5%	80.3%	33	6.2%	13.3%
Proposed	68.7%	81.0%	46	8.2%	**9.9%**

TABLE II
Quantitative comparison between proposed method and different approaches on TUD-Stadtmitte dataset.

Method	MOTP (↑)	MOTA (↑)	ID (↓)	FPR (↓)	FNR (↓)
Andriyenko [7]	**65.8%**	60.5%	**7**	-	-
DT-MTT [8]	61.6%	56.2%	15	-	-
Riahi [9]	57.2%	**67.0%**	22	6.0%	26.0%
GSDL [19]	61.7%	62.0%	9	7.7%	30.1%
Proposed	62.0%	65.7%	22	**3.5%**	28.8%

Tables I and II show the quantitative comparisons with other state-of-the-art tracking methods. Likewise, Fig. 3 depicts some example qualitative tracking results produced by the proposed method on both datasets. For the PETS2009 dataset, the proposed method delivers a better performance by ranking the second best in MOTP and MOTA among all methods listed here. In addition, our method achieves the best FNR when compared with the available trackers in Gomez [26], GSDL [19] and Yoon [27]. The reason for the improved performance is because the measurement classification step gives the benefits of removing false detections, and also the weight penalization can better deal with the occlusions and missed detections. As can be seen from the first row of Fig. 3, the proposed method performs well in terms of closely moving targets, particularly in resolving the ambiguities, as it can maintain the target identity and avoid the merged targets during occlusions.

For the TUD-Stadtmitte dataset, our tracker also reports the second best performance regarding the precision score MOTP. In addition, Table II illustrates that our tracker generates fewer missed detections and the fewest false positives, which yields the second highest MOTA score compared to the reported results in Riahi [9] and DT-MTT [8], GSDL [19] and Andriyenko [7]. We can observe along the second row in Fig. 3, some pedestrians with similar appearances that are partially or even almost fully occluded are successfully tracked through our tracker. However, the number of ID switches for both datasets is relatively higher than other methods. This can be attributed by the fact that when a target is missed or out of scene and coming back later, it will be labelled with a new identity. This is the re-identification problem, and it will be further explored in future work to improve the robustness. The tracker also achieves the runtime performance of approximately 20Hz.

IV. Conclusions and Future Work

This paper presents an enhanced GM-PHD filter using CNN-based weight penalization for multiple target tracking in video. By applying the measurement classification, the amount of background clutter and false alarms has been effectively reduced. In order to address the issue of ambiguous targets, we exploited the deep learning method to extract human features, which are used to penalize the weights in the weight matrix. Moreover, an improved track management with occlusion handling has been introduced to correctly estimate target states and eliminate false tracks. The obtained results demonstrate the proposed method achieves competitive tracking performance compared with other approaches. Future work will explore the recurrent neural networks (RNNs) and integrate them to build a robust interaction model, as well as improving the tracker to deal with the re-identification problem.

References

[1] P. Feng, W. Wang, S. Dlay, S. M. Naqvi, and J. Chambers, "Social Force Model-Based MCMC-OCSVM Particle PHD Filter for Multiple Human Tracking," *IEEE Transactions on Multimedia*, vol. 19, no. 4, pp. 725–739, 2017.

[2] A. Ur-Rehman, S. M. Naqvi, L. Mihaylova, and J. A. Chambers, "Multi-Target Tracking and Occlusion Handling with Learned Variational Bayesian Clusters and a Social Force Model," *IEEE Transactions on Signal Processing*, vol. 64, no. 5, pp. 1320–1335, 2016.

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

[4] Y. Bar-Shalom, P. Willett, and X. Tian, *Tracking and Data Fusion: A Handbook of Algorithms*. YBS Publishing, 2011.

[5] 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 Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 9, pp. 1820–1833, 2011.

[6] H. B. Shitrit, J. Berclaz, F. Fleuret, and P. Fua, "Tracking multiple people under global appearance constraints," in *International Conference on Computer Vision (ICCV)*, 2011, pp. 137–144.

[7] A. Andriyenko and K. Schindler, "Multi-target tracking by continuous energy minimization," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011, pp. 1265–1272.

[8] A. Milan, K. Schindler, and S. Roth, "Detection- and trajectory-level exclusion in multiple object tracking," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2013, pp. 3682–3689.

[9] D. Riahi and G.-A. Bilodeau, "Online multi-object tracking by detection based on generative appearance models," *Computer Vision and Image Understanding*, vol. 152, pp. 88–102, 2016.

[10] R. P. S. Mahler, "Multitarget Bayes Filtering via First-Order Multitarget Moments," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 39, no. 4, pp. 1152–1178, 2003.

[11] E. Maggio and A. Cavallaro, *Video Tracking-Theory and Practice*. John Wiley and Sons, 2011.

[12] Z. Fu, P. Feng, S. M. Naqvi, and J. A. Chambers, "Robust Particle PHD Filter with Sparse Representation for Multi-Target Tracking," in *IEEE International Conference on Digital Signal Processing (DSP)*, 2016, pp. 281–285.

[13] X. Zhou, H. Yu, H. Liu, and Y. Li, "Tracking Multiple Video Targets with an Improved GM-PHD Tracker," *Sensors*, vol. 15, no. 12, pp. 30 240–30 260, 2015.

[14] Z. Fu, P. Feng, S. M. Naqvi, and J. A. Chambers, "Particle PHD Filter based Multi-Target Tracking using Discriminative Group-Structured Dictionary learning," in *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2017, pp. 4376–4380.

[15] B.-N. Vo and W. K. Ma, "The Gaussian Mixture Probability Hypothesis Density Filter," *IEEE Transactions on Signal Processing*, vol. 54, no. 11, pp. 4091–4104, 2006.

[16] B.-N. Vo, S. Singh, and D. Arnaud, "Sequential Monte Carlo methods for multitarget filtering with random finite sets," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 51, no. 4, pp. 1224–1245, 2005.

[17] K. Panta, D. E. Clark, and B. N. Vo, "Data Association and Track Management for the Gaussian Mixture Probability Hypothesis Density Filter," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 45, no. 3, pp. 1003–1016, 2009.

[18] Y. Wang, H. Meng, Y. Liu, and X. Wang, "Collaborative Penalized Gaussian Mixture PHD Tracker for Close Target Tracking," *Signal Processing*, vol. 102, pp. 1 – 15, 2014.

[19] Z. Fu, P. Feng, F. Angelini, J. Chambers, and S. M. Naqvi, "Enhanced PHD Filter with Online Group-Structured Dictionary Learning for Multiple Human Tracking," *submitted to IEEE Transactions on Multimedia*, pp. 1–13, 2017.

[20] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems (ANIPS)*, 2012, pp. 1097–1105.

[21] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," *arXiv:1703.07402 [cs.CV]*, pp. 1–5, 2017.

[22] R. Sanchez-Matilla, F. Poiesi, and A. Cavallaro, "Online multi-target tracking with strong and weak detections," in *European Conference on Computer Vision Workshops (ECCVW)*, 2016, pp. 84–99.

[23] I. Goldberg and M. J. Atallah. (2009) Privacy enhancing technologies. [Online]. Available: http://ftp.pets.rdg.ac.uk/pub/PETS2009/Crowd_PETS09_dataset/a_data/a.html

[24] M. Andriluka, S. Roth, and B. Schiele, "People-tracking-by-detection and people-detection-by-tracking," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1–8.

[25] K. Bernardin and R. Stiefelhagen, "Evaluating Multiple Object Tracking Performance: The CLEAR MOT Metrics," *EURASIP Journal on Image and Video Processing*, vol. 2008, pp. 1–10, 2008.

[26] D. Gerónimo Gomez, F. Lerasle, and A. M. López Peña, "State-driven particle filter for multi-person tracking," in *Advanced Concepts for Intelligent Vision Systems: 14th International Conference, ACIVS(2012).*, 2012, pp. 467–478.

[27] J. H. Yoon, M. H. Yang, J. Lim, and K. J. Yoon, "Bayesian multi-object tracking using motion context from multiple objects," in *IEEE Winter Conference on Applications of Computer Vision*, 2015, pp. 33–40.