GM-PHD FILTER BASED ONLINE MULTIPLE HUMAN TRACKING USING DEEP DISCRIMINATIVE CORRELATION MATCHING

Zeyu Fu, Federico Angelini, Syed Mohsen Naqvi, Jonathon A. Chambers

Intelligent Sensing and Communications Research Group, Newcastle University, UK
Emails: {z.fu2, f.angelini2, mohsen.naqvi, jonathon.chambers}@newcastle.ac.uk

ABSTRACT

In this paper, we propose deep discriminative correlation matching within the Gaussian Mixture Probability Hypothesis Density (GM-PHD) filter for online multiple human tracking. In this matching scheme, we mainly exploit the Convolutional Neural Network (CNN) based Discriminative Correlation Filter (DCF) as a target-specific classifier to discriminate the desired target from background and remaining targets. DCFs are learned through the extracted features obtained from the outputs of the last convolutional layers which are capable to encode the target appearances with better discriminativity and robustness to appearance changes. Moreover, we present a hybrid likelihood function that fuses the spatio-temporal relation and correlation matching score to collaboratively enhance the PHD association step. Experimental results on the MOT17 Challenge benchmark [1] confirm the improved performance of our proposed method as compared with other state-of-the-art techniques.

Index Terms— Multiple human tracking, GM-PHD filter, discriminative correlation filter, CNN

1. INTRODUCTION

Tracking multiple human targets in video has been a fundamental and crucial task in many applications such as intelligent video surveillance, autonomous driving and human behavior analysis [2–4]. The main objective of multiple human tracking is to locate a number of human targets, retrieve their trajectories, and recognise their identities from a stream of noisy images. This task becomes more challenging especially in complex scene conditions with background clutter, long-term occlusions, and illumination changes.

Recently, tracking-by-detection with data association driven by the advancements in object detection has become a commonly-used framework for video-based multiple human tracking [5–8]. These methods can be categorized into online and off-line tracking modes. Off-line trackers [5, 7, 8] utilize both past and future detection responses to address a global optimization problem, but such non-causal systems are difficult to be applied in real-time applications. Alternatively, online trackers [6, 9–11] only rely on the detections given up to the present time, which is more suitable for real-time processing. The PHD filter [12] as an effective online multi-target Bayesian filtering method has been extensively applied in video-based multiple human tracking [4, 9, 10, 13, 14]. This technique has the advantage of dealing with varying number of targets, reducing missed detections, and mitigating spatial noise.

Discriminative correlation filters have recently been shown to achieve high computational efficiency and robustness in visual tracking applications [15, 16], particularly in adopting the deep features extracted from a pre-trained network. These deep features have outperformed heuristic, hand-crafted features in learning a strong discriminative appearance model on a wide range of computer vision tasks [17]. To track multiple target in video, Park et al. [18] developed a two-step data association combining the confidence-based relative motion network with a correlation filter, which is generally considered as a verifying step after the motion network. Kutschbach et al. [9] proposed to utilize the kernelized correlation filter to perform an extended update step for the PHD filter, the main goal of which is to adopt the label tree [19] as a reference for the correlation filter to correct the state prediction. Nevertheless, both approaches may be easily prone to false detections when the correlation filter is performed with unreliable references or labels. In this work, we cast the appearance modelling as a person re-identification problem within the GM-PHD filtering framework. This is accomplished by exploiting the CNN-based discriminative correlation filter as an independent target-specific classifier to discriminate the corresponding target from background and remaining targets in the next frame. Furthermore, we present a hybrid likelihood function that combines the spatio-temporal relation with the correlation matching response to compete for performing the target association. The overall block diagram of the proposed method is shown in Fig.1.

2. THE PROPOSED TRACKING ALGORITHM

For a video-based multiple human tracking system, the state of a target m is represented by a six dimensional vector $\mathbf{x}_k^m = [x_k^m, y_k^m, v_{x,k}^m, v_{y,k}^m, w_k^m, h_k^m]^T$ and contains the actual 2D image location, velocity and the bounding box size of the tar-

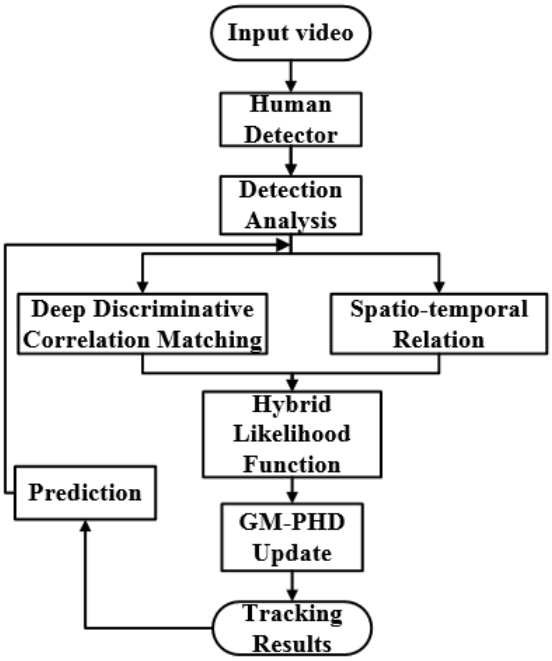


Fig. 1: Overview of proposed method for multiple human tracking

get respectively, where $m = 1, \ldots, M_k$, and M_k denotes the number of targets at time k.

2.1. The Gaussian Mixture PHD Filter

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 $\mathbf{Z}_k = \{\mathbf{z}_k^1, ..., \mathbf{z}_k^{N_k}\}$, where $\mathbf{z}_k^n = [\bar{x}_k^n, \bar{y}_k^n, \bar{w}_k^n, \bar{h}_k^n]^T$, $n = 1, \ldots, N_k$, and N_k is the number of measurements at time k. The GM-PHD filter proposed by Vo and Ma [20] 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. 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 $\mathcal{N}(\cdot; \mathbf{m}, \mathbf{P})$ represents Gaussian components with mean $\mathbf{m}$ and covariance $\mathbf{P}$. The GM-PHD prediction can also be represented by a Gaussian mixture at time k [20],

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

where $\nu_{k|k-1}^s(\mathbf{x}) = e_{k|k-1} \sum_{j=1}^{J_{k-1}} w_{k-1}^j \mathcal{N}(\mathbf{x}; \mathbf{F}\mathbf{m}_{k-1}^j, \mathbf{Q} + \mathbf{F}\mathbf{P}_{k-1}^j\mathbf{F}^T)$ denotes the predicted intensity of survival targets and $\gamma_k(\mathbf{x}) = \sum_{j=1}^{J_{\gamma,k}} w_{\gamma,k}^j \mathcal{N}(\mathbf{x}; \mathbf{m}_{\gamma,k}^j, \mathbf{P}_{\gamma,k}^j)$ is the predicted intensity of new-born targets, $\mathbf{F}$ is the state transition matrix and $\mathbf{Q}$ is the process noise covariance. The spawned targets are treated as new-born targets in this work. The predicted intensity of the GM-PHD filter can be modelled as,

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

Once the new set of observations is available, the GM-PHD update at time k can be given as [20],

$$\nu_k(\mathbf{x}) = p_M \nu_{k|k-1}(\mathbf{x}) + \sum_{\mathbf{z} \in \mathbf{Z}_k} \sum_{j=1}^{J_{k|k-1}} w_k^j(\mathbf{z}) \mathcal{N}(\mathbf{x}; \mathbf{m}_{k|k}^j(\mathbf{z}), \mathbf{P}_{k|k}^j) \tag{4}$$

where

$$w_k^j(\mathbf{z}) = \frac{(1 - p_M) w_{k|k-1}^j q_k^j(\mathbf{z})}{\kappa_k(\mathbf{z}) + (1 - p_M) \sum_{i=1}^{J_{k|k-1}} w_{k|k-1}^i q_k^i(\mathbf{z})} \tag{5}$$

$$q_k^j(\mathbf{z}) = \mathcal{N}(\mathbf{z}; \mathbf{H}\mathbf{m}_{k|k-1}^j, \mathbf{R} + \mathbf{H}\mathbf{P}_{k|k-1}^j\mathbf{H}^T) \tag{6}$$

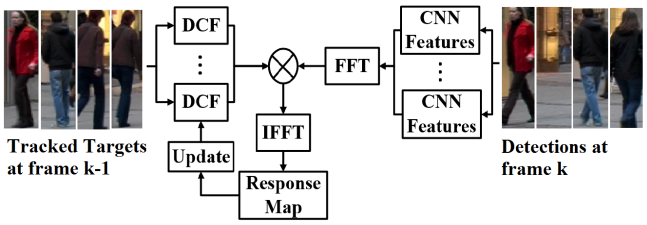
where $\mathbf{H}$ is the observation matrix, $\mathbf{R}$ is the observation noise covariance, p_M denotes the missing detection probability, and κ_k denotes the clutter density. The above work underpins the GM-PHD filter [20] as the basis of the proposed tracking framework.

2.2. Detection Analysis and Spatio-Temporal Relation

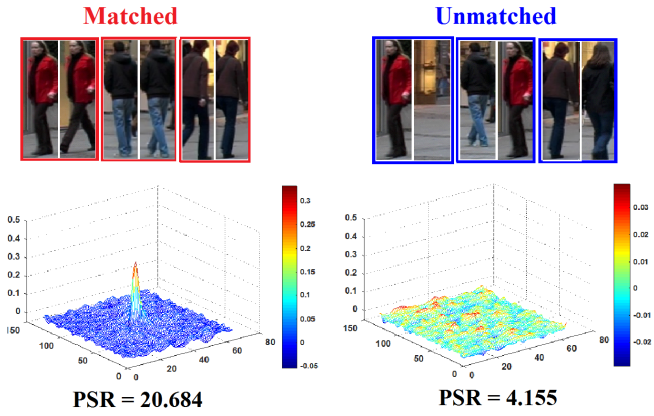
The observable measurements $\mathbf{Z}_k$ can be obtained by an object detector at each time step k. Due to the imperfections in the object detector, there is much potential uncertainty in the original detection results, which increases the inefficiency of the PHD update and birth prediction. In order to build a robust measurement model for target association, these noisy measurements can be addressed as three subsets: duplicate measurement, spurious measurement, and reliable measurement. Firstly, the non-maxima suppression [10] is utilized to merge duplicate detections on the same target into a single detection, thereby forming a combined measurement set $\mathbf{Z}_k^\star$. Then, 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 strictly discarded, because correlation filters are easily affected by false positives, and c_{th} defines the value of confidence threshold. Therefore, a reliable measurement set is achieved by $\mathbf{Z}_k^+ = \mathbf{Z}_k^\star \setminus \mathbf{\Gamma}_k$. The spatio-temporal relation has been commonly used in multi-target tracking systems [10, 11, 19]. Given the j-th current measurement $\mathbf{z}_k^j \in \mathbf{Z}_k^+$ and the i-th predicted state $\mathbf{x}_{k|k-1}^i$, a spatio-temporal cost matrix $\mathbf{W}_{k,st} \in \mathbb{R}^{M \times N^+}$ for target association can be achieved by computing each pairwise similarity score,

$$S(\mathbf{x}_{k|k-1}^i, \mathbf{z}_k^j) = \frac{1}{(2\pi\sigma_s)^{1/2}} \exp\left(- \frac{|\mathbf{H}\mathbf{x}_{k|k-1}^i - \mathbf{z}_k^j|}{2\sigma_s^2} \right) \tag{7}$$

where $|\cdot|$ denotes the Euclidean distance, σ_s^2 represents the variance and N^+ denotes the cardinality of $\mathbf{Z}_k^+$.



(a)



(b)

Fig. 2: Illustration of deep discriminative correlation matching scheme using CNN-based DCF classifier. Subfigure (a) demonstrates an example workflow of using DCF-based target-specific classifiers for corresponding targets matching (b) shows the outputs from the response map, where matched pairs are denoted with high Peak-to-Sidelobe Ratio (PSR).

2.3. Convolutional Features

We adopt the deep neural network [21] trained on a large-scale person re-identification dataset [22] for feature generation. In this work, we remove the fully-connected layer and employ the outputs of the last convolutional layer *conv4-3* as desired features, since the last layers of CNNs preserve structural and semantic information of targets and their outputs are robust to appearance variations [16]. The employed network uses the same input size of 128×64 RGB image patch for feature extraction. These convolutional features are learned offline from the appropriate target region $\mathbf{z}_k^+ \in \mathbf{Z}_k^+$ before tracking implementation. To mitigate the boundary effect, the extracted features are multiplied with a cosine window [16].

2.4. Deep Discriminative Correlation Matching

In this section, we propose a deep discriminative correlation matching scheme to link the newly-detected pedestrians with previously tracked ones in image data. The key idea is to exploit the discriminative correlation filter with convolutional features as a target-specific classifier to discriminate the desired target from noisy background and other appearing targets. Fig. 2 shows the implementation details of the proposed deep discriminative correlation matching. We train the individual DCF for each tracked target at image frame $k-1$. Given valid detections with generalized features in the next frame, we compute the pairwise correlation response maps between the tracked targets and newly detected targets to find the matched pairs with high PSRs.

2.4.1. *Training Phase*

In this work, the discriminative correlation filter is trained on a feature map $\mathbf{f}$ of size $A \times B \times D$ extracted from an image patch, where A, B, and D denote the width, height, and the number of channels, respectively. This feature map $\mathbf{f} \in \mathbb{R}^{A \times B \times D}$ can be achieved from the outputs of the last convolutional layer. Training samples for discriminative correlation filters are generated from all circular shifts $\mathbf{f}_{a,b}, (a,b) \in \{0, ..., A-1\} \times \{0, ..., B-1\}$. Each shifted sample has a desired output $g(a,b) = \exp(-\frac{(a-A/2)^2+(b-B/2)^2}{2\sigma_c^2})$ to form a Gaussian label matrix $\mathbf{g} = \{g(a,b)|(a,b) \in \{0, ..., A-1\} \times \{0, ..., B-1\}$, where σ_c is the kernel width. The discriminative correlation filter $\mathbf{c}$ with the same size of $\mathbf{f}$ can be learned by minimizing the following loss [15],

$$\arg\min_{\mathbf{c}} \sum_{a,b} \left\| \sum_{d=1}^{D} (\mathbf{c}_{a,b}^d)^T \mathbf{f}_{a,b}^d - \mathbf{g} \right\|_2^2 + \lambda \|\mathbf{c}\|_2^2 \tag{8}$$

where λ is the regularization parameter. We follow the literature of training the DCF in [15] which is to perform the fast Fourier transform (FFT) in the frequency domain. Therefore, the solution of (8) on the d-th ($d \in 1, ..., D$) channel can be written as,

$$\hat{\mathbf{c}}_{k-1}^d = \frac{\hat{\mathbf{g}} \odot (\hat{\mathbf{f}}^d)^\dagger}{\sum_{d=1}^{D} \hat{\mathbf{f}}^d \odot (\hat{\mathbf{f}}^d)^\dagger + \lambda} \tag{9}$$

where the hat stands for FFT operator, and the dagger represents complex conjugation operation. The operator $\odot$ defines the Hadamard (element-wise) product.

2.4.2. *Correlation Matching*

Let $\mathbf{y}_k \in \mathbb{R}^{A \times B \times D}$ denote the feature map of a new image patch cropped out at frame k, the correlation response map can be calculated by [15],

$$\mathbf{r}_k = \mathscr{F}^{-1} \left\{ \sum_{d=1}^{D} \hat{\mathbf{c}}_{k-1}^d \odot (\hat{\mathbf{y}}_k^d)^\dagger \right\} \tag{10}$$

where $\mathscr{F}^{-1}$ denotes the inverse fast Fourier transform (IFFT). Given the response map $\mathbf{r} \in \mathbb{R}^{A \times B}$, the PSR as an effective measurement is used to quantify the affinity between pairs: $PSR = \frac{\max(\mathbf{r}) - \mu}{\sigma_r}$, where μ and σ_r denote the mean value and the standard deviation of the sidelobes. Different from prior works [9] and [18], in which PSRs above a certain threshold indicated the correlation result is taken to correct the predicted state, whereas there was nothing to change at the association step, we then propose to use a generalized *sigmoid* function to compute a pairwise matching score between a valid detection $\mathbf{z}_k^j \in \mathbf{Z}_k^+$ and the predicted target $\mathbf{x}_{k|k-1}^i$. This function $sigmoid(x) = \frac{1}{1+e^{-(\alpha x+\beta)}}$ squashes the PSRs to a range of $[0, 1]$, because we consider PSRs < 10 [24] as unmatched pairs, and two coefficients α and

Table 1: Quantitative comparison with other state-of-the-art methods on the MOT2017 Challenge benchmark. The best results are shown in bold. Evaluation measures with (↑) indicate that higher is better, and with (↓) denote lower is better.

Method	Mode	MOTA↑	MOTP↑	MT↑	ML↓	FP↓	FN↓	IDS↓	Frag↓	Hz↑
PHD-DCM	**Online**	**46.5**	**77.2**	**16.9%**	**37.2%**	23859	**272,430**	**5649**	9,298	1.6
GMPHD-KCF [9]	**Online**	40.3	75.4	8.6 %	43.1%	47,056	283,923	5,734	**7,576**	3.3
GM_PHD [13]	**Online**	36.2	76.1	4.2%	56.6%	**23,682**	328,526	8,025	11,972	**38.4**
FWT [7]	Offline	**51.3**	77.0	21.4 %	**35.2%**	24,101	247,921	2,648	4,279	0.2
EDMT17 [8]	Offline	50.0	**77.3**	**21.6%**	36.3%	32,279	**247,297**	**2,264**	**3,260**	0.6
IOU17 [23]	Offline	45.5	76.9	15.7%	40.5%	19,993	281,643	5,988	7,404	**1522.9**
DP_NMS [5]	Offline	43.7	76.9	12.6%	46.5%	**10,048**	302,728	4,942	5,342	137.7

β are set to 0.2 and -2 respectively for generating the decision boundary. Finally, these scores are formulated as a cost matrix $\mathbf{W}_{k,dcm} \in \mathbb{R}^{M \times N^+}$ for data association. In addition, it is necessary to update the DCFs of matched targets during tracking for handling the appearance variations. Therefore, the update mechanism in [16] is adopted for the model update.

2.5. Hybrid Likelihood Function

In this section, we present a hybrid likelihood function fused with the spatio-temporal realtion $\mathbf{W}_{k,st}$ and discriminative correlation matching $\mathbf{W}_{k,dcm}$, as a result, the total association cost can be computed as, $\mathbf{W}_{k,h} = \mathbf{W}_{k,st} \odot \mathbf{W}_{k,dcm}$. The advantage of this design is that using the combination of matrices can compensate for unreliability present in the individual matrices, especially when targets ambiguities occur in either motion dynamics or visual content. The optimal association step is accomplished by the Hungarian algorithm [25] which is also intended to provide identity assignment for the PHD filter. To mitigate the sensitivity of correlation filters to false detections, we only add a new-born target and simultaneously initialise a discriminative correlation filter for its appearance modelling, if it can be tracked in the next frame.

3. EXPERIMENTS

In this section, we validate the proposed method denoted by PHD-DCM on the test set of the MOT2017 Challenge Benchmark [1]. The following parameters are utilized to implement the tracker, including the missed detection probability $p_M = 0.1$, the survival probability $e = 0.95$, the clutter intensity $\kappa = 10^{-4}$, the kernel width $\sigma_c = 0.1$, the regularization parameter of (8) $\lambda = 10^{-4}$, the variance for the similarity score σ_s^2 is empirically set to 30.

In order to achieve a fair comparison between methods, we use the same public detections for all sequences and the centralized evaluation tool provided by the website of the MOTChallenge Benchmark[1]. This evaluation tool entails the widely accepted CLEAR MOT metrics [26] including Multiple Object Tracking Precision (MOTP), Multiple Object Tracking Accuracy (MOTA), the total number of False Negatives (FN), the total number of False Positives (FP), the total number of Identity Switches (IDS). Other tracking metrics are also computed: Mostly Track targets (MT), Mostly Lost targets (ML), and the total number of times a trajectory is fragmented (Frag).

Table 1 shows the quantitative comparisons between the proposed method (PHD-DCM) and a number of state-of-the-art tracking methods presented in the MOT2017 Challenge benchmark. These include online methods: GMPHD-KCF [9] and GM_PHD [13], and other offline methods: DP_NMS [5], FWT [7], EDMT17 [8], and IOU17 [23]. The proposed method achieves the best performance compared with other state-of-the-art online methods on most evaluation measures, and even outperforms some offline methods in [5] and [23] which utilize the future detections. In fact, off-line methods based on global association techniques usually achieve better performance than online counterparts. The proposed method records the highest MOTA score which indicates the most important metric for performance analysis, and particularly improves 6.2% compared with the similar approach [9] which yields inefficiency in false positives and missed detections. In turn, our approach has successfully reduced a large amount of false positives and missed-detections, as well as providing more reliably stable tracks with high MT and low ML. Overall, evaluations above verify the proposed tracker has the merits of mitigating background clutter, maintaining the track continuity, and improving occlusions handling. In addition, average runtime performance comparisons with other approaches are also listed in Table 1, where our method runs at approximately 1.6Hz with most time consumed on the deep discriminative correlation matching scheme.

4. CONCLUSION

We have proposed a unified tracking algorithm that incorporates deep discriminative correlation matching with the GM-PHD filter for online multiple human tracking. Furthermore, a hybrid likelihood function fused with motion dynamics and visual content has been presented to enhance the target association of the GM-PHD filter. Results on MOT17 Challenge demonstrate the effectiveness of the proposed method.

[1] https://motchallenge.net/

5. REFERENCES

[1] A. Milan, L. Leal-Taixe, I. Reid, S. Roth, and K. Schindler, "MOT16: A Benchmark for Multi-Object Tracking," *arXiv:1603.00831 [cs.CV]*, pp. 1–13, 2016.

[2] P. Feng, W. Wang, S. M. Naqvi, S. Dlay, and J. A. Chambers, "Social Force Model Aided Robust Particle PHD Filter for Multiple Human Tracking," in *ICASSP*, 2016, pp. 4398–4402.

[3] 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.

[4] 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 *ICASSP*, 2017, pp. 4376–4380.

[5] H. Pirsiavash, D. Ramanan, and C. C. Fowlkes, "Globally-optimal greedy algorithms for tracking a variable number of objects," in *CVPR*, 2011, pp. 1201–1208.

[6] J. H. Yoon, C. R. Lee, M. H. Yang, and K. J. Yoon, "Online Multi-object Tracking via Structural Constraint Event Aggregation," in *CVPR*, 2016, pp. 1392–1400.

[7] R. Henschel, L. Leal-Taix, D. Cremers, and B. Rosenhahn, "A Novel Multi-Detector Fusion Framework for Multi-Object Tracking," *arXiv:1705.08314 [cs.CV]*, pp. 1–10, 2017.

[8] J. Chen, H. Sheng, Y. Zhang, and Z. Xiong, "Enhancing Detection Model for Multiple Hypothesis Tracking," in *CVPRW*, 2017, pp. 2143–2152.

[9] T. Kutschbach, E. Bochinski, V. Eiselein, and T. Sikora, "Sequential Sensor Fusion Combining Probability Hypothesis Density and Kernelized Correlation Filters for Multi-Object Tracking in Video Data," in *AVSS*, 2017, pp. 1–5.

[10] R. Sanchez-Matilla, F. Poiesi, and A. Cavallaro, "Online Multi-target Tracking with Strong and Weak Detections," in *ECCVW*, 2016, pp. 84–99.

[11] Z. Fu, P. Feng, F. Angelini, J. Chambers, and S. M. Naqvi, "Particle PHD Filter based Multiple Human Tracking using Online Discriminative Group-Structured Dictionary Learning," *submitted to IEEE Access*, 2017.

[12] 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.

[13] V. Eiselein, D. Arp, M. Ptzold, and T. Sikora, "Real-Time Multi-human Tracking Using a Probability Hypothesis Density Filter and Multiple Detectors," in *AVSS*, 2012, pp. 325–330.

[14] Z. Fu, S. M. Naqvi, and J. A. Chambers, "Enhanced GM-PHD Filter Using CNN-Based Weight Penalization for Multi-Target Tracking," in *SSPD*, 2017, pp. 1–5.

[15] C. Ma, J. B. Huang, X. Yang, and M. H. Yang, "Hierarchical Convolutional Features for Visual Tracking," in *ICCV*, 2015, pp. 3074–3082.

[16] M. Danelljan, G. Hger, F. S. Khan, and M. Felsberg, "Convolutional Features for Correlation Filter Based Visual Tracking," in *ICCVW*, 2015, pp. 621–629.

[17] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *NIPS*, 2012, pp. 1097–1105.

[18] S. H. Park, K. Lee, and K. J. Yoon, "Robust online multiple object tracking based on the confidence-based relative motion network and correlation filter," in *ICIP*, 2016, pp. 3484–3488.

[19] 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.

[20] 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.

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

[22] L. Zheng, Z. Bie, Y. Sun, J. Wang, C. Su, S. Wang, and Q. Tian, "MARS: A video benchmark for large-scale person re-identification," in *ECCV*, 2016, pp. 868–884.

[23] E. Bochinski, V. Eiselein, and T. Sikora, "High-Speed Tracking-by-Detection Without Using Image Information," in *AVSS*, 2017, pp. 1–6.

[24] M. Savvides, B. Kumar, and P. Khosla, "Face verification using correlation filters," in *AIAT*, 2002.

[25] H. W. Kuhn, *The Hungarian method for the assignment problem*. Naval Research Logistics Quarterly, 1955.

[26] 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.