PARTICLE PHD FILTER BASED MULTI-TARGET TRACKING USING DISCRIMINATIVE GROUP-STRUCTURED DICTIONARY LEARNING

Zeyu Fu, Pengming Feng, Syed Mohsen Naqvi, Jonathon A. Chambers

Communications, Sensors, Signal and Information Processing Group, Newcastle University, UK
Emails: {z.fu2, p.feng2, mohsen.naqvi, jonathon.chambers}@newcastle.ac.uk

ABSTRACT

Structured sparse representation has been recently found to achieve better efficiency and robustness in exploiting the target appearance model in tracking systems with both holistic and local information. Therefore, to better simultaneously discriminate multi-targets from their background, we propose a novel video-based multi-target tracking system that combines the particle probability hypothesis density (PHD) filter with discriminative group-structured dictionary learning. The discriminative dictionary with group structure learned by the hierarchical K-means clustering algorithm implicitly associates the dictionary atoms with the group labels, simultaneously enforcing the target candidates from the same group (class) to share the same structured sparsity pattern. Furthermore, we propose a new joint likelihood calculation by relating the discriminative sparse codes with the maximum voting technique to enhance the particle PHD updating step. Experimental results on two publicly available benchmark video sequences confirm the improved performance of our proposed method over other state-of-the-art techniques in video-based multi-target tracking.

Index Terms— Dictionary learning, group-structured sparsity, particle PHD filter, multitask, multi-target tracking

1. 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 interaction interface [1]. Many researchers seek higher-level tracking systems to locate a number of targets, retrieve their trajectories, and recognise their identities from some video sequences. However, there still exist many challenging problems caused by complicated environments such as the presence of noise, occlusions resulting in targets having similar appearance, background clutter and illumination changes [2] [3]. In recent years, the random finite set (RFS) based probability hypothesis density (PHD) filter originated from radar tracking has been successfully explored in video-based multi-target tracking [2] [4]. This technique 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 . The PHD filter based tracking scheme effectively avoids the difficulties in data association techniques and thus provides a computationally tractable alternative [5].

Sparse representation based classification has already achieved great success in many research areas such as computer vision and visual tracking applications [6]. However, recent literature shows the structured sparse representation that incorporates the structure information in terms of both group and multi-task level in the learning process provides improved performance in pattern recognition and single object visual tracking [7] [8]. Therefore, we extend this method with learning a discriminative structured dictionary to the particle PHD filter framework so as to address the challenges in video-based multi-target tracking. Different from previous methods, we employ the hierarchical K-means clustering method to learn a discriminative dictionary with group structure information instead of training the reconstructive dictionaries, since the learned discriminative dictionary has shown better performance in tracking and some other applications [9]. The collaborative hierarchical Lasso (C-HiLasso) approach is adopted to address this multi-task group-structured sparse representation thereby strengthening the discriminability of sparse coefficients at group level [10]. Based on the discriminative sparse coefficients, a novel joint likelihood calculation is proposed to further improve the particle PHD updating step using the maximum voting technique. Experimental results on multi-target tracking demonstrate improved performance of our method comparing with other state-of-the-art methods.

2. PRELIMINARIES

2.1. The Particle PHD Filter

Based upon the concept of random finite set (RFS), the PHD filter recursively propagates the first-order moment

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.

of the multi-target posterior. The target states and measurements at time k can be denoted as $\mathbf{X}_k = \{\mathbf{x}_k^1, ..., \mathbf{x}_k^{M_k}\}$ and $\mathbf{Z}_k = \{\mathbf{z}_k^1, ..., \mathbf{z}_k^{N_k}\}$ respectively, where M_k denotes the number of targets, and N_k is the number of measurements. The state of a target m is $\mathbf{x}_k^m = [p_{x,k}^m, p_{y,k}^m, v_{x,k}^m, v_{y,k}^m, w_k^m, h_k^m]^T$ contains the image location, 2D velocity and the size of the target [11]. The PHD prediction equation is :

$$\Phi_{k|k-1}(\mathbf{X}_k|\mathbf{Z}_{k-1}) = \int [e_{k|k-1}(\xi) f_{k|k-1}(\mathbf{x}_k^m|\xi) + \beta_{k|k-1}(\mathbf{x}_k^m|\xi)] \times \Phi_{k-1|k-1}(\xi) d(\xi) + \Upsilon_k \tag{1}$$

where Υ_k is the intensity function of the new-born target RFS, $f_{k|k-1}(\cdot)$ is the multi-target transition density, $e_{k|k-1}(\xi)$ is the probability that the target still exists at time k and $\beta_{k|k-1}(\mathbf{x}_k^m|\xi)$ is the intensity of the RFS that a target is spawned from the state ξ. The PHD update step is defined as [5]:

$$\Phi_{k|k}(\mathbf{X}_k|\mathbf{Z}_k) = \left[p_M(\mathbf{x}_k^m) + \sum_{\mathbf{z}_k \in \mathbf{Z}_k} \frac{\psi_{k,\mathbf{z}_k}(\mathbf{x}_k^m)}{\kappa_k + \langle \psi_{k,\mathbf{z}_k}(\mathbf{x}_k^m), \Phi_{k|k-1} \rangle} \right] \times \Phi_{k|k-1}(\mathbf{X}_k|\mathbf{Z}_{k-1}) \tag{2}$$

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

In our work, we adopt the sequential Monte Carlo method to approximate the PHD filter with a set of weighted random samples $\{\tilde{w}_{k-1}^i, \tilde{\mathbf{x}}_{k-1}^i\}_{i=1}^{i=(M_{k-1})\times N}$. The PHD prediction at time k can be represented with a set of weighted particles including both survived targets and birth targets, $\{\tilde{w}_{k|k-1}^i, \tilde{\mathbf{x}}_k^i\}_{i=1}^{i=(M_{k-1}+J_k)\times N}$ where J_k denotes the expected number of new-born targets at time k, and N is the number of particles for each target. Once the new set of observations is available, we can substitute the approximation of $\Phi_{k|k-1}(\mathbf{X}_k|\mathbf{Z}_{k-1})$ into (2) and the weights of each particle are updated as

$$\tilde{w}_k^i = \left[p_M(\tilde{\mathbf{x}}_k^i) + \sum_{\mathbf{z}_k \in \mathbf{Z}_k} \frac{\psi_{k,\mathbf{z}_k}(\tilde{\mathbf{x}}_k^i)}{\kappa_k + C_k(\mathbf{z}_k)} \right] \tilde{w}_{k|k-1}^i \tag{3}$$

where

$$C_k(\mathbf{z}_k) = \sum_{i=1}^{(M_{k-1}+J_k)\times N} \psi_{k,\mathbf{z}_k}(\tilde{\mathbf{x}}_k^i) \tilde{w}_{k|k-1}^i \tag{4}$$

and $M_k = \sum_{i=1}^{(M_{k-1}+J_k)\times N} \tilde{w}_k^i$. The above particle PHD filter has been used extensively in multi-target tracking [5]. In this paper, we also employ this framework to track multiple targets in video.

2.2. Structured Sparse Representation

Recently, structured sparse representation has been proved to provide better efficiency and robustness than the simple sparsity [9] in single object visual tracking and recognition applications, the success of which is attributed to exploiting the block structure in sparse representation and considering prior information in the predefined structure of the dictionary [12]. By giving a dictionary as $\mathbf{D} = [\mathbf{d}_1, ...\mathbf{d}_n] \in \mathbb{R}^{m\times n}$, the input signal $\mathbf{y} \in \mathbb{R}^m$ can be approximated by the linear combination with the dictionary, i.e.

$$\mathbf{y} \approx \mathbf{D}\mathbf{a} = \alpha_1\mathbf{d}_1 + \alpha_2\mathbf{d}_2 + \cdots + \alpha_n\mathbf{d}_n \tag{5}$$

The dictionary $\mathbf{D}$ can be formulated as a concatenation of p blocks that have the same length of l, $\mathbf{D} = [\mathbf{D}[1], ..., \mathbf{D}[p]]$ where $\mathbf{D}[i] = [\mathbf{d}_1, ..., \mathbf{d}_l] \in \mathbb{R}^{m\times l}$ specifically represents the i-th block of the dictionary and $n = ql$. Accordingly, the sparse coefficient vector $\mathbf{a} \in \mathbb{R}^n$ can be denoted as $\mathbf{a} = [\mathbf{a}^T[1], \mathbf{a}^T[2], ..., \mathbf{a}^T[p]]^T$, where $\mathbf{a}[i] = [\alpha_1, ..., \alpha_l] \in \mathbb{R}^l$ is the i-th block of the sparse vector. It is known that seeking a solution for sparse code $\mathbf{a} \in \mathbb{R}^n$ corresponding to $\mathbf{y}$ is NP-hard. The Group Lasso [10] can be considered as an efficient way to solve the ℓ_2-regularized least-squares problem,

$$\min_{\mathbf{a}\in\mathbb{R}^n} \frac{1}{2}\|\mathbf{y} - \mathbf{D}\mathbf{a}\|_2^2 + \lambda \sum_{i=1}^{p} \|\mathbf{a}_i\|_2 \tag{6}$$

where the λ is the regularization parameter. In our proposed tracking scheme, we introduce within-class C-HiLasso method [10] to enhance the differentiation between the multi-targets and background clutter. The hierarchical K-means clustering method is used to learn a discriminative dictionary to acquire the in-group structured sparsity in multi-task level.

3. THE PROPOSED TRACKING ALGORITHM

3.1. Dictionary Construction with Group Structure Information

Feature extraction is necessary for target appearance modelling to be applied in the training and testing process. Human features can be extracted with training data from each image patch in the target region, including the grey-scale histogram of oriented gradients (HOG) [13] and colour histogram. We form the color feature vectors as a matrix $\mathbf{F}_c = [\mathbf{c}_1, \mathbf{c}_2, ..., \mathbf{c}_n] \in \mathbb{R}^{d_c\times n}$, where n denotes the total number of feature vectors in the training data, and d_c is the dimensionality of the color feature. Likewise the vectorized HOG features are represented by a matrix $\mathbf{F}_h = [\mathbf{h}_1, \mathbf{h}_2, ..., \mathbf{h}_n] \in \mathbb{R}^{d_h\times n}$, where d_h is the dimensionality of HOG features. For simplicity, the HOG and color features can be concatenated to a combined feature set, $\mathbf{F} = [\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_n] \in \mathbb{R}^{(d_c+d_h)\times n}$.

Different from imposing data directly to a dictionary, we employ the unsupervised learning method - hierarchical K-means clustering algorithm [14] to learn a discriminative group-structured dictionary. This method allows the dictionary atoms in each class to be well clustered, and results in a large within-class similarity. For example, the same tracked target in different image frames under different illumination and pose conditions can be clustered into the same group (class). Furthermore, the learned dictionary with group structure enforces the label consistency between sub-dictionaries

and training data [8]. Consequently, the resulting discriminative dictionary comprised of independent sub-dictionaries $\mathbf{D} = [\mathbf{D}_{[g_1]}, \mathbf{D}_{[g_2]}, ..., \mathbf{D}_{[g_q]}] \in \mathbb{R}^{d \times n}$ belonging to different groups is transformed from the original feature template $\mathbf{F}$, where the group structure is defined as $G = \{g_1, ..., g_q\}$, and g_q is the sub-dictionary index. We assume that the group structure G has q groups with the same number of l sub-dictionary atoms in each group. We treat the particles as the observation signals, which are randomly sampled from the current predicted states of the multi-targets at time k. Typically, we crop the observed target region $\mathbf{z}_k$ in the current frame as well as extracting the human features, where the observed target vector $\mathbf{y} \in \mathbb{R}^d$ in this application is refined from a particle $\tilde{\mathbf{x}}_k^i$ at time instant k containing the localization. In fact, learning the representation for each particle can be viewed as an individual task, while we exploit similarities between particles and the group-structured dictionary in a multi-task approach, which yields an observation matrix $\mathbf{Y} = [\mathbf{y}_1, ..., \mathbf{y}_h] \in \mathbb{R}^{d \times h}$, where the number of columns j denotes the total number of predicted particles. This approach renders the particle representations to be jointly sparse, and only a few groups of atoms should be used to represent all the corresponding particles at each frame [15].

Since the dictionary-based tracking methods suffer from computational complexity, we adopt the well-known principal component analysis (PCA) method [16] for reducing the dimension of our learned dictionary. The advantage of this technique is to reduce the dictionary dimensions to be more robust for the classification performance. Our learned dictionary $\mathbf{D} \in \mathbb{R}^{d \times n}$ is efficiently reduced to a small-sized dictionary $\mathbf{D}_s \in \mathbb{R}^{d_s \times n}$. Moreover, dimensionality reduction is also applied in the testing process, which yields a new observation matrix $\mathbf{Y}_s \in \mathbb{R}^{d_s \times h}$.

3.2. Discriminative Group-Structured Dictionary Learning for Multi-Target Tracking (DGSDL-MTT)

According to the simple structured sparsity, the sparse coding solution $\mathbf{a}_j$ for each test target $\mathbf{y}_j$ can be performed separately, since different tasks are capable of choosing the dictionary atoms independently. However, multiple test targets from the same category associated with dictionary atoms would share the same sparsity pattern at the group level, which can be achieved by the within-class multi-task group-structured sparsity model [17]. The sparsity pattern is shown in Fig.1, which is effective and suitable to perform multi-class classification for multi-target tracking. Given by the input signals $\mathbf{Y}_s \in \mathbb{R}^{d_s \times h}$ and the learned dictionary $\mathbf{D}_s \in \mathbb{R}^{d_s \times n}$, the sparse coefficients matrix $\mathbf{A} = [\mathbf{a}_1, ..., \mathbf{a}_h] \in \mathbb{R}^{n \times h}$ can be accomplished by the following C-HiLasso model [10],

$$\min_{\mathbf{A} \in \mathbb{R}^{n \times h}} \frac{1}{2} \|\mathbf{Y}_s - \mathbf{D}_s \mathbf{A}\|_F^2 + \lambda_2 \sum_{g \in G} \|\mathbf{A}^g\|_F + \lambda_1 \sum_{j=1}^{h} \|\mathbf{a}_j\|_1 \quad (7)$$

where $\mathbf{A}^g$ is the sub-matrix consisting of all the rows belonging to the group g, and $\|\cdot\|_F$ denotes the Frobenius norm. In addition, the selection of λ_1 and λ_2 is dependent of the application and data, such parameters can be obtained by cross validation. With the employment of this sparsity solution, the group stucture could enforce the sparse coefficients for different classes to deal with different subspaces, so the sparse coefficients in our system would be further strengthened to simultaneously discriminate the candidate targets from the background clutter.

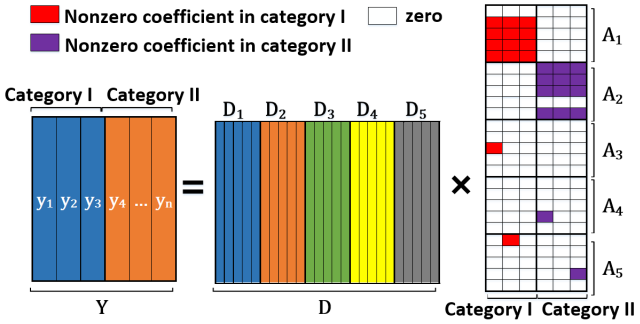


Fig. 1: Illustration of multi-task structured sparsity solution in particle representation induced by the C-HiLasso model. The dictionary $\mathbf{D}_s$ consists of sub-dictionaries for five different groups, $\mathbf{D}_{s_1}, ..., \mathbf{D}_{s_5}$, with five atoms in each group. Input signals $\mathbf{Y}_s$ could contain different categories such as category I and category II, and each category consists of the number of N columns. All the signals within the same class are forced to reveal the same group-sparsity structure $\mathbf{A}_1, ..., \mathbf{A}_5$, while the within-group sparsity pattern can be varied because of different samples in the same group.

3.3. Joint Likelihood Calculation with Maximum Voting

In general, the nonzero sparse codes in each category are concentrated on the sub-matrix $\mathbf{A}^g$ including all the rows belonging to the group g, as depicted in Fig.1. However, when candidate targets are outliers and out of the dictionary, the nonzero coefficients tend to scatter among groups instead of centralizing in some single group [6]. So then we define the following condition with maximum voting to remove the foreigners,

$$\mu_k(\tilde{\mathbf{x}}_k^i) = \begin{cases} 0 & r^{max} < \varepsilon, \text{if outliers} \\ \exp^{-(\gamma \times \eta)} & r^{max} \geq \varepsilon \end{cases} \quad (8)$$

where γ is the regularized parameter, ε denotes the threshold value, and the η is determined by the selected sparse codes using average pooling. The detailed algorithm is summarized in Algorithm 1. Furthermore, background subtraction results containing the localization $\mathbf{b}_k = [b_{x,k}, b_{y,k}]^T$ are also used to compute another likelihood function for each candidate target.

$$p(\mathbf{b}_k | \tilde{\mathbf{x}}_k^i) = \exp^{-\frac{(\mathbf{\Omega}_k^i - \mathbf{b}_k)^T (\mathbf{\Omega}_k^i - \mathbf{b}_k)}{\sigma^2}} \quad (9)$$

where $\mathbf{\Omega}_k^i = [p_{x,k}^i, p_{y,k}^i]^T$ denotes the position of the targets taken from the particle $\tilde{\mathbf{x}}_k^i$, and σ is the standard deviation in the observation model. The combined likelihood function for the PHD update step can be determined by the product of background subtraction results and the structured sparseness,

$$p(\mathbf{z}_k | \tilde{\mathbf{x}}_k^i) = \mu_k(\tilde{\mathbf{x}}_k^i) p(\mathbf{b}_k | \tilde{\mathbf{x}}_k^i) \quad (10)$$

By feeding(10) in to (3), the updated weights of the particle PHD filter are obtained.

Algorithm 1: Weight Calculation by Maximum Voting

Input : Current frame at k; All h predicted particles $\tilde{\mathbf{x}}_k^i$; The matrix $\mathbf{A} \in \mathbb{R}^{n \times h}$ with C categories and N columns for each category; the group structure $G = \{g_1, ..., g_q\}$.

Output: Weight function μ_k using maximum voting

Initialization: $\mathbf{e}_0$ and $\mathbf{e}_1$ are the vectors of all ones; the q-th group of the sparse vector $\mathbf{a}[q]$ with the same length l.

for *each category* C **do**

  Initialize the similarity ratio $r_g = \emptyset$;

  Compute the ratio $r_g = \frac{\mathbf{e}_0^T A_c^g \mathbf{e}_0}{\mathbf{e}_1^T A_c \mathbf{e}_1}$, $g = g_1, ..., g_q$;

  Maximum voting method : $\theta = \arg\max_g(r_g)$, $r^{max} = \max(r_g)$;

  for *test target* $i = 1, ..., N$ **do**

    Calculate the selected sparse codes using average pooling method: $\eta = \frac{1}{l}\sum_{j=1}^{l} \mathbf{a}_i[\theta]$;

    Compute the weight function μ_k by solving (8).

  end

end

4. EXPERIMENTS

In this section, we evaluate both the effectiveness and strength of our proposed tracking method via implementing it on the video sequences from the well-known CAVIAR [18] and PETS2009 [19] datasets. In this work, 100 particles are employed to represent each target in the particle PHD filter framework. The missing detection probability $p_M = 0.01$, survival probability $e = 0.99$, the new birth intensity is $\Upsilon = 0.9$ and the clutter intensity $\kappa = 0.01$. The group structure G has 6 groups with 5 sub-dictionary atoms in each group. Besides, the regularization parameters for DGSDL-MTT are $\lambda_1 = 0.1$, $\lambda_2 = 0.01$, $\gamma = 10$ and $\sigma = 25$.

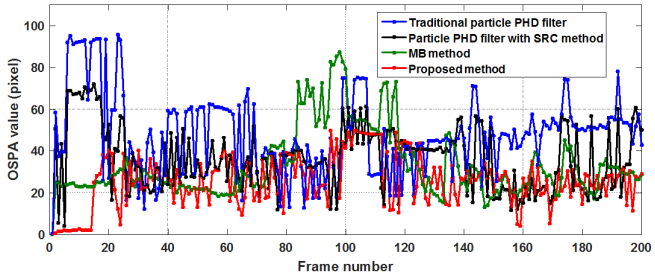
Table 1: Quantitative comparison between proposed method and other state-of-the-art methods on the CAVIAR dataset

Method	**Proposed method**	PHD filter method [21]	PHD-SRC method [22]	MB [20] method
OSPA(pixel)	25.59	48.26	34.39	33.71
AEE(pixel)	19.71	32.24	26.62	25.46

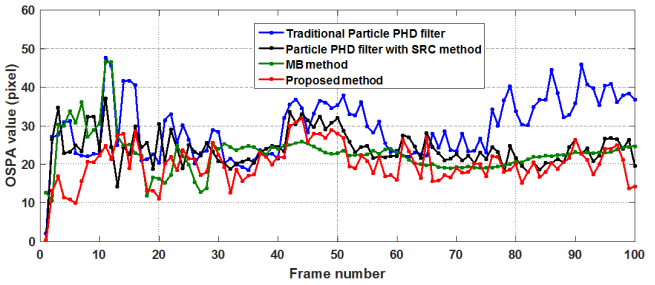
Table 2: Quantitative comparison between proposed method and other state-of-the-art methods on the PETS2009 dataset

Method	**Proposed method**	PHD filter method [21]	PHD-SRC method [22]	MB [20] method
OSPA(pixel)	19.51	32.54	24.16	23.06
AEE(pixel)	12.17	22.89	17.57	15.01

The optimal subpattern assignment (OSPA) metric [23] and average Euclidean error (AEE) are both utilized as the performance measure to evaluate our proposed tracking system. Tables 1 and 2 summarize the quantitative results in both



(a) OSPA comparison in CAVIAR dataset



(b) OSPA comparison in PETS2009 dataset

Fig. 2: Comparison in OSPA value for two video sequences between the our method and other three methods. Subfigure (a) demonstrates the comparison with the CAVIAR dataset and (b) is corresponding result for the PETS2009 dataset.

video sequences respectively, indicating our proposed tracking system effectively improves the tracking performance in terms of the average OSPA and AEE measures in comparison with other three state-of-the-art methods. Furthermore, the improved accuracy can be visually seen from Fig.2, the OSPA value of our proposed method is shown to be lower than the baseline methods in most frames. Overall, both comparative results demonstrate that our proposed tracker improves the ability to eliminate the background noise and false alarms. Besides, more recent state-of-the-art methods and available video sequences are being applied to further evaluate the proposed tracking scheme.

5. CONCLUSION

In this paper, we proposed a novel multi-target tracking method incorporating the particle PHD filter with discriminative group-structured dictionary learning. We explored the properties of group-structured dictionary learning to improve the discriminative power of sparse coding. A new joint likelihood calculation based on the collaborative structured sparsity was applied to overcome the challenging tracking problems of false alarms and background clutter caused by the noisy measurements. The results were shown to demonstrate the proposed method performs significantly better than the baseline methods. Future work will integrate an online approach to update our group-structured dictionary, this updated dictionary will be dealing with the appearance changes of the target in order to further improve the accuracy.

6. REFERENCES

[1] C. H. Kuo, C. Huang, and R. Nevatia, "Multi-target Tracking by On-line Learned Discriminative Appearance Models," in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, 2010, pp. 685–692.

[2] P. Feng, W.Wang, S. M. Naqvi, S. Dlay, and J. Chambers, "Social Force Model based MCMC-OCSVM Particle PHD Filter for Multiple Human Tracking," *IEEE Transactions on Multimedia*, vol. PP, no. 99, pp. 1–15, 2016.

[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, 2015.

[4] X. Zhou, Y. Li, B. He, and T. Bai, "GM-PHD-Based Multi-Target Visual Tracking Using Entropy Distribution and Game Theory," *IEEE Transactions on Industrial Informatics*, vol. 10, no. 2, pp. 1064–1076, 2014.

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

[6] W. Z. Lu, C. Bai, K. Kpalma, and J. Ronsin, "Multi-Object Tracking using Sparse Representation," in *2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2013, pp. 2312–2316.

[7] T. Zhang, B. Ghanem, S. Liu, and N. Ahuja, "Robust Visual Tracking via Structured Multi-Task Sparse learning," *International Journal of Computer Vision*, vol. 101, no. 2, pp. 367–383, 2013.

[8] Y. Suo, M. Dao, U. Srinivas, V. Monga, and T. Tran, "Structured dictionary learning for classification," *arXiv:1406.1943 [cs.CV]*, pp. 1–14, 2014.

[9] B. Liu, J. Huang, C. Kulikowski, and L. Yang, "Robust Visual Tracking Using Local Sparse Appearance Model and K-Selection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 12, pp. 2968–2981, 2013.

[10] P. Sprechmann, I. Ramirez, G. Sapiro, and Y. C. Eldar, "C-HiLasso: A Collaborative Hierarchical Sparse Modeling framework," *IEEE Transactions on Signal Processing*, vol. 59, no. 9, pp. 4183–4198, 2011.

[11] 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 *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2016, pp. 4398–4402.

[12] X. Yun and Z. Jing, "Kernel joint visual tracking and recognition based on structured sparse representation," *Neurocomputing*, vol. 193, pp. 181 – 192, 2016.

[13] N. Dalal and B. Triggs, "Histograms of Oriented Gradients for Human Detection," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR05)*, vol. 1, 2005, pp. 886–893.

[14] M. Barnard, P. Koniusz, W. Wang, J. Kittler, S. M. Naqvi, and J. Chambers, "Robust Multi-Speaker Tracking via Dictionary Learning and Identity Modelling," *IEEE Transactions on Multimedia*, vol. 16, no. 3, pp. 864–880, 2014.

[15] T. Zhang, S. Liu, C. Xu, S. Yan, B. Ghanem, N. Ahuja, and M. H. Yang, "Structural Sparse Tracking," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2015*, 2015, pp. 150–158.

[16] I. T. Jolliffe, *Principal Component Analysis*. Springer-Verlag New York, Inc, 2002.

[17] Y. Xu, Y. Sun, Y. Quan, and B. Zheng, "Discriminative structured dictionary learning with hierarchical group sparsity," *Computer Vision and Image Understanding*, vol. 136, pp. 59 – 68, 2015.

[18] R. Fisher. (2003) Caviar test case scenarios. [Online]. Available: http://homepages.inf.ed.ac.uk/rbf/CAVIARDATA1/

[19] 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

[20] R. Hoseinnezhad, B. N. Vo, and B. T. Vo, "Visual Tracking in Background Subtracted Image Sequences via Multi-Bernoulli Filtering," *IEEE Transactions on Signal Processing*, vol. 61, no. 2, pp. 392–397, 2013.

[21] Y. Wang, J. Wu, A. A. Kassim, and W. Huang, "Tracking a variable number of human groups in video using probability hypothesis density," in *18th International Conference on Pattern Recognition (ICPR'06)*, vol. 3, 2006.

[22] Z. Fu, P. Feng, S. M. Naqvi, and J. A. Chambers, "Robust Particle PHD Filter with Sparse Representation for Multi-Target Tracking," *Accepted in IEEE International Conference on Digital Signal Processing (DSP)*, pp. 1–5, 2016.

[23] D. Schuhmacher, B.-T. Vo, and B.-N. Vo, "A consistent metric for performance evaluation of multi-object filters," *IEEE Transactions on Signal Processing*, vol. 56, no. 8, pp. 3447–3457, 2008.