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

Zeyu Fu, Pengming Feng, Syed Mohsen Naqvi and Jonathan A. Chambers
ComS[2]IP Research Group, School of Electrical and Electronic Engineering, Newcastle University, UK.
Emails: {z.fu2, p.feng2, mohsen.naqvi, jonathon.chambers}@newcastle.ac.uk

Introduction

- A novel video-based multi-target tracking system is proposed by combining the particle PHD filter with discriminative group-structured dictionary learning.
- Learn a discriminative dictionary with group structure information.
- The collaborative hierarchical Lasso (C-HiLasso) model is used to compute this multi-task group-structured sparse representation.
- A novel joint likelihood calculation aims to further improve the particle PHD updating step using the maximum voting technique.

Particle PHD Filter Framework

- The target state model at time k can be denoted as $\mathbf{X}_k = \{\mathbf{x}_k^m, m = 1, \dots, M_k\}$, and $\mathbf{Z}_k$ is the measurement set. The PHD prediction equation is given as:

$$\upsilon_{k|k-1}(\mathbf{X}_k|\mathbf{Z}_{k-1}) = \int \phi_{k|k-1}(\mathbf{x}_k^m, \xi)\upsilon_{k-1|k-1}(\xi)d(\xi) + \Upsilon_k$$

- The analogue of the state transition probability in the single target case is defined as:

$$\phi_{k|k-1}(\mathbf{x}_k^m, \xi) = e_{k|k-1}(\xi) f_{k|k-1}(\mathbf{x}_k^m|\xi) + \beta_{k|k-1}(\mathbf{x}_k^m|\xi)$$

- Given a set of particles $\{\tilde{\mathbf{x}}_k^i, \tilde{\omega}_{k-1}^i\}_{i=1}^{i=M_{k-1}\times N}$ at time k, the PHD prediction at time k can be represented with a set of weighted particles including both survived targets and new-born targets $\{\tilde{\mathbf{x}}_k^i, \tilde{\omega}_{k|k-1}^i\}_{i=1}^{(M_{k-1}+J_k)\times N}$, and the weights represented as:

$$\tilde{\omega}_{k|k-1}^i = \begin{cases} \phi_{k|k-1}(\tilde{\mathbf{x}}_\mathrm{k}^\mathrm{i})\tilde{\omega}_{k-1}^i & \text{survived particles} \\ \dfrac{\gamma_\mathrm{k}}{J_k \times N} & \text{newborn particles} \end{cases}$$

- The update step of the particle PHD filter is given as:

$$\tilde{\omega}_k^i = \left[P_M(\tilde{\mathbf{x}}_k^i) + \sum_{\forall \mathbf{z}_\mathrm{k} \in \mathbf{Z}_\mathrm{k}} \frac{\varphi_{k,z_k}(\tilde{\mathbf{x}}_k^i)}{\kappa_k + C_k(\mathbf{z}_k)}\right]\tilde{\omega}_{k|k-1}^i$$

where

$$\mathrm{C}_\mathrm{k}(\mathbf{z}_\mathrm{k}) = \sum_{\mathrm{j}=1}^{\mathrm{N}} \varphi_{\mathrm{k},\mathrm{z}_\mathrm{k}}(\tilde{\mathbf{x}}_\mathrm{k}^\mathrm{i})\,\tilde{\omega}_{k|k-1}^i$$

$$\varphi_{k,z_k}(\tilde{\mathbf{x}}_k^i) = \left(1 - P_M(\tilde{\mathbf{x}}_k^i)\right)p(\mathbf{z}_k|\tilde{\mathbf{x}}_k^i).$$

- The number of targets is calculated as:

$$M_k = round\left\{\sum_{i=1}^{(M_{k-1}+J_k)\times N} \tilde{\omega}_k^i\right\}$$

- A resampling step that eliminates particles with low importance weight and avoids the degeneracy problem will be performed after the update step.

Dictionary Construction

- We extract two types of features with sufficient training data from each image patch in the target region, the grey-scale histogram of oriented gradients (HOG) and colour histogram,

$$\text{HOG feature}: \mathbf{F_h} = [\mathbf{h}_1, \dots, \mathbf{h}_\mathrm{n}] \in \mathbb{R}^{\mathrm{T_h}\times\mathrm{n}}$$

$$\text{Colour feature}: \mathbf{F_c} = [\mathbf{c}_1, \dots, \mathbf{c}_\mathrm{n}] \in \mathbb{R}^{\mathrm{T_c}\times\mathrm{n}}$$

- For simplicity, the HOG and colour features can be concatenated to a combined vector set,

$$\mathbf{F} = [\mathbf{f}_1, \dots, \mathbf{f}_\mathbf{n}] \in \mathbb{R}^{(\mathrm{T_c}+\mathrm{T_h})\times\mathrm{n}}$$

- Learning a discriminative group-structured dictionary, allows the dictionary atoms in each class to be well clustered, and results in a large within-class similarity.
- The discriminative dictionary comprised of independent sub-dictionaries belonging to different groups is transformed from the original feature template $\mathbf{F}$,

$$\mathbf{D} = \left[\mathbf{D}_{[g_1]}, \dots, \mathbf{D}_{[g_q]}\right] \in \mathbb{R}^{d\times n}$$

- The group structure is defined as $G = \{g_1, \dots, g_q\}$, g_q is the sub-dictionary index. The group structure G has q groups with the same number of l sub-dictionary atoms in each group.

Discriminative Group-Structured Dictionary Learning for Multi-Target Tracking

- A target $\mathbf{y} \in \mathbb{R}^d$ within the current frame will be approximated by the linear combination of known targets from training samples,

$$\mathbf{y} \approx \mathbf{Da} = \alpha_1\mathbf{d}_1 + \alpha_2\mathbf{d}_2 + \cdots + \alpha_\mathrm{n}\mathbf{d}_\mathrm{n}$$

- Given by the input signals $\boldsymbol{Y} \in \mathbb{R}^{d\times h}$ and the learned dictionary $\mathbf{D} \in \mathbb{R}^{d\times n}$, the sparse coefficients matrix $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_h] \in \mathbb{R}^{n\times h}$ can be accomplished by the following C-HiLasso model,

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

- Multiple test targets from the same category associated with dictionary atom share the same sparsity pattern at the group level, which can be achieved by the within-class multi-task group structured sparsity model.
- The group structure could enforce the sparse coefficients for different classes to deal with different subspaces, so the sparse coefficients would be further strengthened to simultaneously discriminate the candidate targets from the background clutter.

Weight Calculation by Maximum Voting

- **Initialization**: $\boldsymbol{e}_0$ and $\boldsymbol{e}_1$ are the vectors of all ones; the qth group of the sparse vector $\mathbf{a}[q]$ with the same length l.
- **For** each category C **do**

  Compute the ratios: $r_g = \dfrac{\boldsymbol{e}_0^T \boldsymbol{A}_c^g \boldsymbol{e}_0}{\boldsymbol{e}_1^T \boldsymbol{A}_c \boldsymbol{e}_1}, g = g_1, \dots, g_q$;

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

- **For** test target $i = 1, \dots, N$ **do**

  Calculate the average of the selected sparse code:

  $\eta = \frac{1}{l}\sum_{j=1}^{l}\{\mathbf{a}_i[\theta]\}_j$

  Compute the weight function: $\mu_k(\tilde{\mathbf{x}}_k^i) = \begin{cases} 0 & r^{max} < \varepsilon, \text{if outlier} \\ \exp^{-(\gamma\times\eta)} & r^{max} \geq \varepsilon \end{cases}$

  End

End

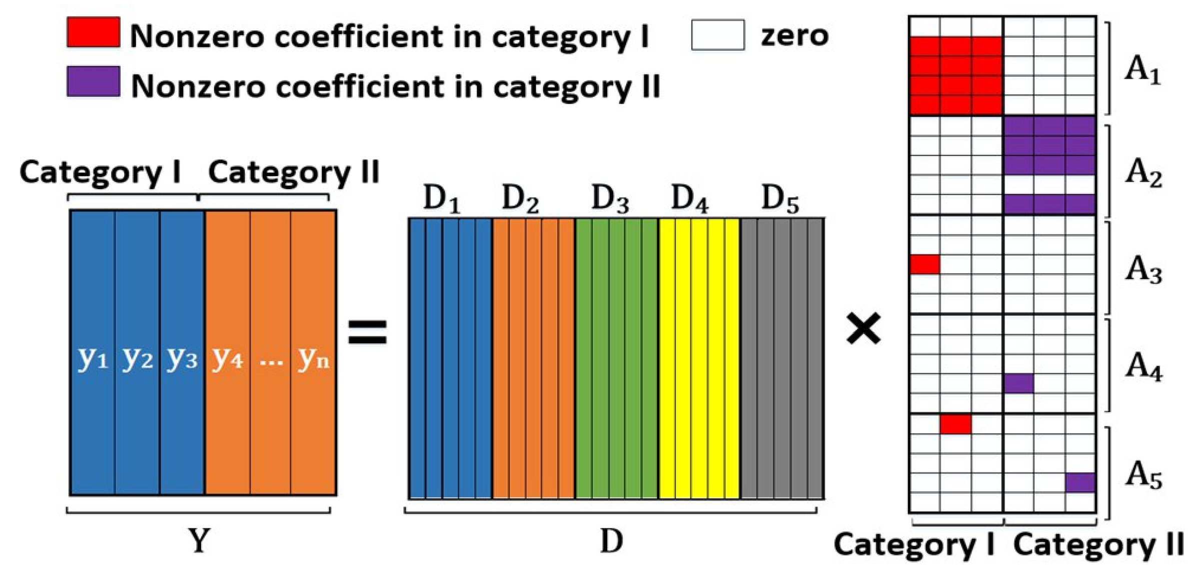

Fig. 1: Illustration of multi-task structured sparsity solution in particle representation induced by the C-HiLasso model.

Experimental Results

- We evaluate both the effectiveness and strength of our proposed tracking method via implementing it on the video sequences from the well-known CAVIAR and PETS2009 datasets.

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

(a)

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

(b)

Table I: Quantitative comparison of different method (a) CAVIAR (b) PETS2009 dataset

- The optimal subpattern assignment (OSPA) metric and average Euclidean error (AEE) both serve as the performance measure, in order to evaluate and compare our proposed tracking system with other state-of-the-art tracking method.

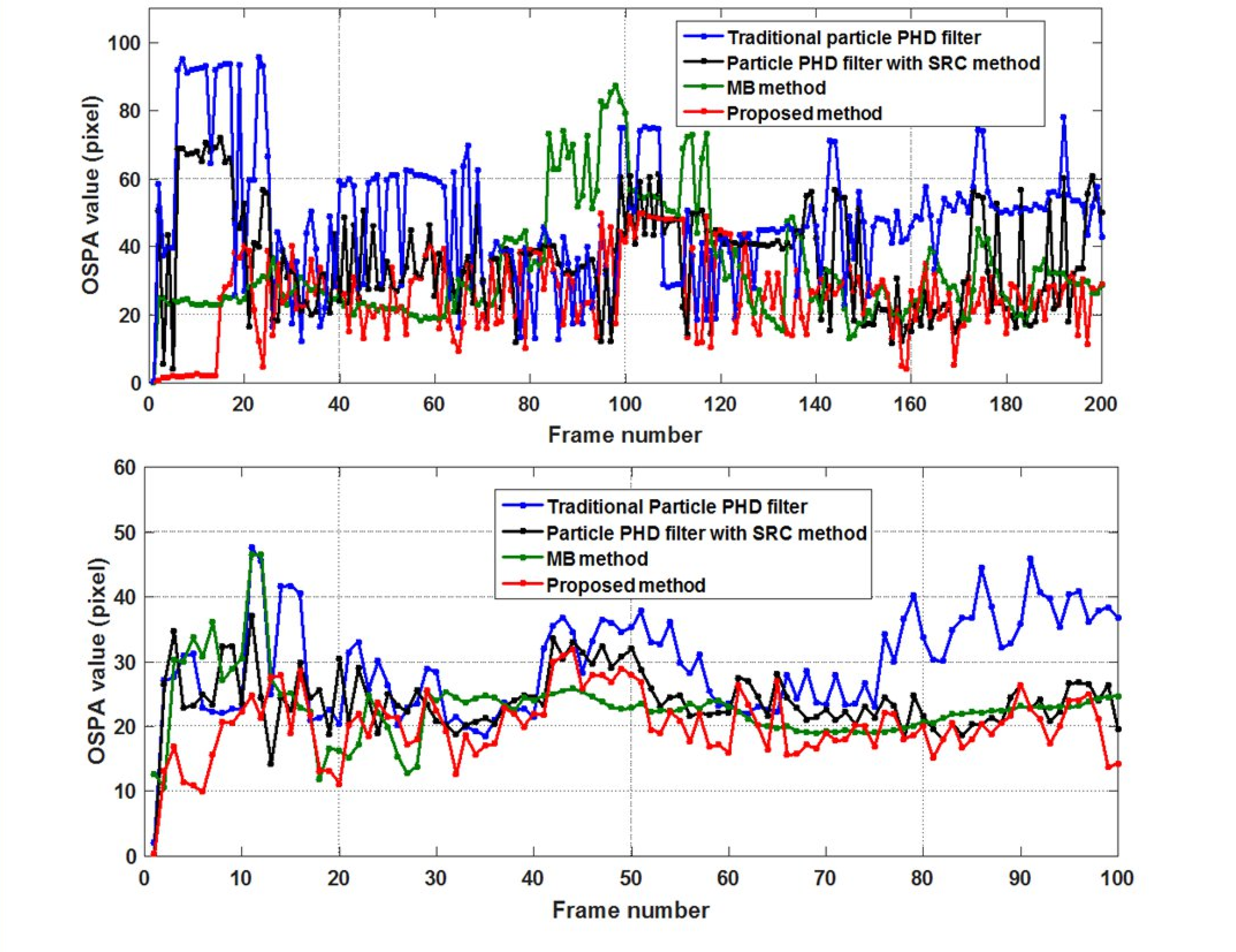

Fig. 2: OSPA comparison of different tracking systems

Conclusion and Future Work

- Proposed a novel multi-target tracking method incorporating the particle PHD filter with discriminative group-structured dictionary learning.
- 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.
- 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.

References

[1] 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.
[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, 2015.
[3] 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.
[4] 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.