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

Zeyu Fu, Syed Mohsen Naqvi and Jonathon A. Chambers

ISC Research Group, School of Engineering, Newcastle University, UK

Introduction

Video Surveillance Behaviour Analysis Assisted Living Homeland Security

- The main objective of video-based multiple human tracking is to locate a number of human targets, retrieve their trajectories with 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.
- The Gaussian mixture-probability hypothesis density (GM-PHD) filter [15] as an effective online state estimation technique has the ability to deal with varying number of targets, reduce missed detections, and mitigate spatial noise.
- **Our approach is to exploit the convolutional neural network (CNN) based weight penalization to enhance the GM-PHD filter for tracking multiple targets in video.**

Baseline Method

- **The Gaussian Mixture PHD Filter**
  The GM-PHD filter [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.
  - Initialization:

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

  - Prediction:

$$\nu_{k|k-1}(\mathrm{x}) = \nu_{k|k-1}^{s}(\mathrm{x}) + \gamma_k(\mathrm{x}) \quad \mathcal{I}_{\gamma,k} = \{I_{\gamma,k}^{1}, ..., I_{\gamma,k}^{J_{\gamma,k}}\} \tag{2}$$

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

  - Update:

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

  where

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

Measurement Classification

- **Detection Analysis Eliminates the False Measurements**
  - Use the detection confidence score $c_k \in [0, 1]$ associated with each detection to categorize the spurious measurement set $\Gamma_k = \{\mathrm{z}_{k,f} : c_k < c_{th}\}$ that will be discarded.
  - A real measurement set is obtained by $\mathrm{Z}_{k,r} = \mathrm{Z}_k \setminus \Gamma_k$.
- **Adaptive Gating Technique Initialises New-Born Targets**
  - An adaptive gating method [19] based on spatio-temporal relation is used to further extract the birth measurement set $\mathrm{Z}_{k,b}$ and the survival measurement set $\mathrm{Z}_{k,s}$ from the real measurement set $\mathrm{Z}_{k,r} = \mathrm{Z}_{k,b} \cup \mathrm{Z}_{k,s}$.
  - Target Birth: each measurement $\mathrm{z}_{k,b}$ in $\mathrm{Z}_{k,b}$ will be initialised as a new target trajectory with a new identity, and it will be eventually infused with the birth prediction.

References

- 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.
- 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.
- 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.

CNN-based Weight Penalization

- **Weight Matrix $\mathrm{W}_k \in \mathbb{R}^{J_{k|k-1} \times N_k}$ Initialization**

$$\begin{array}{c c} & \begin{array}{ccccc} n=1 & 2 & \cdots & N_k \end{array} \\ \begin{array}{c} j=1 \\ 2 \\ \vdots \\ J_{k|k-1} \end{array} & \begin{bmatrix} w_k^{(1,1)} & w_k^{(1,2)} & \cdots & w_k^{(1,N_k)} \\ w_k^{(2,1)} & w_k^{(2,2)} & \cdots & w_k^{(2,N_k)} \\ \vdots & \vdots & \vdots & \vdots \\ w_k^{(J_{k|k-1},1)} & w_k^{(J_{k|k-1},2)} & \cdots & w_k^{(J_{k|k-1},N_k)} \end{bmatrix} \end{array}$$

- **CNN-based Feature Extraction**
  - We adopt the deep neural network [21] trained on a largescale person re-identification dataset that contains over 1,100,000 images of 1,261 pedestrians for feature generation.
  - The global feature map of dimensionality 128 is to construct feature vectors of f_k and d_k.
- **Weight Penalization**
  - Bhattacharyya distance is utilized to calculate the following similarity score in terms of feature space between the j-th predicted target and n-th measurement at time k,

$$\theta_k(j, n) = \frac{1}{\sqrt{2\pi\sigma_\theta^2}} \exp\left(-\frac{\{S_k(j,n)\}^2}{2\sigma_\theta^2}\right) \tag{6}$$

  where

$$S_k(j, n) = \sqrt{1 - (\mathrm{f}_k^j)^T \mathrm{d}_k^n}. \tag{7}$$

  - Weight Refinement:

$$w_k^{(j,n)} = w_k^{(j,n)} \times \theta_k(j, n). \tag{8}$$

Improved Track Management

We adapt the method in [17] to select targets with the maximum weights as a collection of possible tracks. The maximum weights for each row of the penalized weight matrix W_k^p can be computed as, $\tilde{n} = \arg\max_{n=1:N_k}(w_k^{(j,n)})$.

- **Target Confirmation**: targets can be confirmed with $w_k^{(j,\tilde{n})} \geq w_{th}$ and labelled with the same identity as that in prediction.
- **Target Termination**: The rest of the targets which fail to reach w_{th} are tentatively eliminated after a certain value of T_{miss} frames.
- **Merged Target Segmentation**: Search for the ambiguous weights in W_k^p with the same value of $\tilde{n}$, and select the targets with smaller weights among them as covered targets, which will remain unchanged during the update step.

Experiments

- CLEAR MOT metrics [25] are employed to evaluate the tracking performance.
- Quantitative comparison between proposed method and different approaches on PETS2009 and TUD-stadtmitte datasets.

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%**

Method	MOTP (↑)	MOTA (↑)	IDS (↓)	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%

Conclusions

- Presented an enhanced GM-PHD filter using CNN-based weight penalization for multiple target tracking in video.
- Exploited the deep learning method to extract human features, which are used to penalize the weights in the weight matrix.
- An improved track management has been introduced to correctly estimate target states and eliminate false tracks.