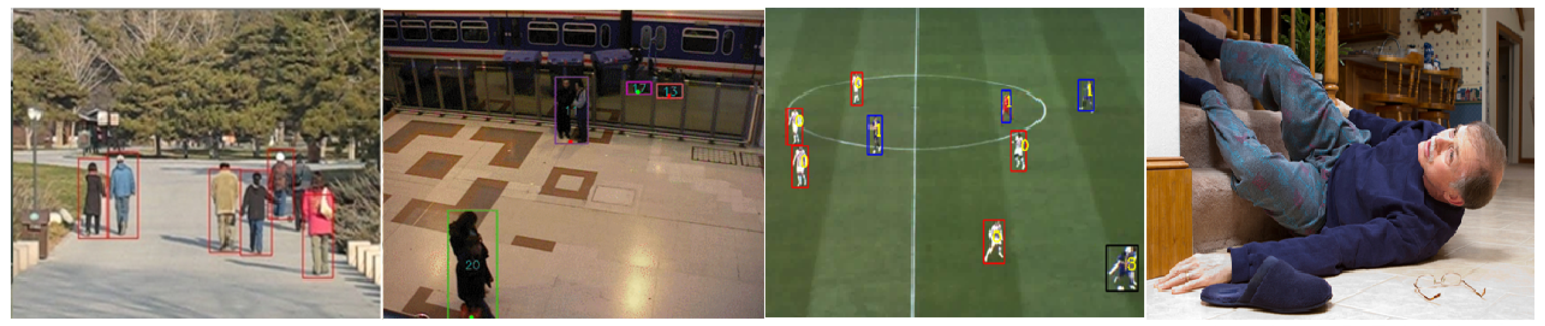
GM-PHD Filter Based Online Multiple Human Tracking using Deep Discriminative Correlation Matching

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

Intelligent Sensing and Communications Research Group, Newcastle University, UK

1. Introduction



Video Surveillance | Homeland Security | Sports Analysis | Assisted Living

- **Challenges:**
  Variable number of targets, Targets moving in close proximity, False alarms, and Long-term occlusions.
- **Our contributions:**
  - Develop individual target-specific classifiers built on the CNN-based discriminative correlation filter (DCF) to discriminate the desired targets from noisy background and other appearing targets.
  - Present a hybrid likelihood function to address the target ambiguity.

2. Baseline Method

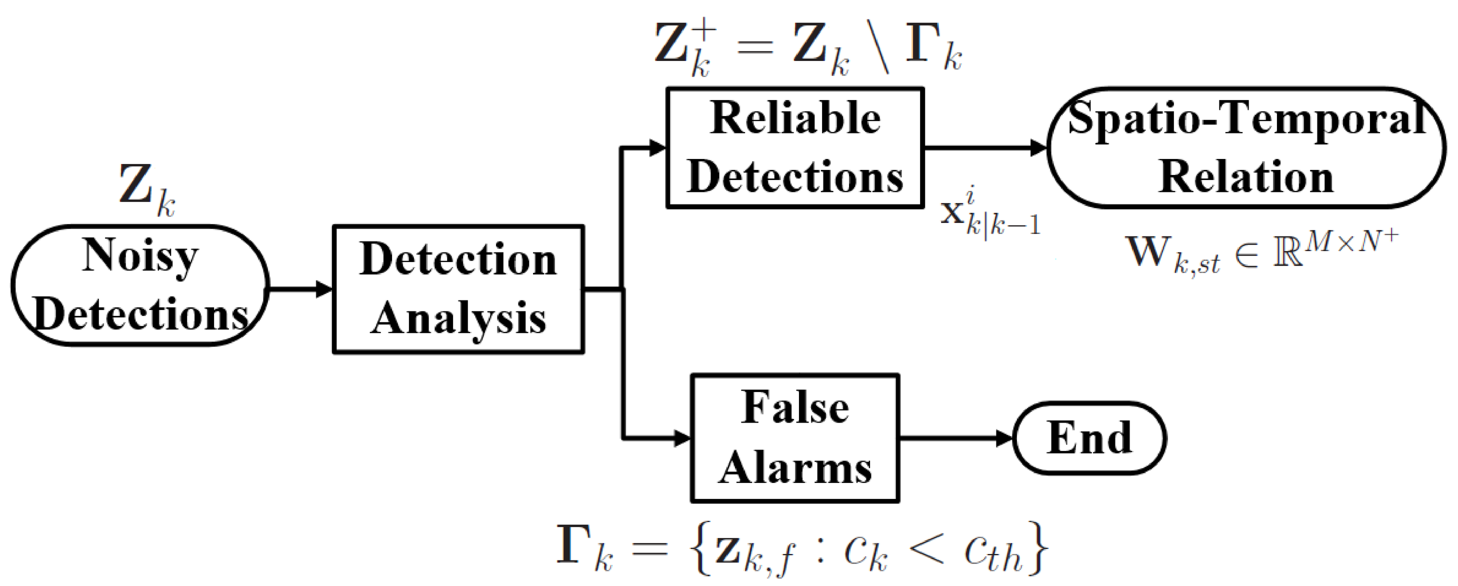
- **The Gaussian Mixture PHD Filter**
  The PHD filter with the GM implementation [1] is much more efficient than its SMC counterpart. The posterior PHD intensity function can be represented by a sum of weighted Gaussian components that are propagated analytically in time.
  - Prediction:

$$\nu_{k|k-1}(\mathbf{x}) = \sum_{j=1}^{J_{k|k-1}} w^j_{k|k-1}\mathcal{N}(\mathbf{x}; \mathbf{m}^j_{k|k-1}, \mathbf{P}^j_{k|k-1}) \quad (1)$$

  - Update:

$$\nu_k(\mathbf{x}) = p_M \nu_{k|k-1}(\mathbf{x}) + \sum_{\mathbf{z}\in\mathbf{Z}^+_k}\sum_{j=1}^{J_{k|k-1}} w^j_k(\mathbf{z})\mathcal{N}(\mathbf{x}; \mathbf{m}^j_{k|k}(\mathbf{z}), \mathbf{P}^j_{k|k}) \quad (2)$$

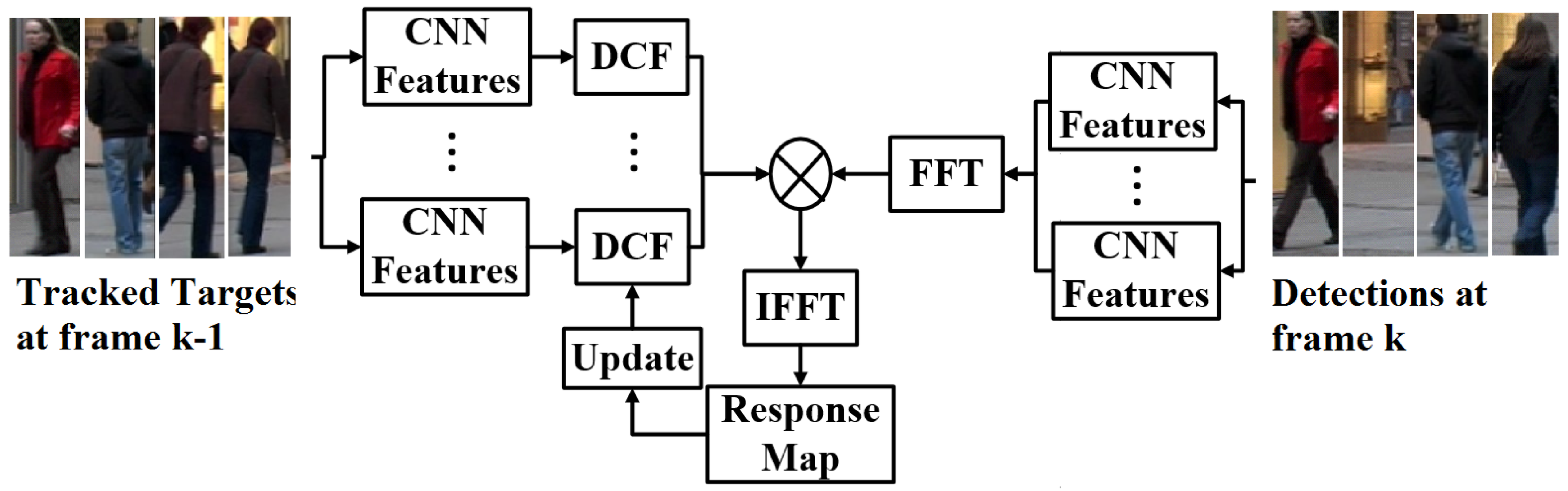
3. Detection Analysis and Spatio-Temporal Relation



- A spatio-temporal cost matrix $\mathbf{W}_{k,st} \in \mathbb{R}^{M\times N^+}$ for target association :

$$S(\mathbf{x}^i_{k|k-1}, \mathbf{z}^j_k) = \frac{1}{(2\pi\sigma_s)^{1/2}}\exp\left(-\frac{|\mathbf{H}\mathbf{x}^i_{k|k-1} - \mathbf{z}^j_k|}{2\sigma_s^2}\right) \quad (3)$$

4. Proposed Approach



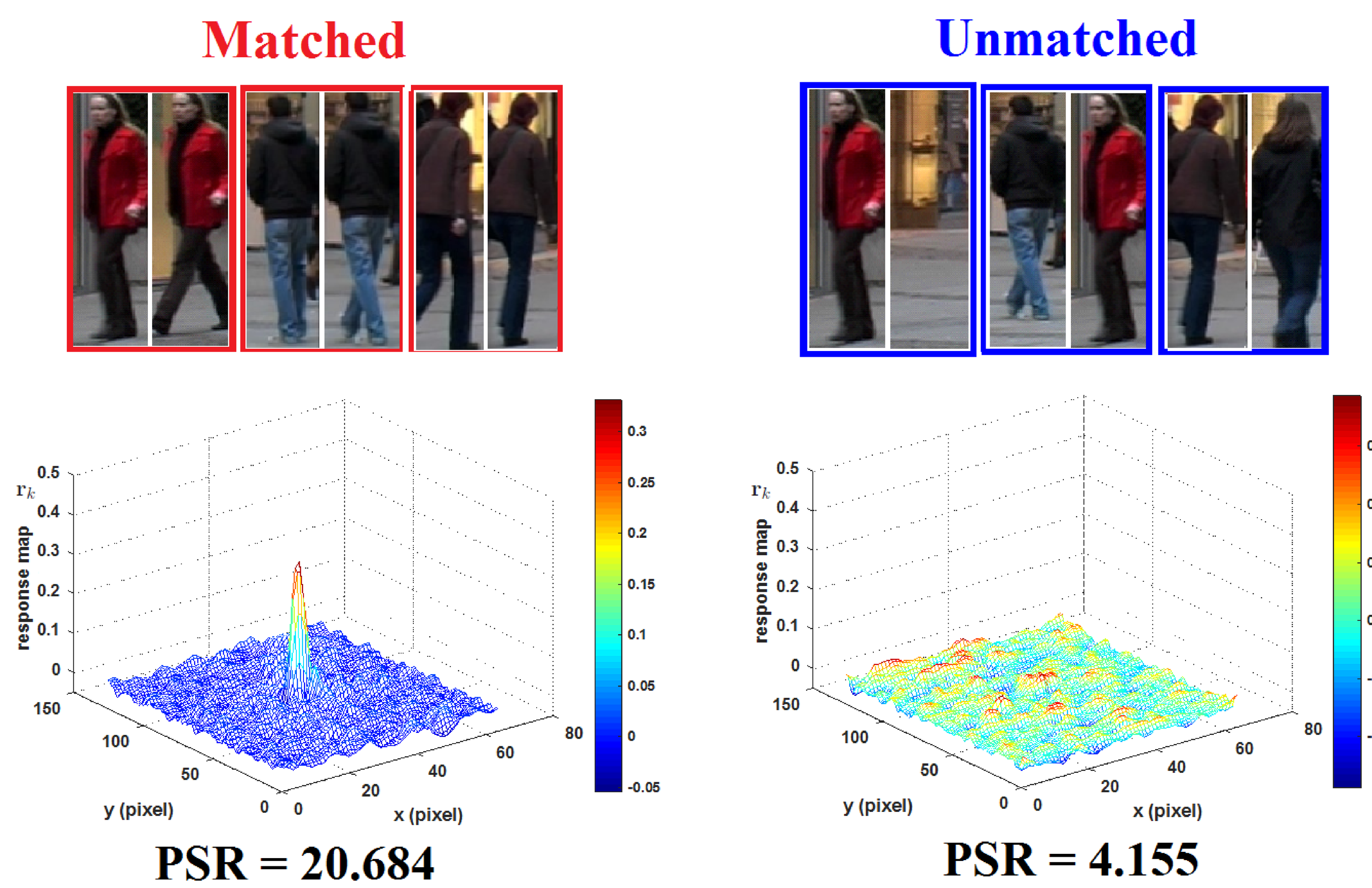
- **Training Phase:** Perform the fast Fourier transform (FFT) in the frequency domain with CNN features $\mathbf{f}$ and Gaussian label matrix $\mathbf{g}$ [2],

$$\hat{\mathbf{c}}^d_{k-1} = \frac{\hat{\mathbf{g}}\odot(\hat{\mathbf{f}}^d)^\dagger}{\sum_{d=1}^{D}\hat{\mathbf{f}}^d\odot(\hat{\mathbf{f}}^d)^\dagger + \lambda} \quad (4)$$

- **Correlation Matching:**
  - Response Map:

$$\mathbf{r}_k = \mathscr{F}^{-1}\left\{\sum_{d=1}^{D}\hat{\mathbf{c}}^d_{k-1}\odot(\hat{\mathbf{y}}^d_k)^\dagger\right\} \quad (5)$$

  - Pairwise matching score: $sigmoid(x) = \frac{1}{1+e^{-(\alpha x+\beta)}}$ squashes the PSRs to a range of $[0, 1]$. These scores form a cost matrix $\mathbf{W}_{k,dcm} \in \mathbb{R}^{M\times N^+}$.



- **Model Update**: Update the DCFs of matched targets during tracking for handling the appearance variations.
- **Hybrid likelihood function:**

$$\mathbf{W}_{k,h} = \mathbf{W}_{k,st}\odot\mathbf{W}_{k,dcm} \quad (6)$$

  Advantage: Compensate for unreliability present in the individual likelihood functions, especially when targets ambiguities occur in either motion dynamics or visual content.
- **Target initialization:** 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.

5. Experiments

- We evaluate on the MOTChallenge Benchmark [3] using the standard detections for all sequences.
- CLEAR MOT metrics are employed to evaluate the tracking performance.

Method	Mode	MOTA(↑)	MOTP(↑)	FP(↓)	FN(↓)	IDS(↓)
Proposed	**Online**	**46.5**	**77.2**	23,859	**272,430**	**5,649**
GMPHD-KCF	**Online**	40.3	75.4	47,056	283,923	5,734
GM PHD	**Online**	36.2	76.1	**23,682**	328,526	8,025
FWT	Offline	**51.3**	77.0	24,101	247,921	2,648
EDMT17	Offline	50.0	**77.3**	32,279	**247,297**	**2,264**
IOU17	Offline	45.5	76.9	19,993	281,643	5,988
DP_NMS	Offline	43.7	76.9	**10,048**	302,728	4,942

- **Quantitative results:**
  - Competitive performance compared to other state-of-the-art methods on the leaderboard.
  - Best performance amongst GM-PHD filtering methods.
- **Visual results:** MOT Benchmark 2017.

MOT17-03 #209 | MOT17-06 #183 | MOT17-07 #499 | MOT17-08 #171 | MOT17-14 #473 | MOT17-12 #03 | MOT17-01 #181

6. Conclusions and Future Work

- Developed a unified tracking algorithm that incorporates deep discriminative correlation matching with the GM-PHD filter for online multiple human tracking.
- Experimental Results on MOT17 Challenge demonstrate the effectiveness of the proposed method.
- We plan to integrate an interaction model to further address the occlusions.

7. References

[1] 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.
[2] C. Ma, J. B. Huang, X. Yang, and M. H. Yang, “Hierarchical Convolutional Features for Visual Tracking”, in ICCV, 2015, pp. 3074–3082.
[3] A. Milan, L. Leal-Taixe, I. Reid, S. Roth, and K. Schindler, “MOT16: A Benchmark for Multi-Object Trackings”, arXiv:1603.00831 [cs.CV], pp. 1–13, 2016.
[4] Z. Fu, P. Feng, F. Angelini, J. Chambers, and S. M. Naqvi, “Particle PHD Filter Based Multiple Human Tracking using Online Group-Structured Dictionary Learning”, IEEE Access, vol. 6, pp. 14764 - 14778, 2018.

VIDEO Tracking Results

VIDEO Multimodal Signal and Information Processing Team