A HIERARCHICAL FEATURE MODEL FOR MULTI-TARGET TRACKING

Mohib Ullah[1], Mohammed Ahmed [1], Faouzi Alaya Cheikh [1], Zhaohui Wang[2]

[1] Faculty of Computer Science and Media Technology,
Norwegian University of Science and Technology, Gjøvik, Norway.
[2] University of Hainan, Hainan, China.

NTNU

The Norwegian **Colour and Visual Computing** Laboratory

COLOURLAB.NO

ABSTRACT

- We propose a novel Hierarchical Feature Model (HFM) for multi-target tracking.
- Traditional approaches use local or global hand-crafted features to model the appearance of a target.
- In this work, we investigate deep features for modeling the appearance of the targets.
- Deep features are sparse coded for computational efficiency and a Bayesian filter is used to track the targets.

INTRODUCTION

- Multi-target tracking is a classical problem in computer vision. Huge effort is dedicated to this problem from the vision community and many breakthroughs have been achieved in the past few years. Nevertheless, it is still an open problem due to a number of challenges.
- Modeling the appearance of a target is one of the fundamental building blocks of tracking pipeline. We explored Google's architecture of a convolution neural network [1] to extract the feature for the target tracking. Moreover, the extracted features are sparse coded for the computational efficiency.
- Bayesian filtering is adopted for target tracking and combinatorial optimization is used for target association.

Applications:

- Video Surveillance
  (Security, crowd management, anomaly detection)
- Robotics
  (Autonomous vehicle, industrial automation)
- CGI & Gaming
  (Match moving)

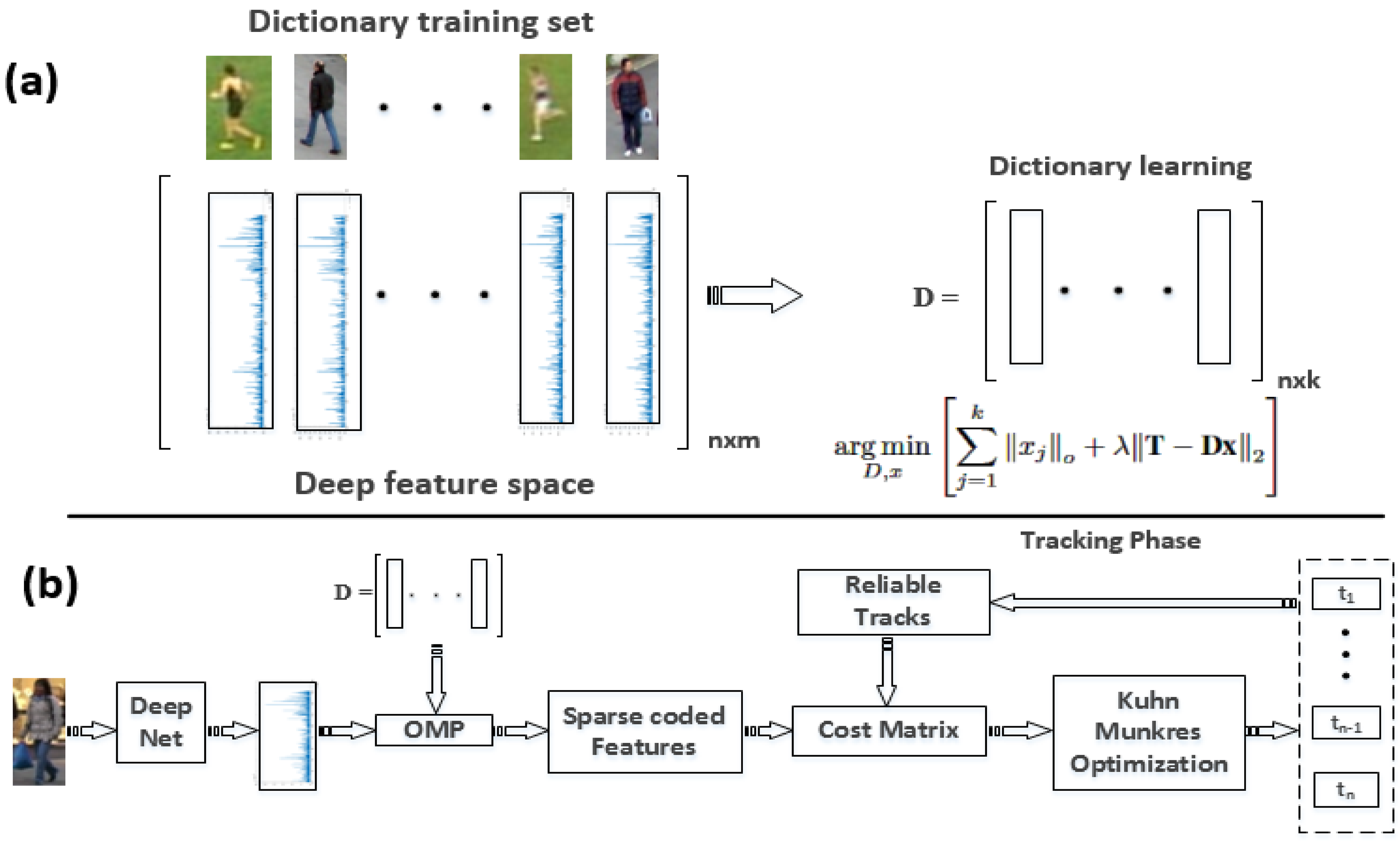
Challenges:

- Occlusions and background clutter
- Light intensity variation
- Complex target shape & articulation
- Unknown number of targets

METHOD

(a). For training the dictionary, targets are chosen in the first 100 frames of each dataset randomly. Selected target's feature vectors are used for dictionary learning.

(b). Feature vector of a target is sparse coded using the learnt dictionary and OMP algorithm. It is compared with all the targets in the previous frame and correct association is established through Kuhn Munkres algorithm. Afterwards, it is used as the observation model of the corresponding Bayesian filter.



RESULT

The proposed algorithm is tested on 4 benchmark video datasets and results are compared against four state-of-the-art methods. Standard performance matrices of MOTA and MOTP are used for calculating quantitative results.

$$MOTA = 1 - \frac{\sum_t c_f(fp_t) + c_m(fn_t) + c_s(mme_t)}{\sum_t g_t}$$

$$MOTP = \frac{\sum_i^t d_t^i}{\sum_t c_t}$$

Datasets	Methods	MOTA	MOTP
Pets2009	Berclaz et al. [2]	82.0%	56.0%
	Dehghan et al. [4]	90.4%	63.12%
	Andriyenko et al. [5]	89.3%	56.4%
	Proposed HFM	94.8%	74.6%
TUD-crossing	Dehghan et al. [3]	91.9%	70.0%
	Dehghan et al. [4]	92.9%	69.2%
	Proposed HFM	93.0%	72.8%
TUD-Stadmitt	Berclaz et al. [2]	45.8%	73.9%
	Dehghan et al. [3]	82.4%	73.9%
	Andriyenko et al. [5]	61.8%	63.2%
	Proposed HFM	89.1%	84.5%
AFL1	Proposed HFM	92.3%	87.7%

(a) Frame 182 (b) Frame 329 (c) Frame 559
(d) Frame 39 (e) Frame 92 (f) Frame 118
(g) Frame 24 (h) Frame 66 (i) Frame 118
(j) Frame 140 (k) Frame 171 (l) Frame 200

REFERENCES

1. Christian Szegedy et al. "Going deeper with convolutions," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1–9
2. Jerome Berclaz et al. "Multiple object tracking using k-shortest paths optimization," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 33, no. 9, pp. 1806–1819, 2011.
3. Afshin Dehghan et al. "Gmmcp tracker: Globally optimal generalized maximum multi clique problem for multiple object tracking," in IEEE Conference on Computer Vision and Pattern Recognition, 2015, pp. 4091–4099.
4. Afshin Dehghan et al, "Target identity-aware network flow for online multiple target tracking," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 1146–1154.
5. Anton Andriyenko et al, "Discrete-continuous optimization for multi-target tracking," in IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2012, pp. 1926–1933.

mohib.ullah@ntnu.no