Statistical and Spatial Consensus Collection for Detector Adaptation

Enver Sangineto

DISI, University of Trento, Italy
Enver.Sangineto@unitn.it

1. Motivations

Outcome of a generic pedestrian detector

Problem
- Unsupervised adaptation of a pedestrian detector to a new scenario

Drawbacks of Existing Methods
- Drifting
- Ad hoc heuristics

Contributions
- Transforming the outlier rejection problem in a classifier selection problem
- Spatially-dependent consensus collection of an ensemble of detectors

2. Main Idea

- A generic pedestrian detector is run on the target videos collecting candidate pedestrian samples (*T*)
- The target sample outlier rejection problem is transformed in a classifier selection problem
- Different classifiers are trained with different random subsets of *T*
- *S* (the source dataset) is used to estimate the accuracy of the classifiers
- The best performing classifiers are selected to form an ensemble
- The final ensemble decision is based on nearby detections of different classifiers

.

3. Ranking Candidate Pedestrians

- A generic detector is used to collect candidate target samples *B*
- *B* is ranked according to:

$$s(b) = \sum_{p \in S_P} ||f(b) - f(p)||_2^2$$

- *T* is obtained by discarding half of the ranked samples

4. Target Sample Selection

Given T, we look for the best T_G s.t.:

$$T_G = \arg\min_{T_i \subseteq T} E(C_{T_i}, \mathcal{D}^t)$$

Which is approximated by:

$$T_G = \arg\min_{T_i \subseteq T} L(C_{T_i}, S),$$

subject to:

$$C_{T_i} = \arg\min_{C \in \mathcal{C}} \mathcal{R}(C) + \theta\lambda(T_i, N_i)$$

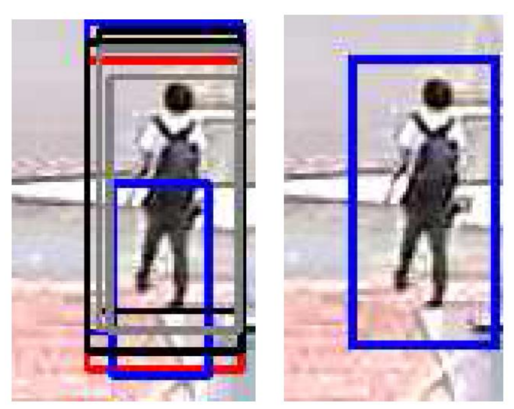
where N_i is a set of random negatives on D^t and:

$$L(C_i, S) = 1 - \frac{TP}{TP + FP}$$

5. RANSAC-like Solution

- T is split in randomly, partially overlapping subsets $T_1, \ldots, T_M$
- Each T_i is used together with a random set N_i to train a different weak classifier C_i
- Let $V = \{C_i\}$
- We associate each C_i in V with an error estimate $e_i = L(C_i, S)$
- Using $\{e_i\}$ we can rank V and select the k lowest error classifiers to build an ensemble E

6. Collecting Spatially-dependent Consensus

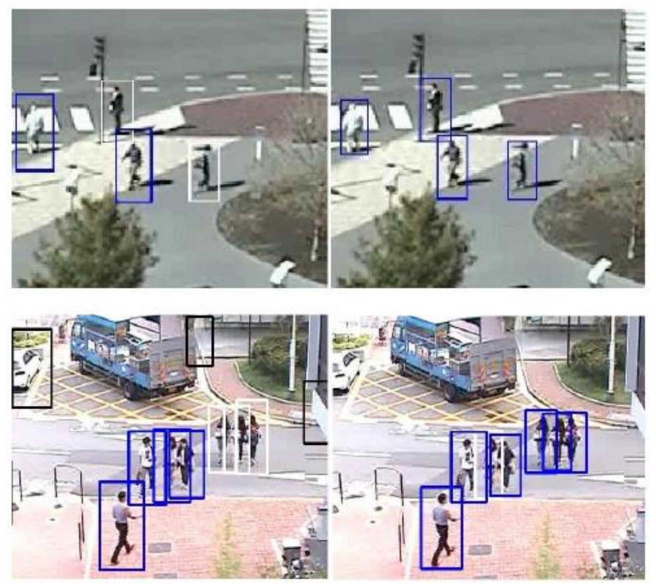


- Cluster all the positive windows of all the classifiers
- For each cluster G apply:

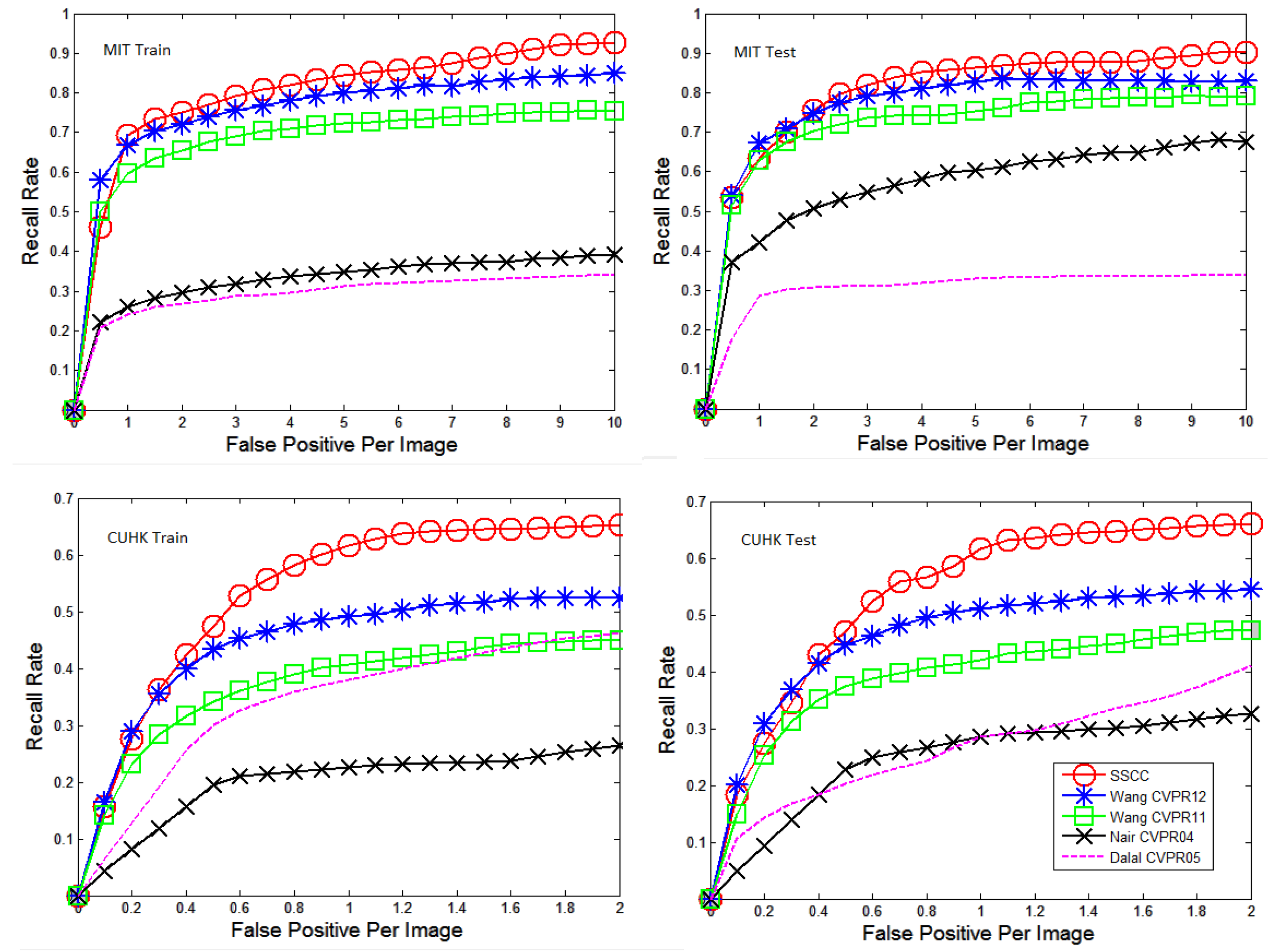
$$\mathcal{E}(G) = \mathbf{1}_{||\mathbf{v}_G||_0 > k/2}(G)$$

- Compute the average rectangle over G

7. Experiments



Some detection results of our system (right) and the Dalal's method (left) on the MIT Traffic dataset and the CUHK Square dataset.



Comparison of our system with other state-of-the-art pedestrian adaptation methods and with the generic detector

SEDR5	SSCC1	SSCC3	SSCC5	SSCC7	SSCC9	SSCC11
0.4845	0.4959	0.5118	0.5184	0.5183	0.5172	0.5175

AP over the CUHK Test with different ensemble cardinalities and decision rules

Precision	Recall	Error	Random	Random-1
0.5184	0.4750	0.4789	0.4494 (0.0232)	0.4382 (0.0408)

AP over the CUHK Test using different loss functions for classifier selection

References

[Wang CVPR12] Wang, M., Li, W., Wang, X.: Transferring a generic pedestrian detector towards specific scenes. In: *CVPR*, pp. 3274–3281 (2012)
[Wang CVPR11] Wang, M., Wang, X.: Automatic adaptation of a generic pedestrian detector to a specific traffic scene. In: *CVPR*, pp. 3401–3408 (2011)
[Nair CVPR04] Nair, V., Clark, J.J.: An unsupervised, online learning framework for moving object detection. In: *CVPR*, pp. 317–325 (2004)
[Dalal CVPR05] Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: *CVPR*, pp. 886–893 (2005)