Real-time Pedestrian Detection via Random Forest

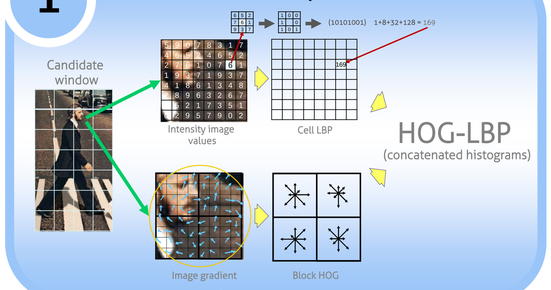
Javier Marín*, Sebastian Ramos*, David Vázquez*, Antonio M. López*, Jaume Amores*, Bastian Leibe[I] and Germán Ros*

*Advanced Driver Assistance Systems Group,
Universitat Autònoma de Barcelona, Spain

[I] Computer Vision Group,
RWTH Aachen University, Germany

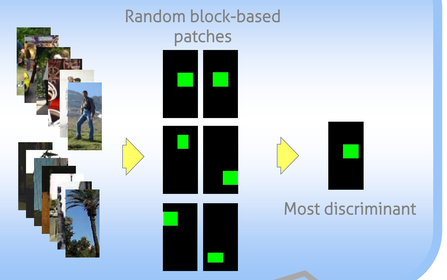
Random Forest Based on Local Experts [1]

1 Block-based representation



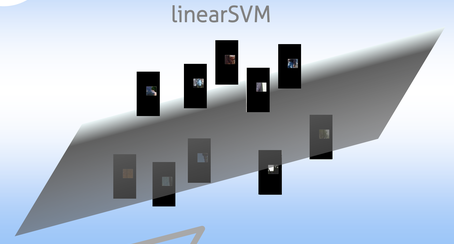
We use a richer representation than usual such as HOG-LBP [2].

2 Local patch selection



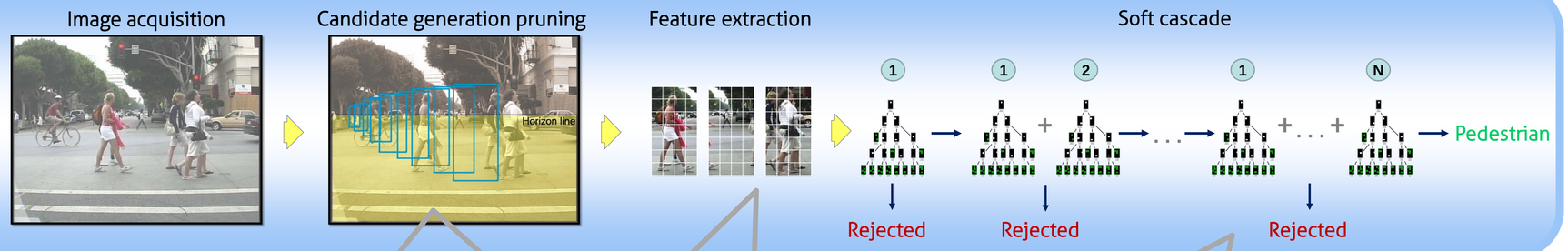
Each node of the tree automatically selects the local patch that is found to be the most discriminant one, based on the subset of samples received.

3 Weak learner optimization



The optimization at each node is carried out by a discriminant optimizer such as the linear SVM learner.

Real-Time Pedestrian detection

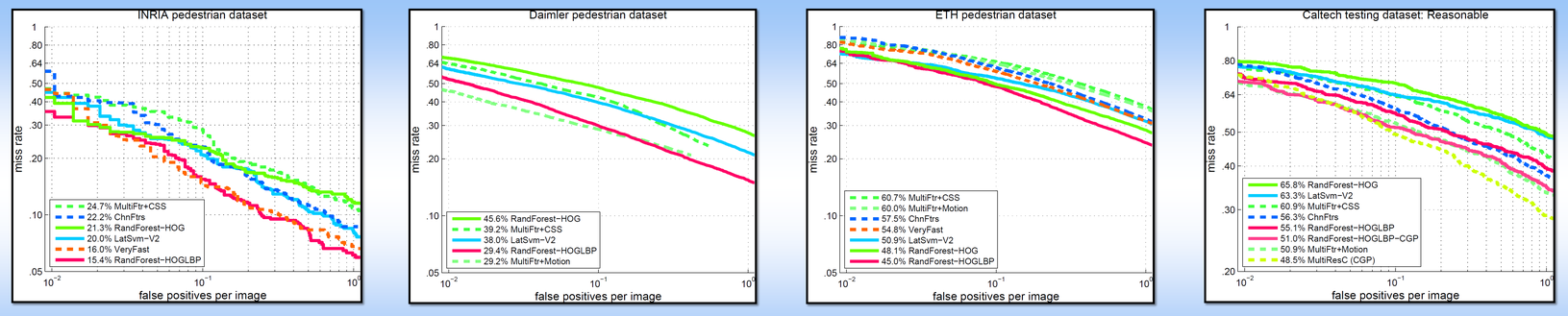


We remove all the candidate windows whose relationship vertical position vs scale is not physically plausible [3].

The block-based descriptor is reused by all the experts ot the ensemble.

Efficiency through a soft cascade architecture [4].

Performance Evaluation



Miss rate versus false positive per image curves in the INRIA, Daimler, ETH and Caltech testing. For the Caltech testing dataset we show results for the reasonable subset. We use the Caltech evaluation toolkit [5] to evaluate and compare our method with the state-of-the-art.

Summary

- A new approach based on local experts by means of a random forest
  - Efficiency comparable with the state-of-the-art
  - Ranked in the top positions in terms of accuracy in several datasets
  - Robust to partial occlusions
- Advantages:
  - No manual labeling is needed for the body parts
  - No semantic spatial components are required
  - No additional data coming from motion, multi-resolution or stereo
  - Can be easily extended to other objects

Time Comparison

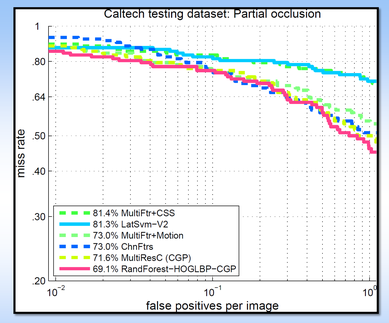
-x4 -x6-8

>=50 pixels	Baseline	SoftCascade (SC)	SC+CGC
HOG	0,15 fps	0,60 fps	1,23 fps
HOGLBP	0,09 fps	0,45 fps	0,93 fps

>= 96 pixels	Baseline	SoftCascade (SC)	SC+CGP
HOG	0,75 fps	2,51 fps	4,01 fps
HOGLBP	0,53 fps	1,88 fps	3,17 fps

Time comparison using Caltech dataset with multi-resolution pyramid set to 1.05 scale stride (C++ implementation). Using AVX instructions we reach 5,6 fps (HOG) and 4,6 fps (HOGLBP).

Partial Occlusions



Miss rate versus false positive per image curves in Caltech dataset results for the partial occlusion subset.

References

[1] Random Forests of Local Experts for Pedestrian Detection. In ICCV, 2013.
[2] An HOG-LBP Human Detector with Partial Occlusion Handling. In ICCV, 2009.
[3] Multiresolution models for object detection. In ECCV, 2010.
[4] Robust Object Detection via Soft Cascade. In CVPR, 2005.
[5] Pedestrian Detection: An Evaluation of the State of the Art. TPAMI, 2012.