Efficient 3D Object Detection using Multiple Pose-Specific Classifiers

Michael Villamizar[1]
mvillami@iri.upc.edu

Helmut Grabner[2]
grabner@vision.ee.ethz.ch

Juan Andrade-Cetto[1]
cetto@iri.upc.edu

Alberto Sanfeliu[1]
sanfeliu@iri.upc.edu

Luc Van Gool[2,3]
vangool@vision.ee.ethz.ch

Francesc Moreno-Noguer[1]
fmoreno@iri.upc.edu

[1] Institut de Robòtica i Informàtica Industrial, CSIC-UPC
Barcelona, Spain

[2] Computer Vision Laboratory
ETH Zurich
Switzerland

[3] ESAT - PSI / IBBT
K.U. Leuven
Belgium

We propose an efficient method for object localization and 3D pose estimation. We study the problem more closely on the detection of cars from multiple views. To this end, a two-step approach is used. In the first step, a pose estimator is evaluated in the input images to come up with potential object locations and pose candidates. These candidates are then validated in a second step, by the corresponding pose-specific classifier. The result is a detection approach that avoids the inherent and expensive cost of testing the complete set of pose-specific classifiers over the entire image. A further speedup is achieved from feature sharing. Features are computed only once and are then used to evaluate the pose estimator as well as all specific classifiers. These characteristics yield remarkable efficiency while keeping high detection rates. Fig. 1 depicts an overview of the method, whereas Fig. 2 shows some detection results and the corresponding estimated poses.

The method uses as features Random Ferns (RFs) [4] over local histograms of oriented gradients. These features contain simple comparisons with binary outputs that encode objects appearance and can be computed extremely fast. HOG-based RFs are shared among object classes and used both for pose estimation as well as for pose-specific classification. Unlike prior works that use Hough-based approaches as object classifiers [1, 2, 3], We propose Hough-RFs for building an efficient and robust 3D pose estimator. The pose estimator uses the Hough transform to learn and map the local appearances of objects (encoded by HOG-based RFs) into probabilistic votes for the object center. The methodology overcomes previous works which also compute rough estimators or predict the object size first [5, 7].

The proposed method has been validated on two public datasets for the problem of detecting cars under several views. The results show that our method yields high detection rates efficiently. In particular, the estimator and specific classifiers can be learned in a couple of minutes, whereas the object detection is performed in about 1 second, using non-optimized Matlab code. Fig. 3 shows detection rates over the UIUC car dataset [6]. They are comparable and even better than existing approaches.

[1] O. Barinova, V. Lempitsky, and P. Kohli. On detection of multiple object instances using hough transform. In *CVPR*, 2010.

[2] J. Gall and V. Lempitsky. Class-specific hough forests for object detection. In *CVPR*, 2009.

[3] S. Maji and J. Malik. Object detection using a max-margin hough transform. In *CVPR*, 2009.

[4] M. Ozuysal, P. Fua, and V. Lepetit. Fast keypoint recognition in ten lines of code, . In *CVPR*, 2007.

[5] M. Ozuysal, V. Lepetit, and P. Fua. Pose estimation for category specific multiview object localization, . In *CVPR*, 2008.

[6] S. Savarese and L. Fei-Fei. 3d generic object categorization, localization and pose estimation. In *ICCV*, 2007.

[7] M. Villamizar, F. Moreno-Noguer, J. Andrade-Cetto, and A. Sanfeliu. Efficient rotation invariant object detection using boosted random ferns. In *CVPR*, 2010.

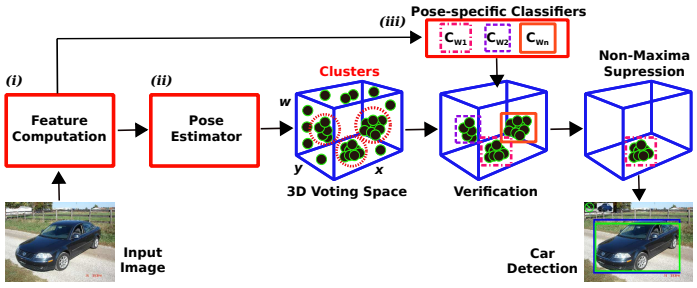

Figure 1: Overview of the proposed approach. To detect the 3D pose, given an input image we initially compute a set of shared RFs (Feature Computation). We then apply the pose estimator to generate several object/pose hypotheses which are verified by the pose-specific classifiers. Non-maximal potential detections are finally filtered out.

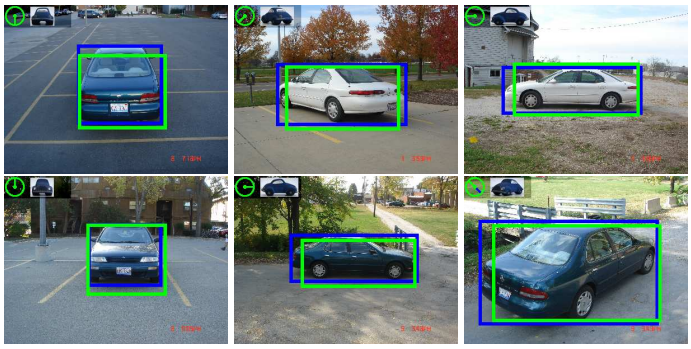

Figure 2: Efficient localization and pose estimation in the UIUC database [6]. Our approach allows localizing cars and estimating their pose despite large inter-class variations in about 1 second. Correct detections are depicted by green rectangles, and ground truth labels are indicated by blue rectangles. The circle and car toy located in the top left corner of each frame indicate the computed pose.

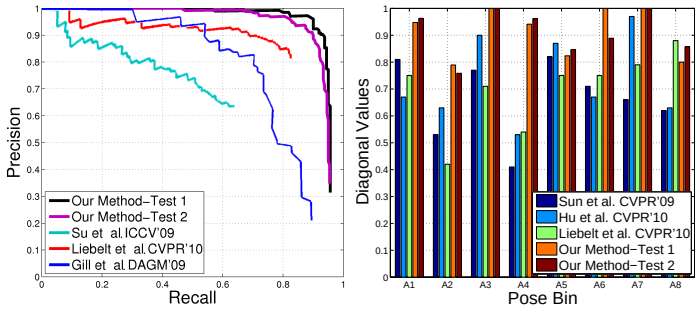

Figure 3: UIUC Dataset. Comparison against state of the art. **Left:** Detection rates using recall-precision plots. **Right:** Comparison in terms of the diagonal values of the confusion matrix.