Discriminative Hough Forests for Object Detection

Paul Wohlhart
wohlhart@icg.tugraz.at
Samuel Schulter
schulter@icg.tugraz.at
Martin Köstinger
koestinger@icg.tugraz.at
Peter M. Roth
pmroth@icg.tugraz.at
Horst Bischof
bischof@icg.tugraz.at

Institute for Computer Graphics and Vision
Graz University of Technology
Austria

Motivation and Method

Object detection models based on the Implicit Shape Model (ISM) [3] use small, local parts that vote for object centers in images. Since these parts vote completely independently from each other, this often leads to false-positive detections due to random constellations of parts. Thus, we introduce a verification step, which considers the activations of all voting elements that contribute to a detection. The levels of activation of each voting element of the ISM form a new description vector for an object hypothesis, which can be examined in order to discriminate between correct and incorrect detections.

In particular, we observe the levels of activation of the voting elements in Hough Forests [2], which can be seen as a variant of ISM. In Hough Forests, the voting elements are all the positive training patches used to train the Forest. Each patch of the input image is classified by all decision trees in the Hough Forest. Whenever an input patch falls into the same leaf node as a patch from training, a certain amount of weight is added to the detection hypothesis at the relative position of the object center, which was recorded when cropping out the training patch. The total amount of weight one voting element (offset vector) adds to a detection hypothesis (the total *activation*) can be calculated by summing over all input patches and trees in the forest. Stacking the activations of all elements gives an *activation vector* for a hypothesis.

We learn classifiers to discriminate correct and wrong part constellations based on these *activation vectors* and thus assign a better confidence to each detection. We use linear models as well as a histogram intersection kernel SVM. In the linear classifier, one weight is learned for each voting element. We additionally show how to use these weights, not only as a post processing step, but directly in the voting process. This has two advantages: First, it circumvents the explicit calculation of the activation vector for later reclassification, which is computationally more demanding. Second, the non-maxima suppression is performed on cleaner Hough maps, which allows for reducing the size of the suppression neighborhood and thus increases the recall at high levels of precision.

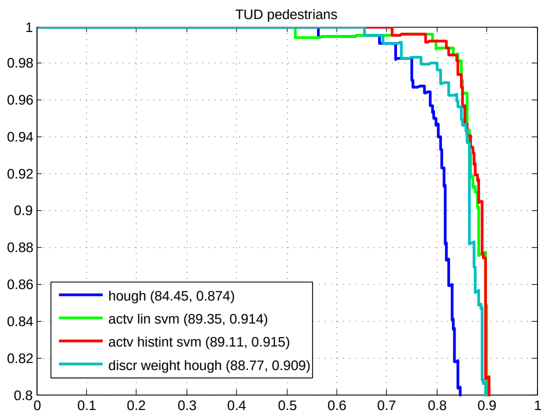

Figure 1: Precision/Recall curves on *TUD pedestrians* for standard Hough Forests [2] (*hough*), linear SVM on the activation vector (*actv lin svm*), histogram intersection kernel SVM on the activation vector (*actv histint svm*), and Hough voting with learned discriminative weights (*discr weights hough*).

Figure 2: Hough maps for an example test image from the TUD pedestrian dataset (top), with discriminative (middle) and uniform (bottom) weights

Results

The experiments on two different object classes, namely pedestrians [1] and cars [4], show significant improvements over the baseline. Visual inspection of the voting maps created with discriminatively learned voting weights (as shown for one test image in Figure 2) shows much cleaner backgrounds and clearly sharpened and pronounced peaks for correct locations. This is also reflected in the detection scores (see Figure 1 for results on *TUD pedestrians*).

This work was supported by the Austrian Science Foundation (FWF) project Advanced Learning for Tracking and Detection in Medical Workflow Analysis (I535-N23) and by the Austrian Research Promotion Agency (FFG) project SHARE in the IV2Splus program.

[1] M. Andriluka, S. Roth, and B. Schiele. People-tracking-by-detection and people-detection-by-tracking. In *Proc. CVPR*, 2008.

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

[3] B. Leibe, A. Leonardis, and B. Schiele. Combined object categorization and segmentation with an implicit shape model. In *Proc. ECCV Workshop on Statistical Learning in Computer Vision*, 2004.

[4] B. Leibe, N. Cornelis, K. Cornelis, and L. van Gool. Dynamic 3D scene analysis from a moving vehicle. In *Proc. CVPR*, 2007.