3D Pose-by-Detection of Vehicles via Discriminatively Reduced Ensembles of Correlation Filters

Yair Movshovitz-Attias[1]
www.cs.cmu.edu/~ymovshov
Vishnu Naresh Boddeti[2]
vishnu.boddeti.net
Zijun Wei[2]
hzwzijun@gmail.com
Yaser Sheikh[2]
www.cs.cmu.edu/~yaser/

[1] Computer Science Department
Carnegie Mellon University
Pennsylvania, USA

[2] Robotics Institute
Carnegie Mellon University
Pennsylvania, USA

Accurate estimation of the pose of a 3D model in an image is a fundamental operation in many computer vision and graphics applications, such as 3D scene understanding, inserting new objects into images, and manipulating current ones. One class of approaches to pose estimation is correspondence-based: individual parts of the object are detected, and a pose estimation algorithm (e.g., perspective-*N*-point) can be used to find the pose of the 3D object in the image. When the parts are visible, these methods produce accurate continuous estimates of pose. However, if the size of the object in the image is small or if the individual parts are not detectable (e.g., due to occlusion, specularities, or other imaging artifacts), the performance of such methods degrades precipitously. In contrast to correspondence-based approaches, pose-by-detection methods use a set of view-specific detectors to classify the correct pose; these methods have appeared in various forms such as filter banks, visual sub-categories, and exemplar classifier ensembles. While such approaches have been shown to be robust to many of the short-comings of correspondence-based methods, their primary limitation is that they provide discrete estimates of pose and as finer estimates of pose are required, larger and larger sets of detectors are needed.

Reduced representations are attractive because of their statistical and computational efficiency. Most approaches reduce the set of classifiers via the classic notion of minimizing the reconstruction error of the original filter set. Such a reduction does not directly guarantee optimal preservation of *detection* performance. This is particularly problematic in the case of viewpoint discrimination, as filters of proximal pose angles are similar. Reduction designed to minimize reconstruction error often results in a loss of view-point precision as the distinctive differences in proximal detectors are averaged out by the reduction.

In this paper, we present a pose-by-detection approach that uses an ensemble of correlation filters for precise viewpoint discrimination, by using a 3D CAD model of the vehicle to generate renders from viewpoints at the desired precision. A key contribution of this paper is a training framework that generates a discriminatively reduced ensemble of exemplar correlation filters by explicitly optimizing the detection objective. As the ensemble is estimated jointly, this approach intrinsically calibrates the ensemble of exemplar classifiers during construction, precluding the need for an after-the-fact calibration of the ensemble. The result is a scalable approach for pose-by-detection at the desired level of pose precision.

While our method can be applied to any object, we focus on 3D pose estimation of vehicles since cheap, high quality, 3D CAD models are readily available. We demonstrate results that outperform the state-of-the-art on the Weizmann Car View Point (WCVP) dataset, the EPFL car multi-view car dataset, and the VOC2007 car viewpoint dataset. We also report results on a new data-set based on the CMU-car dataset [1]. for precise viewpoint estimation and detection of cars. Fiigure 1 shows example results of our system on the WCVP dataset. Each row shows input images (top) and overlaid pose estimation results (bottom). Fiigure 2 These results demonstrate that pose-by-detection based on ensemble of exemplar correlation filters can achieve and exceed the level of precision of correspondence based methods in real datasets; and that discriminative reduction of an ensemble of exemplar classifiers allows scalable performance at higher precision levels.

[1] V. N. Boddeti, T. Kanade, and B. V. K. Vijaya Kumar. Correlation filters for object alignment. In *Computer Vision and Pattern Recognition*. IEEE, 2013.

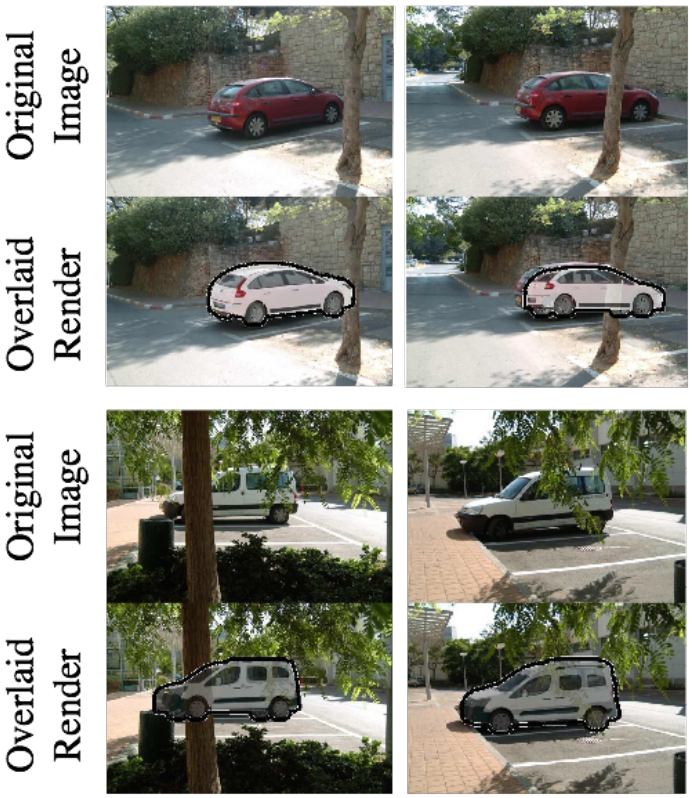

Figure 1: Example results. Each row shows input images (top) and overlaid pose estimation results (bottom).

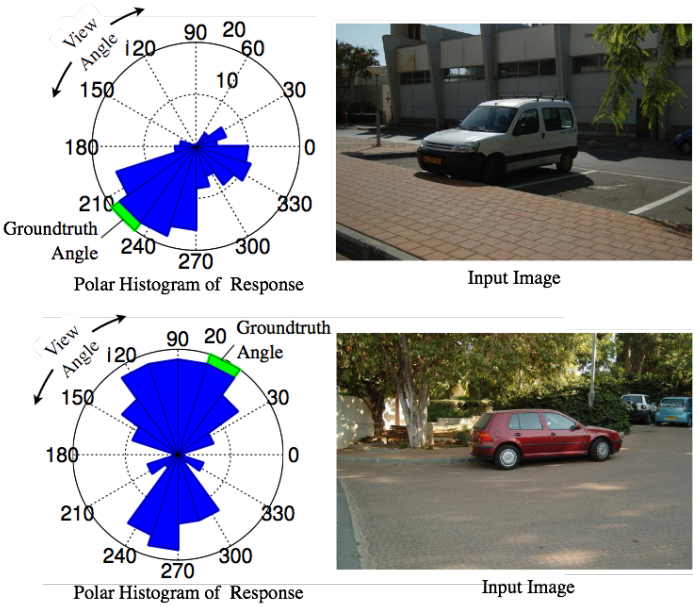

Figure 2: Polar histogram of scores. The left example shows a van at an oblique angle, with little ambiguity in the distribution of responses. The right example shows a side view with a distinctive symmetric ambiguity.