Scene-driven Cues for Viewpoint Classification of Elongated Object Classes

José Oramas M.
http://homes.esat.kuleuven.be/~joramas

Tinne Tuytelaars
http://homes.esat.kuleuven.be/~tuytelaa

KU Leuven, ESAT-PSI, iMinds
Leuven, Belgium

Motivation

Object viewpoint classification, also referred to as object pose estimation, is a task of interest for several applications. However, since the early days of computer vision, it has been addressed from a very "local" perspective. This perspective focuses on learning from the features on the object itself, e.g. color, texture, or gradients [1, 2], to identify the different viewpoints in which an object may appear in an image. Lately, this trend has been extended from reasoning about local visual properties of the object in the image space to properties in the 3D scene [3, 4, 5]. Despite the effectiveness of the mentioned methods, they have the weakness of ignoring scene-related cues that can assist the classification process.

Contributions

We complement existing work by exploiting scene-driven cues for object viewpoint classification. The main contributions of this work are:

- We exploit the orientation of the elongation of the object as a cue to estimate its viewpoint. For example, in Fig. 1a. even when we have no direct access to the local features of the object, we are able to predict, up to some level, the orientation of the object (Fig. 1b).
- We enforce scene-consistency in the viewpoint classification process by exploring specific regions of the scene that are more likely to host certain objects with particular features such as class, orientation or size. For example, note how the orientation of the objects in Fig. 1c is closely related to the regions of the scene in which they occur.

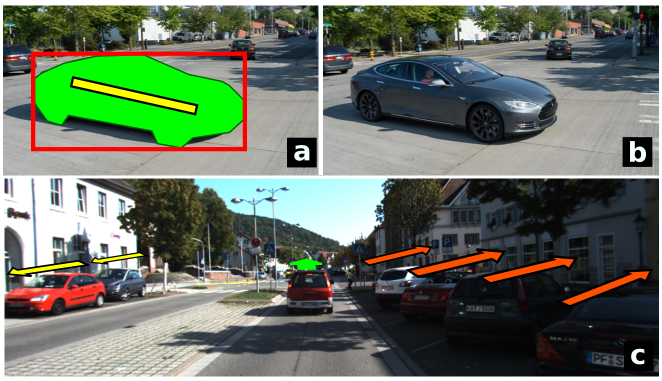

Figure 1: Note how the shape (a,b) and the location (c) of the bounding box of an object is related to its viewpoint.

Proposed method

Our method can be summarized in five steps (Fig. 2): First, we run a viewpoint-aware object detector to collect a set of hypotheses o_i. Then, we generate a set of scene-driven object proposals o'_i, Third, we estimate a correspondence descriptor d_i between each hypothesis o_i and its matching proposal o'_i. Then, we estimate the elongation orientation of the hypothesis o_i via multiclass classification of the descriptor d_i. Finally, the viewpoint of the objects is estimated by the fusion of the responses of the local detector and the elongation orientation classifier.

Figure 3: Viewpoint classification results encoded in jet scale. Continuous line, local detector prediction; Dashed line, scene-driven object proposals. Circle, ground-truth viewpoint (Best viewed in color).

Findings

Experiments on the KITTI object detection dataset show that:

- Considering scene-driven object elongation orientations brings improvements over purely appearance-based viewpoint-aware object detectors on the task of viewpoint classification (see Fig. 3).
- Our results based on 3D object proposals confirms the emerging consensus that coarse 3D scene-level reasoning, apart from context, is specially beneficial for these problems.
- This work complements very recent work, by sending the message that there are relatively simple cues in the scene that can bring improvements for the task of object viewpoint classification.

References

[1] R. J. Lopez-Sastre, T. Tuytelaars, and S. Savarese. Deformable part models revisited: A performance evaluation for object category pose estimation. In *ICCV WS*, 2011.

[2] M. Ozuysal, V. Lepetit, and P. Fua. Pose estimation for category specific multiview object localization. In *CVPR*, 2009.

[3] B. Pepik, M. Stark, P. Gehler, and B. Schiele. Teaching 3d geometry to deformable part models. In *CVPR*, 2012.

[4] Y. Xiang and S. Savarese. Object detection by 3d aspectlets and occlusion reasoning. In *3ddr@ICCV*, 2013.

[5] Z. Zia, M. Stark, and K. Schindler. Are cars just 3d boxes? - jointly estimating the 3d shape of multiple objects. In *CVPR*, 2014.

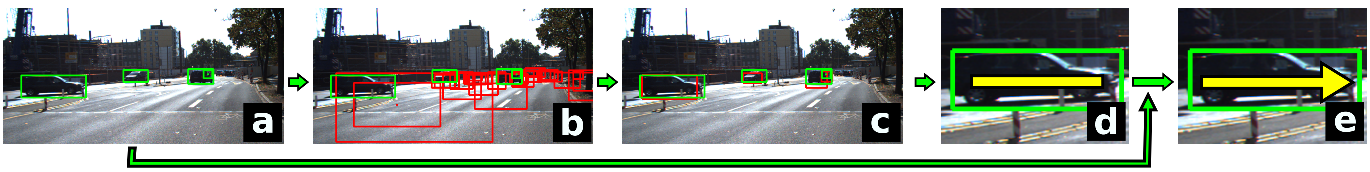

Figure 2: Algorithm Pipeline: a) Object Detection, b) Scene-driven Object Proposal Generation c) Object-hypotheses - Object-Proposal Matching, d) Elongation Classification, and e) Viewpoint Classification.