Class-Specific Object Pose Estimation and Shape Recovery

Arun CS Kumar
Department of Computer Science, The University of Georgia
{aruncs@uga.edu}

Abstract

Object Detection and Reconstruction, two of the core computer vision problems, despite significant advances, are still far from being solved. Owing to the ill-posed nature of the problem, approaches in the recent times try to solve them jointly, so that one can be used to resolve the ambiguities of the other. My research departs from the beaten path of using regression based models to a purely geometric approach to reason the object (as a set of parts) in 3D, and simultaneously detect, estimate continuous pose and recover the underlying 3D shape of the instance.

1. Introduction

Formally stated, given an image containing an object of a specific class, the goal is to simultaneously detect and reason the object in 3D, to recover its shape, from the given 2D appearance cues. The ill-posed nature of the problem leads to a complex solution landscape with several local minima. There are two general ways of tackling this problem, *part-based* approaches, where they model the object's shape as a set of part positions in 3D, or wholistic reasoning [3], where they train strong appearance classifiers (regressors) to categorize the viewpoint and shape of the object by modeling viewpoint and shape using a bank of appearance templates. Most of the recent approaches fall into one of these two categories. My research is focused on part based reasoning, which allows us to model the intra-class appearance and shape variation better, and reason the object in 3D, instead of using a bunch of 2D templates/representations. My work differs from existing part-based approaches in two ways. (1) Most part-based approaches [1, 2] use publicly available 3D CAD models to reason the object's part geometry. But manually generating CAD models for objects is expensive, painstaking to design, and are limited in their capability to capture the high levels of intra-class variation in terms of the object's appearance and shape geometry. Instead, we learn SfM based class-specific shape and appearance models from real image sequences. (2) Even existing part-based approaches view the problem of viewpoint/pose estimation a fine-grained categorization problem. Hejrati et. al. [4] reasons the underlying 3D shape, but uses a brute force approach to search object's pose and shape using appearance regressors, whereas [1] extends *Deformable Parts* model to 3D, and uses a cost function to penalize shape deformations, ultimately binning them into a fixed pose interval. Alternatively, we geometrically model the object, by accurately reasoning the physical projection of the underlying shape, to estimate the continuous viewpoint of the object.

2. Learning

Given a set of image sequences taken around an object of a specific class, our objective is to learn a deformable shape model that captures the object's appearance and 3D shape geometry using a set of 3D parts. Parts are, in general, semantically meaningful and salient regions (in 2D images), obtained via manual annotations or automated techniques. In [5], we use images taken around instances of *cars*, to compute 3D shape instances using SfM techniques. Using the 3D shapes (SfM) allows us to learn object's shape (as a set of 3D part positions) with a subspace spanned by basis shapes, along with the visibility of parts with respect to viewpoint, and appearances of parts (conditioned on viewing angle) from corresponding 2D images.

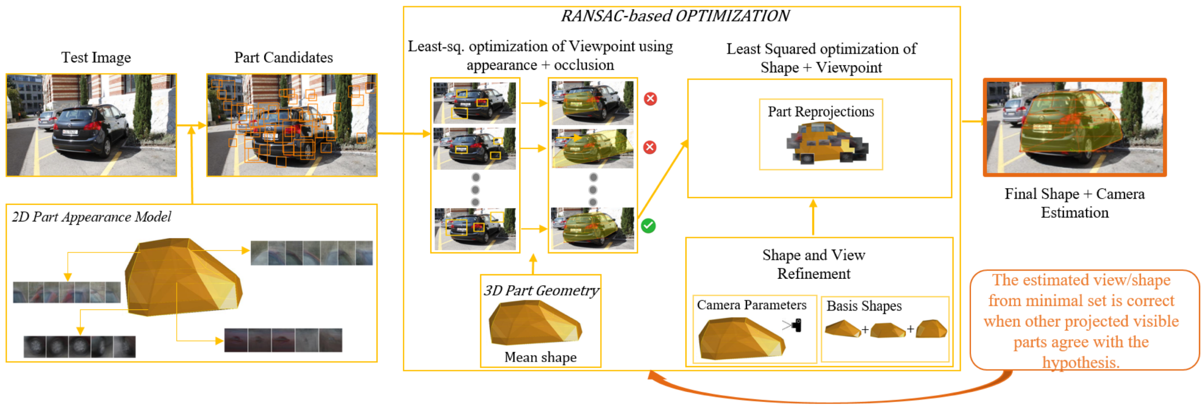

Figure 1. Pipeline of proposed object pose estimation and shape recovery technique from [5].

3. Pose Estimation and Reconstruction

At test time, we estimate the best deformation of the 3D object based on the 2D observations in the query image, using the learned appearances and shape deformations (as shown in Figure 1). For shape deformation, we assume any object shape is a linear combination of the basis shapes. For the query image, we use the learned appearance based part detectors (trained on deep features, using SVM classifiers) to jointly fit the deformable shape and compute the camera matrix, such that the fitted 3D shape model projects to corresponding part observations in the image. In [5], we follow a simple *Ransac*-based approach, where we first compute coarse shape and pose estimations using a minimal set of part detections, followed by refining the pose and shape iteratively by adding support. We also have extended this approach to learn parts automatically (in 3D) from 2D images, and have extended the Ransac-based approach to refine the 3D shape and viewpoint estimate, using other mid-level cues from the query image, than just part appearances (under submission).

4. Future Direction

The ultimate goal of my research is, to build an end-to-end pipeline, that learns a complete 3D representation of objects purely from 2D images, and for given a query image, geometrically reason the object and faithfully render a complete reconstruction. Also currently I use deep learning only for feature extraction, by fine-tuning pre-trained models for viewpoint classification tasks. In the future, I intend to develop end-to-end ConvNets combining image and 3D shape cues to learn a joint representation, for reconstructing objects at test time. Also, I aim to extend the ConvNets from discrete output prediction framework to more continuous viewpoint estimation framework, by reasoning the object's 3D shape and camera position.

References

[1] Pepik, B., Gehler, P., Stark, M., and Schiele, B. 3D^2PM - 3D deformable part models. *Proc. ECCV*, 2012. 1

[2] Pepik, B., Stark, M., Gehler, P., and Schiele, B. Teaching 3D geometry to deformable part models. *Proc. IEEE CVPR*, 2012. 1

[3] Tulsiani, S., Malik, J. Viewpoints and keypoints. *Proc. IEEE CVPR*, 2015. 1

[4] Hejrati, M., and Ramanan, D. Analysis by synthesis: 3D object recognition by object reconstruction. *Proc. IEEE CVPR*, 2014. 1

[5] Kumar, A.C.S., Bodis-Szomoru, A., Bhandarkar, S.M., & Prasad, M. Class-specific object pose estimation and reconstruction using 3D part geometry *Proc. ECCVW, Geometry Meets Deep Learning*, 2016. 1, 2