The Related Work of 6D Object Pose Estimation

Cheng Guan

May 31, 2018

1 Related work

The topic of object detection and pose estimation has been widely researched in the past decade. In the brief review below, we focus only on recent works and split them into three categories. We will omit the methods [2] since they were already discussed in the previous section.

Sampling-Based Methods. Sparse feature based methods [3] have shown good results for accurate pose estimation. They extract points of interest and match them based on a RANSAC sampling scheme. With the shift of the application scenario into robotics their popularity declined since they rely on texture. Shotton *et al.* [5] addressed the task of camera re-localization by introducing the concept of scene coordinates. They learn a mapping from camera coordinates to world coordinates and generate camera pose hypotheses by random sampling. Most recently Phillips *et al.* [1] presented a method for pose estimation and shape recovery of transparent objects where a random forest is trained to detect transparent object contours. Those edge responses are clustered and random sampling is employed to find the axis of revolution of the object. Instead of randomly selecting individual pixels we will use the entirety of the image to find pose hypotheses.

Non-Sampling-BasedMethods. An alternative to random sampling of pose hypotheses are Hough-voting based methods where all pixels cast a vote into a quantized prediction space (e.g. 2D object center and scale). The cell with the majority of votes is taken as the winner. Template have also been applied to the task of pose estimation. To find the best match the template is scanned across the image and a distance metric is computed at each position. Those methods are harmed by clutter and occlusion which disqualifies them to be applied toourscenario. In our approach each pixel is processed,but instead of them voting individually we find pose-consistent pixel-sets by global reasoning.

Pose Estimation using Graphical Models. In an older piece of work the pose of object categories was found in images either in 2D or in 3D. They also use the key concept of discretized object coordinates for object detection and pose estimation. The MRF-inference stage for finding pose-consistent pixels is closely related to ours. Foreground pixels are accepted when the layout consistency constraint (where layout consistency means that neighboring pixels should be long to the same part)is satisfied. However since the shape of the object is unknown, the pairwise terms are not as strong as in our case. The closest related work to ours is Bergholdt *et al.* [4]. They use the same strategy of discriminatively modeling the local appearance of object parts and globally inferring the geometric connections between them. To detect and find the pose of articulated objects (faces, human spines, human poses) they extract feature points locally and combine them in a probabilistic, fully-connected, graphical model. However they rely on a exact solution to the problem while a partial optimal solution is sufficient in our case. We therefore employ a different approach to solve the task.

References

[1] C.J.Phillips, M.Lecce, and K.Daniilidis. Seeing glassware: from edge detection to pose estimation and shape recovery. In *Robotics: Science and Systems*, 2016. 1

[2] E.Brachmann, A.Krull, F.Michel, J.Shotton, S.Gumhold, and C.Rother. Learning 6d object pose estimation using 3d object coordinates. 2014. 1

[3] I.Gordon and D.GLowe. *What and Where: 3D Object Recognition with Accurate Pose*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2006. 1

[4] M.Bergtholdt, J.Kappes, S.Schmidt, and C.Schnörr. A study of parts-based object class detection using complete graphs. *International journal of computer vision*, 87(1):93, 2009. 1

[5] J. Shotton, B. Glocker, C. Zach, S. Izadi, A. Criminisi, and A. Fitzgibbon. Scene coordinate regression forests for camera relocalization in rgb-d images. 2013. 1