The Method Overview of 6D Object Pose Estimation

Cheng Guan

June 3, 2018

1 Method Overview

Before we describe our work in detail, we will introduce the task of 6D pose estimation formally and provide a high-level overview of our method. The objective is to find the 6D pose $H_c = [R_c|t_c]$ of object$c$, with R_c (3×3 matrix) describing a rotation around the object center and t_c (31 vector) representing the position of the object in camera space. The pose H_c transforms each point in object coordinate space $y \in \chi \subseteq \mathbb{R}^3$ into a point in camera space $x \in \chi \subseteq \mathbb{R}^3$.

Our algorithm consists of three stages(see Fig. 1). In the first stage we densely predict object probabilities and object coordinates using a random forest. Instead of randomly sampling pose hypotheses as e.g. in [1] we use a graphical model to globally reason about hypotheses inliers. This second stage is described roughly and later in detail. In the final stage werefineand rank our pose hypotheses to determine the best estimate.

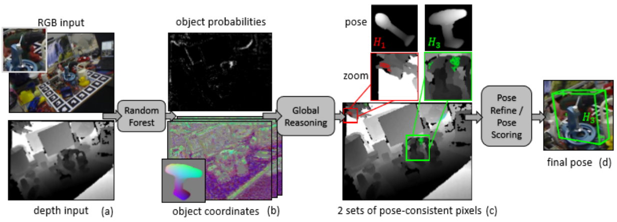

Figure 1: Our pipeline: Given an RGB-D image (a) a random forest provides two predictions: object probabilities and object coordinates (b). In a second stage our novel, fully-connected CRF infers pose-consistent pixel-sets (see zoom) (c). In the last stage, pose hypotheses given by pose-consistent pixels of the CRF are refined and scored by an ICP-variant. The pose with the lowest score is given as output (d).

1.1 Random Forest

We use the random forests from Brachmann *et al.* [1]. Each tree T of the forest τ predicts for each pixel an object probability and an object coordinate. As mentioned above, an object coordinate corresponds to a 3D point on the surface of the object. In our case we have $T = 3$. As in [1] the object probabilities from multiple trees that are combined to one value using Bayes rule. This means that for a pixel i and object c we have the object probability $p_c(i)$. The object probabilities can be seen as a soft segmentation mask.

1.2 Global Reasoning

In general, to estimate the pose of a rigid object, a minimal set of three correspondences between 3D points on the object and in the 3D scene is required . The 3D points on the object, *i.e.* in the object coordinate system, are predicted by the random forest. One possible strategy is to generate such triplets randomly by RANSAC [2], as proposed in [1]. However, this approach has a serious drawback: the number of triples which must be generated by RANSAC in order to have at least a correct triple with the probability of 95%, is very high. Assuming that n out of N pixels contain correct correspondences, the total number of samples is $\frac{\log(1-0.95)}{\log\left(1-(1-n/N)^3\right)}$. For $\frac{n}{N} = 0.05$, which corresponds to a state-of-the-artlocalclassifier,this constitutes 24.000.000 RANSAC iterations. Therefore, we address this problem with a different approach. Our goal is to assign to each pixel either one of the possible correspondence candidates, or an outlier label. We achieve this by formalizing a graphical model where each pixel is connected to every other pixel with a pairwise term. The pairwise term encodes a geometric check which is defined later.

1.3 Refinement and Hypothesis Scoring

The output of the optimization of the graphical model is a collection of pose-consistent pixels where each of those pixels has a unique object coordinate. The collection is clustered into sets. In the example in Fig. 1(c) there are two sets (red, green). Each set provides one pose hypothesis. These pose hypotheses are refined and scored using our ICP-variant. In order to be robust to occlusion we only take the pose-consistent pixels within the ICP [3] for fitting the 3D model.

References

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

[2] M. A. Fischler and R. C. Bolles. *Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography*. Elsevier, 1987. 1

[3] R. B. Rusu and S. Cousins. 3D is here: point cloud library. In *ICRA*, 2011. 1