Global Hypothesis Generation for 6D Object Pose Estimation

Cheng Guan

May 30, 2018

Abstract

This paper addresses the task of estimating the 6D pose of a known 3D object from a single RGB-D image. Most modern approaches solve this task in three steps: i) Compute local features; ii) Generate a pool of pose-hypotheses; iii) Select and rene a pose from the pool. This work focuses on the second step. While all existing approaches generate the hypotheses pool via local reasoning, e.g. RANSAC or Hough-voting, we are the rst to show that global reasoning is benecial at this stage. In particular, we formulate a novel fully-connected Conditional Random Field (CRF)thatoutputsaverysmallnumberofpose-hypotheses. Despite the potential functions of the CRF being nonGaussian, we give a new and efcient two-step optimization procedure, with some guarantees for optimality. We utilize our global hypotheses generation procedure to produce results that exceed state-of-the-art for the challenging Occluded Object Dataset.

1 Introduction

Thetaskofestimatingthe6Dposeoftexture-lessobjects has gained a lot of attention in recent years. From an application perspective this is probably due to the growing interestinindustrialrobotics, and invarious forms of augmented reality scenarios. From an academic perspective the dataset of Hinterstoisser *et al.* [4] marked a milestone, since researchers started to benchmark their efforts and progress in research started to be more measurable. In this work we focus on the following task. Given an RGB-D image of a 3Dscene, in which a known 3D object is present,i.e.its3D shape and appearance is known, we would like to identify the 6D pose (3D translation and 3D rotation) of that object.

Let us consider an exhaustive-search approach to this problem. We generate all possible 6D pose hypotheses,and for each hypothesis we run a robust ICP algorithm [2] to estimate a robust geometric t of the 3D model to the underlying data. The nal ICP score can then be used as the objective function to select the nal pose. This approach has two great advantages: (i)It considers all hypotheses;(ii) It uses a geometric error to prune all incorrect hypotheses. Obviously,this approach is in feasible from a computational perspective, hence most approaches generate rst a pool of hypotheses and use a geometrically motivated scoring function to select the right pose, which can be rened with robust ICP if necessary. Table 1 lists ve recent works with different strategies for hypotheses generation and geometric selection. The rst work by Drost *et al.* [1], and recently extended by Hinterstoisser *et al.* [3], has no geometric selection process,and generate savery large number of hypotheses. The pool of hypotheses is put into a Houghspace and the peak of the distribution is found as the nal pose.Despite its simplicity, the method achieves very good results, especially on the Occluded Object Dataset1, i.e. where objects are subject to strong occlusions. We conjecture that the main reason for its success is that it generates hypotheses from all local neighborhoods in the image. Especially for objects that are subject to strong occlusions, it is important to predict poses from as local information as possible. The other three approaches use triplets, and are all similar in spirit.

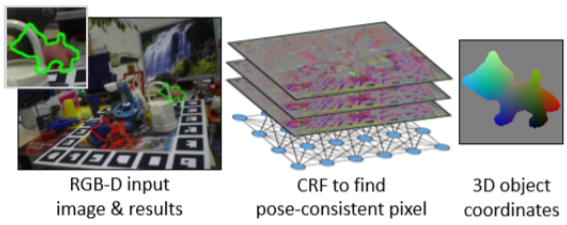

Figure 1: Given an RGB-D input image(left) we aim at nding the 6D pose of a given object, despite it being strongly occluded (see zoom). Here our result (green) is correct, while outputs an incorrect pose (red). The key concept of this work is to have a global, and hence powerful, geometric check, in the beginning of the pose estimation pipeline. This is in stark contrast to local geometric checks performed by all other methods.In a rst step, a random forest predicts for each pixel a set of three possible object coordinates, i.e. dense continuous part labeling of the object (middle). Given this, a fully-connected pairwise Conditional Random Field (CRF) infers globally those pixels which are consistent with the 6D object pose. We refer to those pixels as pose-consistent. The nal pose is derived from these pose-consistent pixels via an ICP-variant

In a first step they compute for everypixelone,or more, so-called object coordinates,a 3D continuous part-label on the given object (see Fig. 1 right). Then they collect locally triplets of points, in [33] these are all local triplets and they are randomly sampled with RANSAC. For each triplet of object coordinates they rst perform a geometry consistency check , and if successful, they compute the 6D object pose, using the Kabsch algorithm.

Method	Intermediate Representation	Hypotheses Generation	Average Number of Hypotheses	Hypotheses Selection	Hypotheses Renement	Run Time
Drost*et al.* [1]	Dense Point Pair Features	All local pairs	20.000	Sub-optimal search	ICP	0.4s
our	multiple object coordinates	Fully-connected CRF with geometric check	0-10	Optimal w.r.t. ICP variant	ICP variant	1-3s

Table 1: A broad categorization of six different 6D object pose estimation methods with respect to four different computational steps: (a) Intermediate representation, (b) Hypotheses generation, (c) Hypotheses selection, (d) Hypotheses renement, (e) Runtime.

References

[1] B.Drost, M.Ulrich, N.Navab, and S.Ilic. Model globally, match locally: Efcient and robust 3d object recognition. 2010. 1, 2

[2] P.J.Besl and N.D.McKay. A method for registration of 3-d shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239–256, 1992. 1

[3] S.Hinterstoisser, V.Lepetit, N.Rajkumar, and K.Konolige. Going further with point pair features. 2016. 1

[4] S.Hinterstoisser, V.Lepetit, S.Ilicand S.Holzer, G.R.Bradski, K. Konolige, and N. Navab. Model based training, detectionand poseestimationoftexture-less 3dobjectsinheavily cluttered scenes. 2012. 1