Detection and 3D Pose Estimation of New Objects in Color Images without Retraining

Giorgia Pitteri
giorgia.pitteri@u-bordeaux.fr
LaBRI, University of Bordeaux
Bordeaux, France

Vincent Lepetit
vincent.lepetit@u-bordeaux.fr
LaBRI, University of Bordeaux
Bordeaux, France

Slobodan Ilic
slobodan.ilic@siemens.com
Siemens AG
Munich, Germany

ABSTRACT

We present a novel approach to the detection and 3D pose estimation of industrial objects in color images that does not require a training phase for new objects. Only their untextured CAD models are required.

KEYWORDS

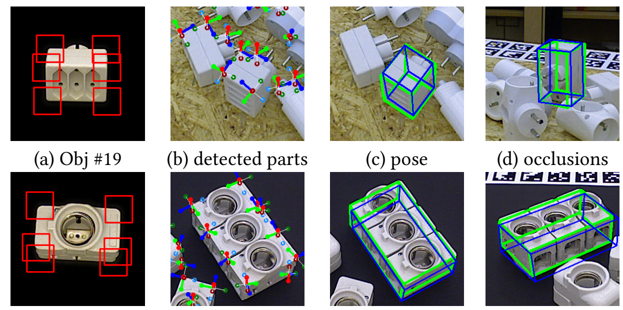
3D pose estimation, convolutional neural networks, CAD models.

1 INTRODUCTION

3D object detection and pose estimation are of primary importance for tasks such as robotic manipulation and augmented reality. Methods relying on depth data acquired by depth cameras are robust but unfortunately, active depth sensors are power hungry or sometimes it is not possible to use them. It is therefore often desirable to rely on color images, and many methods to do so have been proposed recently [3]. However, the success of these methods can be attributed to supervised Machine Learning approaches, and for each new object, these methods have to be retrained on many different images of this object. Even if domain transfer methods allow for training such methods with synthetic images instead of real ones, at least to some extent, such training sessions take time, and it is highly desirable to avoid them in practice.

2 METHOD

We propose a method that does not require additional learning nor training images for new objects. We consider a scenario where CAD models for the target objects exist, but not necessarily training images. We rely on a database of 3D parts which we learn to detect and estimate the 3D poses during an offline stage. Our approach focuses on industrial objects that are often made of similar parts. Among others, the most frequent parts shared among many objects are corners. However, they are arranged in different geometric configurations. Detecting these parts and determining their poses is the basis for our approach. We follow a deep learning approach [1] and train a network on training images depicting the parts of some objects from different viewpoints (Fig. 1 (a) (e)) to regress the 3D pose of each part in the form of a set of 2D reprojections of 3D virtual points (Fig. 1 (b) (f)). Given these 2D-3D correspondences for each part, the 3D pose of the object can then be computed by solving a P*n*P problem.



(a) Obj #19 (b) detected parts (c) pose (d) occlusions
(e) Obj #06 (f) detected parts (g) pose (h) occlusions

Figure 1: Qualitative results on TLESS test scenes #10 and #2. Our method can detect and estimate the 3D pose of new objects with very good precision even in case of occlusions (d) (h).

Problems may occur when dealing with industrial-type objects: Many ambiguities happen when trying to predict the 3D poses of a part from its appearance because of the symmetries of the parts, and similarities between parts from the same object or from different objects. Because of these ambiguities, it is difficult to know how to associate the 3D control points to the predicted reprojections. We solve this problem by handling them using a RANSAC-like robust pose estimation algorithm to iteratively find the good correspondences to get the best pose among all the hypotheses.

3 RESULTS

To the best of our knowledge, the problem of pose estimation of new objects that have not been seen at training time has not been addressed yet. To evaluate our method, we considered the challenging T-Less dataset [2]. We trained our network first on Object #19 and tested it on Object #20 in Scene #10, and on Object#6 and tested it on Object #7 in Scene #2, as shown in Fig. 1.

Conference'17, July 2017, Washington, DC, USA

ACM ISBN 978-x-xxxx-xxxx-x/YY/MM...$15.00
https://doi.org/10.1145/nnnnnnn.nnnnnnn

REFERENCES

[1] A. Crivellaro, M. Rad, Y. Verdie, K. Moo Yi, P. Fua, and V. Lepetit. 2018. Robust 3D Object Tracking from Monocular Images Using Stable Parts. (2018).
[2] T. Hodan, P. Haluza, S. Obdrzalek, J. Matas, M. Lourakis, and X. Zabulis. 2017. T-LESS: An RGB-D Dataset for 6D Pose Estimation of Texture-less Objects. In *IEEE Winter Conference on Applications of Computer Vision*.
[3] M. Rad and V. Lepetit. 2017. BB8: A Scalable, Accurate, Robust to Partial Occlusion Method for Predicting the 3D Poses of Challenging Objects Without Using Depth.