Technical Demonstration on Model Based Training, Detection and Pose Estimation of Texture-Less 3D Objects in Heavily Cluttered Scenes

Stefan Hinterstoisser[1], Vincent Lepetit[2], Slobodan Ilic[1], Stefan Holzer[1],
Kurt Konolige[3], Gary Bradski[3], and Nassir Navab[1]

[1] Department of Computer Science, CAMP,
Technische Universität München (TUM), Germany
{hinterst,slobodan.ilic,holzers,navab}@in.tum.de
[2] Ecole Polytechnique Federale de Lausanne (EPFL),
Computer Vision Laboratory, Switzerland
vincent.lepetit@epfl.ch
[3] Industrial Perception Inc., USA
{kurt,gary}@industrial-perception.com

Abstract. In this technical demonstration, we will show our framework of automatic modeling, detection, and tracking of arbitrary texture-less 3D objects with a Kinect. The detection is mainly based on the recent template-based LINEMOD approach [1] while the automatic template learning from reconstructed 3D models, the fast pose estimation and the quick and robust false positive removal is a novel addition.

In this demonstration, we will show each step of our pipeline, starting with the fast reconstruction of arbitrary 3D objects, followed by the automatic learning and the robust detection and pose estimation of the reconstructed objects in real-time. As we will show, this makes our framework suitable for object manipulation e.g. in robotics applications.

1 Introduction

Many current vision applications, such as pedestrian tracking, dense SLAM [2], or object detection [1], can be made more robust through the addition of depth information. In this work, we focus on object detection for Robotics and Machine Vision, where it is important to efficiently and robustly detect objects and estimate their 3D poses, for manipulation or inspection tasks. Our approach is based on LINEMOD [1], an efficient method that exploits both depth and color images to capture the appearance and 3D shape of the object in a set of templates covering different views of an object. Because the viewpoint of each template is known, it provides a coarse estimate of the pose of the object when it is detected.

However, the initial version of LINEMOD [1] has some disadvantages. First, templates are learned online, which is difficult to control and results in spotty

A. Fusiello et al. (Eds.): ECCV 2012 Ws/Demos, Part III, LNCS 7585, pp. 593–596, 2012.

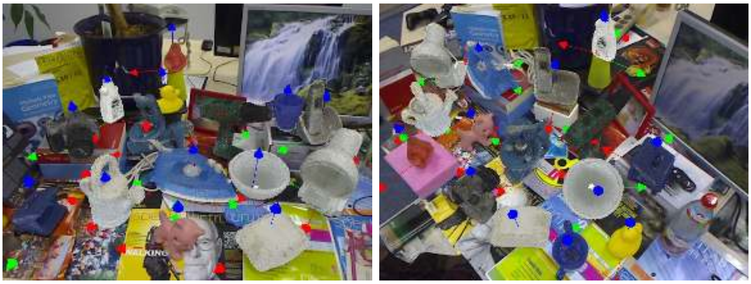

Fig. 1. 15 different texture-less 3D objects are simultaneously detected with our approach under different poses on heavy cluttered background with partial occlusion. Each detected object is augmented with its 3D model. We also show the corresponding coordinate systems. **See supplemental video.**

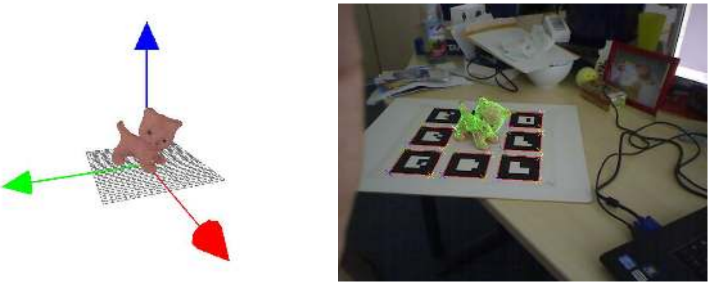

Fig. 2. In this figure, we show the simple reconstruction of the "cat" model. On the left hand side, we see the reconstructed 3D model whereas on the right hand side we augment the real object with the reconstruction.

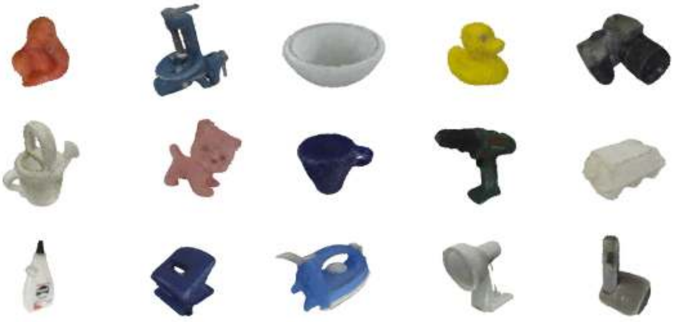

Fig. 3. The reconstructed 3D models of the 15 different texture-less 3D objects detected in Figs. 1 and 4.

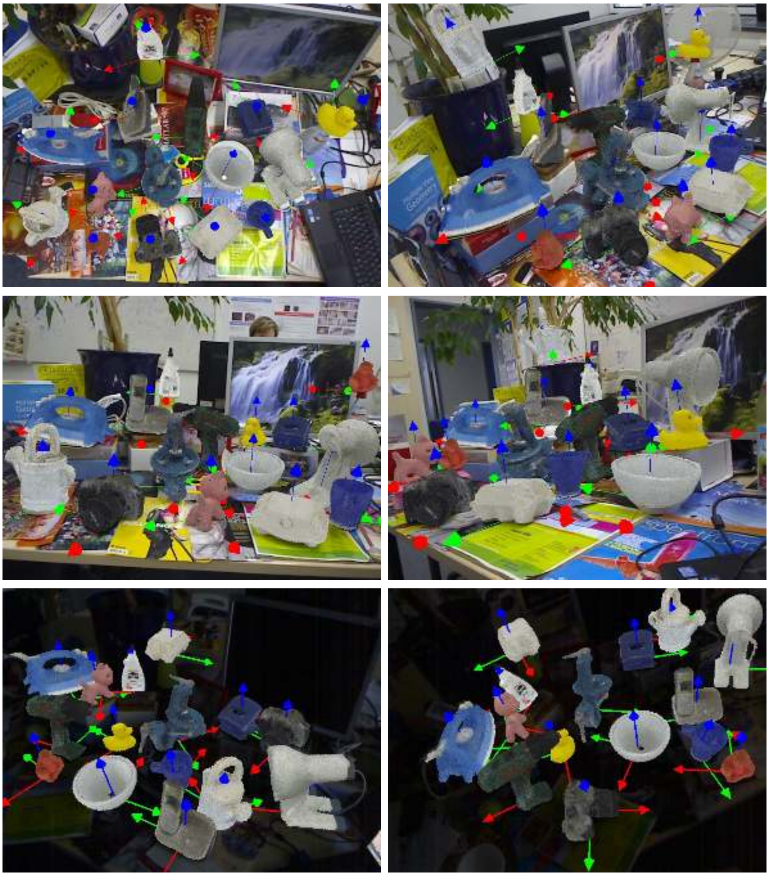

Fig. 4. 15 different texture-less 3D objects are simultaneously detected under different poses on heavy cluttered background with partial occlusion and illumination changes. Each detected object is augmented with its 3D model. We also show the corresponding coordinate systems. **See also the supplemental video.**

coverage of viewpoints. Second, the pose output by LINEMOD is only approximately correct, since a template covers a range of views around its viewpoint. And finally, the performance of LINEMOD, while extremely good, still suffers from the presence of false positives.

In this technical demonstration, we show the result of our most recent work [3] where we overcome these disadvantages, and create a system based on LINEMOD for the automatic modeling, detection, and tracking of 3D objects with RGBD sensors. Our main insight is that a 3D model of the object can be exploited to remedy these deficiencies. Note that accurate 3D models can now be created very quickly [2,4,5,6], and requiring a 3D model beforehand is not a disadvantage anymore. For industrial applications, a detailed 3D model often exists before the real object is even created. In this demonstration, we will use our own model creation framework to show that an appropriate 3D model can be created very easily and very quickly.

Given such a 3D model of an object, we will show the automatic template learning where templates are generated which cover a full view hemisphere of regularly sampled viewpoints of the 3D model. During this learning, the templates are defined only with the most useful appearance and depth information, which allows us to speed up the template detection stage. In addition, we will also show in this demonstration that the 3D model can be used to obtain a fine estimate of the object pose, starting from the one provided by the templates. Together with a simple test based on color, this allows us to remove false positives, by checking if the object under the recovered pose aligns well with the depth map. The end result is a system that significantly improves the original LINEMOD implementation in performance, while providing accurate pose for applications.

To show the computational efficiency of our framework we will perform our technical demonstration on a standard notebook with an Intel i7-2820QM processor with 2.3 GHz and 8 GB of RAM.

In short, we will demonstrate a 3D reconstruction, detection and pose estimation framework that is easy to deploy, reliable, and fast enough to run in real-time.

References

1. Hinterstoisser, S., Holzer, S., Cagniart, C., Ilic, S., Konolige, K., Navab, N., Lepetit, V.: Multimodal Templates for Real-Time Detection of Texture-Less Objects in Heavily Cluttered Scenes. In: ICCV (2011)
2. Newcombe, R.A., Izadi, S., Hilliges, O., Molyneaux, D., Kim, D., Davison, A.J., Kohli, P., Shotton, J., Hodges, S., Fitzgibbon, A.: KinectFusion: Real-Time Dense Surface Mapping and Tracking. In: ISMAR (2011)
3. Anonymous, Authors: Anonymous Title. In: submitted to ACCV (2012)
4. Pan, Q., Reitmayr, G., Drummond, T.: ProFORMA: Probabilistic Feature-based On-line Rapid Model Acquisition. In: BMVC (2009)
5. Weise, T., Wismer, T., Leibe, B., Van Gool, L.: In-hand Scanning with Online Loop Closure. In: International Workshop on 3-D Digital Imaging and Modeling (2009)
6. Newcombe, R.A., Lovegrove, S.J., Davison, A.J.: DTAM: Dense Tracking and Mapping in Real-Time. In: ICCV (2011)