

BOP: Benchmark for 6D Object Pose Estimation

Tomáš Hodaň¹, Frank Michel², Eric Brachmann³, Wadim Kehl⁴
Anders Glent Buch⁵, Dirk Kraft⁵, Bertram Drost⁶, Joel Vidal⁷, Stephan Ihrke²
Xenophon Zabulis⁸, Caner Sahin⁹, Fabian Manhardt¹⁰, Federico Tombari¹⁰
Tae-Kyun Kim⁹, Jiří Matas¹, Carsten Rother³

¹CTU in Prague, ²TU Dresden, ³Heidelberg University, ⁴Toyota Research Institute
⁵University of Southern Denmark, ⁶MVTec Software, ⁷Taiwan Tech
⁸FORTH Heraklion, ⁹Imperial College London, ¹⁰TU Munich

Abstract. We propose a benchmark for 6D pose estimation of a rigid object from a single RGB-D input image. The training data consists of a texture-mapped 3D object model or images of the object in known 6D poses. The benchmark comprises of: i) eight datasets in a unified format that cover different practical scenarios, including two new datasets focusing on varying lighting conditions, ii) an evaluation methodology with a pose-error function that deals with pose ambiguities, iii) a comprehensive evaluation of 15 diverse recent methods that captures the status quo of the field, and iv) an online evaluation system that is open for continuous submission of new results. The evaluation shows that methods based on point-pair features currently perform best, outperforming template matching methods, learning-based methods and methods based on 3D local features. The project website is available at bop.felk.cvut.cz.

1 Introduction

Estimating the 6D pose, i.e. 3D translation and 3D rotation, of a rigid object has become an accessible task with the introduction of consumer-grade RGB-D sensors. An accurate, fast and robust method that solves this task will have a big impact in application fields such as robotics or augmented reality.

Many methods for 6D object pose estimation have been published recently, e.g. [34,24,18,2,36,21,27,25], but it is unclear which methods perform well and in which scenarios. The most commonly used dataset for evaluation was created by Hinterstoisser et al. [14], which was not intended as a general benchmark and has several limitations: the lighting conditions are constant and the objects are easy to distinguish, unoccluded and located around the image center. Since then, some of the limitations have been addressed. Brachmann et al. [1] added ground-truth annotation for occluded objects in the dataset of [14]. Hodaň et al. [16] created a dataset that features industry-relevant objects with symmetries and similarities, and Drost et al. [8] introduced a dataset containing objects with reflective surfaces. However, the datasets have different formats and no standard evaluation methodology has emerged. New methods are usually compared with only a few competitors on a small subset of datasets.

Authors have been leading the project jointly.

