

Occlusion Dataset

March 20, 2018

All data regarding our ECCV 14 paper can be downloaded from our project page: <https://hci.iwr.uni-heidelberg.de/vislearn/research/scene-understanding/pose-estimation/#ECCV14>. If you run into problems contact: eric <dot> brachmann <at> tu-dresden.de.

1 Overview

This dataset contains additional annotations for the dataset of Hinterstoisser et al.[1]. The original dataset contained RGB-D images of multiple scenes. In each scene various objects were present, but only ground truth poses for one object were given. We took the images of one scene (*Benchwise*) and annotated the poses of 8 additional objects. These objects are heavily occluded in some images of the scene. Annotation was done by manual registration of a 3D model of the object with the image, followed by ICP. Poses were propagated through the sequence using the original ground truth. Sometimes the scene layout changes because objects are moved. In this case, we repeated manual registration.

We introduced this dataset in our ECCV14 paper[2]. If you use this dataset, please cite the aforementioned paper.

2 Structure

The dataset is structured as follows: At the top level there are 8 folders, one for each additional object we annotated. Each object folder contains one sub-folder named `info` with the annotation data. Each annotation data item is named after the following scheme:

```
info_<image number>.txt
```

Each file is a text file and comes in two variants depending on whether the object is visible in the associated image or completely occluded. If the object is completely occluded, the annotation file has the following content:

```
image size  
<iw> <ih>
```

If the object is visible the annotation file looks like this:

```

image size
<iw> <ih>
<sequence name>
rotation:
<r1> <r2> <r3>
<r4> <r5> <r6>
<r7> <r8> <r9>
center:
<t1> <t2> <t3>
extent:
<ow> <oh> <od>

```

Image width `<iw>` and image height `<ih>` are measured in pixels, and are always 640 resp. 480. The rotation and center entries are combined into transformation $T_{o \rightarrow c}$ in the following way:

$$T_{o \rightarrow c} = \begin{bmatrix} r1 & r2 & r3 & t1 \\ r4 & r5 & r6 & t2 \\ r7 & r8 & r9 & t3 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (1)$$

$T_{o \rightarrow c}$ maps 3D coordinates in the object coordinate system (i.e. *object coordinates*) to 3D coordinates in the camera coordinate system. This is the object pose. All coordinates are assumed to be measured in meters. The last three entries `<ow>`, `<oh>` and `<od>` represent object width, object height and object depth, respectively. They are measured in meters.

References

- [1] Hinterstoisser, S., Lepetit, V., Ilic, S., Holzer, S., Bradski, G., Konolige, K., , Navab, N.: Model based training, detection and pose estimation of texture-less 3d objects in heavily cluttered scenes. (2012)
- [2] Brachmann, E., Krull, A., Michel, F., Gumhold, S., Shotton, J., Rother, C.: Learning 6d object pose estimation using 3d object coordinates. In Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T., eds.: Computer Vision – ECCV 2014. Volume 8690 of Lecture Notes in Computer Science. Springer International Publishing (2014) 536–551