Occlusion Dataset

July 31, 2015

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 where present, but only ground truth poses for one object were given. We took the images of one scene (Benchvise) 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. Poses where 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 the our ECCV14 paper[2]. If you use this dataset, please cite the aforementioned paper.

2 Strucutre

The dataset contains the following folders:

- models - 3D models of the kinematic chains
- poses - ground truth pose annotations
- RGB-D - color and depth frames

2.1 Models

We provide the models of the 8 additional objects present in the Benchvise scene of [1]. At the top level of the models folder are 8 subfolders. Each subfolder contains the Wavefront OBJ file of the object. The coordinates of the mesh are measured in meters and the center is in the middle of the bounding volume. The texture is attachted as a RGB-triple per vertex.
We also provide point cloud files (*.xyz) of each object. Each line contains `<x> <y> <z>` of one object vertex. The coordinates are also measured in meters with the center lying in the middle of the bounding volume.

2.2 Poses

We provide the ground truth pose information for 8 additional objects within the Benchvise scene. At the top level of the poses folder there are 8 subfolders, one for each additional object. The text files contain both the pose and the object size which are named after the following scheme:

```
<info>_<image number>_<object index>
```

The data is structured as follows:

```
image size
<iw> <ih>
<object index>
rotation:
<r1> <r2> <r3>
<r4> <r5> <r6>
<r7> <r8> <r9>
center:
<t1> <t2> <t3>
extent:
<oh> <ow> <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 to transformation $\mathrm{T}_{o\to c}$ in the following way:

$$\mathrm{T}_{o\to c}= \begin{bmatrix} r1 & r2 & r3 & t1 \\ r4 & r5 & r6 & t2 \\ r7 & r8 & r9 & t3 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

$\mathrm{T}_{o\to c}$ maps 3D coordinates in the object coordinate system (i.e. object coordinates) to 3D coordinates in the camera coordinate system. Note that the camera viewing direction is the negative Z-axis. All coordinates are assumed to be measured in meters. The last three entries <ow>, <oh> and <od> represent object with, object height and object depth, respectively. They are measured in meters.

2.3 RGB-D

We provide the color and depth frames of the Benchvise scene.

2.3.1 depth_noseg

These folders contain depth images. Each image is a 1 channel 16 bit (unsigned short) PNG. The depth values are stored in millimeters. A depth value of 0

means missing depth.

2.3.2 rgb_noseg

These folders contain rgb images. Each image is a 3 channel 8 bit (unsigned char) PNG.

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