Tracking Dataset

April 7, 2015

All data regarding our ACCV 14 paper can be downloaded from our project page: `http://cvlab-dresden.de/research/scene-understanding/pose-estimation/#ACCV14`. If you run into problems contact: alexander <*dot*> krull <*at*> tu-dresden.de.

1 Overview

We recorded and annotated six sequences showing three different objects. For each frame in each sequence we provide: a depth image, a rgb image, and the pose of the object. Additionally we provide a 3D model of the objects.

We introduced this dataset in the our paper [1]. If you use this dataset, please cite the aforementioned paper.

2 Structure

The dataset is structured as follows: At the top level there are six folders corresponding to the six sequences. Each each sequence folder contains three subfolders named: `rgb_noseq` containing rgb images, `depth_noseq` containing depth images, and `info` containing the pose annotation. Additionally we provide in every sequence folder a mesh file as well as a point cloud file of the object.

3 RGB Images

The sub folders `rgb_noseq` contain rgb images. Each image is a 3 channel 8 bit (unsigned char) PNG file.

4 Depth Images

The sub folders `depth_noseq` contain depth images. Each image is a single channel 16 bit (unsigned short) PNG file. The value at a pixel corresponds to the depth in millimetres. A depth value of 0 means missing depth.

5 Pose Annotation

The files in the info sub folders are named info_<image number>.txt.

Their contents 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 $T_{o \to c}$ in the following way:

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

$T_{o \to c}$ 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.

5.1 Mesh File (.obj)

This Wavefront OBJ file contains a mesh of the object. The coordinates of the mesh file are measured in meters, and live in the object coordinate system.

5.2 Point Cloud File (.xyz)

This file contains a point cloud of the object. Each line contains <x> <y> <z> of one 3D point. The coordinates of the points are measured in meters and live in the object coordinate system.

6 Calculation of Additional Information

This section explains how to derive additional information from depth data.

6.1 Camera Coordinates

For each pixel of an RGB-D image you can calculate the X_c, Y_c, Z_c-coordinates of the pixel in the camera coordinate system in meters:

$$X_c = (x - \frac{w_i}{2})\frac{d}{1000f} \tag{2}$$

$$Y_c = -(y - \frac{h_i}{2})\frac{d}{1000f} \tag{3}$$

$$Z_c = \frac{-d}{1000} \tag{4}$$

Where x and y are pixel coordinates measured from the top left image corner, and d is the associated depth in millimetres. Image width w_i and image height h_i are 640 and 480, respectively. We use a focal length f of 575.816px. The factor 1000 in all three equations converts millimetres in meters.

References

[1] Alexander Krull, Frank Michel, Eric Brachmann, , Stefan Gumhold, Stephan Ihrke, and Carsten Rother. 6-dof model based tracking via object coordinate regression. In *Proceedings of the Asian Conference on Computer Vision*, ACCV, 2014.