Articulated Objects Dataset

July 31, 2015

This document provides information regarding the dataset released with our Pose Estimation of Kinematic Chain Instances via Object Coordinate Regression, BMVC 2015 paper. The data can be downloaded from our project page: http://cvlab-dresden.de/research/scene-understanding/pose-estimation/#BMVC15. Should issues arise contact: frank <dot> michel <at> tu-dresden.de.

1 Overview

Our dataset consists of 4 articulated objects. Each of them can be modeled as a kinematic chain consisting of parts and interconnecting joints. The joints are limited to 1 degree of freedom (revolute, prismatic) which allows 1 DOF rotations and 1 DOF translations. We provide 3D models for each of the 11 individual parts of the articulated objects and text files describing the topology of the underlying kinematic chain structure. We recorded 8 test sequences (2 for each articulated object) using a standard RGB-D (Microsoft Kinect) camera. The articulation of the objects change between the sequences but stays constant within one sequence.

If you use this dataset, please cite the aforementioned paper.

2 Strucutre

The dataset contains the following folders:

- configs - text files describing the topology of the kinematic chains
- models - 3D models of the kinematic chains
- test - test sequences including pose annotations
- train - poses used to render training data

2.1 Configurations

Each subfolder contains topological information of one kinematic chain object. Within each subfolder exists one text file per part. This text file provides the following information:

```
Object
<name>
Extent
<ow> <oh> <od>
Pivot_point
<pX_pre> <pY_pre> <pZ_pre>
Articulate_rotation
<rX_pre> <rY_pre> <rZ_pre>
Articulate_translation
<tX_pre> <tY_pre> <tZ_pre>
Articulation_min_max
<art_min_pre> <art_max_pre>
Pivot_point
<pX_suc> <pY_suc> <pZ_suc>
Articulate_rotation
<rX_suc> <rY_suc> <rZ_suc>
Articulate_translation
<tX_suc> <tY_suc> <tZ_suc>
Articulation_min_max
<art_min_suc> <art_max_suc>
```

The Extent describes the width, height and the depth of the object. To keep the meta information consistent each part of the kinematic chain has two joints, one for the predecessor and one for the successor. Those joints are represented as follows. The pivot point, <pX_> <pY_> <pZ_>, describes the position of the joint within the coordinate system of the part. The parameters Articulated_rotation and Articulate_translation determine the type and the axis of the articulation. Since we only consider 1 DOF joints only one of those 6 values is set to one and all the others are set to 0.
Finally the Articulation_min_max represents the minimum and maximum articulation for the joint. The parameters Extent and Pivot_point are given in meters. Articulation_min_max is given in meters for prismatic joints and in degrees for revolute joints. In case of one part being the first or the last part of the kinematic chain one joint is not existent which is expressed by setting all 6 (Articulated_rotation, Articulate_translation) values of that joint to 0.

2.2 Models

Each subfolder contains the Wavefront OBJ files of the articulated objects. The coordinates of the mesh are measured in meters and the center is in the middle

of the bounding volume. For all models except the Toy train we provide high resolution models with a higher point density which where used for training only. Those models do also have a color 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.

2.3 Test Sequences

The test sequences are structured als follows: At the top level of the test folder there are 8 folders, one for each sequence recorded (4 articulated objects x 2 sequences per object). The sequence folders are named according to the following sceme:

```
<object_name>_Seq_<sequence_number>
```

Each sequence folder contains 3 sub-folders with the actual sequence data. Each sub-folder is explained below.

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.

2.3.3 info

These folders contain the annotated ground truth information for all parts of the articulated object. The Annotations are text files containing both the pose and the object size which are named after the following scheme:

```
info_<image number>_<part 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.

2.4 Training data

We provide the pose information we used to create the training data. 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:

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

$\mathrm{T}_{o\rightarrow c}$ 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.