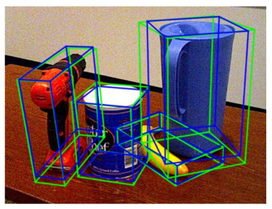
Large-scale Synthetic Domain Randomized 6DoF Object Pose Estimation Dataset

Mona Jalal[1], Josef Spjut[2], Ben Boudaoud[2] , David Luebke[2] , Margrit Betke[1]

[1] Boston University, [2] NVIDIA

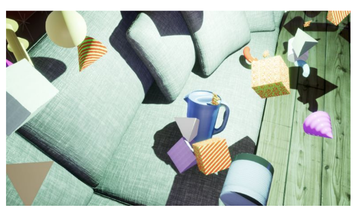
Abstract


Object pose estimation is a very important problem in domains such as robotics manipulation and augmented reality; however, 3D bounding box annotation of objects is a very expensive and laborious task and ground-truth labels are often just an approximation. In this work, we have created a large-scale object pose estimation dataset which makes use of **domain randomization** techniques such as use of extreme lighting and flying distractors for both single objects as well as multiple object interactions. The ground truth annotations are precise and are created by an proprietary version of the NVIDIA Deep Learning Data Synthesizer.

NVIDIA Deep Learning Data Synthesizer

- A plugin for **Unreal Engine 4** to create synthetic dataset. It is very fast (50-100HZ) and provides the **following annotations**:
  - 3D object pose via 3D bounding box for each annotated object
  - Projected 3D bounding boxes
  - 2D bounding boxes
  - Instance segmentation mask
  - Class segmentation mask
  - Depth map
  - Percentage of object visibility from camera



Extreme Lighting and Flying Distractors Example

It provides **various randomization** such as:

- Random camera movement
- Random object pose
- Random texture
- Camera path following via anchor points

Our Dataset Annotations

Synthetic Objects:

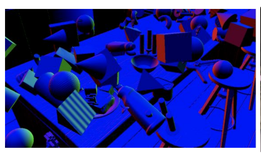
21 synthetic objects used in creating the dataset from YCB

Left Image

Left Class Segmentation

Left Object Mask

Left Depth Map



Left Surface Normal Map

Right Image

Our dataset captured by a stereo camera has ground-truth maps for class segmentation, object mask, 8bit and 32bit depth, and surface normal in PNG format for both left and right images and also includes JSON file for synthetic object settings, camera settings as well as each frame annotation for 2D/3D bounding box, 3D projected bounding box, and orientation of objects.

Consider the following Domino Sugar Box

JSON Annotations

```
"bounding_box":
{
  "top_left": [ 196.25709533691406, 646.3275146484375 ],
  "bottom_right": [ 374.356201171875, 778.64520263671875 ]
},
```

```
"cuboid": [
    [ 17.469699859619141, -4.7862000465393066, 69.09429931640625 ],
    [ 26.722200393676758, -4.5423002243041992, 68.623497009277344 ],
    [ 26.770500183105469, 10.703000068664551, 77.468399047851563 ],
    [ 17.517999649047852, 10.459099769592285, 77.939201354980469 ],
    [ 17.72760009765625, -7.0480999946594238, 72.991500854492188 ],
    [ 26.980199813842773, -6.8042001724243164, 72.520698547363281 ],
    [ 27.028400421142578, 8.4411001205444336, 81.365600585937500 ],
    [ 17.775899887084961, 8.197199821472168, 81.836402893066406 ]
],
```

Visualizing the 3D cuboids using NVIDIA Dataset Utilities toolset

Dataset	# objects	# frames	depth	stereo	3D pose	Extreme lighting	segmentation	BBox coord	Flying distractors
LINEMOD	15	18k	✓		✓				
T-LESS	30	10k	✓		✓				
YCB Video	21	134k	✓		✓		✓	✓	
Falling Things	21	60k	✓	✓	✓	✓	✓	✓	
Ours	21	112k	✓	✓	✓	✓	✓	✓	✓

Comparison with other Object Pose Estimation Datasets