PRINCETON VISION GROUP

Sliding Shapes for 3D Object Detection in Depth Images

Shuran Song Jianxiong Xiao

Code+Data Available
http://slidingshapes.cs.princeton.edu

Motivation

Object detection is difficult due to various reasons :

1. Shape variance

Use a data-driven approach to leverage a collection of CG models that cover the space of shape variance.

2. Viewpoint variance

Densely render different viewpoints to cover all typical viewing angles.

3. Inter-object occlusion

Use depth to identify occluded part, and use 3D sliding window to separate the object from the occluder.

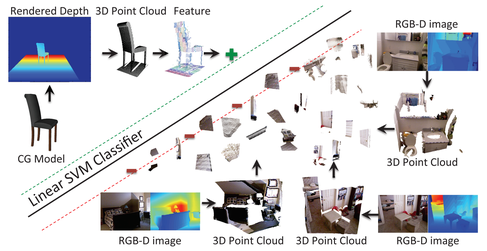
4. Clutter

Use a 3D mask to indicate which parts should be considered during classification.

5. Texture and illumination variance

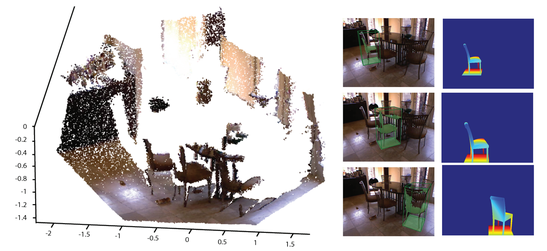
Use depth maps instead of color images.

Training



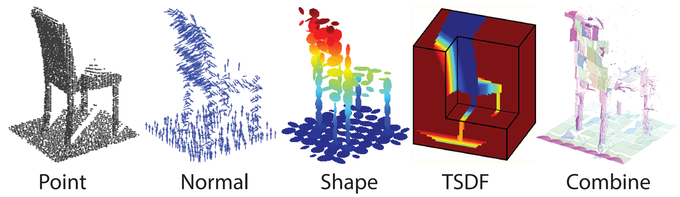
For each object category, we collect the CAD models from the Internet, and for each CAD model, we render depth map from hunderds of viewpoints. For each rendering, a feature vector is extracted from the point cloud and used to train an exemplar SVM using negative data from a RGB-D data set.

Testing



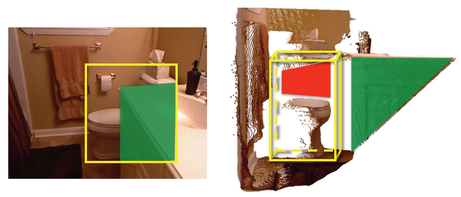
During testing we exhaustively classify each box in the 3D space as chair vs. nonchair, using the Exemplar-SVMs.

Features



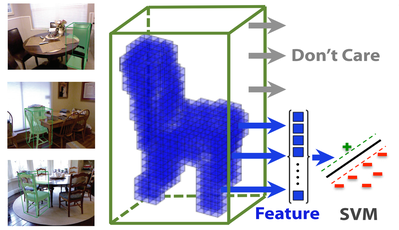
To capture properties of 3D objects we design the above features and combine all of them, forming a discriminative descriptor.

Occlusion



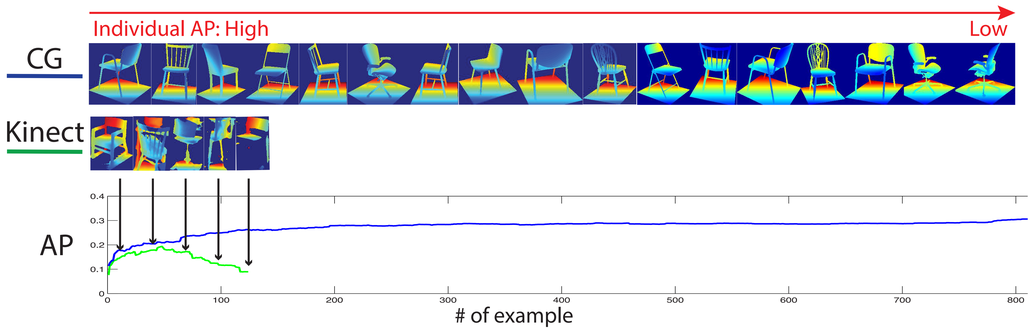
Using depth, we can know which part is occluded (red). In 2D, the occluder (green) will be inside the box and confuse the detector, while in 3D, the object is naturally separated from the occluder.

Occupation Mask



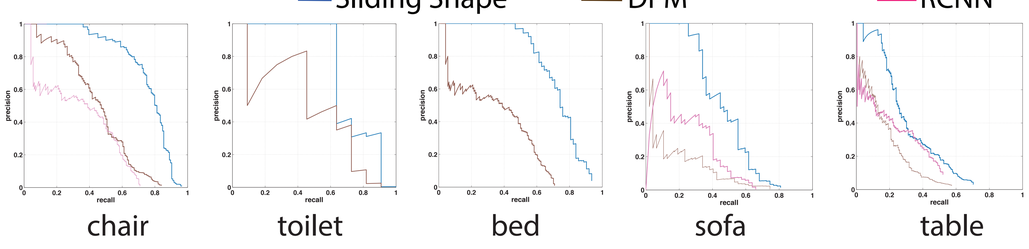
To handle clutter we use an occupation mask for each training CG model and only use the features inside the mask to do classification.

Data: CG vs. Kinect



We sort the Exemplar-SVMs by their individual AP. Adding more exemplars trained from CG rendering always help to improve the performance, but adding more exemplars trained from Kinect data increases the AP at first, then AP starts to drop when bad data are added.

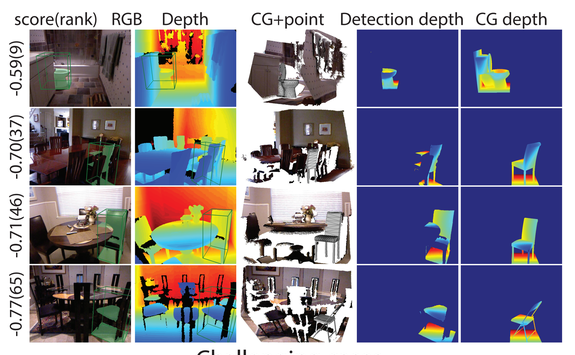
Evaluation



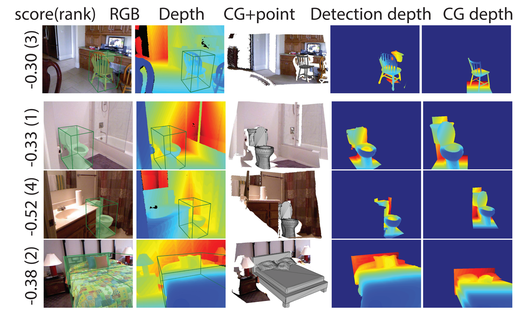
We evaluate our algorithm on five object categories on NYU depth v2. Compare to the best of DPM and RCNN, we achieve around x1. 7 improvement on AP.

Results

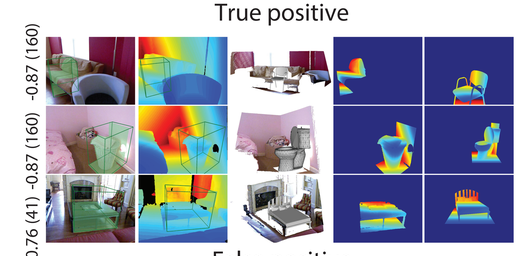
Multiple detections in one image



Challenging cases



True positive



False positive

Miss