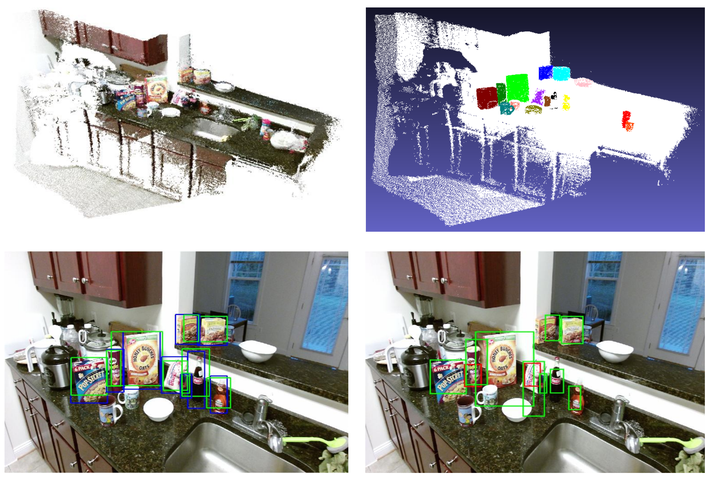
Multiview RGB-D Dataset for Object Instance Detection

Georgios Georgakis, Md. Alimoor Reza, Arsalan Mousavian,
Phi-Hung Le and Jana Kosecka
George Mason University

3DV 2016
International Conference on 3DVision
Stanford University, California, USA
October 25th - 28th, 2016

Overview



Contributions:

• A new RGB-D dataset of cluttered kitchen scenes, annotated in both 2D and 3D, for detection and recognition of hand-held objects in realistic settings. Some objects were taken from the BigBird dataset [1].
URL: ***http://cs.gmu.edu/~robot/gmu-kitchens.html***

•A multiview object proposal generation method which uses only 3D information.

• Detection baselines that investigate how different training strategies can affect the performance of CNNs.

Kitchen Scenes Dataset

Procedure:

• Collected the scenes with Kinect V2 (1920x1080).
• Sparse reconstructions are created with the latest structure from motion (SfM) software COLMAP.
• Dense point clouds are created using the estimated camera poses to project all points to the world coordinate frame.

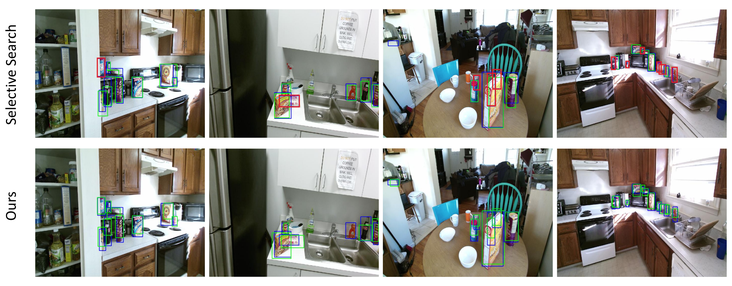
Contents:

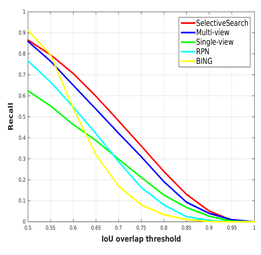
• 9 RGB-D kitchen video sequences (6735 images).
• 10-15 object instances per scene, with 23 instances in total.
• Bounding box annotations for all objects.
• 3D point labeling for each scene.

Multiview Object Proposals

Steps:

1) Removal of large planar surfaces from the dense point cloud.
2) Mean-shift clustering of remaining points in multiple ranges.
3) Cuboid fitting for removing outlier points.





Object Detection

Baselines training:

1) Turntable: Cropped object images from BigBird[1].
2) Turntable background: Same as (1) augmented with images superimposed on random backgrounds.
3) HMP Folds: Scenes are split into three training-test folds and HMP[2] is used.
4) CNN Folds: Same as (3) but we train a CNN instead of HMP. Baselines (1),(2),(4), train a CNN.

	coca cola	coffee mate	honey bunches	hunts sauce	mahatma rice	nature valley 1	nature valley 2	palmolive orange	pop secret	pringles bbq	red bull	Background	mAP
Turntable	1.0	25.7	8.6	2.8	17.2	21.2	24.0	7.6	40.1	1.2	2.6	89.5	20.1
Turntable Background	0.1	33.0	15.9	17.9	18.0	19.9	26.0	10.8	32.5	3.2	3.3	89.5	22.5
HMP Scene Folds	0.0	26.8	22.8	13.2	2.7	33.7	17.2	4.1	14.3	11.5	8.0	86.5	20.1
CNN Scene Folds	**3.5**	**48.8**	**50.0**	**27.6**	**27.9**	**52.4**	**48.1**	**18.8**	**53.6**	**46.9**	**32.7**	**90.6**	**41.7**

Table 3: Average precision (%) results for the object detection baselines on the kitchen scenes dataset.

Comparison to WRGB-D[2]

Dataset	Recall(%) / No. Proposals
WRGB-D [2]	89.3 / 17
Our Kitchen Scenes	62.4 / 2989

Table 2: Performance of the single-view proposal generation algorithm on the WRGB-D[2] and our Kitchen Scene dataset.

WRGB-D [2]	bowl	cap	cereal box	coffee mug	soda can	background	mAP
Turntable	42.2	71.7	72.3	49.0	57.7	82.5	62.6

Table 4: Object detection results for the WRGB-D [2] following the Turntable baseline.

Conclusions

• Multiview 3D object proposals outperform singleview 3D proposals and are comparable to established proposal techniques.
• Training on similar backgrounds as the test set leads to much better performing detectors, however that data are hard to acquire. Training on random backgrounds helps just slightly, which suggests that more sophisticated approaches are needed.
• Comparative experiments on the WRGB-D [2] show that the Kitchen scenes dataset is more challenging.

References

1. A. Singh, J. Sha, K. Narayan, T. Achim, and P. Abbeel. A large scale 3D database of object instances . (ICRA). 2014.
2. K. Lai, L. Bo, and D. Fox. Unsupervised feature learning for 3d scene labeling. (ICRA). 2014.
3. J. Uijlings, K. Sande, T. Gevers, and A. Smeulders. Selective search for object recognition. (IJCV). 2013.
4. S. Ren, K. He, R. Girshick, and J. Sun. Faster R-CNN: Towards real-time object detection with region proposal networks (NIPS). 2015.
5. M.M. Cheng, Z. Zhang, W.Y. Lin, and P. Torr. BING: Binarized normed gradients for objectness estimation at 300fps. (CVPR). 2014

Acknowledgments: We acknowledge support from NSF NRI grant 1527208.