Differential Geometry boosts Convolutional Neural Networks for Object Detection

Chu Wang and Kaleem Siddiqi

School of Computer Science and Center for Intelligent Machines, McGill University.

Introduction

3D object detection and recognition become popular because of the emergence of Kinect like sensors. However the manner in which depth information from Kinect sensors has been used thus far is ad-hoc. Together with the absence of applying higher order shape cues in the literature, we are motivated to abstract surface differential geometry for 3D object detection in a principled manner: we use zero-th order (depth), first-order (surface normal) and second-order (curvature) estimated from point cloud data, as input to the state-of-the-art deep learning framework for object detection. To prove that higher order shape features are critical for object detection and recognition, we provide benchmarking results on NYUv2 and UW RGBD datasets.

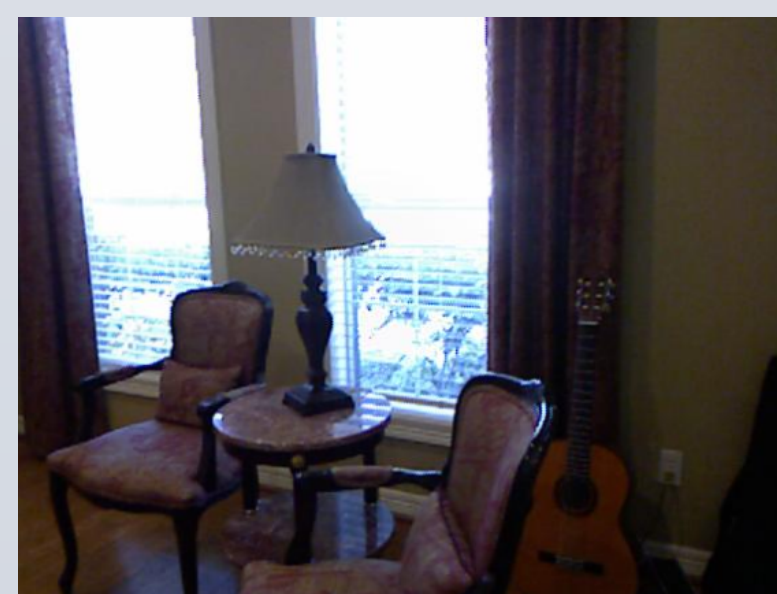## Datasets

We introduce two datasets that we experiment on.

1. NYUv2 dataset [1] : *1449* indoor scenes with registered color and depth images. We perform object detection on *19* most represented household object classes in this data set. Training/testing criteria follows exactly the same as Gupta et al. [3].

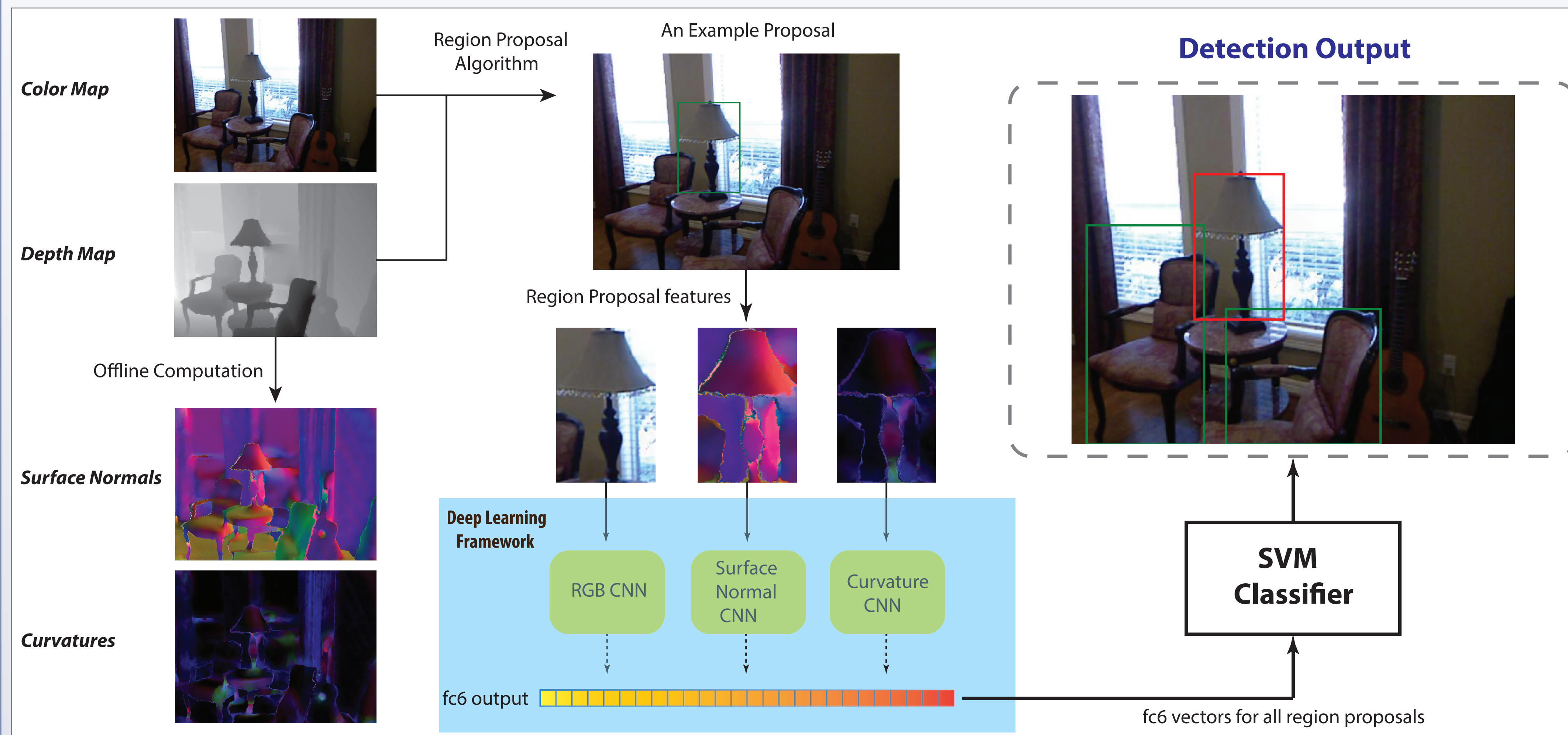Scene 1 Color Map   Scene 1 Depth Map

Scene 1294 Color Map   Scene 1294 Depth Map

2. UW RGBD dataset [2] : *300* everyday household object instances spanning *51* categories. Contains totally *42k* images for training and testing. Train/test data split used the leave-object-out splits provided by [2].

UW Point Cloud Data Showcase

R-CNN Framework [3]

Color Map — Depth Map — Surface Normals — Curvatures

Region Proposal Algorithm — An Example Proposal — Detection Output

Region Proposal features

Offline Computation

Deep Learning Framework: RGB CNN, Surface Normal CNN, Curvature CNN

fc6 output — fc6 vectors for all region proposals

SVM Classifier

• Shape feature offline computation does not provide overhead to R-CNN framework's real time detection.

Experimental Results

Table 1 NYUv2 Benchmark Result Summary

Input Features	RGB + HHA [3]	RGB + Shape	RGB + GaussD + Shape	RGB + HHA + Shape
Mean Accuracy	35.51	35.23	37.02	**39.30**

Table 2 UW Benchmark Result Summary

Input Features	RGB + HHA [3]	RGB + HHA + Shape	RGB + GaussD + Shape	Bo et al. [2]
Mean Accuracy	82.98	88.37	**88.67**	87.5

Detection Output Showcase

Shape Features

GaussD : depth based mixture attention map. Separate depth data into foreground, middel range and background maps.


Color Map   Raw Depth   GaussD

• Use EM to estimate the following Gaussian Mixture Model based on raw depth data.

$$p(d \mid \Theta) = \alpha_f \mathcal{N}_f(d; \mu_f, \sigma_f) + \alpha_m \mathcal{N}_m(d; \mu_m, \sigma_m) + \alpha_b \mathcal{N}_b(d; \mu_b, \sigma_b)$$

• Process raw depth with learned mixture models.

$$\mathcal{D}_i = \mathcal{N}_i(\mathcal{D}_{raw}; \mu_i, \sigma_i) \quad, \quad i = f, m, b$$

Surface Normals and Curvatures : first and second order shape features. Computed using eigen system analysis on covariance matrices constructed from point cloud data.
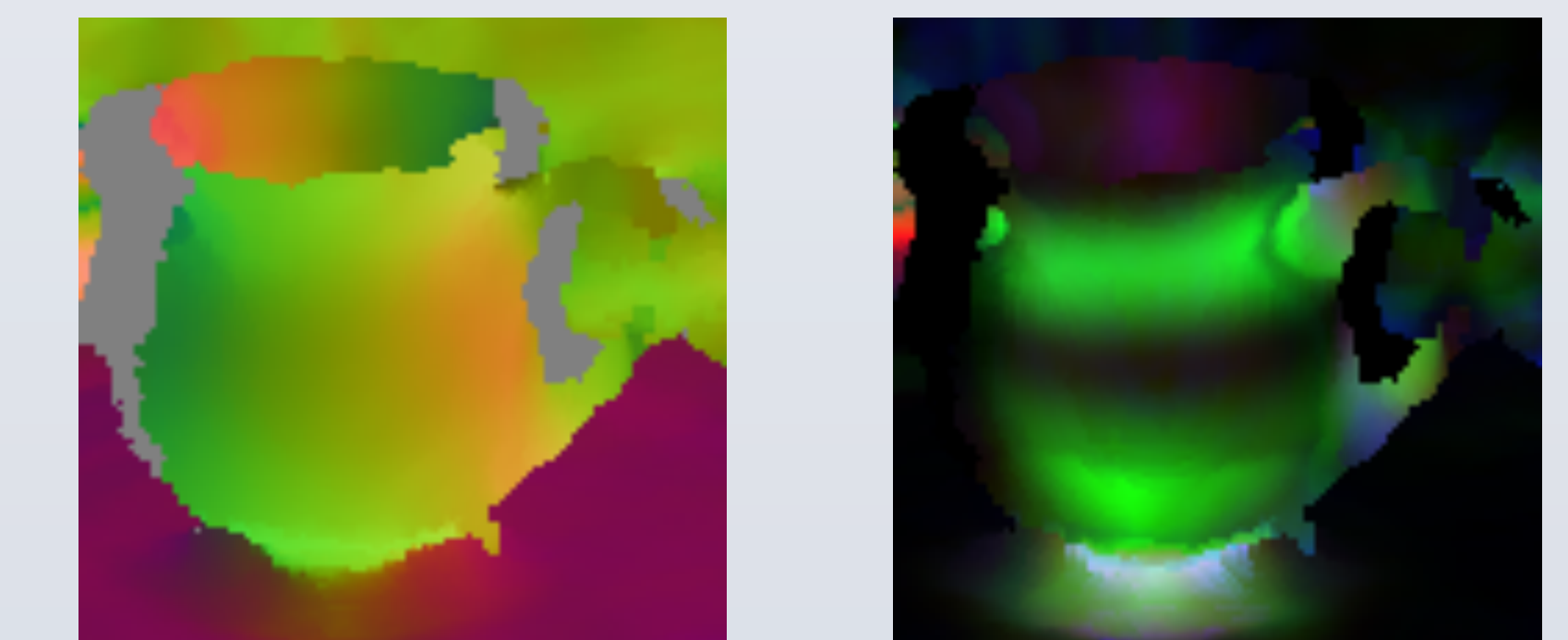
Surface Normals   Curvatures

• The eigenvector associated with smallest eigenvalue is the surface normal

$$\mathcal{C} = \frac{1}{N} \sum_{i=1}^{N} (p_i - \overline{p})(p_i - \overline{p})^T$$

• The max and min eigenvalue of the following covariance matrix is proportional to the first and second principle curvatures

$$\mathcal{C}_s = \frac{1}{N} \sum_{j=1}^{N} (s'_j - \overline{s'})(s'_j - \overline{s'})^T \quad, \quad s'_j = P s_j$$

Conclusion

• Elementary differential geometric features boost CNN for object detection and recognition.
• From depth data itself, CNN does not possess the power to abstract higher order shape features.

References

[1] P. K. Nathan Silberman, Derek Hoiem and R. Fergus. Indoor segmentation and support inference from RGBD images. In ECCV, 2012.
[2] L. Bo, X. Ren, and D. Fox. Unsupervised feature learning for RGB-D based object recognition. In Experimental Robotics, pages 387-402. Springer, 2013.
[3] S. Gupta, R. Girshick, P. Arbeláez, and J. Malik. Learning rich features from RGB-D images for object detection and segmentation. In Computer Vision-ECCV 2014, pages 345-360. Springer, 2014.