Semantic Classification of Boundaries of an RGBD Image

Nishit Soni[1], Anoop M. Namboodiri[1], C. V. Jawahar[1], Srikumar Ramalingam[2]

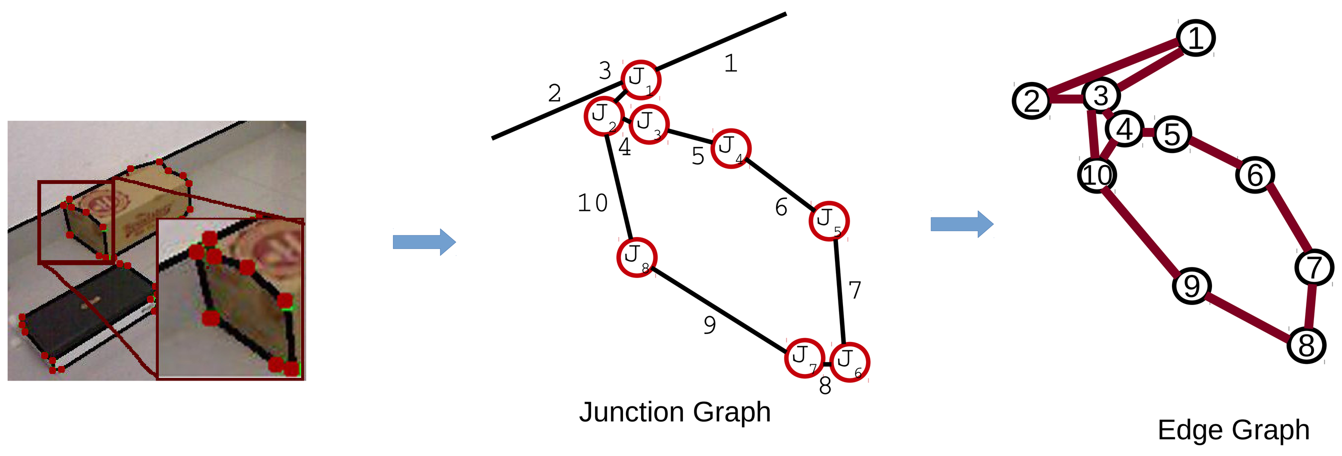
[1]CVIT, IIIT Hyderabad, [2]Mitsubishi Electric Research Lab (MERL), Cambridge, USA.

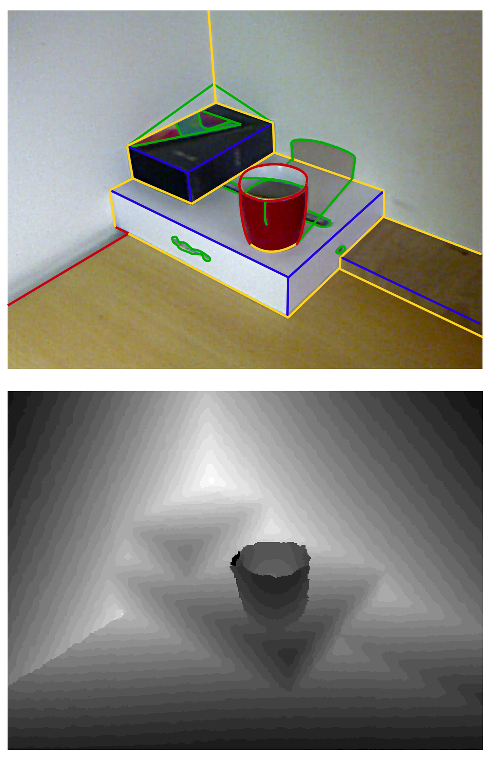
Objective

- **Goal:** Classify edge pixels in an image into occluding, convex, concave and planar entities using RGBD data.
- Occluding edges result from depth discontinuities and convex/concave edges result from normal discontinuities. Planar edges may result from shadows, reflection, specularities and albedo variations.

Approach

- We use both image and depth cues to infer the labels of edge pixels.
- Given a set of edge pixels from an edge detection algorithm, the goal is to assign one of the four labels to each of these edge pixels.
- Each edge pixel is uniquely mapped to one of the contour segments. Contour segments are sets of linked edge pixels.
- We formulate the problem as an optimization on a graph constructed using contour segments.
- Unary potentials are comptued from a Random forest pixel classifier. The feature vector uses simple yet robust geometric depth comparisons.
- We use a simple Potts model for pairwise potentials.





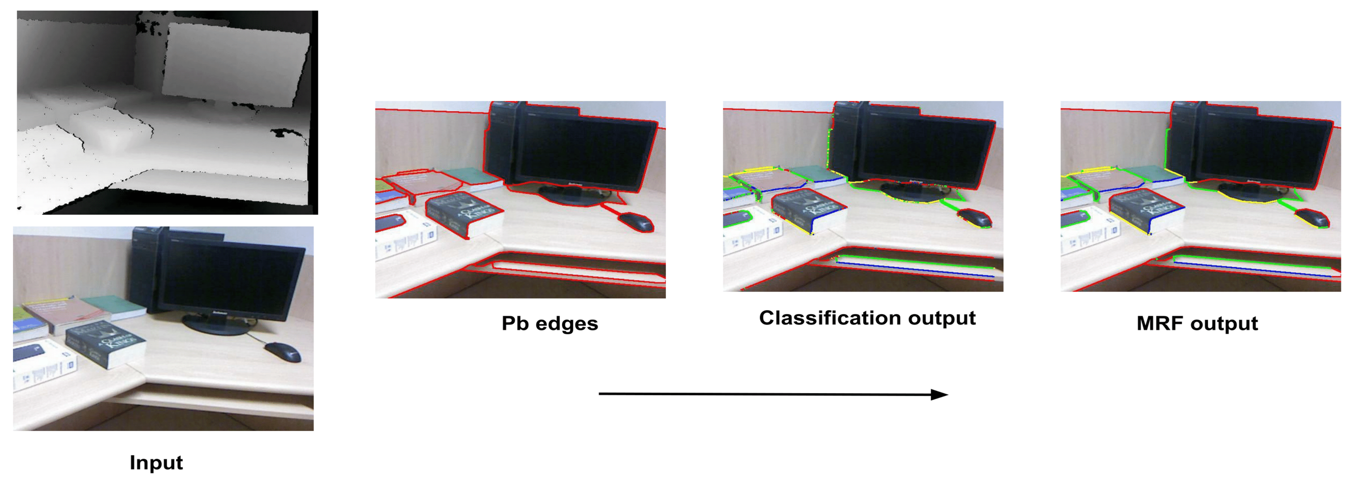


Figure 1 : This figure summarizes the pipeline of our approach. It shows RGB and depth maps as input (1st image set), with Pb edge detection [3] (2nd image). The classification and MRF outputs are shown in the last two images respectively. Color code: red (occluding), green (planar), blue (convex), yellow (concave).

	Occluding	Planar	Convex	Concave
Recall	**0.85**	**0.92**	**0.70**	**0.78**
Gupta et al. [1] Recall	0.70	0.84	0.52	0.67
Our Recall on NYU	0.76	0.85	0.56	0.69
Precision	**0.86**	**0.81**	**0.93**	**0.89**
Gupta et al. [1] Precision	0.71	0.75	0.72	0.71
Our Precision on NYU	0.79	0.80	0.77	0.71
F-measure	**0.86**	**0.86**	**0.80**	**0.83**
Gupta et al. [1] F-measure	0.71	0.79	0.61	0.69
Our F-measure on NYU	0.77	0.83	0.65	0.70

Table 1 : Precision, Recall and F-measure for each edge type on our and NYU datasets. 1st and 2nd rows of each set gives the results of our approach and comparison with [1]. The 3rd row in each set shows the results of our approach on NYU dataset.

	Occluding	Planar	Convex	Concave
Pixel Recall	0.82	0.87	0.69	0.75
Final Recall	**0.85**	**0.92**	**0.70**	**0.78**
Pixel Precision	0.84	**0.85**	0.90	0.86
Final Precision	**0.86**	0.81	**0.93**	**0.89**
Pixel F-measure	0.83	0.86	0.78	0.80
Final F-measure	**0.86**	**0.86**	**0.80**	**0.83**

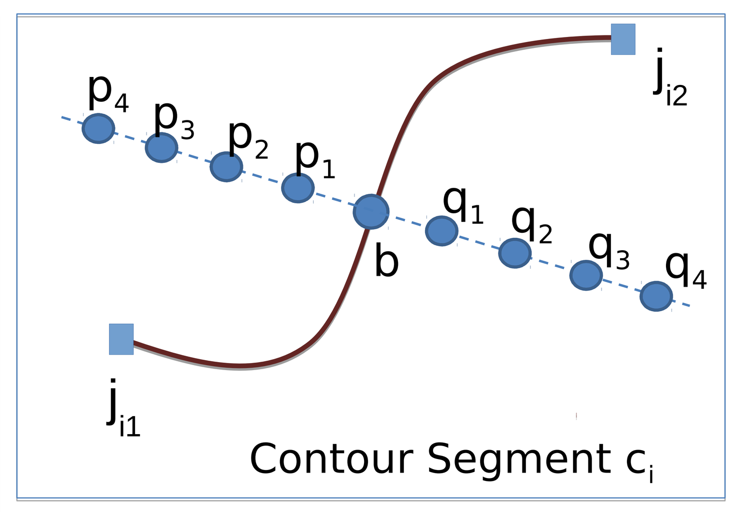
Table 2 : Precision, recall and F-measure for each edge type without and with pairwise potentials.

Experiments

- Annotated dataset of 500 RGBD images of varying complexities. Train to test ratio is 3:2. Algorithm tested on 100 images of NYU dataset [2].
- Recall, precision and F-measure used to evaluate the performance of the labeling algorithms (see Table 1).
- Table 2 shows the effect using pair-wise terms in classification of edge pixels and edge contour segments.
- We achieve an average F-score of 0.82 on edge classification. Use of smoothness constraints in the MRF improves it to 0.84 on our dataset. On the NYU dataset, we get an F-score of 0.74.
- Comparison of results from Gupta et al. [1] is done by computing their results on our dataset of annotated edges (see Table 1).

Classifier and MRF

- Features are extracted at each edge pixel and consists of simple yet robust geometric computations on the neighborhood pixels.
- A random forest classifier (30 trees) is used to assign the likelihood of each edge pixel for the four classes.
- Edge labeling is formulated as an inference problem in a graph, where the nodes take different labels or states.
- Contour segments form the nodes and their junctions provide the connectivity.
- The liklehood scores of edge pixels provide the unary potential. Pairwise term is based on a simple Potts model.



A set of 8 points on either side of an edge pixel is considered while computing the features.

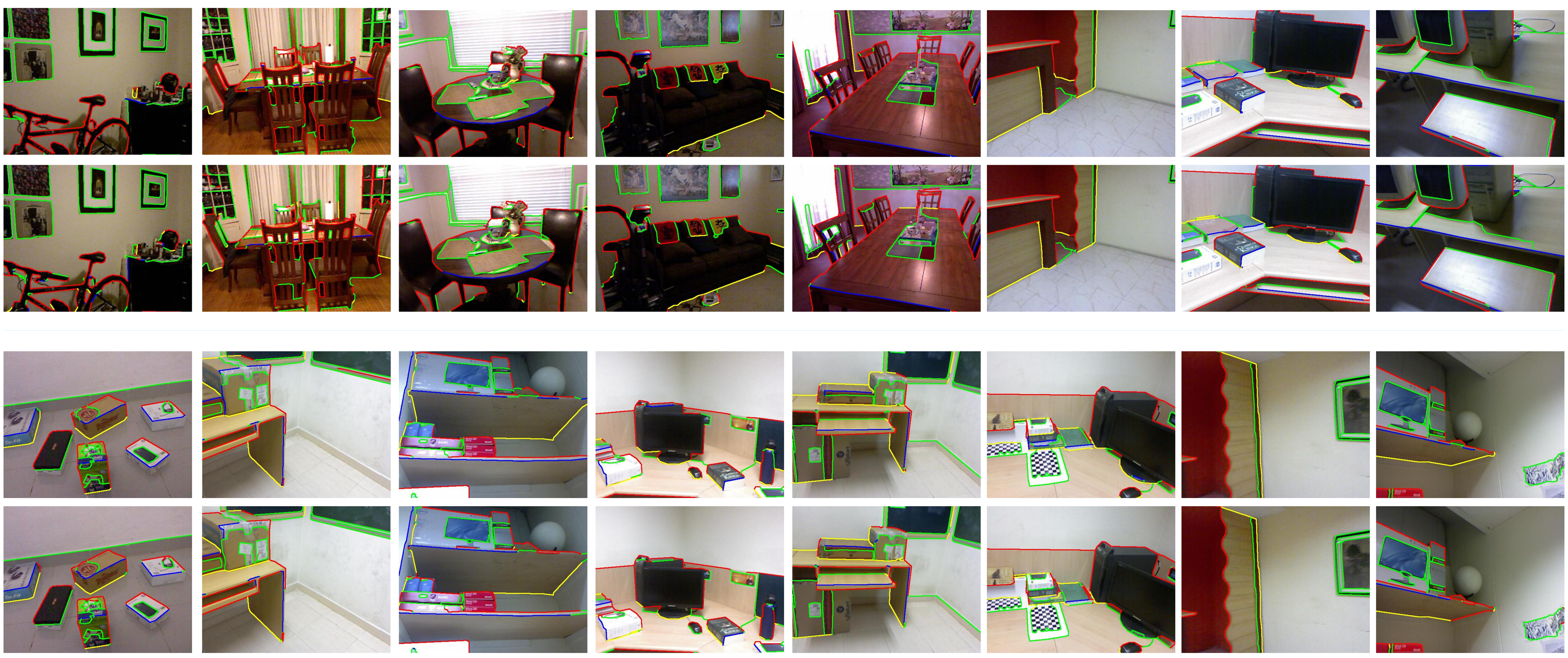


Figure 2 : Ground truths (above) and the corresponding results from our approach (below). Color code: red (occ), green (pln), blue (cvx), yellow (ccv).

Discussion

- We achieve high precision for each type of edge. Recalls are also high except for convex and concave edges. This is primarily a result of poor depth quantization or depth registration around the edge pixel.
- We are able to correctly classify complex convex/concave edges even with narrow regions having steep slope on either sides of the edge, provided the depth map is good.
- The primary causes of errors in our approach were found to be :
  - i. missing depth values from Kinect
  - ii. very small depth differences for occluding edges

While the first problem may be solved using better sensors and using image based potentials, the second would require a higher-level understanding of the scene and objects.

References

- [1] S. Gupta, P. Arbelaez, and J. Malik. *Perceptual organization and recognition of indoor scenes from RGBD images*. In CVPR, 2013.
- [2] Nathan Silberman, Derek Hoiem, Pushmeet Kohli, and Rob Fergus. *Indoor segmentation and support inference from RGBD images*. In ECCV, 2012.
- [3] D. Martin, C. Fowlkes, and J. Malik. *Learning to detect natural image boundaries using local brightness, color, and texture cues.* PAMI, 2004.

Code and dataset available at : http://cvit.iiit.ac.in/projects/semanticBoundaries

BMVC 2015