

# Semantic Classification of Boundaries of an RGBD Image

Nishit Soni<sup>1</sup>  
nishit.soni@research.iiit.ac.in  
Anoop M. Namboodiri<sup>1</sup>  
anoop@iiit.ac.in  
C. V. Jawahar<sup>1</sup>  
jawahar@iiit.ac.in  
Srikumar Ramalingam<sup>2</sup>  
ramalingam@merl.com

<sup>1</sup> International Institute of Information Technology,  
Hyderabad, India  
<sup>2</sup> Mitsubishi Electric Research Lab (MERL),  
Cambridge, USA

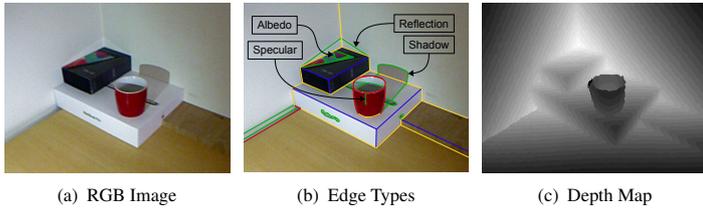


Figure 1: The edges in (a) are marked in (b) with the color code: red (occluding), green (planar), blue (convex), and yellow (concave). Planar edges are caused by different phenomena as marked. (c) shows the Kinect depth map (note the depth quantization artifacts).

**Abstract:** The problem of labeling the edges present in a single color image as convex, concave, and occluding entities is one of the fundamental problems in computer vision [4]. It has been shown that this information can contribute to segmentation, reconstruction and recognition problems. Recently, it has been shown that this classification is not straightforward even using RGBD data. This makes us wonder whether this apparent simple cue has more information than a depth map? In this paper, we propose a novel algorithm using random forest for classifying edges into convex, concave and occluding entities. We release a data set with more than 500 RGBD images with pixel-wise ground labels. Our method produces promising results and achieves an F-score of 0.84 on the data set.

**Introduction:** Edges in an image often correspond to depth discontinuities at object boundaries (occlusion edges) or normal discontinuities (convex or concave edges). In addition, there could be planar edges that are within planar regions. Figure 1 shows an image containing different types of edges. Note that planar edges may result from shadows, reflection, specularities and albedo variations. In classical line labeling with synthetic line drawings, we do not have any planar edges as the purpose of edge labeling has always been to classify the depth edges as occluding, convex and concave. However, in real images planar edges occur more frequently than others. This paper studies the problem of classifying boundaries from RGBD data. In many 3D models obtained using RGBD sensors or multi-view reconstruction techniques, we typically have very noisy 3D point cloud near the boundaries. This is because most stereo reconstruction algorithms and structured light techniques are known to provide noisy reconstruction near the boundaries. This makes the labeling problem challenging.

**Algorithm:** We use both image and depth cues to infer the labels of edge pixels. We start with a set of edge pixels obtained from an edge detection algorithm and 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. We obtain unary features using pixel classifier based on Random forest. We design a feature vector with simple geometric depth comparison features. We use a simple Potts model for pairwise potentials. The individual steps in the algorithm is shown in Figure 2.

**Experiments:** For quantitative evaluation of the method, we have created an annotated dataset of 500 RGBD images of varying complexity. Train to test ratio is 3:2. Our dataset consists of objects such as tables, chairs, cupboard shelves, boxes and household objects in addition to walls and floors. We also annotate 100 images from NYU [3] dataset, which include varying scenes from bed-room, living-room, kitchen, bathroom and so on with different complexities.

We compare our approach with Gupta *et al.* [1] and show that our

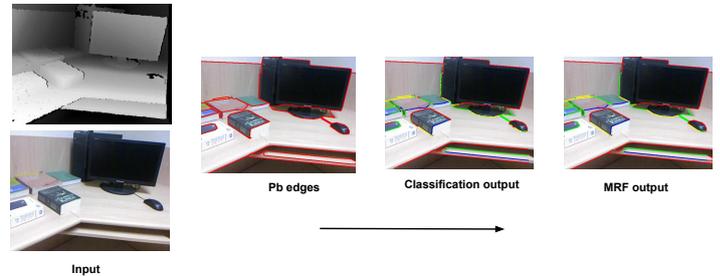


Figure 2: This figure summarizes the pipeline of our approach. It shows RGB and depth maps as input (1st image set), with Pb edge detection [2] (2nd image). The classification and MRF outputs are shown in the last two images respectively. Color code: red (occ), green (pln), blue (cvx), yellow (ccv).

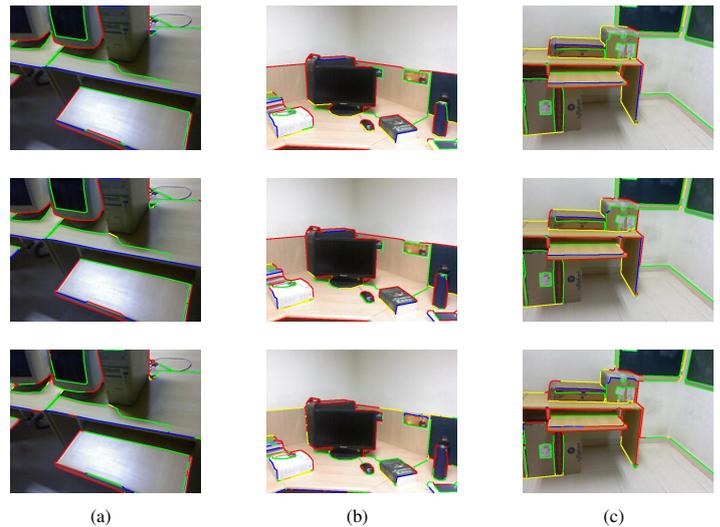


Figure 3: Ground truths (above) and the corresponding results from our approach (2<sup>nd</sup> row) and Gupta *et al.* [1] (3<sup>rd</sup> row). Color code: red (occ), green (pln), blue (cvx), yellow (ccv).

approach provides better results. Figure 3 shows a qualitative comparison between our approaches. The approach that we present here provides good labels for most pixels with high precision. We get an average F-score of 0.82 on the classification results for our data set. The use of smoothness constraints in the MRF achieves an F-score of 0.84. The NYU dataset contains complex scenes containing glass windows and table heads. We achieve an average F-score of 0.74 for the NYU dataset.

- [1] S. Gupta, P. Arbelaez, and J. Malik. Perceptual organization and recognition of indoor scenes from rgb-d images. In *CVPR*, 2013.
- [2] D. Martin, C. Fowlkes, and J. Malik. Learning to detect natural image boundaries using local brightness, color, and texture cues. *PAMI*, 2004.
- [3] Nathan Silberman, Derek Hoiem, Pushmeet Kohli, and Rob Fergus. Indoor segmentation and support inference from rgb-d images. In *ECCV*, 2012.
- [4] K. Sugihara. *Machine Interpretation of Line Drawings*. MIT Press, 1986.