Learning to identify depth edges in real-world images with 3D ground truth

Krista A. Ehinger[1], Kevin T. Joseph[1], Wendy J. Adams[2], Erich W. Graf[2], James H. Elder[1,3]

[1]Centre for Vision Research, York University [2]University of Southampton [3]Vision: Science to Applications (VISTA) Program, York University

Introduction

Edges in natural images have multiple causes: they may be produced by object boundaries or a change in surface normal, color, or illumination. The human capacity to distinguish these different types of edge in natural scenes seems effortless, but little is known about how we do it.

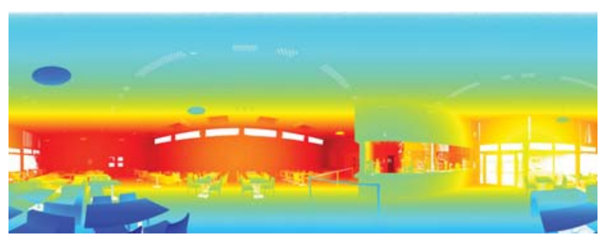
Previous investigations [1] used small hand-labeled datasets that may be subject to bias. Here we employ the new SYNS 3D dataset (syns.soton.ac.uk) to automatically and objectively label image edges as depth/non-depth, and use the resulting ground truth dataset to better understand the image cues that could underlie this visual discrimination.

Objective ground truth for depth edge classification

Stimuli

Southampton-York Natural Scenes (SYNS) database [2]: Spherical HDR and LiDAR range data from 72 randomly-sampled locations (60 outdoor, 12 indoor).


Spherical HDR image

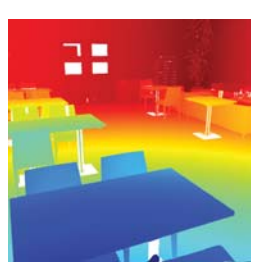

LiDAR range data

Method

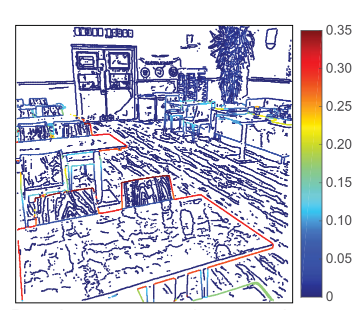
1. Project over a uniform sampling of the view sphere
2. Use multi-scale edge detector [3] to detect luminance edges in the image and depth edges in the range map.
3. Match image edges to range edges to infer ground truth depth contrast at each image edge.


Projected HDR image


Image edges (noise parameter = 3)


Projected LiDAR map


Depth edges (noise parameter = 1 mm)

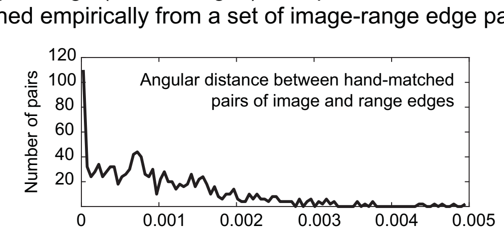

Depth contrast at image edges

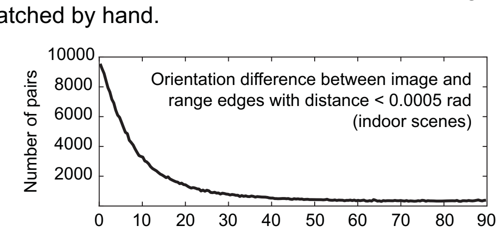
$$\text{Depth contrast} = \frac{d_1 - d_2}{d_1 + d_2}$$

Image-range map edge matching:
Each image edge is matched to the range edge i maximizing:

$$\frac{p(match|x_i,\theta_i)}{p(nonmatch|x_i,\theta_i)} = \frac{p(x_i|match)p(\theta_i|match)p(match)}{p(x_i|nonmatch)p(\theta_i|nonmatch)p(nonmatch)}$$

Where x_i = angular distance to range edge i and θ_i = orientation difference between image and range edges
If no match is above a threshold probability of 0.05 is found, the image edge is left unmatched and labeled as a "non-depth" edge (28% of edges). The priors, threshold, and distance/orientation distributions for matched edges were determined empirically from a set of image-range edge pairs matched by hand.





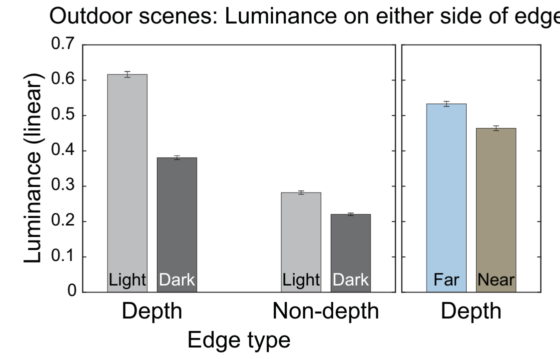
Defining "depth" edges:
Depths d_1 and d_2 on either side of the edge were estimated by averaging three point samples on either side of edge. Edges with depth contrast > 0.1 were labeled as "depth" edges.

What image cues distinguish depth edges?

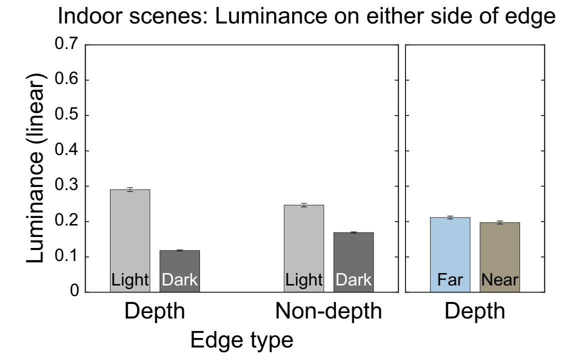
Luminance & Contrast

Consistent with prior work [1], we find that depth edges have higher contrast than non-depth edges. In outdoor scenes, depth edges have higher overall luminance than non-depth. In both outdoor and indoor scenes, the more distant side of an edge has higher luminance.

Outdoor scenes: Luminance on either side of edge



Indoor scenes: Luminance on either side of edge

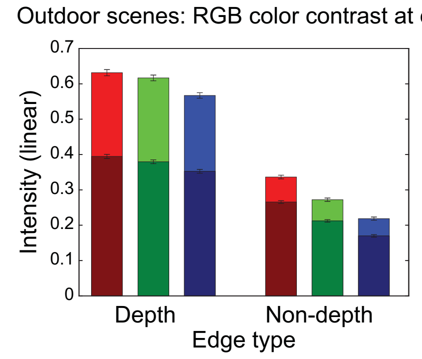


Mean HDR luminance in a 3 x 3 pixel patch on either side of edge
Linear luminance computed using CIE 1931 color space = 0.2126*R + 0.7152*G + 0.0722*B

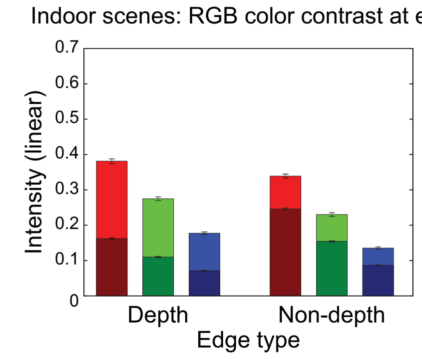
Color

The color of the image tends to be slightly warmer (more red, less blue) at a non-depth edge.

Outdoor scenes: RGB color contrast at edge



Indoor scenes: RGB color contrast at edge



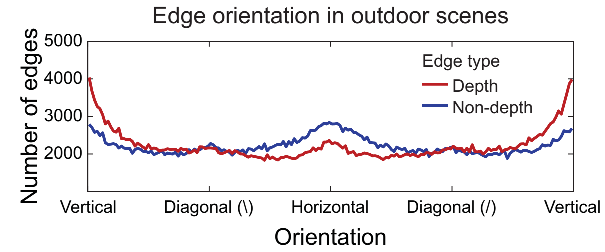
Light side of edge
Dark side of edge

Mean red, green, blue values in a 3 x 3 pixel patch on either side of edge

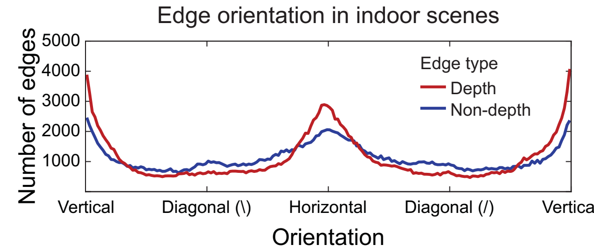
Orientation

In both indoor and outdoor scenes, vertical edges are more likely than other orientations to be depth edges. In indoor scenes, horizontal edges are also more likely to be depth edges.

Edge orientation in outdoor scenes



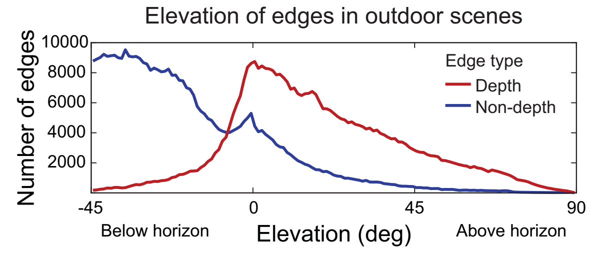
Edge orientation in indoor scenes



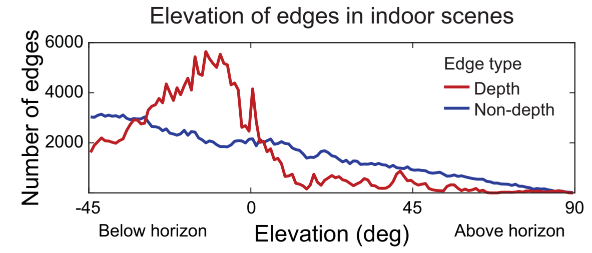
Elevation

In outdoor scenes, depth edges generally appear above the horizon; in indoor scenes depth edges are usually below the horizon. This may partially explain differences in mean luminance and color at depth and non-depth edges.

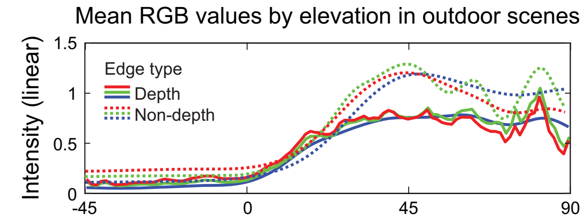
Elevation of edges in outdoor scenes



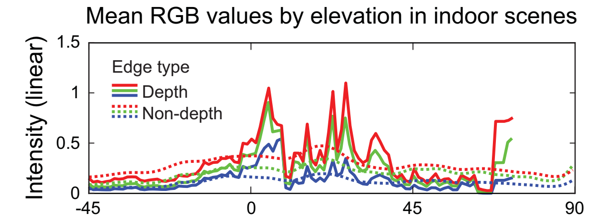
Elevation of edges in indoor scenes



Mean RGB values by elevation in outdoor scenes



Mean RGB values by elevation in indoor scenes



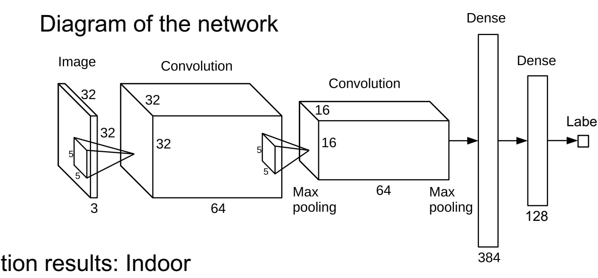
Classifying edges as depth vs. non-depth

Convolutional neural network

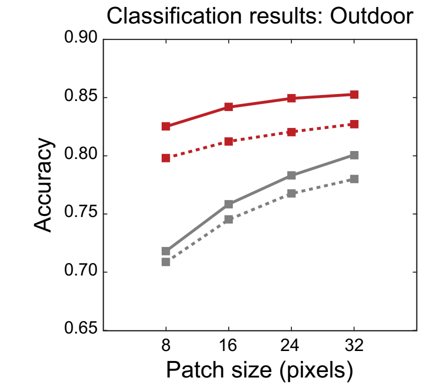
Task: Given a small image patch centered on an edge, classify as depth or non-depth.
Indoor and outdoor scenes trained separately.

Outdoor: 41 training scenes, 19 test (5,000 edges per scene)
Indoor: 9 training scenes, 3 test (10,000 edges per scene)

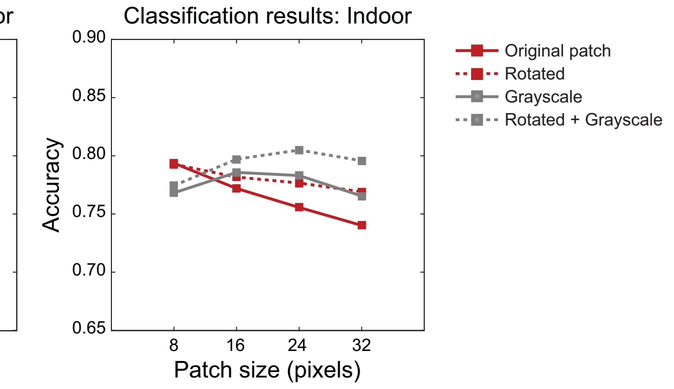
Diagram of the network



Classification results: Outdoor

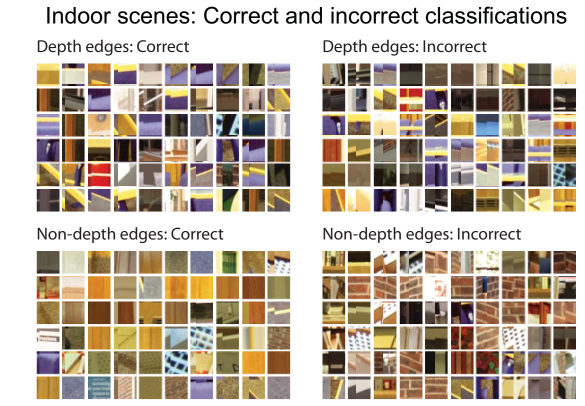


Classification results: Indoor



Outdoor scenes: Correct and incorrect classifications

Indoor scenes: Correct and incorrect classifications



Other classifiers

Using individual edge cues

Naïve Bayes classifier (KDE)	Outdoor	Indoor
Luminance contrast (Michelson)	0.66	**0.73**
Red-green contrast (Michelson)	0.68	0.66
Blue-yellow contrast (Michelson)	0.69	0.69
Edge orientation	0.50	0.52
Edge elevation	**0.78**	0.59
All cues except elevation	0.70	0.72
All cues	0.71	**0.73**

Using 32 x 32 pixel image patches

	Outdoor	Indoor
Logistic regression	0.71	0.68
K-Nearest Neighbors (K = 10)	0.58	0.69
Linear SVM	0.72	0.73
Quadratic SVM	**0.76**	**0.75**

Conclusions

Outdoor scenes:
- Classification accuracy as high as 85% is attainable. Color cues and edge orientation are informative.
- Accuracy of 78% is still attainable without these cues, perhaps based on texture, junctions, and/or shape cues.

Indoor scenes:
- Results may not generalize due to limited dataset (12 scenes).

References

[1] Vilankar, K.P., Golden, J.R., Chandler, D.M., & Field, D.J. (2014). Local edge statistics provide information regarding occlusion and nonocclusion edges in natural scenes. *Journal of Vision*, 14, 13.
[2] Adams, W.J., Elder, J.H., Graf, E.W., Leyland, J., Lugtigheid, A.J., & Muryy, A. (2016). The Southampton-York Natural Scenes (SYNS) dataset: Statistics of surface attitude. *Scientific Reports*, 6, 35805.
[3] Elder, J. H., & Zucker, S. W. (1998). Local scale control for edge detection and blur estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20, 699–716.

This research was supported by an NSERC Discovery Grant and the NSERC CREATE Training Program in Vision Science & Applications.