DOI: 10.2352/ISSN.2470-1173.2016.21.3DIPM-044

Shadow Detection on 3D Point Cloud

Shuyang Sheng and B. Keith Jenkins
Signal and Image Processing Institute, Ming Hsieh Dept. of Electrical Engineering,
University of Southern California, Los Angeles, CA, US

Abstract

Shadow detection is undergoing active research because it plays an important role in scene understanding, and has a wide range of applications including household robots and autonomous cars. In this effort, we present a novel approach to detect cast shadows on 3D point clouds. Point Cloud Library (PCL) is used to perform plane detection on point clouds. A Markov Random Field (MRF) is then constructed on the detected plane region, with an energy term that combines plane labels, depth cues and brightness cues. The resulting system is tested against USC Shadow, a dataset we collected in a controlled environment, as well as selected scenes from NYU Depth, a dataset that contains 1449 RGB-D images of various indoor scenes. Our system shows very stable performance even on complicated scenes and heavily textured planes.

Introduction

Shadow detection plays an important role in scene understanding; with correctly detected shadows, one can infer lighting conditions, and improve object recognition results.

Most previous efforts have based their work on 2D images. Most recently, [1] built a Markov Random Field (MRF) on segmented regions (super-pixels), and used various segment based features such as intensity, chromatic alignment, texture histogram and normalized distance. [5] and [6] combined an edge based method with convolutional neural network (CNN), in which a neighborhood patch is extracted for each pixel on detected edges, and fed into a CNN that classifies between shadow and non-shadow edges.

Although such methods achieved some degree of success, the problem still remains ill posed due to the limited information provided by 2D images. Shadow is inherently a result of complex interactions among light sources and multiple objects in 3D space, and hence cannot be completely characterized with the information contained in 2D images. The methods mentioned before all tried to engineer some higher level features from images, either explicitly [1] or implicitly [5][6]; however, there's no guarantee that these features will always be present in a scene. For example, some scenes in Figure 8 and Figure 9 are typical counter examples where a majority area of the scene lacks texture, and objects with drastically different albedo are present. In addition, region based features can be unstable if soft shadows are present.

3D sensing and scene understanding is an area under active research recently, and has enabled a wide range of successful applications including household robots and autonomous cars.

Developing a shadow detection algorithm in 3D can benefit from the extra depth information provided by 3D sensors. The problem becomes less ill posed with the extra information: planes can be extracted, scene structure can be estimated, and the complex interaction among lights and objects can be simulated in a physically meaningful manner. The improved algorithm in turn, could enhance various 3D sensing and scene understanding applications.

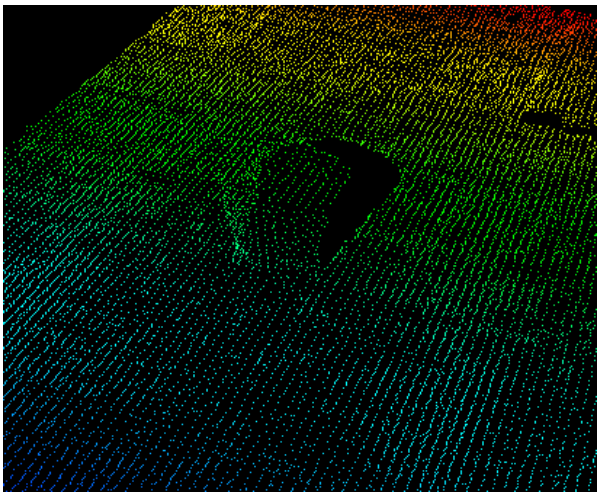


Figure 1: A typical PCD file. Y-axis coordinates are shown as pseudocolor for visualization purposes. The view shown is rotated relative to the original camera view.

Some initial success has already been observed in 3D shadow detection. [2] moved one step beyond 2D: its algorithm inputs 2D image data; 3D geometry information was provided in the form of a manually input bounding box around the object of interest; and this 3D information was used to supplement a traditional 2D algorithm. More recently, [3] first used RGB-D images in shadow detection, using the depth channel as an extra image to enhance traditional 2D shadow detection results.

For all its promise, shadow detection in 3D presents its own challenges: methods need to be re-thought, and directly extending existing 2D approaches to 3D does not take full advantage of the 3D shape information.

The goal of this project is to develop a shadow detection system that performs inference in a 3D world coordinate system. The system should be fast and fit well in robotics and other 3D sensing applications.

Method

Our current system comprises two modules, a plane detection module followed by an MRF based shadow detection module. The plane detection module separates plane area from objects and empty space; the MRF module performs shadow detection on the detected plane area.

DOI: 10.2352/ISSN.2470-1173.2016.21.3DIPM-044

Plane Detection

Images captured by 3D sensors (e.g. Kinect) are normally represented as color and depth image pairs; most publicly available 3D datasets use this format as well. So, we first convert each image pair into a point cloud, or PCD file, the 3D format proposed by Point Cloud Library (PCL) [4], and perform some preprocessing such as outlier removal.

Unlike 2D images that are always represented as rectangular 2D arrays, PCD (Figure 1) is a 1D array of points that have their own x, y and z positions, as well as additional fields such as RGB color, curvature or custom labels. The PCD format combines color and depth channels into a single representation.

The plane detection module uses the Random Sample and Consensus (RANSAC) algorithm available in PCL to find the largest set of points that represents a plane, and returns an array of plane labels (Figure 2), as well as the normal and position of the plane. Depending on the application, if we assume that orientation of the camera is known, we may further restrict the orientation of planes we are looking for (e.g., horizontal floors or vertical walls).

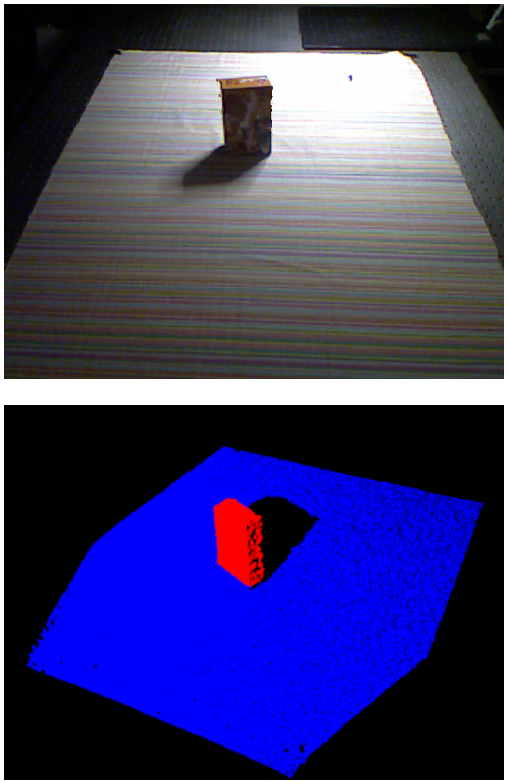


Figure 2: RGB channel and detected plane labels of a point cloud. Top: RGB channel of original point cloud. Bottom: Detected plane labels (blue denotes points on the plane, red denotes points not on the plane). Side view is shown for better illustration of 3D shape.

Shadow Detection with MRF

An undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is then constructed with the preprocessed point cloud and detected plane label array. Instead of constructing the graph as a standard rectangular 2D lattice, we discard all graph nodes (i.e. points in the point cloud) that are not on the plane (either part of objects or those do not have valid depth information), and disconnect graph edges (i.e. links between nodes) that are considered a discontinuity in the original point cloud, e.g. resulting in Figure 3.

It is worth pointing out that although $\mathcal{G}$ is not rectangular, it is still pixed based and 4-connected. We chose to construct a pixel based graph instead of a region based graph for the following reasons:

- On a pixel based graph, a smooth inferred energy map can be obtained, which, with appropriate thresholding, can stably handle soft shadows.
- A pairwise graph structure can be kept, so inference can be performed efficiently in polynomial time.
- Region segmentation is a time consuming process. Popular algorithms such as mean shift or watershed can take seconds, sometimes minutes to finish.

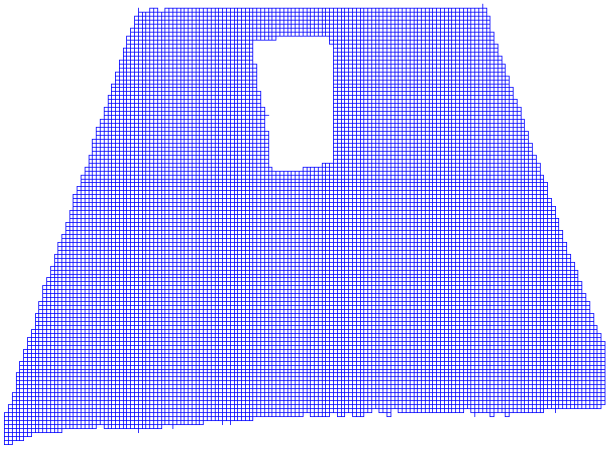


Figure 3: Constructed MRF graph of the point cloud in Figure 2

A MRF model is then constructed on the graph $\mathcal{G}$, with a binary state variable (0 for shadow and 1 for non-shadow) assigned to each node on the graph. The MRF minimizes an objective function of two terms:

- Unary energy, which is a function of pixel brightness (that encourages darker nodes to be shadows) and object proximity (that encourages nodes connected to objects to be shadows).
- Pairwise energy, following a modified version of the Ising model that encourages neighboring nodes to have the same label.

In the following subsections, we derive the energy terms of our model.

Maximum a Posteriori Estimation on MRF

MRF models the posterior probability $P(\boldsymbol{y}|\boldsymbol{X},\boldsymbol{\theta})$ using Bayes' theorem:

$$P(\boldsymbol{y}|\boldsymbol{X},\boldsymbol{\theta}) \propto P(\boldsymbol{X}|\boldsymbol{y},\boldsymbol{\theta})P(\boldsymbol{y}|\boldsymbol{\theta}) \quad (1)$$

where each element of the unrolled vector $\boldsymbol{y}$ is the shadow/non-shadow label of a pixel, each row of the matrix $\boldsymbol{X}$ is the feature vector of a pixel, and $\boldsymbol{\theta}$ is the set of parameters of the model.

DOI: 10.2352/ISSN.2470-1173.2016.21.3DIPM-044

With a given $\boldsymbol{X}$ and $\boldsymbol{\theta}$ we estimate the state variables $\boldsymbol{y}$ by minimizing the negative log posterior as an energy function:

$$E(\boldsymbol{y}|\boldsymbol{X},\boldsymbol{\theta}) = -\log P(\boldsymbol{y}|\boldsymbol{X},\boldsymbol{\theta}) \tag{2}$$

which can be expressed as a sum of unary and pairwise terms:

$$E(\boldsymbol{y}|\boldsymbol{X},\boldsymbol{\theta}) = \sum_{i\in V} log\Phi(y_i,\boldsymbol{x_i},\boldsymbol{w_v}) + \sum_{\varepsilon(i,j)\in E} log\Psi(y_i,y_j,\boldsymbol{x_i},\boldsymbol{x_j},\boldsymbol{w_e}) \tag{3}$$

Unary Energy

The feature vector $\boldsymbol{x_i}$ contains two terms, and we take an augmented notation:

$$\boldsymbol{x_i} = [1,\ x_{brightness},\ x_{boundary}] \tag{4}$$

where $x_{brightness}$ is the brightness of the pixel in HSV color space, $x_{boundary}$ is a binary variable that is set to 1 if the pixel is connected to an object pixel.

The unary term is defined as follows:

$$\Phi(\boldsymbol{x_i},\boldsymbol{w_v}|y_i=0) = exp(w_{v,0} + w_{v,1}\cdot x_{brightness} + w_{v,2}\cdot x_{boundary}) \tag{5}$$

where $\boldsymbol{w_v}$ is a three dimensional vector. The global offset $w_{v,0}$ is added to reflect the influence of the global brightness of the scene. $\Psi(\boldsymbol{x_i},\boldsymbol{w_v}|y_i=1)$ is simply set to a constant. Note that the unary term can be calculated before the inference process.

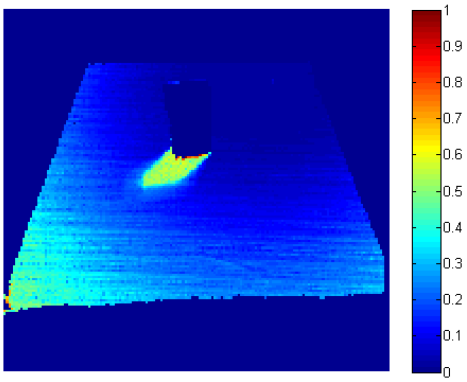


Figure 4: Node likelihood of the point cloud in Figure 2

Figure 4 shows that the node likelihood (maximized when the unary energy is minimized) of the point cloud in Figure 2 is larger in darker area, as well as on the edge at the bottom of the object. Note that pixels adjacent to the sides of the object are not considered object edges because they are disconnected on the depth map, which is clear in Figure 2.

Pairwise Energy

The pairwise term is defined as follows:

$$\Psi(y_i,y_j,\boldsymbol{x_i},\boldsymbol{x_j},\boldsymbol{w_e}) = exp(w_{e,0} + \frac{w_{e,1}}{1+(\boldsymbol{x_i}-\boldsymbol{x_j})^2})\cdot|y_i-y_j| \tag{6}$$

where $\boldsymbol{w_e} = [w_{e,0}, w_{e,1}]$. $w_{e,0}$ is a global term that imposes a penalty whenever $y_i \neq y_j$, hence encourages neighborhood nodes to have same label regardless of the similarity between $\boldsymbol{x_i}$ and $\boldsymbol{x_j}$. On the other hand $w_{e,1}$ is an adaptive term that further encourages similar connected nodes to have same label.

Compared to the original Ising model, our pairwise term not only depends on y_i and y_j, but also depends on x_i and x_j, which utilizes the intuition that the connected nodes with similar features are even more likely to have same label than the connected nodes that are less similar.

Model Tuning

Combining equations (5) and (6) and denoting $\boldsymbol{\theta} = [\boldsymbol{w_v},\boldsymbol{w_e}]$, we have 5 parameters in our model. A grid search is performed to find the set of parameters that gives the best qualitative result. Mean Cut, an approximate inference algorithm is used in the optimization.

Results

Our algorithm shows stable performance on both USC Shadow images, which contain textured surfaces and very uneven lighting conditions; as well as on NYU Depth images, which contain very complicated scenes.

USC shadow currently contains RGB-D images of 4 different scenes on a plane: a box, a noodle bowl, a stuffed toy and a combination of the three objects with occlusions. Each scene is recorded under a combination of 4 different light source directions and 2 global illumination levels. Total number of images is 32. Figure 7 shows that our results on USC Shadow are very stable for different scenes under different lighting conditions. The first two columns specifically show that our method is robust against global illumination changes, without any manual offsetting.

(a) (b) (c)

Figure 5: Details of Figure 9. (a) original image, (b) shadow removal results in [3], (c) our detection results. In (c) red mask indicates detected shadow, green mask indicates manually labeled shadow, yellow mask hence indicates agreement (overlap of green and red areas).

Our results on NYU Depth images (Figure 8) are at least on par with state of the art [3] (note that [3] focused on shadow removal, from which shadow labels can be estimated), and show better performance in certain areas. Figure 9 shows a sample comparison between our results and that of [3]. It can be seen that [3] failed to detect the shadow area under the black storage bench against the right wall (enlarged in Figure 5), which our algorithm successfully detected. This shadow area is difficult to detect in 2D even with human eyes, but is more obvious to humans when viewed as PCD (Figure 9d), especially from a side view (Figure 9e).

It is worth pointing out that the manually labeled shadow may not be absolutely accurate, especially when the shape of the

DOI: 10.2352/ISSN.2470-1173.2016.21.3DIPM-044

Figure 6: Ambiguous cases. Top: original images. Bottom: our detection results, red mask indicates detected shadow. Manual label is not shown here because there are more than one way to annotate shadow area in these cases.

shadow is difficult to trace or the shadow boundaries are fuzzy. For example, in Figure 7 we would expect some variation in shadow boundaries that are manually labeled by different people, caused by fuzzy edges in the original shadows.

Compared to 5 to 7 minutes reported by [3], our pipeline (including all preprocessing steps and reading/writing intermediate images) converges in under 90 seconds on 640 x 480 organized point clouds (native resolution of Kinect), and in typically 5 seconds when down-sampled to 160 x 160 in constructing the MRF, without visible degradation in detection accuracy. All results shown herein used the down-sampled MRF.

Conclusions and Future Work

In this effort we presented the first shadow detection algorithm on point clouds. Our system is innovative in various ways:

- It presents (to our knowledge) the first shadow detection algorithm that solves the problem in a 3D world coordinate system. Our algorithm is hence compatible with existing 3D sensing techniques and may benefit 3D applications like object shape detection, scene segmentation, 3D scene reconstruction with varying illumination, and robotics.
- It redefines the originally ill-posed shadow detection problem by distinguishing between cast shadows and shading effects on objects.
- It is tested on a new RGB-D dataset (USC Shadow) for shadow detection; novel features of this dataset include labeled shadow regions for performance evaluation, and a controlled environment with known source locations. (Source locations were not used for the work described in this paper).

Due to the extra depth information, our method can detect shadow areas that are otherwise impossible to identify with single 2D images; however, it still fails in various difficult (e.g. middle column of Figure 7) or ambiguous cases.

Some ambiguous cases are shown in Figure 6, where more than one shades of shadow exist on the floor. These cases suggest that in complicated scenes, a binary shadow/non-shadow label might not suffice, and that a hierarchical or soft labeling scheme might be necessary.

In the future, this work will be further extended by combining it with other 3D scene understanding tasks such as light source detection and semantic segmentation, and will be improved to address the issues that are mentioned above.

References

[1] Guo R, Dai Q, Hoiem D., Single-Image Shadow Detection and Removal Using Paired Regions, Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on. IEEE, 2011.

[2] Panagopoulos, A., Chaohui Wang, Samaras, D., Paragios, N., Simultaneous Cast Shadows, Illumination and Geometry Inference Using Hypergraphs, Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol.35, no.2, pp.437,449, Feb. 2013.

[3] Xiao Y, Tsougenis E, Tang C K., Shadow Removal From Single RGB-D Images, Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on. IEEE, 2014.

[4] Radu Bogdan Rusu, Steve Cousins, 3D is here: Point Cloud Library (PCL), IEEE International Conference on Robotics and Automation (ICRA), May 2011.

[5] Shen, Li, Teck Wee Chua, and Karianto Leman, Shadow Optimization From Structured Deep Edge Detection, Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2015.

[6] Khan, Salman Hameed, et al., Automatic feature learning for robust shadow detection, Computer Vision and Pattern Recognition (CVPR), 2014 IEEE Conference on. IEEE, 2014.

Author Biography

Shuyang Sheng is a PhD student in the Optics and Machine Learning for Perception (OMLP) Group, Ming Hsieh Department of Electrical Engineering, University of Southern California, working with Prof. B. Keith Jenkins. He received his B.S. in Electrical Engineering from Dalian University of Technology, China, and the M.S. from University of Southern California. His research interests lie in machine learning and computer vision, especially in scene understanding with point clouds.

B. Keith Jenkins received the B.S. in Applied Physics from Caltech (1977), and the M.S. (1979) and Ph.D. (1984) in Electrical Engineering from University of Southern California, where he is currently Professor of Electrical Engineering and Director of the OMLP Group. His research activities have included 3-D photonic computing systems, optical and computer-generated holography, 3D displays, neural networks, pattern and object recognition, and machine learning. His memberships include OSA, SPIE, IEEE, SID, and ASEE.

DOI: 10.2352/ISSN.2470-1173.2016.21.3DIPM-044

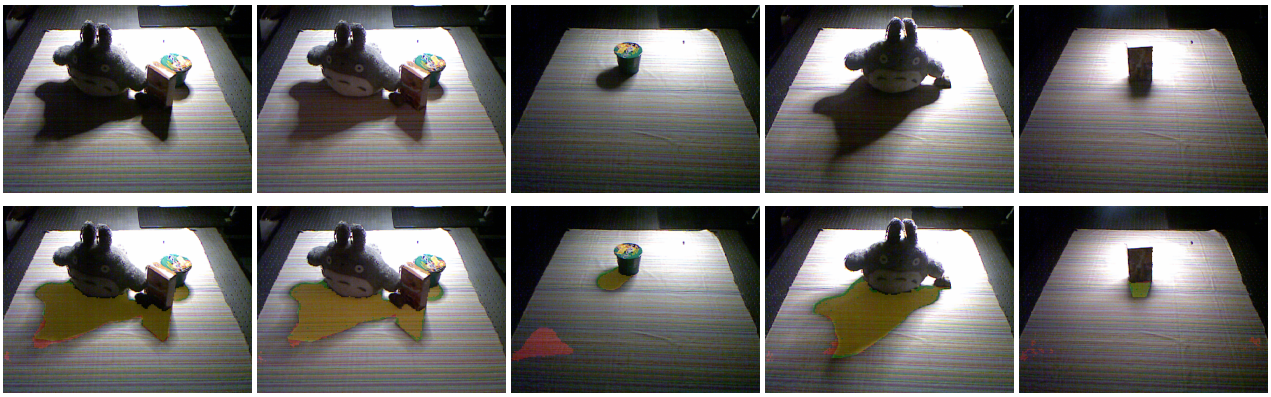
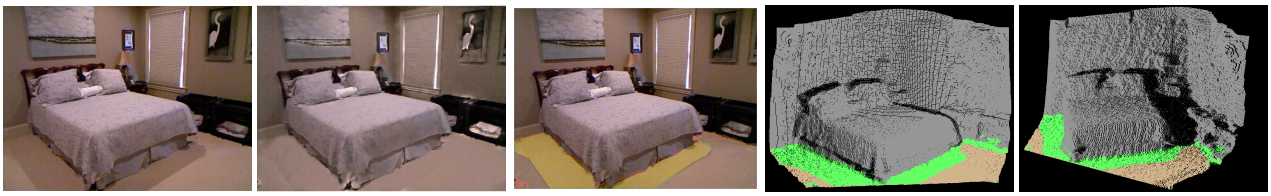


Figure 7: Shadow detection results on part of USC Shadow Dataset. Top row: RGB channels of original images. Bottom row: detection results. In the bottom row, red mask indicates detected shadow, green mask indicates manually labeled shadow, yellow mask indicates agreement.

Figure 8: Shadow detection results on part of NYU Depth Dataset. In the bottom row, red mask indicates detected shadow, green mask indicates manually labeled shadow yellow mask indicates agreement.



(a) (b) (c) (d) (e)

Figure 9: Sample comparison between the shadow removal result in [3] and the shadow detection result of our method. (a) original bedroom scene (b) shadow removal result in [3] (c) our result compared with our manual label (d) our result in point cloud, front view. (e) our result in point cloud, side view. In (d) and (e) green area indicates detected shadow.