

Efficient Depth Edge Detection Using Structured Light

Jiyoung Park¹, Cheolhwon Kim¹, Juneho Yi¹, and Matthew Turk²

¹ School of Information and Communication Engineering, Sungkyunkwan University, Korea
Biometric Engineering Research Center
{jiyp, ani4one, jhyi}@ece.skku.ac.kr

² Computer Science Department, University of California,
Santa Barbara, CA 93106
mturk@cs.ucsb.edu

Abstract. This research features a novel approach that efficiently detects depth edges in real world scenes. Depth edges play a very important role in many computer vision problems because they represent object contours. We strategically project structured light and exploit distortion of the light pattern in the structured light image along depth discontinuities to reliably detect depth edges. Distortion along depth discontinuities may not occur or be large enough to detect depending on the distance from the camera or projector. For practical application of the proposed approach, we have presented methods that guarantee the occurrence of the distortion along depth discontinuities for a continuous range of object location. Experimental results show that the proposed method accurately detects depth edges of human hand and body shapes as well as general objects.

1 Introduction

Object contour is valuable information in image analysis problems such as object recognition and tracking. Object contours can be represented by depth discontinuities (a.k.a. depth edges). However, the use of traditional Canny edges cannot distinguish between texture edges and depth edges. We describe a structured light based framework for reliably capturing depth edges in real world scenes without dense 3D reconstruction.

Depth edges directly represent shape features that are valuable information in computer vision [1, 2]. Unfortunately, few research results have been reported that provide only depth discontinuities without computing 3D information at every pixel in the input image of a scene. On the other hand, most effort has been devoted to stereo vision problems in order to obtain depth information. In fact, stereo methods for 3D reconstruction fail in textureless regions and along occluding edges with low intensity variation [3]. Recently, the use of structured light was reported to compute 3D coordinates at every pixel in the input image [4, 5]. However, the fact that this approach needs a number of structured light images makes it hard to be applicable in realtime. One notable technique was reported recently for non-photorealistic rendering [6]. They capture a sequence of images in which different light sources illuminate the scene from various positions. Then they use shadows in each image to assemble a depth edge map. This technique was applied to finger spelling recognition [7]. Although very attractive, it only works where shadows can be reliably created. In

contrast, our method is shadow free. In addition, by a slight modification of the imaging system so that it can capture white and structured images at the same time, it can be easily applied to dynamic scenes where the camera moves.

The remaining of this paper is organized as follows. In section 2, we describe the procedure to detect depth edges in a patterned image. Section 3 presents our methods that guarantee the occurrence of the distortion along depth discontinuities for a continuous range of object location. We report our experimental results in section 4. Finally, conclusions and future work are discussed in section 5.

2 Detecting Depth Edges

We detect depth edges by projecting structured light onto a scene and exploiting distortion of light pattern in the structured light image along depth discontinuities. Fig. 1 illustrates the basic method to compute depth edges. First, as can be seen in Fig. 1 (a), we project a white light and structured light consecutively onto a scene where depth edges are to be detected. Second, we extract horizontal patterns simply by differencing the white light and structured light images. We call this difference image the ‘patterned image’ (see Fig. 1 (b)). Third, we exploit distortion of light pattern in the structured light image along depth edges. We use 2D Gabor filtering that is known to be useful in segregating textural regions. The amplitude response of the Gabor filter is very low where distortion of light pattern occurs. We then accurately locate depth edges using edge information from the white light image. Fig. 1 (c) illustrates this process and a final depth edge map is shown in Fig. 1 (d).

However, distortion along depth discontinuities may not occur or be sufficient to detect depending on the distance from the camera or projector. Fig. 2 shows an example situation. Along the depth edges between objects A and B and between objects C and D, the distortion of pattern almost disappears. This makes it infeasible to detect these depth edges using a Gabor filter. For practical application of the proposed approach, it is essential to have a solution that guarantees the occurrence of the distortion along depth discontinuities irrespective of object location.

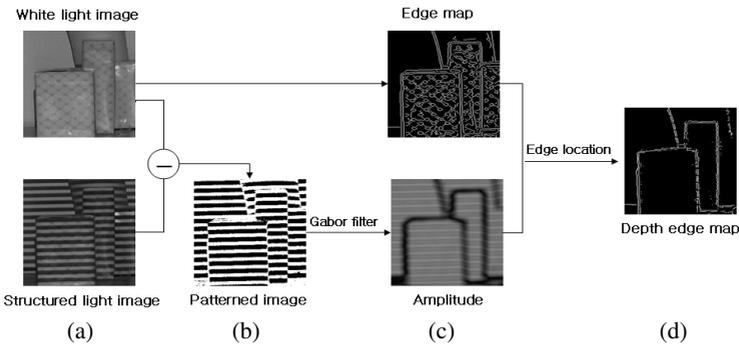


Fig. 1. The basic idea to compute a depth edge map: (a) capture of a white light and structured light image, (b) patterned image, (c) detection of depth edges by applying a Gabor filter to the patterned image with edge information from the white light image, (d) final depth edge map

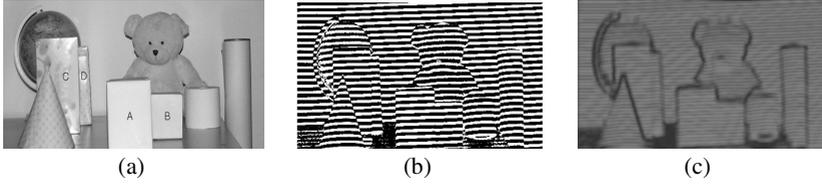


Fig. 2. Problem of disappearance of distortion along depth edges depending on the distance of an object from the camera and projector: (a) white light image (b) patterned image (c) amplitude response of Gabor filter applied to the patterned image

3 Detectable Range of Depth Edges

We have described that we can easily detect depth edges by exploiting the distortion along depth discontinuities in the patterned image. However, as previously mentioned, the distortion may not occur or be sufficient to detect depending the distance of depth edges from the camera or projector. In this section, we present methods to guarantee the occurrence of the distortion for a continuous range of object location.

3.1 Reliably Detectable Distortion

In order to compute the exact range where depth edges are detectable, we have modeled the imaging geometry of the camera, projector and object as illustrated in Fig. 3. The solid line represents a light ray from the projector. When structured light is projected onto object points A and B, they are imaged at different locations in the image plane due to different depth values. That is, distortion of horizontal pattern occurs along the depth discontinuity. The amount of distortion is denoted by Δ . Note that the width of horizontal stripes projected onto object locations A and B are the same in the image plane although they have different depth values. This is because the perspective effect of the camera and projector cancel each other out. From this model, we can derive the following equation using similar triangles:

$$\Delta = fd \left(\frac{1}{a} - \frac{1}{b} \right) = \left(\frac{fdr}{a(a+r)} \right) \tag{1}$$

In order for a depth edge to be detectable by applying a Gabor filter, the disparity of the same horizontal stripe, Δ , in the image plane should be above a certain amount. We have confirmed through experiments that an offset of at least 2/3 of the width of the horizontal stripe (w) is necessary for reliable detection of the distortion. Thus, the range of Δ for reliable detection of pattern distortion can be written:

$$2wk + \frac{2w}{3} \leq \Delta \leq 2wk + \frac{4w}{3}, \quad k = 0,1,\dots \tag{2}$$

From equation (2), we can know that there are ranges where we cannot detect depth edges due to the lack of distortion depending the distance of a depth edge from the camera or projector. Therefore, for practical application of the proposed approach, we need to guarantee that we are operating within these detectable regions.

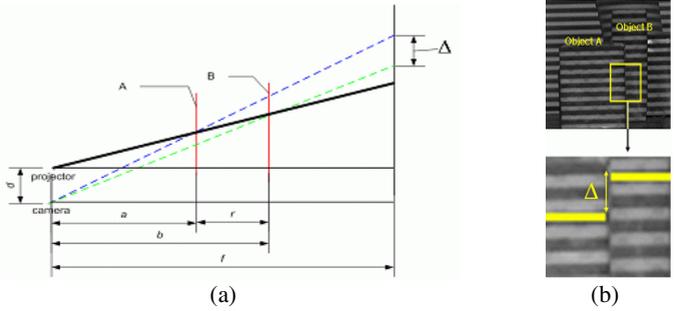


Fig. 3. Imaging geometry and the amount of distortion: (a) spatial relation of camera, projector and two object points viewed from the side (b) the magnitude of pattern distortion, Δ , in a real image

3.2 Extending the Detectable Range of Depth Edges

We propose two methods to extend the range so that detection of distortion is guaranteed. The first method is based on a single camera and projector setup that uses several structured light images with different width of horizontal stripes. We use additional structured light whose spatial period is halved such as, $w_2=2w_1$, $w_3=2w_2$, $w_4=2w_3$, \dots . When n such structured light images are used, the range of detectable distortion, Δ , is as follows.

$$2/3 w_1 < \Delta < (2^n - 2/3) w_1. \tag{3}$$

The second method exploits an additional camera or projector. As illustrated in Fig. 4, this method is equivalent to adding a new curve, $d=d_2$ (the dotted line), that is different from d_1 . Recall that d denotes the distance between the camera and the projector. The new detectable range is created by partially overlapping the ranges by the two curves. A and B represent respectively detectable and undetectable ranges in Δ . They correspond to X and Y regions in a . When $a_1 > a_2$ and $a_4 > a_3$, the undetectable range B in X is overlapped with the detectable range A in Y. Similarly, the undetectable range B in Y is overlapped with the detectable range A in X. Therefore, if we consider both X and Y, we can extend the range where the detection of the distortion is guaranteed. To satisfy the condition $a_1 > a_2$ and $a_4 > a_3$, equation (4) must hold where the range of Δ is $\gamma w k + \alpha w < \Delta < \gamma w k + \beta w$, $k = 0, 1, \dots$.

$$\frac{\alpha + \gamma + \gamma k}{\beta + \gamma k} d_1 < d_2 < \frac{\beta + \gamma k}{\alpha + \gamma k} d_1 \tag{4}$$

3.3 Computation of the Detectable Range of Depth Edges

The detectable range of depth edges $[a_{\min}, a_{\max}]$ is computed in the following two steps:

- Step 1: Determination of the width of a stripe, w , in the structured light
- First, we set a_{\max} to the distance from the camera to the farthest background. Given r_{\min} , w can be computed by equation (5) which is derived from equation (1).

$$w = \frac{3fd_1 r_{\min}}{2a_{\max}(a_{\max} + r_{\min})} \tag{5}$$

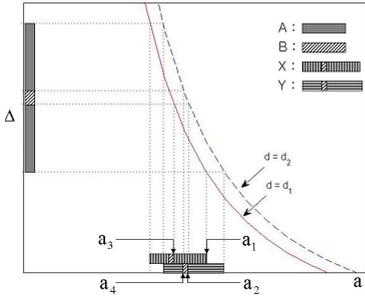


Fig. 4. Extending the detectable range using an additional camera or projector

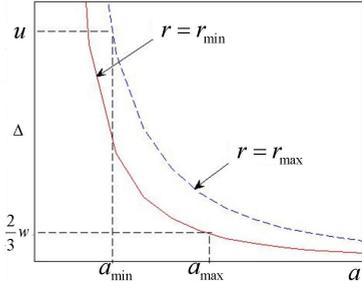


Fig. 5. Computation of the detectable range of depth edges

Thus, given a_{max} and r_{min} , we can compute the ideal width of stripes of the structured light. Using this structured light, we can detect depth edges of all object points that are located in the range, $a = [0, a_{max}]$, and apart from each other no less than r_{min} .

- Step 2: The minimum of the detectable range, a_{min}

Given w from step 1, we can compute a_{min} that corresponds to the upper limit of Δ , u , as shown in Fig. 5. We have described two methods in section 3.2 for extending the detectable range of depth edges. The expression for u is different depending on which method is used. After determining u and r_{max} , a_{min} can be computed by equation (1). r_{max} denotes the maximum distance between object points in the range $[a_{min}, a_{max}]$ that guarantees the occurrence of the distortion along depth discontinuities. Clearly, the distance between any two object points is bounded by $(a_{max} - a_{min})$. Therefore, when a_{min} and r_{max} satisfy equation (6), we are guaranteed to detect depth edges of all object points located in the range $[a_{min}, a_{max}]$, and apart from each other no less than r_{min} and no more than r_{max}

$$a_{max} - a_{min} = r_{max} \tag{6}$$

In this case, u , a_{min} and r_{max} have the following relationship.

$$u = \frac{fd_1 r_{max}}{a_{min} (a_{min} + r_{max})} \tag{7}$$

Then r_{max} becomes:

$$r_{max} = \frac{a_{max}^2 u}{fd_1 + a_{max} u} \tag{8}$$

Substituting equation (8) into equation (6), we obtain the following equation.

$$a_{min} = \frac{fd_1 a_{max}}{fd_1 + ua_{max}} \tag{9}$$

This way, we can employ structured light of optimal spatial resolution that is most appropriate for given application. Furthermore, we can use this method in an active way to collect information about the scene.

4 Experimental Results

For capturing structured light images, we have used a HP xb31 DLP projector and Canon IXY 500 digital camera. In this section, we present experimental results for two different experimental setups.

A. 1 camera and 1 projector

In order to extend detectable range of depth edges, this setup uses the first method (as explained in section 3.2) that just employs additional structured light whose spatial frequency is halved. Fig. 6 shows the result of depth edge detection using three structured light images with different width of horizontal stripes. Setting $f=3m$, $a_{max}=3m$, $d=17.3cm$ and $r_{min}=10cm$, w_1 and a_{min} are determined as $0.84cm$ and $2.33m$, respectively. Each Gabor amplitude map (Fig. 6 (c)~(e)) shows that we cannot detect all the depth edges in the scene using a single structured light image. However, combining the results from the three cases, we can obtain the final Gabor amplitude map as in Fig. 6 (f). The result Fig. 6 (g) shows that this method is capable of detecting depth edges of all the objects located in the detectable range. We have also compared the result with the output of the traditional Canny edge detector Fig. 6 (h).

B. 1 camera and 2 projectors

We can apply both extension methods when using a single camera and two projectors. The detectable range can be extended more than setup A when the same number of the structured light images is used. Fig. 7 shows the experimental result when two

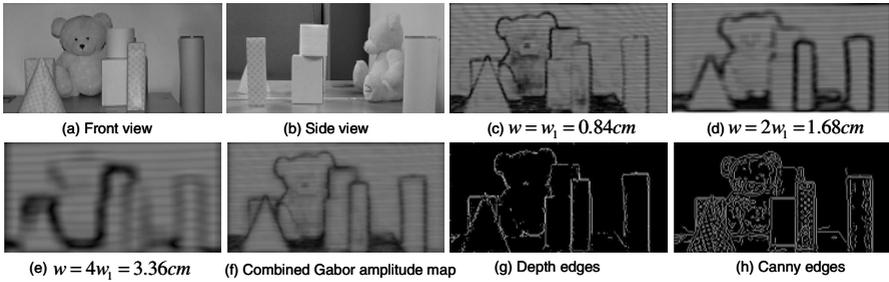


Fig. 6. Detecting depth edges using a single camera and projector

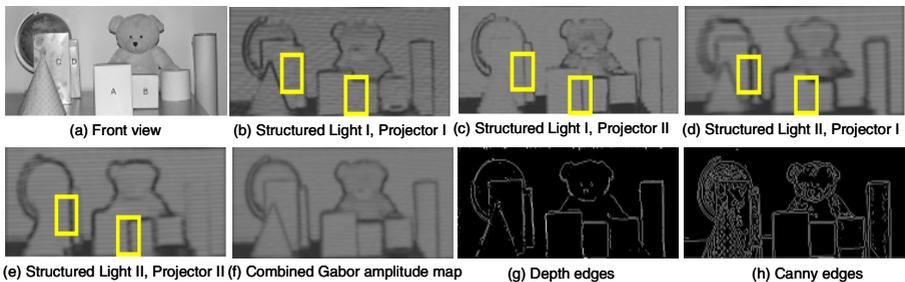


Fig. 7. Detecting depth edges using a single camera and two projectors

structured lights are used for each projector. When $f = 3\text{m}$, $a_{max} = 3\text{m}$, $d_1 = 17.3\text{cm}$, $d_2 = 20.7\text{cm}$, $r_{min} = 10\text{cm}$ are used, w_1 and a_{min} are determined as 0.84cm and 1.94m , respectively. Fig. 7 (b)~(e) show how the four structured light images play complementary roles to produce a final depth edge map.

Fig. 8 shows the result of detecting of hand and human body contours. Our method accurately detects depth edges by eliminating inner texture edges by using only a single camera and one structured light image. The result shows that our method is effectively applicable to gesture recognition.



Fig. 8. (a) Detection of depth edges in the case of hand gestures for fingerspelling, (b) Detecting human body contours for gesture recognition, from left to right: white light image, Gabor amplitude map, depth edges and canny edges

5 Conclusions

We have proposed a new approach using structured light that efficiently computes depth edges. Through a modeled imaging geometry and mathematical analysis, we have also presented two setups that guarantee the occurrence of the distortion along depth discontinuities for a continuous range of object location. These methods enable the proposed approach to be practically applicable to real world scenes. We have demonstrated very promising experimental results.

We have also observed that infrared projectors show the same distortion characteristics in patterned images. This makes us directly apply the same analysis from the LCD projectors to infrared projectors. By bypassing dense 3D reconstruction that is computationally expensive, our methods can be easily extended to dynamic scenes as well. We believe that this research will contribute to great improvement of many computer vision solutions that rely on shape features.

Acknowledgement

This work was supported in part by the Korea Science and Engineering Foundation (KOSEF) through the Biometrics Engineering Research Center (BERC) at Yonsei University.

References

1. T. A.Cass, "Robust Affine Structure Matching for 3D Object Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1264-1265, 1998.
2. I. Weiss and M. Ray, "Model-based recognition of 3D object from single vision," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 116-128, 2001.

3. T. Frohlinghaus and J. M. Buhmann "Regularizing phase-based stereo," Proceedings of 13th International Conference on Pattern Recognition, pp. 451-455, 1996.
4. S. H. Lee, J. M. Choi, D. S Kim, B. C. Jung, J. K. Na, and H. M. Kim, "An Active 3D Robot Camera for Home Environment," Proceedings of 4th IEEE Sensors Conference, 2004.
5. D. Scharstein and R. Szeliski, "High-Accuracy Stereo Depth Maps Using Structured Light," IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Vol. 1, pp. 195-202, 2003.
6. R. Raskar, K. H. Tan, R. Feris, J. Yu, and M. Turk, "Non-photorealistic Camera: Depth Edge Detection and Stylized Rendering Using Multi-Flash Imaging," Proceedings of ACM SIGGRAPH Conference , Vol. 23, pp. 679-688, 2004.
7. R. Feris, M. Turk, R. Raskar, K. Tan, and G. Ohashi, "Exploiting Depth Discontinuities for Vision-based Fingerspelling Recognition," IEEE Workshop on Real-Time Vision for Human-Computer Interaction, 2004.