Available online at www.sciencedirect.com
ScienceDirect
Image and Vision Computing 26 (2008) 1449–1465
www.elsevier.com/locate/imavis

Using structured light for efficient depth edge detection

Jiyoung Park [a], Cheolhwon Kim [b], Jaekeun Na [c], Juneho Yi [c,*], Matthew Turk [d]

[a] Electronics and Telecommunications Research Institute (ETRI), Digital content Research Division, Republic of Korea
[b] Nexteye Co., Ltd, Advanced Technology Institute, Republic of Korea
[c] Sungkyunkwan University, School of Information and Communication Engineering, Republic of Korea
[d] University of California, Santa Barbara, Computer Science Department, USA

Received 27 November 2006; received in revised form 1 November 2007; accepted 20 January 2008

Abstract

This research describes a novel approach that accurately detects depth edges with cluttered inner texture edges effectively ignored. 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. In practice, distortion along depth discontinuities may not occur or be large enough to detect depending on the distance from the camera or projector. We present 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 shapes of human hands and bodies as well as general objects.

Keywords: Depth edge detection; Structured light

1. Introduction

Depth edges play a very important role in many approaches to computer vision problems because they represent object contours [1–3]. Reliable detection of depth edges facilitates tasks of the object and gesture recognition. Fig. 1 shows an example of texture edges and depth edges. The use of traditional edge detection methods 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.

1.1. Overview of our approach

The goal of this research is to produce a depth edge map of the real world scene. We illustrate in Fig. 2 the basic idea for depth edge detection. First, as can be seen in Fig. 2(a), we project a white light and a structured light consecutively onto a scene where depth edges are to be detected. The structured light contains a special light pattern. In this work, we have placed the projector and camera vertically so that we use a pattern comprising simple black and white horizontal stripes of equal width. Vertical stripes can be used with a similar analysis. We capture the white light image and then the structured light image. Second, we extract horizontal patterns by differencing the white light and structured light images and using a robust thresholding method. We call this difference image the "pattern image" (see Fig. 2(b)). Third, we identify depth edges in the pattern image guided by edge information from the white light image. We exploit distortion of the light pattern in the structured light image along depth edges. Since the horizontal pattern can be considered a periodic signal with a specific frequency, we can easily detect candidate locations for depth edges by applying a Gabor filter to the pattern image. The amplitude response of the Gabor filter is very low where distortion of light pattern occurs. Fig. 2(c) illustrates this process. Finally, we accurately locate depth edges using edge information from the white light image, yielding a final depth edge map as in Fig. 2(d).

* Corresponding author. Tel.: +82 31 290 7142.
E-mail address: jhyi@skku.edu (J. Yi).

0262-8856/$ - see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.imavis.2008.01.006

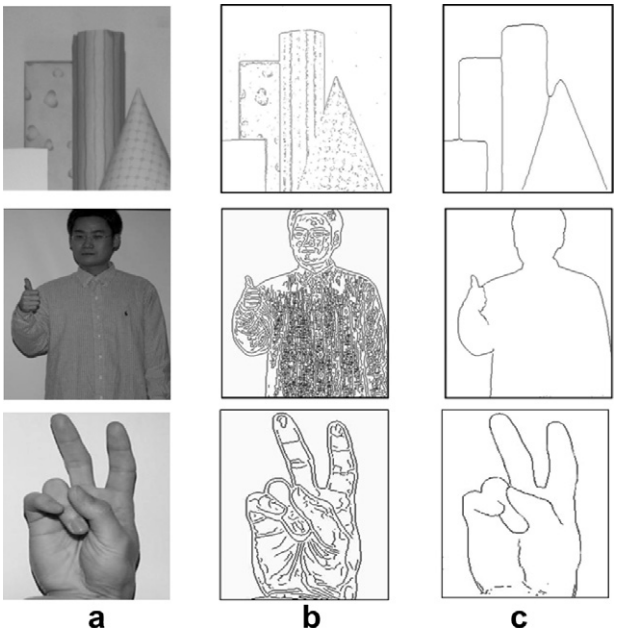


Fig. 1. Texture vs. depth edges: (a) input scenes, (b) Canny edges, (c) depth edges.

In practice, distortion along depth discontinuities may not occur or be sufficient to detect depending on the distance from the camera or projector. Fig. 3 shows an example situation.

Along the depth edges between objects A and B, and between objects C and D, the distortion, i.e., the offset of the pattern, almost disappears. This makes it infeasible to detect these depth edges using a Gabor filter. For successful 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.

We propose methods to guarantee the occurrence of the distortion for a continuous range of object locations. Based on a modeled imaging geometry of camera, projector, object, and its mathematical analysis, we first compute the exact ranges of object location where detection of distortion is not feasible. We present two methods that extend the range where detection of the distortion is guaranteed. The first method is based on a single camera and projector setup that simply uses several structural light images with different widths of horizontal stripes. The use of three gray level stripes instead of black and white ones is also described to cut the number of projections of structured light patterns by half. The other method exploits an additional camera or projector as compared with the first method. We have used a general purpose LCD projector; however, an infrared projector can be employed with the same analysis in order to apply the method to humans. Experimental results have confirmed that the proposed methods work very well for shapes of human hands and bodies as well as for general objects.

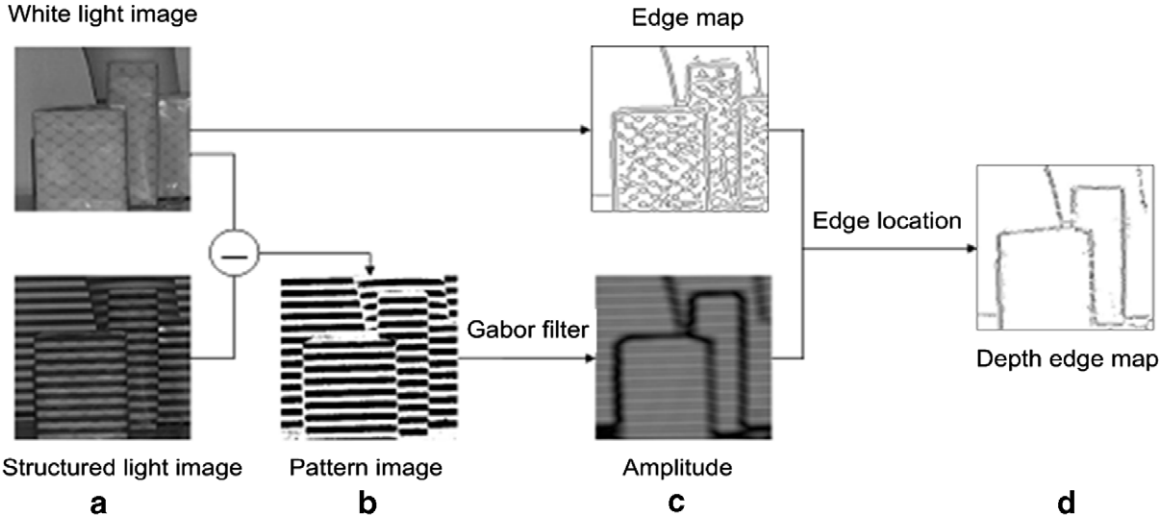


Fig. 2. Illustration of the basic idea to compute a depth edge map: (a) capture of a white light image and structured light image, (b) pattern image, (c) detection of depth edges by applying a Gabor filter to the pattern image with edge information from the white light image, (d) final depth edge map.

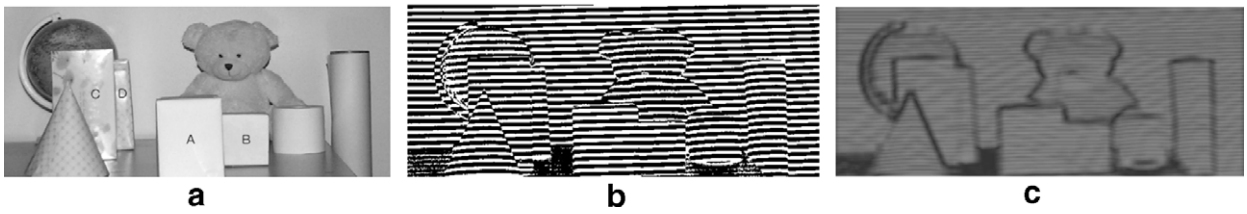


Fig. 3. Problem of disappearance of distortion along depth discontinuities depending on the distance of an object from the camera and projector: (a) white light image, (b) pattern image, (c) amplitude response of Gabor filter. Along the depth edges between objects A and B, and between objects C and D, in the pattern image (b), the distortion of pattern almost disappears. This makes it not feasible to detect these depth edges using a Gabor filter.

1.2. Related work

Depth edges directly represent shape features that are valuable information in computer vision. In order to obtain depth information, considerable efforts have been devoted to stereo vision problems. In fact, most stereo methods for 3D reconstruction without active illumination fail in textureless regions and along occluding edges with low intensity variation [4,5]. On the other hand, structured light systems have been used to reliably recover 3D information of objects. The technique is based on projecting a light pattern and imaging the illuminated scene from a camera. Projecting a single color pattern or a set of binary (or gray) patterns onto the measured scene, they find correspondences between image points and points of the projected pattern. In the case of using a single color pattern [6–9], the shape acquisition result is affected by the variations of the object surface color. Thus, the noisier the application environment is where the technique will be applied, the smaller the number of colors that can be used. When using multiple coded patterns [10–12], some kind of hardware acceleration is employed to switch multiple coded patterns rapidly in a short period of time. Unlike these methods, depth edge detection itself can be much simpler. Instead of having to estimate the full 3D coordinates of points in the scene, and then look for depth discontinuities, our technique reduces the general 3D problem of depth edge recovery to one of filtering.

Unfortunately, there have been few research results that only provide depth discontinuities without computing 3D information at every pixel in the input image of a scene. One notable technique was reported recently for non-photorealistic rendering [13]. In [13], 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 also applied to fingerspelling recognition [14]. Although very attractive, it only works where shadows can be reliably created. Recently, we used structured light patterns to efficiently detect depth edges [15], and presented basic ideas to control key parameters such as stripe width given the detectable depth difference. This paper is an extended and enhanced version of that work. Our method can guarantee the occurrence of the distortion along depth discontinuities for a continuous range of object location. 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 remainder of this paper is organized as follows. In Section 2, we describe imaging geometry that explains the occurrence of strip offsets in a pattern image due to the distortion of light pattern along depth discontinuities. Section 3 gives details of the computation of the pattern image and the application of Gabor filter to detect depth edges. Sections 4 and 5 present our methods to compute and extend the detectable range of depth edges, respectively. We report our experimental results in Section 6.

2. Offset of structured light pattern along depth discontinuities

In order to analyze the relationship between structured light from a projector and an image grabbed by a camera, we have modeled imaging geometry of camera, projector and object as illustrated in Fig. 4(a), which depicts a side view of the imaging setup. The solid line represents a light ray from the projector. When structured light is projected onto objects A and B, they are imaged at different locations in the image plane due to different depth values. The exact amount of offset of the pattern is denoted by Δ_{exact}. From this model, we can derive the following equation using similar triangles:

$$\Delta_{\text{exact}} = fd\left(\frac{1}{a}-\frac{1}{b}\right) = \frac{fdr}{a(a+r)}. \tag{1}$$

a, b, d and f are the distances of object A, object B, projector and virtual image plane from the camera, respectively. r denotes the distance between objects A and B. However, Δ_{exact} may not be measurable because we use simple black and white stripes with equal width and the amount of offset is periodic as it gets large. Thus Δ_{visible} illustrated in Fig. 4(b) is what we actually see in the pattern image.

Note that the width of the 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. In Fig. 5, let us assume that the virtual image plane is at the same depth as object B. When a stripe of the structured light (denoted by solid lines) is projected onto objects A and B, the camera image is created along the dotted line where the widths of the stripes are denoted by x and y, respectively. Using similar triangles, we obtain $x_1 = \frac{b}{a}e$ and $x_2 = \frac{b}{a}g$. This gives $x = y$ as follows.

$$x = x_1 - x_2 = \frac{b}{a}(e-g) = y. \tag{2}$$

This means that the width of a stripe projected onto object points A and B are the same in the image plane irrespective of the distance between the objects. In other words, while the width of a stripe gets bigger at distance, it is imaged at the same width due to the perspective effect of the imaging system.

3. Detecting depth edges

We detect depth edges by projecting structured light onto a scene and exploiting the offset of the pattern in the structured light image along depth discontinuities. In order to detect the offset of the light pattern, we use 2D Gabor filtering that is known to be useful in segregating textural regions [16–18]. We first extract patterns by differencing the white light and structured light images. We then

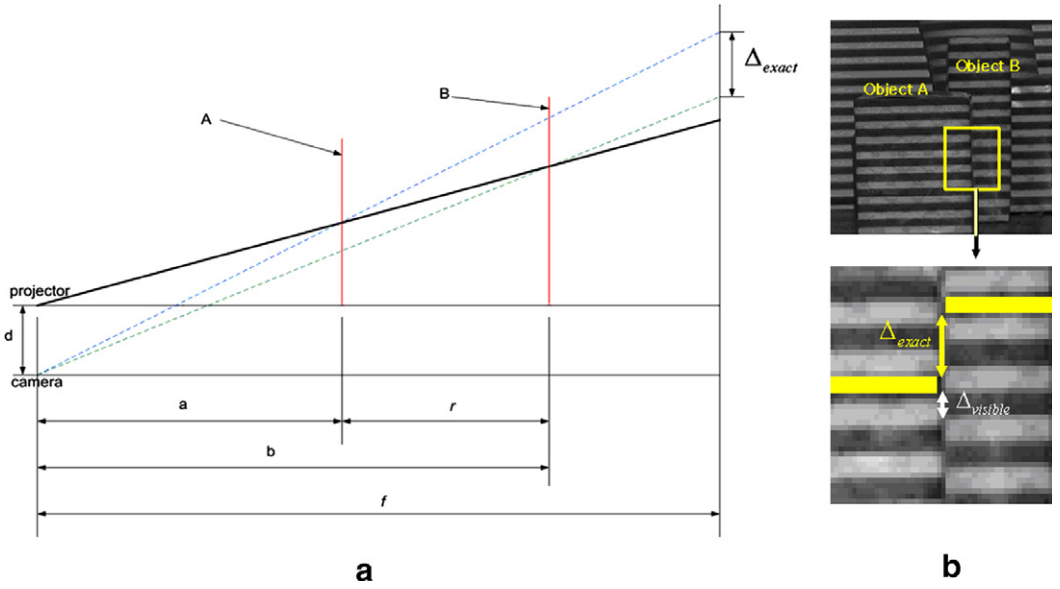


Fig. 4. Imaging geometry and offset of the pattern: (a) spatial relation of camera, projector and two objects viewed from the side, a: distance between camera and object A, b: distance between camera and object B, d: vertical distance between camera and projector, f: distance between camera and virtual image plane, r: distance between objects A and B, Δ_{exact}: offset in the image plane of the same horizontal stripe projected onto different objects, Δ_{visible}: measurable offset of Δ_{exact}, (b) illustration of Δ_{exact} and Δ_{visible} in a real image.

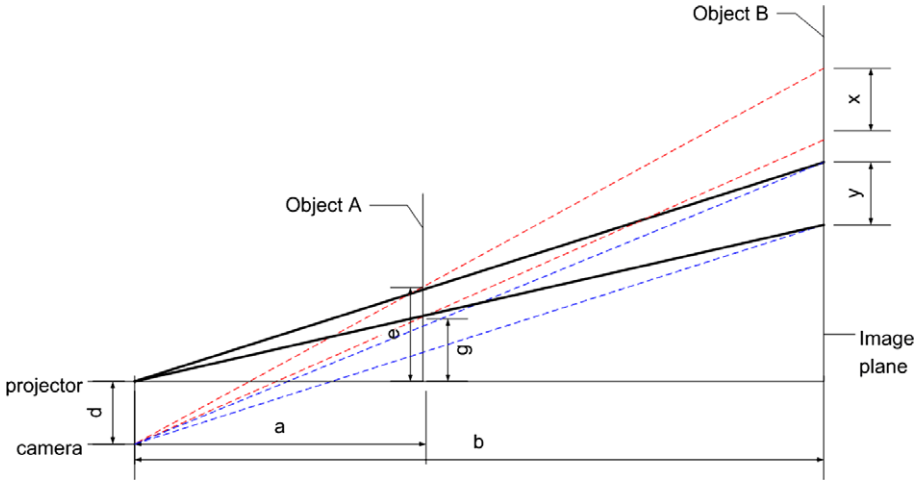


Fig. 5. Cancellation of perspective effect of projector and camera. A stripe projected onto object A has the same width in the image as when projected on the more distant object B.

find candidate depth edges by applying a Gabor filter to the pattern image and accurately locate depth edges using edge information from the white light image.

3.1. Robust computation of pattern images

A pattern image is obtained by differencing the white light and structured light images. This difference image should only yield a horizontal light pattern in order to accurately detect the offsets. Since we use black and white stripes, we apply a thresholding technique to the difference image to get the pattern image. However, thresholding using a single global threshold value does not work well for objects with dark or specular surfaces. Fig. 6(a) and (b) show such examples. The pattern image obtained by a single global threshold does not produce a discernable light pattern in the regions of dark or specular surfaces because there is not much contrast between the white and the structured light images.

To solve this problem, we employ local iterative thresholding. First, we divide the structured light image into small blocks, apply histogram equalization to the blocks, and take differences from the corresponding blocks of the white light images. Second, assuming approximately the same number of black and white pixels in a block, we use iterative thresholding for each block. The iterative thresholding starts with an initial threshold T as the average intensity of the block and then iteratively refines the thresh-

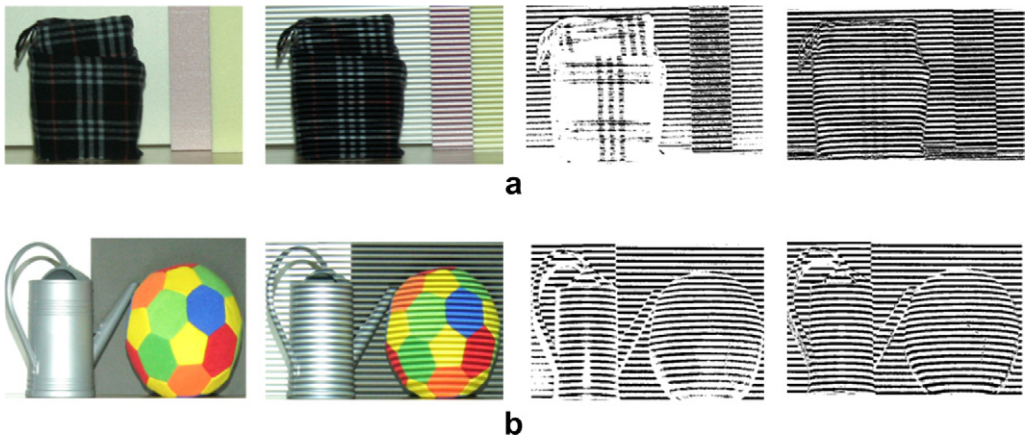


Fig. 6. Robust computation of pattern image: (a) an example of a dark surface object, (b) an example of a specular surface object: from left to right, white light image, structured light image, pattern image from a global thresholding, and pattern image from local iterative thresholding.

old as $T = \frac{1}{2}(\mu_1 + \mu_2)$, where μ_1 and μ_2 are the average intensity values of two partitioned regions. The iteration terminates when μ_1 and μ_2 do not change. As can be seen in the rightmost images of Fig. 6, this local iterative thresholding yields significantly improved pattern images compared to those obtained using a single global threshold.

3.2. The use of Gabor filtering

Since a horizontal pattern can be considered a spatially periodic signal with specific frequency, we can easily identify candidate locations for depth edges by applying a Gabor filter to the pattern image. Fig. 7(a)–(c) show some examples of the pattern image with different instances of offset and their Gabor amplitudes when the horizontal striped pattern is used. By comparing the amplitude responses, the boundaries between patterns differing markedly in their spatial frequency content can be located. Although we use horizontal stripes, it is clear that vertical depth edges can be reliably detected as well as horizontal depth edges.

3.3. Referring to edges in white light image

It is possible to accurately locate depth edges by combining the Gabor filter output and edge information from the white light image. In this work, we use a gradient based technique to detect edges in the white light image, although other methods could also be applied. Fig. 8(b) represents the gradient magnitude of the intensity along the line in

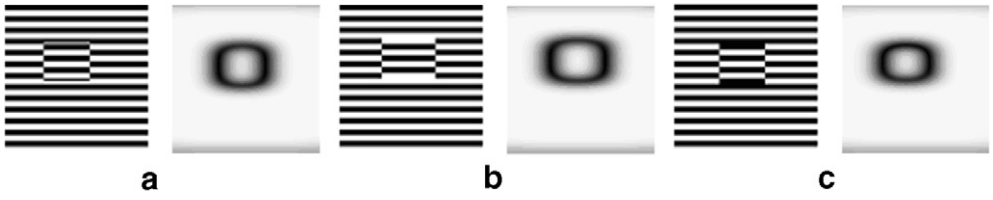


Fig. 7. Examples of Gabor amplitude responses for rectangular depth edges with different offset amounts for upper and lower boundaries: (a) stripes of different colors between upper/lower boundaries and background, (b and c) stripes of the same colors between upper/lower boundaries and background.

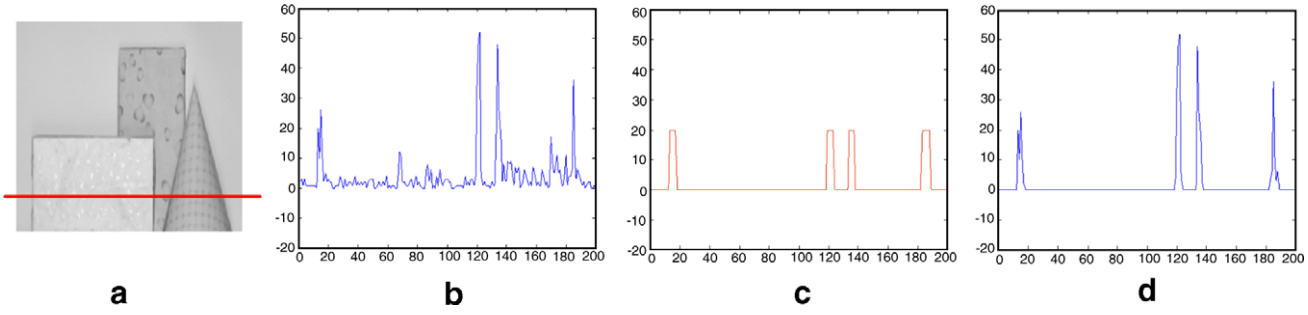


Fig. 8. Location of depth edges using edge information from the white light image: (a) white light image, (b) gradient magnitude along the line in (a), (c) regions of low Gabor amplitude, (d) location of depth edges.

the white light image in Fig. 8(a). We only take the gradient magnitude where Gabor amplitude has low values. Fig. 8(c) illustrates the regions of low Gabor amplitude. The accurate locations of depth edges are obtained by finding peak points of the gradient magnitude in these regions (see Fig. 8(d)).

4. Computation of detectable range of depth edges

We have described that we can easily detect depth edges by exploiting the offset along depth discontinuities in the pattern image. In this section, we set the minimum amount of pattern offset necessary for the reliable detection of depth edges. Then, we compute the width of stripes and detectable range of depth edges given key parameters such as the maximum distance of the detectable range, $a_{\max}$, and the minimum distance between object points, $r_{\min}$. The computation of the width of stripes allows us to employ structured light of optimal spatial resolution that is most appropriate for a given application. Furthermore, by controlling the key parameters, we can collect information about the scene in an active way.

4.1. Reliably detectable offset

As mentioned in Section 2, we can exploit visible offsets of the stripes to detect depth edges using Gabor filtering. In order for an offset to be reliably detectable using a Gabor filter, Δ_{visible} should be above a certain amount. We have plotted the Gabor amplitude response in Fig. 9 as the offset amount, Δ_{visible}, is increased. As expected, the Gabor filter magnitude is low at the boundaries of the offsets. At the offset of $2w/3$, the Gabor filter magnitude decreased markedly. We have set the minimum amount of the pattern offset that is needed to reliably detect depth edges as $2w/3$ where w is the width of a horizontal stripe in structured light image. The threshold value 0.2 is used for the Gabor filter amplitude in the implementation. Additionally, In order to test the sensitivity of Gabor amplitude to oblique angles, we have measured the Gabor amplitude for oblique angles, 10°, 20°, 30°, 40° and 50°. Although the variance of the Gabor amplitudes gets a bit larger about where the oblique plane starts, we have confirmed that the threshold value, 0.2, for the Gabor amplitude can tolerates the variance of Gabor amplitude due to oblique angles from 0° to 50°.

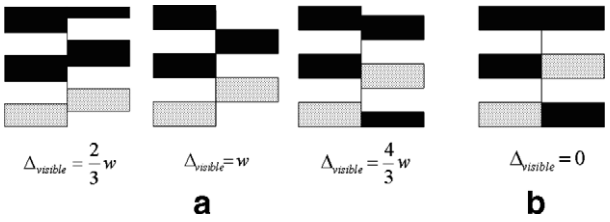


Fig. 10. Examples of Δ_{visible}: For clear display of the offsets, one stripe is particularly highlighted. (a) Detectable Δ from (3): in this case $\Delta_{\text{exact}} = \Delta_{\text{visible}}$, (b) undetectable Δ: in this case, $\Delta_{\text{exact}} = 2w$.

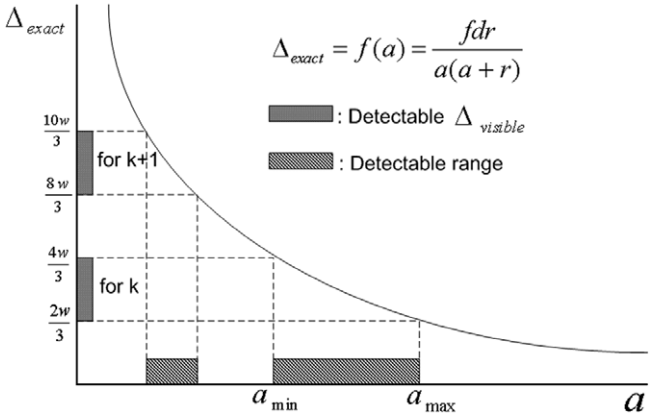


Fig. 11. Detectable range of depth edges according to the distance a to object point A.

The ranges of detectable pattern offset Δ_{visible} can be written as

$$2wk + \frac{2w}{3} \leqslant \Delta_{\text{visible}} \leqslant 2wk + \frac{4w}{3}, \quad k = 0, 1, \cdots \tag{3}$$

Fig. 10 illustrates some instances of detectable and undetectable Δ_{visible}'s when $k = 0$.

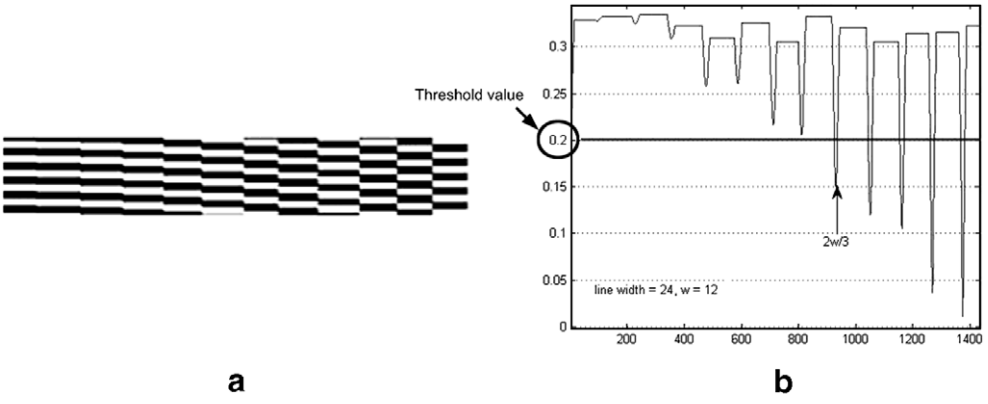


Fig. 9. A period of pattern offset vs. the amplitude response of Gabor filter: (a) a stitched image of horizontal stripes with increased offset, (b) the Gabor amplitude response to the pattern in (a).

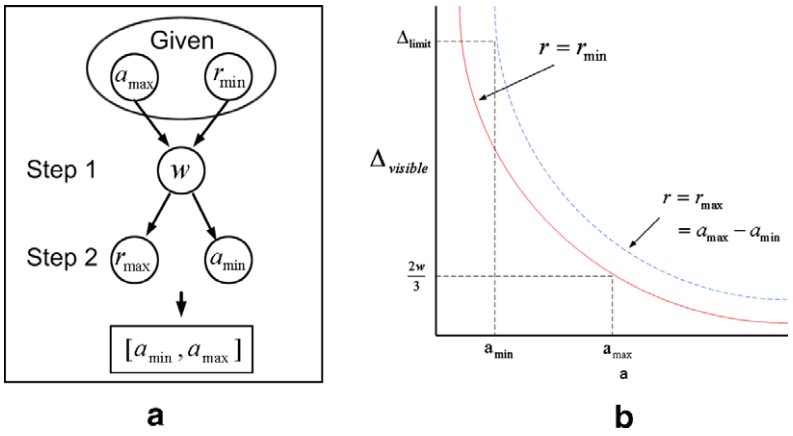


Fig. 12. Computation of the detectable range of depth edges: (a) computation process of $[a_{min}, a_{max}]$, (b) computation of a_{min}.

Given the distance, r, between two object points and the separation, d, between the camera and the projector, Eqs. (1) and (3) produce the exact distance of the foreground object, A, from the camera where reliable detection of offset is guaranteed. Fig. 11 depicts the relationship between Δ_{exact} and a. The marked regions in the horizontal axis, a, represent the ranges of the foreground object point A from the camera that correspond to reliably detectable offset, $\Delta_{visible}$, in the vertical axis. We can see that there are ranges where we cannot detect distortion of the pattern due to the lack of offset, depending on the distance of a depth edge from the camera or projector. Thus, in Fig. 11, $[a_{min}, a_{max}]$ is the longest continuous range where offset detection is guaranteed, that is, depth edges are detectable.

4.2. Detectable range of depth edges: $[a_{min}, a_{max}]$

As shown in Fig. 12(a), the detectable range of depth edges, $[a_{min}, a_{max}]$, is computed in the following two steps:

(1) Setting the maximum distance of the detectable range, a_{max}, and the minimum distance between object points, r_{min}, determines the width of stripes, w, in the structured light image.
(2) This w gives the minimum distance of the detectable range, a_{min}, resulting in the detectable range of depth edges $[a_{min}, a_{max}]$.

Step 1: Determination of the width of a stripe, w, in the structured light

First, we set a_{max} to a distance from the camera to the farthest background. Given r_{min}, w can be computed by (4) which is derived from (1).

$$w = \frac{3fdr_{min}}{2a_{max}(a_{max} + r_{min})}. \tag{4}$$

Thus, given a_{max} and r_{min}, we can compute the ideal stripe width of the structured light.

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

As shown in Fig. 12(b), we can compute a_{min} that corresponds to the upper limit of $\Delta_{visible}$, denoted by Δ_{limit}. The expression for Δ_{limit} is different depending on methods for extending the detectable range of depth edges, which will be discussed in the next section. Let r_{max} denote the maximum distance between object points in the range $[a_{min}, a_{max}]$ that guarantees the occurrence of the distortion along depth discontinuities. Since the distance between any two object points is bounded by $(a_{max} - a_{min})$, we can represent r_{max} in (5).

$$r_{max} = a_{max} - a_{min} \tag{5}$$

Δ_{limit} is obtained when $a = a_{min}$ and $r = r_{max}$ as shown in Fig. 12(b). Thus (1) gives Δ_{limit} as

$$\Delta_{limit} = \frac{fdr_{max}}{a_{min}(a_{min} + r_{max})} = \frac{fd(a_{max} - a_{min})}{a_{min}a_{max}} \tag{6}$$

and

$$a_{min} = \frac{fda_{max}}{fd + \Delta_{limit}a_{max}}. \tag{7}$$

This way, we are guaranteed to detect depth edges of all object points located in the range $[a_{min}, a_{max}]$ and separated in depth no less than r_{min}.

5. Extending the detectable range of depth edges

As previously mentioned, the distortion may not occur or be sufficient to detect depending on the distance of depth edges from the camera or projector. For practical application of the proposed approach, we need to guarantee the occurrence of these regions. Therefore, in this section, we present two methods to guarantee the occurrence of the distortion for a continuous range of object location.

5.1. Using several structured lights

The first method for extending the detectable range of depth edge is based on a single camera and projector setup

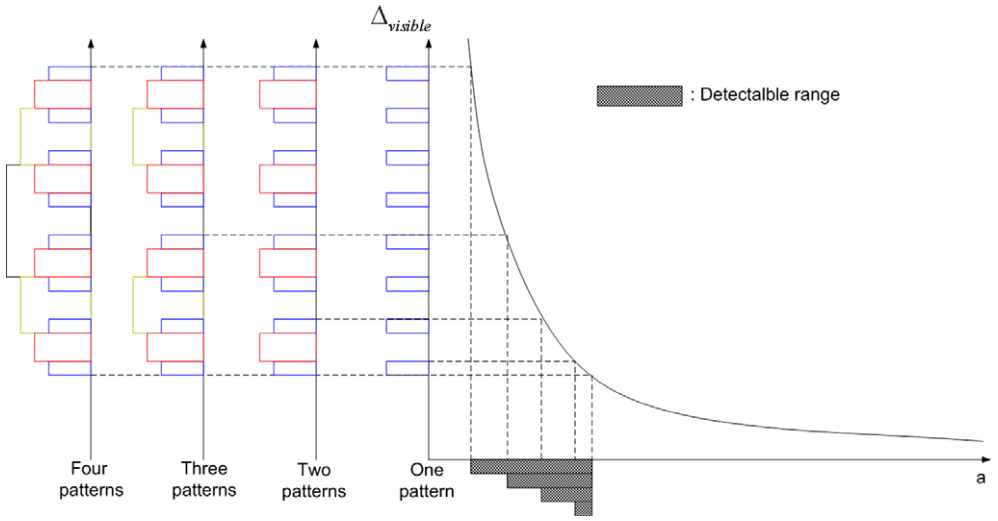


Fig. 13. The detectable range of depth edges can be extended by projecting additional structured light with different stripe widths.

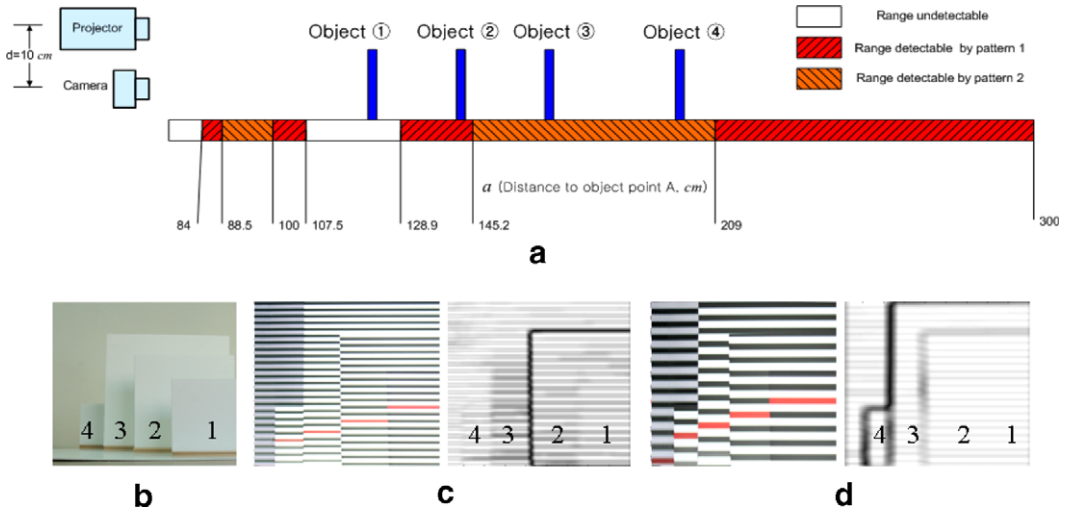


Fig. 14. (a) Detectable and undetectable ranges of depth edges using two structured light, (b) front view of the scene with objects 1–4, (c) result from projecting pattern I, (d) result from projecting pattern II, In (c) and (d), left and right images represent pattern image with offset and Gabor amplitude response, respectively.

that uses several structured light images with different widths of horizontal stripes. As shown in Fig. 13, when we use additional structured light whose spatial period is halved, such as $w_2 = 2w_1$, $w_3 = 2w_2$, $w_4 = 2w_3, \ldots$, the range of detectable offset, Δ_{visible}, is extended. So is the corresponding range, a, of object locations. When n such structured light images are used, the summed range of detectable offset, Δ_{visible}, is expressed as follows:

$$\frac{2}{3}w_1 \leqslant \Delta_{\text{visible}} \leqslant \left(2^n - \frac{2}{3}\right)w_1. \tag{8}$$

Plugging $\Delta_{\text{limit}} = \left(2^n - \frac{2}{3}\right)w_1$ in (7) yields $a_{\min}$ as follows:

$$a_{\min} = \frac{fda_{\max}}{fd + (2^n - 2/3)w_1 a_{\max}}. \tag{9}$$

Fig. 14 shows an experiment that verifies this result when two structured lights, pattern I ($w_1 = 1.8$ cm) and pattern II ($w_2 = 2w_1$) are used. Based on the computed ranges, $[a_{\min}, a_{\max}]$, from the two patterns, we have placed four objects in the following: object 1 in the undetectable range, object 2 in the range detectable by pattern I, and objects 3 and 4 in the range detectable by pattern II. Fig. 14(b) displays the front view of the scene with objects 1–4. Fig. 14(c) and (d) show pattern images (left) and their Gabor amplitude response (right) from the two patterns, respectively. Obviously, the pattern offset for object 1 does not occur in the Gabor amplitude response of Fig. 14(c). However, the offsets along depth discontinuities for objects 2, 3 and 4 are enough to be detectable.

We also show here experimentally that depth edges of all objects located in a continuous range, $[a_{\min}, a_{\max}]$, and apart from each other no less than $r_{\min}$ are detected. We have used three structured light images with different width of horizontal stripes, w_1, $w_2 = 2w_1$ and $w_3 = 2w_2$. Setting

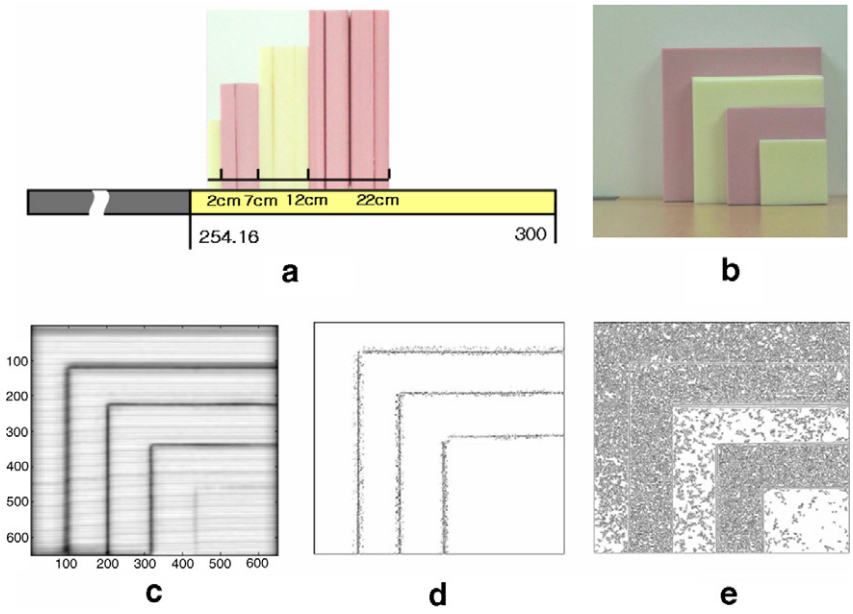


Fig. 15. Detecting depth edges in $[a_{\min}, a_{\max}]$ with depth difference at least $r_{\min}$: (a) continuous detectable range of depth edges, (b) front view of the model, (c) Gabor amplitude map, (d) depth edge map, (e) Canny edges.

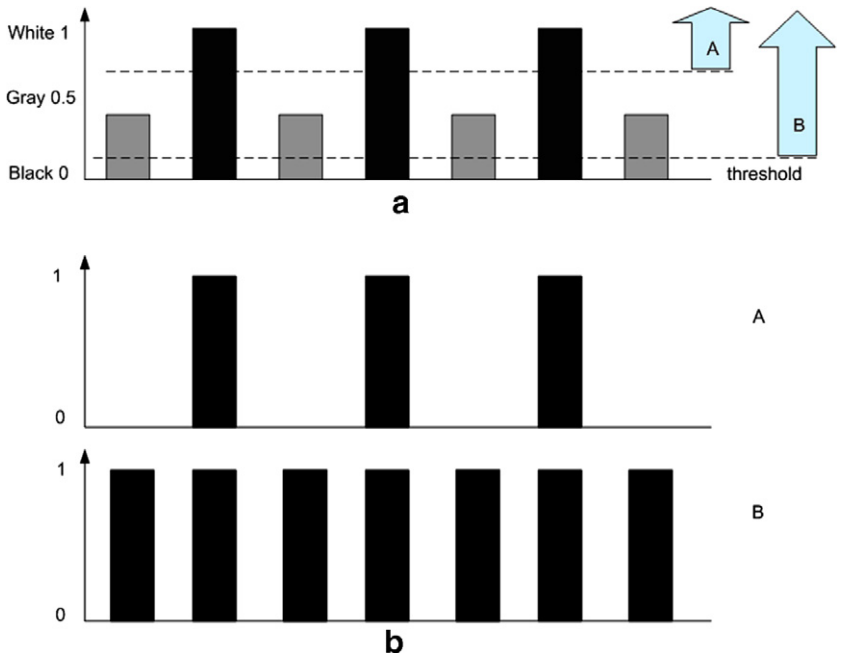


Fig. 16. Two pattern images can be extracted from a three gray level structured light image. (a) Double thresholding of a three gray level structured light image, (b) resulting two pattern images with different width of stripes.

$r_{\min} = 5$ cm and $a_{\max} = 300$ cm, w_1 and $a_{\min}$ are determined as 0.49 cm and 254.16 cm by (4) and (9), respectively. A model object is built of styrofoam to have four step depth discontinuities: 2 cm, 5 cm, 7 cm and 10 cm. The model is located within the range $[a_{\min}, a_{\max}]$ as shown in Fig. 15(a). We can get depth edges as in Fig. 15(d) where no depth edges at least distance $r_{\min}$ in the range [254.16 cm, 300 cm] are missed. We have displayed the output of the traditional Canny edge detector in Fig. 15(e). We have carried out ten more experiments where, for each experiment, the model object is translated away from the camera by 2 cm within the detectable range of depth edges. The results have confirmed that the proposed method is capable of detecting all depth edges with depth difference $\geqslant r_{\min}$ in the detectable range.

The use of many structured light images makes it hard to operate the method in realtime. To get around this problem, we can employ a method to reduce the number of projections by half using a three gray level structural light (black, white, and gray) as shown in Fig. 16(a). Two

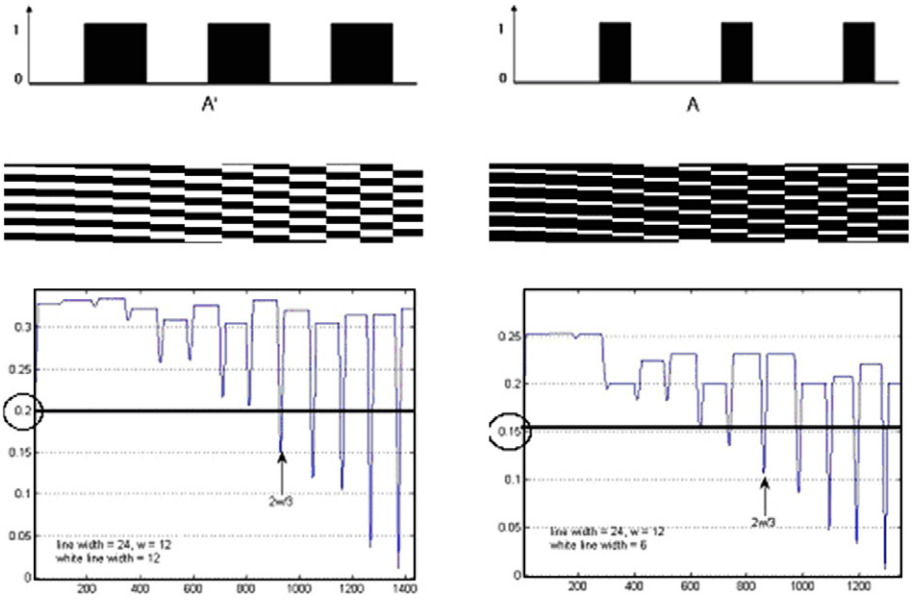


Fig. 17. Gabor amplitude response of pattern images that have the same period but different width of stripes.

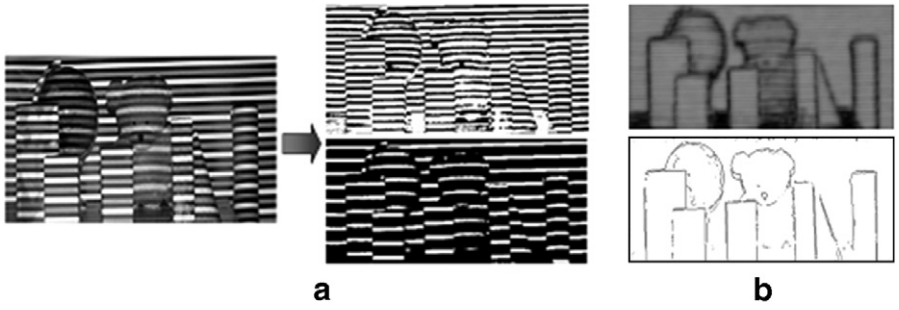


Fig. 18. Detecting depth edges from a three gray level structured light image: (a) two structured light images are extracted from a three gray level structured light image, (b) Gabor amplitude response and depth edge map.

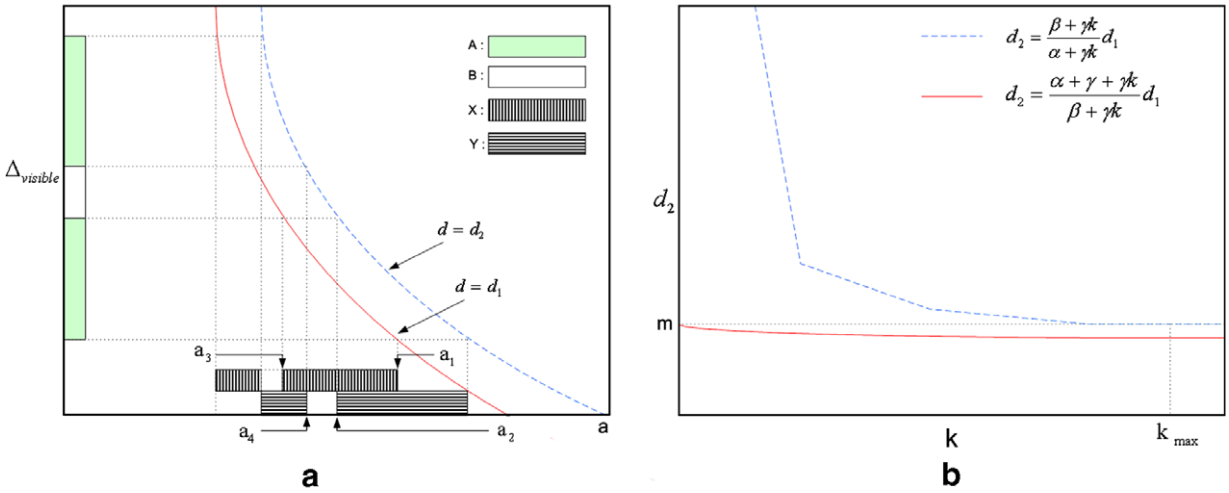


Fig. 19. Extending the detectable range using an additional camera or projector: (a) creating new detectable range by adding a new curve $d = d_2$, (b) determination of the distance between the camera and the projector, d_2.

structured light images A and B that have different width of horizontal stripes can be obtained by applying two threshold values to the three gray level structured light image (Fig. 16(b)).

Let us consider pattern A′ in Fig. 17 that has the same period as pattern A but the same width of white and black stripes. Examining the Gabor amplitude response of A′ and A, the minimum amount of offset, $2w/3$, can still be used by setting the threshold value to 0.15 for pattern A. Hence, we can apply the same analysis to pattern A although the width of white and black stripes is not the same. Fig. 18(a) shows an example of two structured light images extracted from a three gray level structured light image. Using this method, we can reduce the number of projections by half. Fig. 18(b) presents the combined Gabor amplitude response and the final depth edge map that are computed from the two structured light images extracted.

5.2. Using an additional camera or projector

The second method exploits an additional camera or projector. As illustrated in Fig. 19(a), 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 added by partially overlapping with the current range. A and B represent detectable and undetectable ranges in Δ_{visible}; 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 in Y. Similarly, the undetectable range B in Y is overlapped with the detectable range in X. Therefore, if we consider both X and Y, we can extend the range where the detection of the offset is guaranteed. To satisfy the condition $a_1 > a_2$ and $a_4 > a_3$, (11) must hold where the range of Δ_{visible} is expressed as (10):

$$\alpha w_1 + \gamma w_1 k < \Delta_{\text{visible}} < \beta w_1 + \gamma w_1 k, \quad k = 0, 1 \cdots \tag{10}$$

$$\text{where } \alpha = \frac{2}{3}, \quad \beta = \left(2^n - \frac{2}{3}\right), \quad \gamma = 2^n.$$

We can also easily derive the following relation:

$$\frac{\alpha + \gamma + \gamma k}{\beta + \gamma k} d_1 < d_2 < \frac{\beta + \gamma k}{\alpha + \gamma k} d_1. \tag{11}$$

As we mentioned in Section 4.2, in order to minimize $a_{\min}$, Δ_{limit} should be maximized. In addition, in order to maximize Δ_{limit}, k must be maximized where (11) is satisfied. Fig. 19(b) plots (11) when d_1 is constant. As shown in Fig. 19(b), at $k = k_{\max}$, d_2 equals m. This gives d_2 as follows:

$$d_2 = \frac{\alpha + \gamma}{\beta} d_1. \tag{12}$$

Δ_{limit} is the upper limit value of Δ_{visible} where $k = \lceil k_{\max} \rceil$. Thus, we obtain

$$\Delta_{\text{limit}} = \beta w_1 + \gamma \lceil k_{\max} \rceil w_1. \tag{13}$$

Substituting (13) into (7) gives $a_{\min}$ as

$$a_{\min} = \frac{f d a_{\max}}{f d + (\alpha w_1 + \gamma w_1 \lceil k_{\max} \rceil) a_{\max}}. \tag{14}$$

Referring again to Fig. 14, the distance between the camera and the projector $d_1 = 10$ cm. The detectable range can be extended as shown in Fig. 20 by adding an additional camera with $d_2 = 28$ cm. The upper and lower parts of the ranges in Fig. 20 show detectable and undetectable ranges of depth edges using pattern I and II by camera 1 and camera 2, respectively. Undetectable ranges of depth edges become detectable ranges by partially overlapping the ranges from the two curves, $d = d_1$ and $d = d_2$.

6. Experimental results and discussions

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

6.1. Setup 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 5.1 that just employs additional structured light whose spatial frequency is halved. Fig. 21 shows the result of depth edge detection using three structured light images with different width of horizontal stripes.

Fig. 21(a) and (b) display the front and side views of the scene, respectively. All the objects are located within the range of 2.4, 3 m from the camera. Setting $f = 3$ m, $d = 0.173$ m, $a_{\max} = 3$ m and $r_{\min} = 0.1$ m, w_1 and $a_{\min}$ are determined as 0.0084 m and 2.325 m, respectively. That is, the detectable range of depth edges becomes [2.325 m, 3 m] and the length of the range is 0.675 m. Thus,

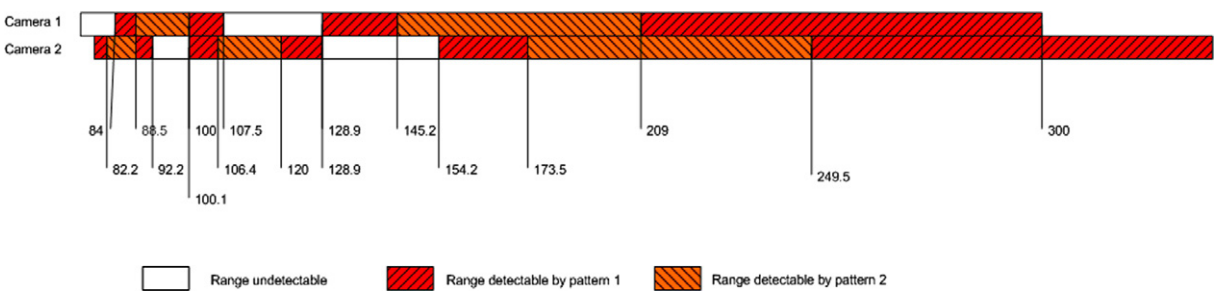
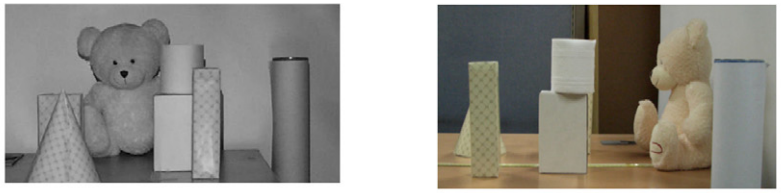
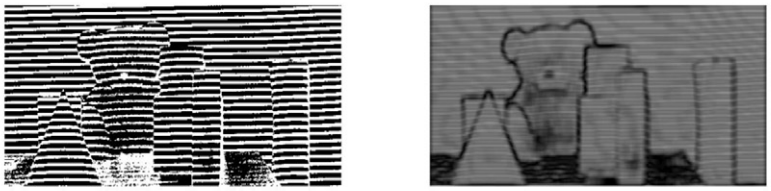


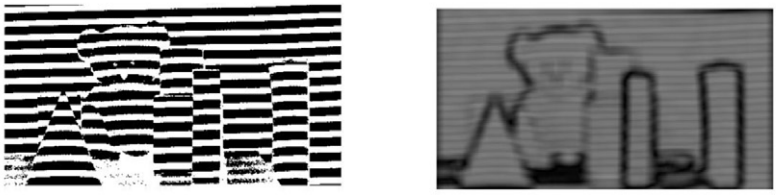
Fig. 20. Extended detectable range of Fig. 14(a) using an additional camera.



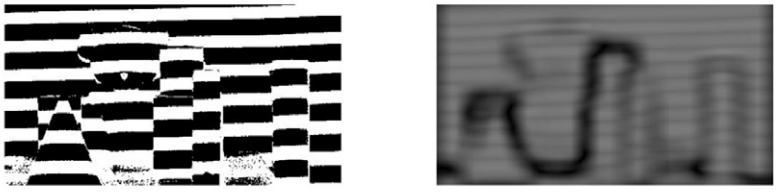
(a) Front view (b) Side view



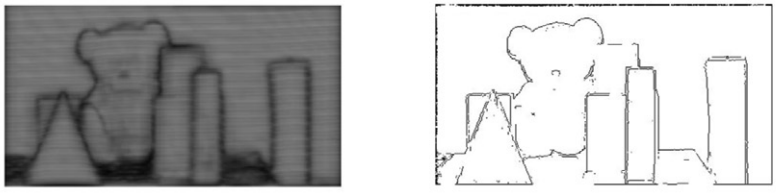
(c) Pattern image and Gabor amplitude map for structured light with $w = w_1 = 0.84\text{cm}$



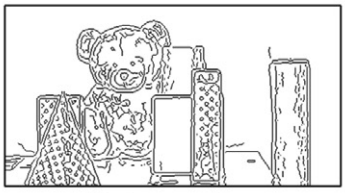
(d) Pattern image and Gabor amplitude map for structured light with $w = 2w_1 = 1.68\ \text{cm}$



(e) Pattern image and Gabor amplitude map for structured light with $w = 4w_1 = 3.36\ \text{cm}$



(f) Combined Gabor amplitude map (g) Depth edges



(h) Canny edges

Fig. 21. Detecting depth edges using a single camera and projector.

the widths of stripes of the three structured lights that guarantee the detection of depth edges in this range are w_1, $2w_1$ and $4w_1$. Fig. 21(c)–(e) show pattern images and their Gabor amplitude maps in the three cases. Each Gabor amplitude map 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. 21(f) where distortion for detection is guaranteed to appear along depth discontinuities in the range of $[2.325\ \text{m}, 3\ \text{m}]$. Finally, we can get the depth edge map as in Fig. 21(g). The result shows that this

method is capable of detecting depth edges of all the objects located in the detectable range. We have also displayed the output of the traditional Canny edge detector for comparison.

6.2. Setup B: 2 cameras and 1 projector

This extends setup A by employing an additional camera and both extension methods. When the same number

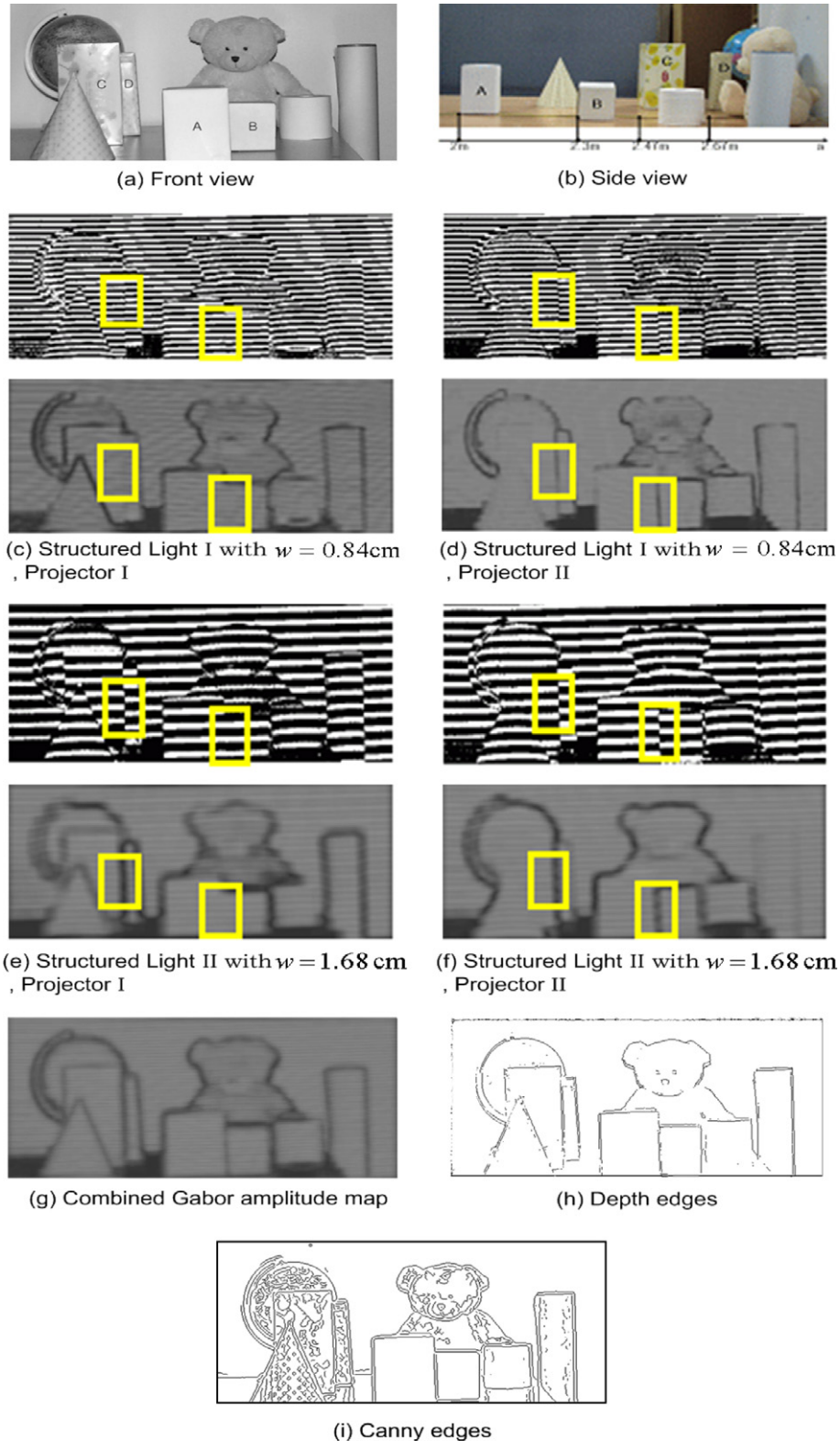


Fig. 22. Detecting depth edges using a single camera and two projectors.

of structured lights is used, we can get a wider range of detecting depth edges than setup A because twice as many pattern images as those in setup A can be created. However, this setup has a disadvantage that the problem of correspondence should be solved in order to combine the results from different cameras (i.e., views). For the result in this case, we can refer to Fig. 22 because setup C that follows yields the same result when the same number of structured lights is used.

6.3. Setup C: 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. Furthermore, there is no correspondence problem because a single camera is used. Fig. 22 shows the experimental result when two structured lights are used for each projector. Let us call these two projectors as Projector I and Projector II, respectively. The two structured lights are denoted as Structured Light I and Structured Light II. Each projector projects two structured lights. This means that we obtain four structured light images all together.

Fig. 22(a) and (b) display the front and side views of the scene. All the objects are located within the range of 1.9–3 m from the camera. When $f = 3$ m, $d_1 = 0.173$ m, $d_2 = 0.207$ m, $a_{\max} = 3$ m, $r_{\min} = 0.1$ m are used, w_1 and $a_{\min}$ are determined as 0.0084 m and 1.9375 m, respectively. Thus, the detectable range of depth edges becomes [1.9375 m, 3 m] and the length of the range is 1.0625 m. Fig. 22(c)–(h) show how the four structured light images play complementary roles to produce a final depth edge map. While Structured Light I from Projector I cannot detect depth edges between objects A and B, Structured Light I from Projector II can. Similarly, as shown in Fig. 22(e) and (f), the depth edges between objects C and D can be detected by Structured Light II from Projector I. These edges cannot be found by Structured Light I from the same projector. Although each case cannot detect all the depth edges of the objects, we can get a complete depth edge map by combining the four Gabor amplitude maps.

In addition, for better visibility of false positives and negatives in Figs. 21 and 22, we have also provided enlarged views in Figs. 23 and 24, respectively. For the given parameters, we have manually marked depth edges in the input image for ground truth. When marking false positives and negatives, location errors within two pixels are ignored.

Most false positives are detected due to two causes. First, regions with dark surfaces or shadow produce errors in detecting a pattern image, resulting in low Gabor amplitude. Thus, depth edges are falsely detected at the locations of large gradient magnitude in these regions of low Gabor amplitude. Second, when strong texture edges are near depth edges, large gradient magnitude due to strong texture can also cause false positives. False negatives (i.e. missing edges) are likely to occur at depth edges of which gradient magnitude is relatively small. In future study for improving the performance of our method, we will look into decision theoretic schemes to overcome these causes.

Fig. 25 shows detection of depth edges in the case of hand gestures for fingerspelling. We have only used a single camera and one structured light image to detect hand configuration.

Fig. 26 compares our method with a commercially available 3D camera, EZSCAN2002 from Solutionix Co. with respect to depth edge detection. EZSCAN is a structured light based 3D camera and uses 10 images for full 3D reconstruction. As can be seen in Fig. 26(d), the regions that are not visible to either the

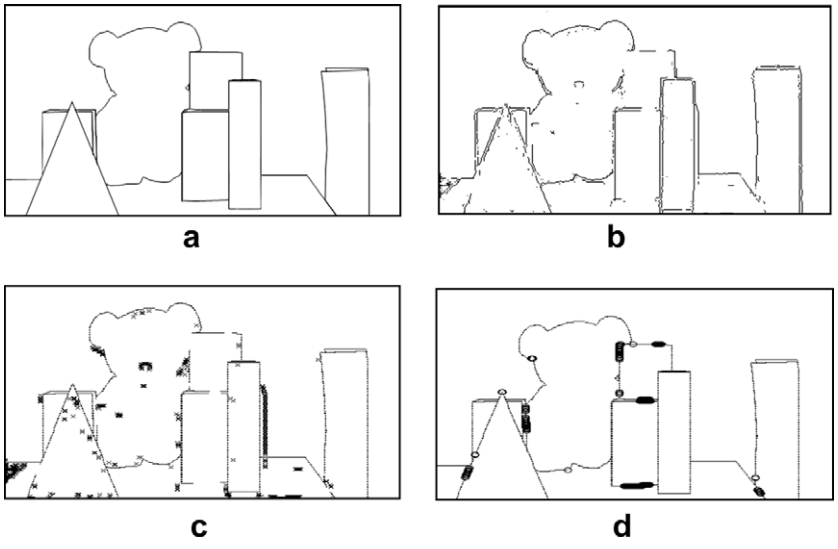


Fig. 23. False positives and negatives: (a) ground truth for the given parameters, (b) depth edges detected, (c) false positives (denoted as 'x') overlaid with ground truth, (d) false negatives (denoted as 'o') overlaid with ground truth.

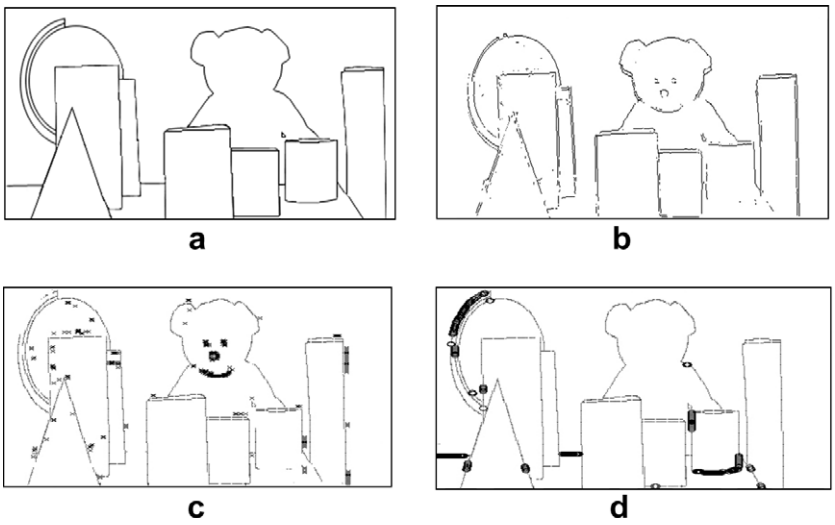


Fig. 24. False positives and negatives: (a) ground truth for the given parameters, (b) depth edges detected, (c) false positives (denoted as 'x') overlaid with ground truth, (d) false negatives (denoted as 'o') overlaid with ground truth.

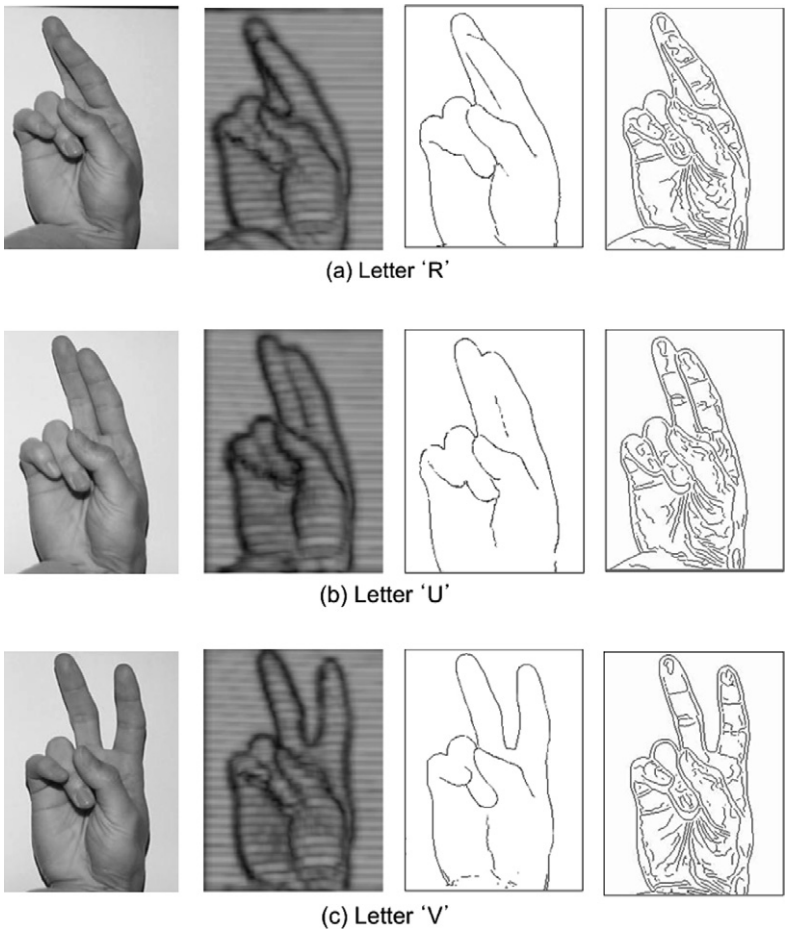


Fig. 25. Detection of depth edges in the case of hand gestures for fingerspelling: from left to right, white light image, Gabor amplitude map, depth edges and Canny edges.

camera or the projector produce holes. Fig. 26(f) shows the result of depth edge detection using a hole-compensated 3D reconstructed image. From the comparison with the result of our method (Fig. 26(c)), we can see that, if the purpose is depth edge detection, our method would be more appropriate.

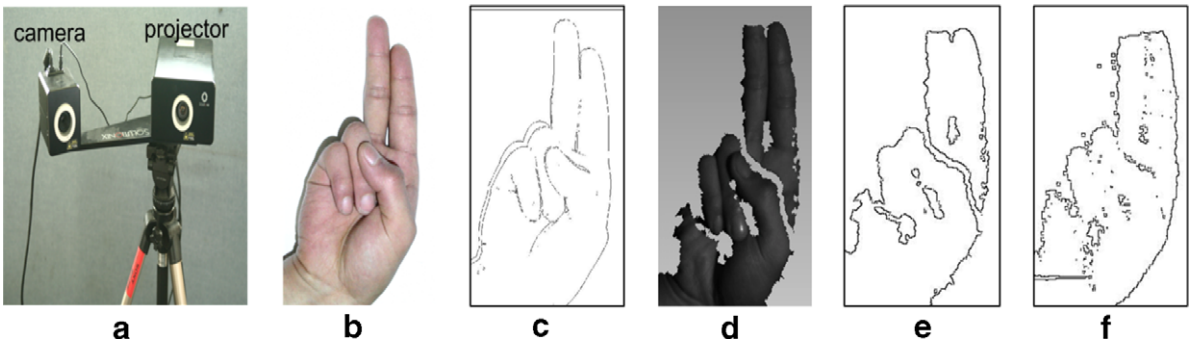


Fig. 26. Comparison of depth edge detection: (a) 3D camera (EZSCAN2002 from Solutionix Co.), (b) hand gesture images for fingerspelling, (c) depth edge detection using our method, (d) depth map from ESCAN2002, (e) depth edge detection using (d) without hole-compensation, (f) depth edge detection using (d) with hole-compensation.

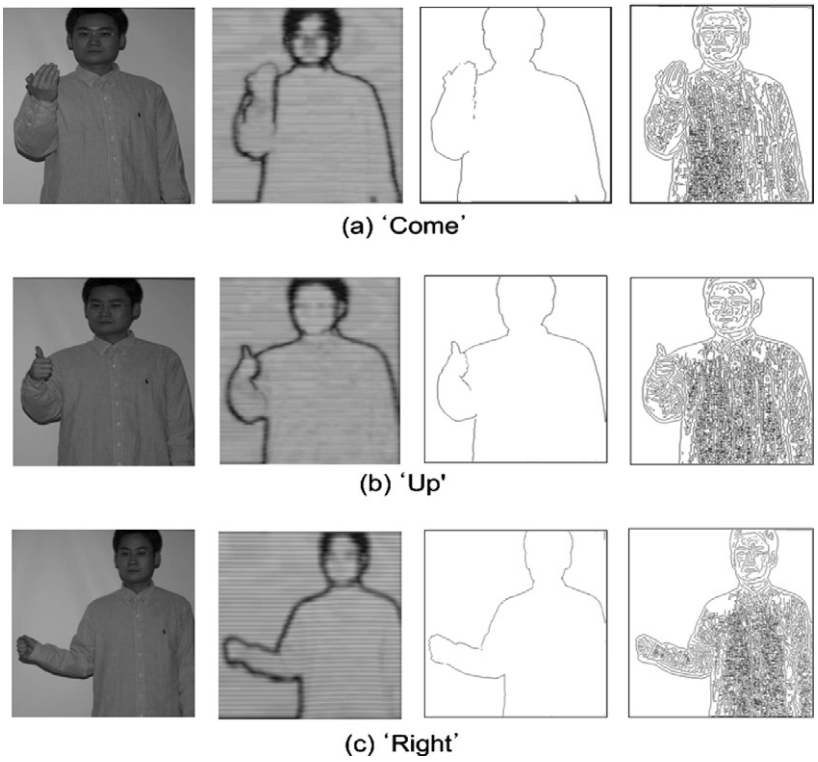


Fig. 27. Detecting human body contours for gesture recognition: from left to right, white light image, Gabor amplitude map, dept edges and Canny edges.

We have also applied our method to human body scenes using setup A. Fig. 27 shows the result of detecting of 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. This result shows that our method can be effectively used to compute silhouettes for human pose analysis [19].

7. 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 three 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.

The setup of one camera and two projectors has advantages of wider distortion range and inexpensive computational cost. We are making efforts to achieve the equivalent performance by implementing it with a single projector and a mirror. In addition, we have observed that infrared projectors show the same distortion characteristics in pattern images. This makes us directly apply the same analysis from LCD projectors to infrared projectors for the application to Human Robot Interfaces. 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.

Acknowledgment

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] I. Weiss, M. Ray, Model-based recognition of 3D object from single images, IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (2001) 116–128.
[2] T.A. Cass, Robust affine structure matching for 3D object recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 20 (1998) 1265–1274.
[3] S. Loncaric, A survey of shape analysis techniques, Pattern Recognition 31 (1998) 983–1001.
[4] T. Frohlinghaus, J.M. Buhmann, Regularizing phase-based stereo, in: The 13th International Conference on Pattern Recognition, vol. 1, 1996, pp. 451–455.
[5] W. Hoff, N. Ahuja, Surfaces from stereo: integrating feature matching, disparity estimation, and contour detection, IEEE Transactions on Pattern Analysis and Machine Intelligence 11 (1989) 121–136.
[6] L. Zhang, B. Curless, S.M. Seitz, Rapid shape acquisition using color structured light and multi-pass dynamic programming, in: The first International Symposium on 3D Data Processing Visualization and Transmission, 2002, pp. 24–36.
[7] D. Caspi, N. Kiryati, J. Shamir, Range imaging with adaptive color structured light, IEEE Transactions on Pattern Analysis and Machine Intelligence 20 (1998) 470–480.
[8] K.G. Harding, Phase grating use for slop discrimination in Moir'e contouring, Proceedings of the SPIE 1614 (1991) 265–270.
[9] M.B. Vieira, L. Velho, A. Sa, P.C. Carvalho, A camera-projector system for real-time 3D video, in: IEEE International Workshop on Projector-Camera Systems PROCAMS, 2005.
[10] S. Zhang, P. Huang, High-resolution, real-time 3D shape acquisition, in: IEEE Computer Vision and Pattern Recognition Workshop on Realtime 3D Sensors and their Uses, vol. 3, 2004, pp. 28–37.
[11] L. Zhang, N. Snavely, B. Curless, S.M. Seitz, Spacetime faces: high resolution capture for modeling and animation, ACM Transactions on Graphics 23 (2004) 548–558.
[12] J.H. Ryu, S.H. Yun, K.G. Song, J.D. Cho, J.M. Choi, S.H. Lee, High speed 3D IR scanner for home service robots, IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences E89-A (3) (2006) 678–685.
[13] R. Raskar, K.H. Tan, R. Feris, J. Yu, M. Turk, Non-photorealistic camera: depth edge detection and stylized rendering using multi-flash imaging, in: ACM SIGGRAPH Conference, vol. 23, 2004, pp. 679–688.
[14] R. Feris, M. Turk, R. Raskar, K. Tan, G. Ohashi, Exploiting depth discontinuities for vision-based fingerspelling recognition, in: IEEE Workshop on Real-Time Vision for Human–Computer Interaction, 2004.
[15] J. Park, C. Kim, J. Yi, M. Turk, Efficient depth edge detection using structured light, in: International Symposium on Visual Computing, LNCS, vol. 3804, 2005, pp. 737–744.
[16] W. Ma, B.S. Manjunath, EdgeFlow: a technique for boundary detection and image segmentation, IEEE Transactions on Image Processing 9 (2000) 1375–1388.
[17] A.C. Bovik, M. Clark, W.S. Geisler, Multichannel texture analysis using localized spatial filters, IEEE Transactions on Pattern Analysis and Machine Intelligence 12 (1990) 55–73.
[18] N. Petkov, P. Kruizinga, Computational models of visual neurons specialized in the detection of periodic and aperiodic oriented visual stimuli: bar and grating cells, Biological Cybernatics 76 (1997) 83–96.
[19] A. Agarwal, B. Triggs, Recovering 3D human pose from monocular images, IEEE Transactions on Pattern Analysis and Machine Intelligence 28 (2006) 44–58.