Active one-shot scan for wide depth range using a light field projector based on coded aperture

Hiroshi Kawasaki, Satoshi Ono, Yuki, Horita, Yuki Shiba
Kagoshima University
Kagoshima, Japan
{kawasaki,ono}@ibe.kagoshima-u.ac.jp

Ryo Furukawa, Shinsaku Hiura
Hiroshima City University
Hiroshima, Japan
{ryo-f,hiura}@hiroshima-cu.ac.jp

Abstract

The central projection model commonly used to model cameras as well as projectors, results in similar advantages and disadvantages in both types of system. Considering the case of active stereo systems using a projector and camera setup, a central projection model creates several problems; among them, narrow depth range and necessity of wide baseline are crucial. In the paper, we solve the problems by introducing a light field projector, which can project a depth-dependent pattern. The light field projector is realized by attaching a coded aperture with a high frequency mask in front of the lens of the video projector, which also projects a high frequency pattern. Because the light field projector cannot be approximated by a thin lens model and a precise calibration method is not established yet, an image-based approach is proposed to apply a stereo technique to the system. Although image-based techniques usually require a large database and often imply heavy computational costs, we propose a hierarchical approach and a feature-based search for solution. In the experiments, it is confirmed that our method can accurately recover the dense shape of curved and textured objects for a wide range of depths from a single captured image.

1. Introduction

One-shot active scanning systems for capturing dynamic scenes have been intensively investigated because of strong demands in various fields, *e.g.*, medical, robotics, games, etc [13, 18]. Previous works mainly used light projectors that radiate a structured light pattern from a single optical center. An advantage of the model is that there is an inherent duality between the geometrical properties of such projectors and the central projection model, *e.g.*, a pinhole or lens camera model, so that the projector in an active stereo system can be formalized as an 'inverse camera' from a passive stereo system. Because of this analogy, the projected pattern image can be treated as an image which is captured by a virtual camera placed at the projector's position, and

most passive stereo algorithms can be directly applied to the image pair (*i.e.*, the pattern image and the image captured by the camera) for stereo reconstruction.

Recently, camera models other than central projection models, *i.e.*, non-central projection models, such as generalized cameras or light-field cameras, have attracted many researchers for their unique possibilities for specific purposes [20, 22, 24, 17]. Especially light-field cameras, which can capture a bundle of incoming light rays from different directions and their intensities for each point of the image plane, are widely studied and commercialized because of their unprecedented capability of controlling the focus at each pixel, which allows to create all-in-focus images [1].

Considering the geometric duality between a camera and a projector, if a pattern projector with a non-central projection model is realized, patterns with novel properties, such as depth-dependent and/or defocus-free projection, are made possible. However, the naive method for constructing a light-field projector in which a large number of projectors are arranged together is known to have several problems, such as optical design, costs and installation [9, 15].

To solve these issues, we propose a novel light field projector that consists of an off-the-shelf video projector with a coded aperture mask attached. The system has the same capability as a densely arranged array of projectors, and the projected patterns change their appearance depending on depth. Such nature is in contrast with a traditional central projection system, where generated patterns are invariant with respect to the depth. By making use of the depth-dependent pattern, the camera and the projector are not required to be set up with a wide baseline, even with no baseline, the depth can be still reconstructed. To leverage both the depth-dependency of the projected pattern and the disparities caused by the baseline between the camera and the projector of our system, we propose a hybrid method fusing light-field projection and active stereo technique.

As for the actual implementation of a light-field projector, we use lines or dots pattern for both an aperture and a pattern for projection. As shown later, by using the pattern, we can generate depth-dependent patterns without blurring

out the high-frequency features. Such defocus free projection is one of our important advantage.

The contributions of the proposed method are:

1. A novel light field projector for defocus free projection is proposed and formalized by combining a coded aperture mask to a standard video projector.

2. A robust depth-dependent high-frequency pattern is designed both for the coded aperture mask as well as the projector, which allows accurate shape recovery.

3. An image-based stereo algorithm which can solve the challenging problem of calibrating the light field projector with complicated distortion characteristics is proposed.

4. Alternative implementation of light field projector using diffractive optical elements (DOEs) is proposed.

2. Related works

In an active stereo system, the video projector is often used as a light source to measure a wide area in a short period of time. To realize fast and accurate acquisition, efficient encoding methods are required and both temporal and spatial approaches have been widely studied [19]. However, as described in the introduction, a light field projector, which cannot be modeled by central projection, has not been utilized for active stereo systems yet. In terms of system setup, several systems have been proposed [9, 15, 7]. Jurik *et al.* proposed a method using a large number of laser projectors to construct a light field directly onto the human retina [9]. Nagano *et al.* extended the technique to make a 2D light field onto a predefined screen [15]. One severe problem of these methods is that they require many laser projectors for the system. Hirsch *et al.* proposed a method using lenticular lenses inside the optics of a video projector [7]. However, the resolution of the system tends to be low and only a narrow angle of light field can be constructed. In this paper, we propose a mask based light field projector. Although a similar idea has been already proposed for light field cameras [23], it has not been applied to projectors yet.

The configuration of the mask-based light field projector is the same as a video projector with a coded aperture mask. In terms of *a video projector with a coded aperture mask*, this has been studied for various purposes, however, there are no previous techniques using coded aperture for active stereo (structured light). For example, Grosse *et al.* put a coded aperture in the video projector to mitigate the defocus effect at the projection [5]. Girod *et al.* used asymmetric aperture to distinguish the forward and backward blur for depth from defocus (DfD) [4]. Moreno-Noguer *et al.* put a small circular aperture to realize DfD [16] and Kawasaki *et al.* put a coded aperture on a video projector to improve accuracy and density on DfD [10, 11]. However, for all the techniques, depth range is limited because defocus blur
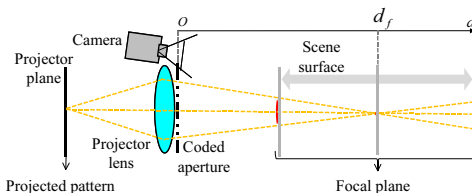

Figure 1. Optical configuration. d_f is a depth of the focal plane.

Figure 2. Actual optical system. (a) Setup with a short baseline and (b) coded aperture installed on the projector lens.

increases rapidly. In contrast, since our system is designed not to defocus, depth range is significantly broadened.

There have been several attempts to achieve the same purpose and to enlarge the depth range in traditional active stereo. The typical solution is to use an aerial laser light source that has been shown to have successful results [13]. However, satisfying requirements of reconstruction density, precision, safety and usability is still an open problem. Mohit *et al.* proposed an active stereo system that decreases the effect of defocus blur by projecting several special patterns based on frequency analysis [6], however, the pattern information degrades rapidly due to defocus, thus making the expansion of possible depth range limited. Masuyama *et al.*, proposed a DfD method that projects multiple patterns along the same optical axis with different focal lengths, which can overcome the above stated problems [8]. However, the complexity of sharing the same optical axis and the decreased contrast of multiple overlapping patterns make practical construction difficult. Zhang *et al.* proposed a method for projecting different patterns and successfully reconstructed a high density depth map by analyzing the captured defocused image set [25]. Achar *et al.* proposed a method projecting a pattern with different foci to enlarge the possible depth range [2]. However, those approaches require multiple images to be captured, which make the application range limited. Our technique can recover the shape from a single image without aforementioned problems.

3. 3D reconstruction using light field projector

3.1. System configuration and algorithm overview

The system setup is similar to the common active stereo setup as shown in Fig. 1 and Fig. 2(a). A projector and a camera are placed with a certain baseline, and a light pattern is projected on the object. The difference from conventional systems is the coded aperture placed over the projector lens to realize a light field projection as shown in Fig. 2(b).

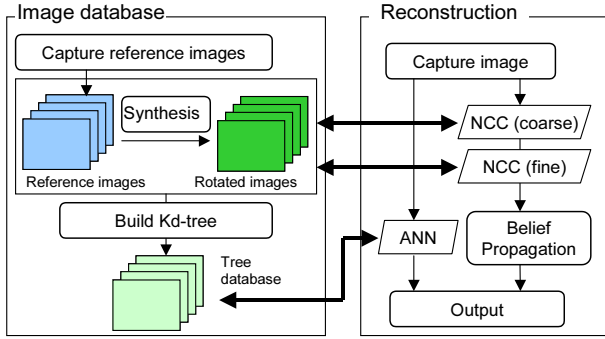As shown in Fig. 2(a), the camera is placed near the pro-

Figure 3. Overview of the reconstruction algorithm. This process is executed for both coarse and fine steps.

Figure 4. Projected patterns with high frequency aperture and circular aperture.

jector lens so that the distance from the camera to the target becomes approximately the same as the distance from the projector. Similar to the projector, the camera has its own depth of field, however, we do not discuss the effect, because camera defocus is typically much weaker than the projector's and it could be generally ignored. However, for applying the technique to wider depth ranges, a solution will be sought in our future research.

Fig. 3 shows the algorithm overview. The technique mainly consists of two parts: image database creation and shape reconstruction. Note that the image database creation is an offline process required only once. To this end, reference images are captured by changing the depth of a planar board with known position on which the specially designed pattern is projected. The reason why we take an image-based approach is explained in the Sec. 4.1. Since the surface normals may not always be parallel to the view direction, stereo matching will fail if the angle difference between the normals on the object surface and reference plane is large. We solve the problem by synthesizing slanted planes from captured images as explained in Sec. 4.2.

During the shape reconstruction phase, we capture the target object by projecting the same pattern and doing stereo matching between the captured image and the image database. To recover the shape of arbitrary objects, small patches from the captured image are compared to the reference images, and the depth which gives the highest correlation is selected. It should be noted that in principle any pattern can be used in active stereo. Since our proposed pattern consists of only vertical lines as described in Sec. 3.3, we use horizontally long rectangular window for matching. In our experiments, we use 8x4, 16x4 and 32x4 pixels windows for evaluation. As for matching algorithm, because of the brightness changes caused by the changing distance to the target surface, materials and normal directions, a scaling invariant technique is required; e.g., normalized cross correlation (NCC), etc. Since NCC computation for all depths requires large memory and computational costs, we introduce two solutions such as hierarchical matching approach and approximate nearest neighbor (ANN) search technique (Sec. 4.3).
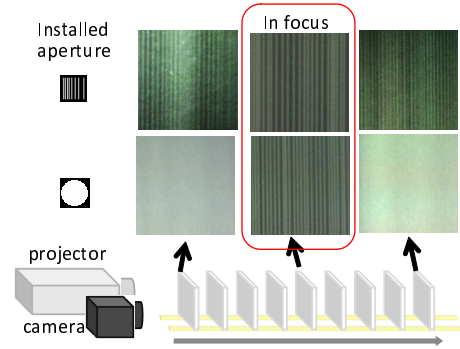
3.2. Light field projection using a coded aperture

In a normal setup with a traditional projector, the projected pattern, which is a convolution of the pattern image and the aperture shape, rapidly blurs out, eliminating high-frequency details. Conversely, we propose to preserve high-frequency patterns while we keep the total amount of light energy as large as possible. Since a convolution of high frequency patterns keeps high frequency, lines or dots for both the pattern on the projector plane and the shape of the coded aperture can be a solution. Furthermore, such configuration has another important feature: the set of rays generated by the convolution of the aperture and the projector pattern forms a light field, realizing depth-dependent pattern projection. Such depth-dependency adds rich new features for depth estimation by altering the patterns depending on the distance. Fig. 4 shows the real patterns generated with a projector with the coded aperture of slit pattern and those of a normal projector with circular aperture for comparison. As shown in the figure, high frequency patterns are preserved with our pattern for all ranges, whereas patterns are rapidly blurred out with circular aperture.

Fig. 5(a) shows an example of how the convolution of high frequency patterns constructs the light field in the space. In the paper, the features of the pattern on the projector plane are composed of lines or dots, which are shown as green points (f_1, f_2 and f_3) in the figure. Similarly, the aperture also consists of lines or dots. In this setup, the projected light becomes a set of sharp rays. The red lines in the figure are rays that are emitted from a point in the projector plane, are then refracted by the lens, go through the aperture mask (h_1 and h_2) and illuminate the target surface. The projected patterns are shown as blue points.

Let the pattern on the projector plane be I_p, the aperture shape be I_a, the distance between the aperture to the focal plane be d_f, and the depth-dependent point spread function (PSF) at depth d be $I_a(d)$, where d is measured from the aperture plane. Theoretically, $I_a(d)$ can be calculated by geometrically scaling I_a by factor of $\frac{d_f - d}{d_f}$ (i.e., $I_a(0) = I_a$ and $I_a(d_f) = \delta(0)$). Then, the projected pattern observed from the optical center of the projector approximately be-
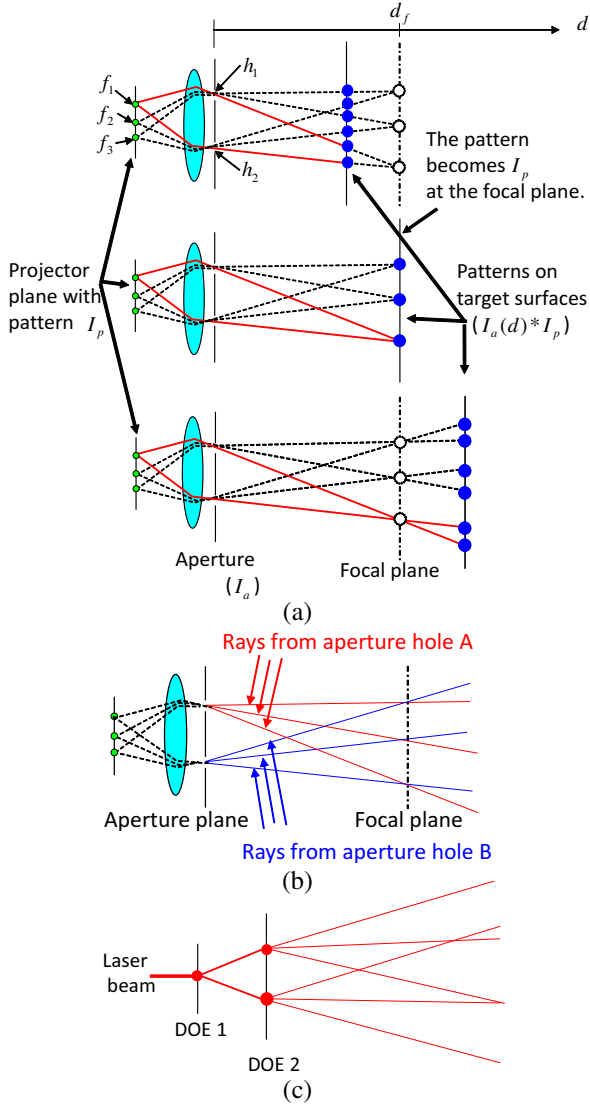
Figure 5. Projection of light field. (a) Pattern projection with coded aperture for different target depths, (b) rays from different aperture holes, and (c) alternative light field projector using DOEs.

comes the convolution $I_a(d) * I_p$ [1]. In designing both of the aperture shape I_a and the projected pattern I_p, we aim to preserve the high-frequency components of the projected pattern, as described in Sec. 3.3.

In Fig. 5(a), there are three features (f_1, f_2 and f_3 in the figure) on the projector plane, and two holes (h_1 and h_2) on the aperture. Then, the number of features on the projected plane becomes generally $3 \times 2 = 6$, except for the overlaps of multiple features. As the distance from the projector to

---

[1] Rigorously, the projected pattern in the 3D space is $I_a(d) * I_p(d)$, where $I_a(d)$ is a geometrically scaled image of I_a by factor of $\frac{d_f - d}{d_f}$, and $I_p(d)$ is a geometrically scaled image of I_p by factor of $\frac{d}{d_f}$. For d near d_f, $I_a(d) * I_p(d)$ can be approximated by $I_a(d) * I_p$ scaled by factor of $\frac{d}{d_f}$, and the observed pattern from the point of $d = 0$ can be approximated by $I_a(d) * I_p$.

the target surface changes, the projected patterns change. Note that, if the target plane is in focus ($d = d_f$), all the rays from each slit will overlap, and the resultant pattern is the same as I_p.

In Fig. 5(b), rays are classified by the aperture holes that the rays go through. This is analogous to placing a small projector at each aperture hole. If the aperture hole A is large, the rays from the hole A will be blurred. Thus, we make the size of the aperture holes small. To compensate the decrease in light intensity, we increase the number of aperture holes to obtain a higher total light energy. From this perspective, using a line aperture pattern, which can be considered continuously aligned dots, is advantageous.

The light field that can be generated by the proposed projector has some constraints, such as, the rays are focused at the focal plane $d = d_f$ so that the projected pattern becomes the same as I_p. Similarly, for the plane $d = 0$, the pattern becomes I_a. The light field should be designed under such constraints.

In this paper, we also show that a similar light field projection is possible using two diffractive optical elements (DOEs) as shown in Fig. 5(c). An advantage of such a device is a smaller energy loss. We experimented on this configuration only for uniform repetitive patterns, and further experiments will be a part of our future work.

3.3. Pattern design of projected pattern and coded aperture

For stable depth estimation, a combination of the projected pattern and coded aperture should be designed carefully to present distinctive features on the target surface as well as considering light energy efficiency. In the field of DfD and deblurring, isotropic 2D broadband patterns are commonly used for aperture design [27, 26, 23, 12, 21]. In contrast with active depth measurement, uniqueness of the projected pattern on the horizontal axis has priority over the vertical information. Moreover, 2D patterns tend to lose contrast because of the law of averages; Specifically, the convolution operator in the defocusing effect $I_a(d) * I_p$ acts as an averaging filter and the more elements we have in the projected and aperture pattens I_p and I_a, the lower contrast we observe on the target surface because of the central limit theorem. Therefore for the depth measurement, we should consider the following conditions:

1. Horizontal spatial frequency of the projected pattern on the target surfaces should be broadband.
2. The number of elements of the projected and aperture patterns for convolution should be small to keep the contrast high.
3. The number of elements of the projected and aperture patterns should be dense to keep a large total energy.
4. The projected pattern on the target surface should be unique on the horizontal axis.

Since some of the conditions are competing, we investigated the trade-offs involved by conducting simulations for multiple cases of I_a and I_p, by generating images of $I_a(d) * I_p$ for various d to find the best parameters. Examples of the generated patterns are shown in Fig. 6 (middle column). From condition **4**, it is desirable that these patterns have distinctive characteristics for different d. To analyze this, the normalized correlations between the simulated patterns are calculated and visualized. The results are shown in Fig. 6 (right column). Large diagonal (*i.e.*, for the images of right column of Fig. 6, the values of diagonal lines that run from the left bottom corners to the right top corners) and low off-diagonal values indicate high autocorrelation and uniqueness at different depths. From the simulation, we found that the randomly arranged vertical stripe patterns for both projected pattern and coded aperture are the best for the depth measurement as shown in Fig. 6 (case D). Thus, in this paper, we use the combination shown in case D. To minimize the effect of texture, we prepare three different random lines patterns for each color channel.

In this combination, since all the projected pattern features become vertical lines, vertical baseline of the camera with respect to the projector does not generate parallax. By positioning the camera with horizontal baseline, the observed patterns include parallax effects.

4. Implementation and method details

4.1. Data sampling for image based approach

There are mainly two reasons why we propose an image-based approach. The first is the difficulty of calibrating light field parameters, since the observed patterns are a convolution of two patterns and decomposition is usually not an easy task. The second is complex lens distortions, which can be usually ignored or approximated by a simple distortion model for passive stereo method, should be taken into account in our method. For example, if a mathematically ideal lens is assumed, PSF is shift-invariant, however, in reality, actual lenses have numerous imperfect characteristics, *e.g.*, field curvature, coma and astigmatism, and they make the PSF not only shift variant but also rotationally asymmetric. Further, the PSF of the frontal defocus differs from the backward defocus when the lens has a spherical aberration. Chromatic aberrations also degrade the PSF intricately. These imperfections of the lens clearly appear and affect the results in our method, since all light rays from the lens are independently utilized with a light field projector. Moreover, both phenomena are mixed together and only an integrated image is observed, so decomposition and parameter estimation become more difficult.

Considering this scenario, an image-based approach is a simple solution because these factor affect equally input and reference images and cancel each other. Moreover, we
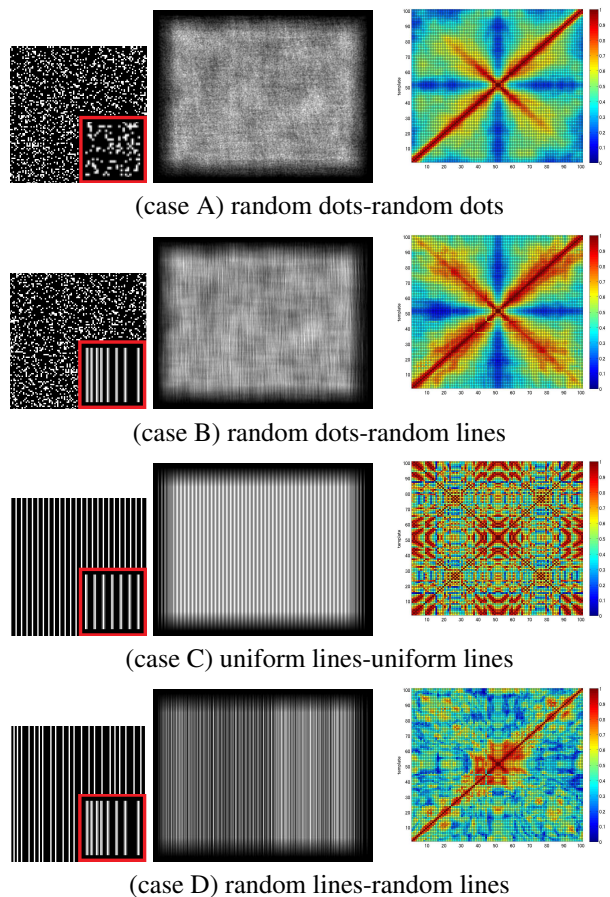

(case A) random dots-random dots

(case B) random dots-random lines

(case C) uniform lines-uniform lines

(case D) random lines-random lines

Figure 6. Selected combinations of simulated patterns I_p and aperture I_a, and visualized correlation matrices between the patterns and depth d. (**Left column**) Projected patterns I_p outside the red square, and coded apertures I_a inside the red square, (**Middle column**) simulated results $I_a(d_s) * I_p$, and (**Right column**) visualized correlations. For the correlations, the horizontal axis is d_x, the vertical axis is d_y, and the origin is at the bottom left corner. Color at (d_x, d_y) means a sampled normalized correlation between $I_a(d_x) * I_p$ and $I_a(d_y) * I_p$. The center of the correlation image corresponds (d_f, d_f).

have a chance to exploit such effects to make the depth estimation more robust. For example, spherical aberration has the potential to disambiguate the frontal and rear defocus.

The actual sampling process is as follows. First, we set a white planar board perpendicular to the projector's optical axis and images are captured by using a motorized stage. Dense sampling is required because our projected pattern drastically changes its appearance with small depth changes. Note that higher precision than the sampling intervals can be naturally realized with our method because reconstruction is based on stereo, which usually achieves sub-pixel accuracy with a window matching approach.

The advantages of our image-based framework are summarized as follows: (1) multiple depth cues such as disparity and defocus can be handled by a unified algorithm, (2) most nonlinear phenomena such as optical aberrations
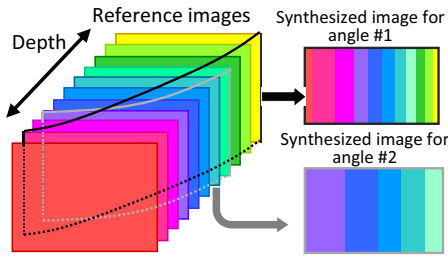
Figure 7. Synthesizing the image with arbitrary rotation angles.

can be naturally canceled, (3) complicated calibrations, for example, PSF modeling or epipolar analysis of defocused images are not necessary.

4.2. Slanted plane adaptation

Since the surface normal of the object is not always parallel to the optical axis of the projector, the projected patterns are naturally distorted. The naive solution is to capture many reference images by rotating the plane for all possible angles, however, it increases capturing time and data storage. As a solution, we synthesize images for the required orientations from the captured image set online. Fig. 7 shows the process of synthesizing images for two different rotation angles. Note that since all the matching steps are applied after rectification, only a single rotation axis is needed to be considered since vertical lines are being considered as features.

4.3. Efficient stereo matching with hierarchical approach and a feature based search technique

Our image-based method requires a large reference image database which consists of captured images and synthesized images for various rotation angles (in our case $-60, -30, 0, 30, 60$ degrees). Furthermore, given such a large database, the computational cost of template matching becomes enormous. In order to solve the problem, this paper proposes two algorithms for depth estimation using a hierarchical template matching and approximate nearest neighbor (ANN) search [14]. The former approach prioritizes depth estimation accuracy rather than the processing speed; the latter reduces the processing time with a little sacrifice in accuracy.

For hierarchical matching, the coarse level solution of NCC matching is searched first with large depth intervals and low spatial resolution, and then, the fine level solution is searched around the low level solution with fine resolution of depth and images. It should be noted that we cannot drastically decrease the number of depth samples for the coarse level because the pattern changes distinctively for small depth changes. In our experiment, 2.5mm and 0.5mm intervals are used for coarse and fine levels respectively. Finally, since the result presents small noise, we apply a global optimization algorithm based on belief propagation to remove the error [3]. The matching costs of NCC for the
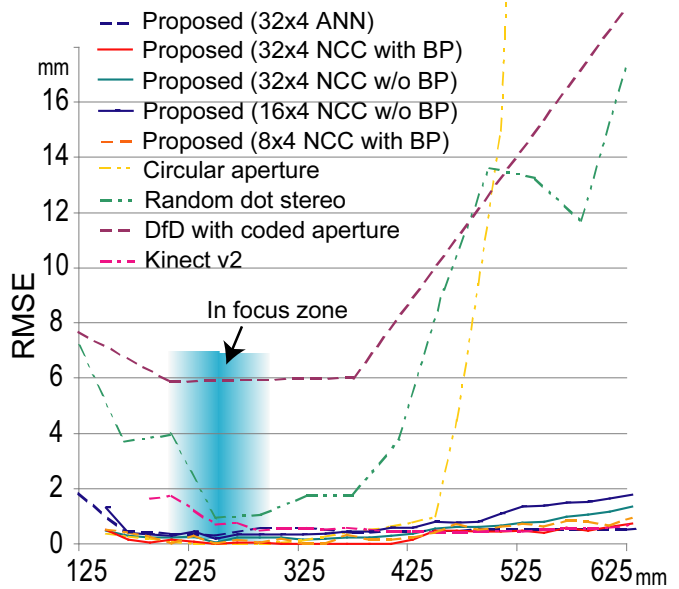

Figure 8. Depth estimation results. Brown broken line represents DfD result [10], green dotted line represents simple random dot stereo result, yellow dotted line represents the result using a circular aperture with the same algorithm of ours and red dotted line represents the result using a time-of-flight sensor (Kinect v2). Note that we make the circular aperture size to be the same as total aperture size of our slit aperture.

searched solution is used for the energy term of belief propagation, with the regularization term for improving spatial smoothness of the solution.

For the ANN approach, first, we make a compact feature vector from a matching window, and then, we build a tree structure from all the feature vectors. To make a compact feature vector representation, intensities of pixels along a vertical line in the matching window are integrated. Note that pixel values are normalized in advance to mitigate the scaling effect. The length of the feature vector can be further shortened by averaging the vector with a certain length. In our experiment, a matching window of 32×4 pixels is first integrated vertically to produce a 32D feature vector, and then, averaged every 4 pixels each to make an 8D feature. Then, those feature vectors are stacked into Kd-tree using depth value as an index. In the reconstruction step, the Kd-tree is searched directly from the compact feature vector constructed from the window of the camera image, and thus, the reconstruction performance is much improved.

5. Experiment

5.1. Plane estimation for evaluation

The first experiment was conducted by using the optical system shown in Fig.2. Images were captured by shifting the target screen placed on a motorized stage. Because of the limitation of the length of the motorized stage, we put a close-up lens to change the scale as to be 1/3 of real length. With this scale, the motion range of the screen is

Table 1. ANN calc. time.

Data creation	108.8
Search	3.2

Table 2. NCC calc. time.

Coarse search	39.1
Fine search	83.7
Total	122.8

(a) Checker board (b) Wooden board (c) Dappled texture (d) News paper

Figure 9. Texture samples used for experiment.

(a) Checker board (b) Wooden board (c) Dappled texture (d) News paper
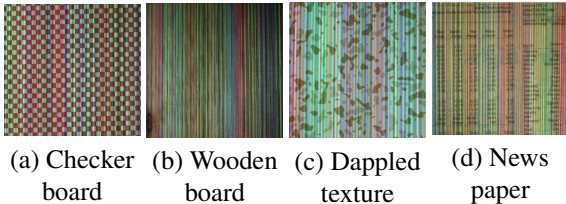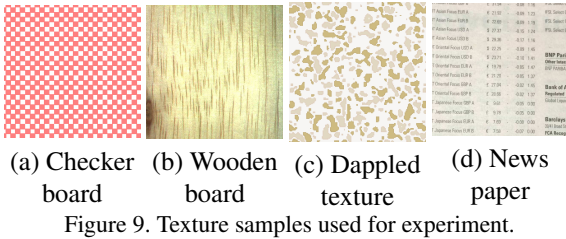
Figure 10. Captured images of the board with various textures.

(a) ANN

(b) NCC

- Reference Board
- Glossy board
- Wooden Board
- Checker pattern
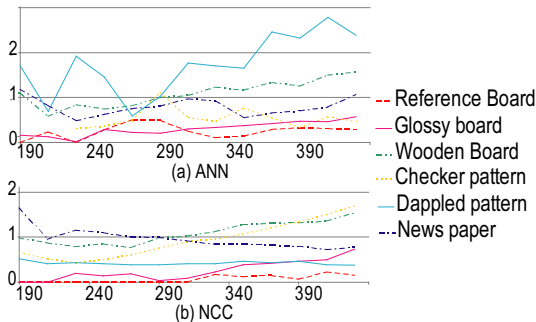- Dappled pattern
- News paper

Figure 11. RMSE of estimated depth of textured planes using (a) ANN and (b) NCC for matching.

150mm-625mm from the projector and the camera, in-focus distance is 250mm ±100mm for the projector, the reference plane capturing interval is 0.5mm, the target plane capturing interval is 10mm, and the matching window sizes were 8x4, 16x4 and 32x4 pixels. The depth value was estimated by using the proposed method and other methods for comparison with RMSE shown in Fig.8. In the graph, we can observe our methods including all the window sizes and time-of-flight sensor can recover the correct depth for all the ranges, whereas others rapidly decrease their accuracies when they enter the defocus range. We can also confirm that even if an accuracy of ANN is almost same as NCC, it drastically reduces the processing time as shown in Table 1 and 2.

5.2. Accuracies on textured object

Next, we evaluated our method on textured objects. Checkerboard pattern, glossy board, wooden board, dappled pattern and newspaper were tested. Sample textures are shown in Fig.9. The configuration of the camera and the projector is the same as described in Sec.5.1. The cap-

tured images in Fig.10 show that the projected patterns are strongly affected by textures. Fig.11 shows the RMSE results of both ANN and NCC. For both technique, the accuracy decreases if there is a texture on the object, however, the proposed method can still estimate depth with high accuracy even if the projected patterns are diminished and some of them are divided into several parts by color differences. We can also confirm that the quality of NCC is slightly better than ANN, and thus, the rest of the experiments are done by hierarchical NCC.

5.3. Accuracies of slanted planes

(a) input image (b) surface direction (c) top view (d) 3D shape

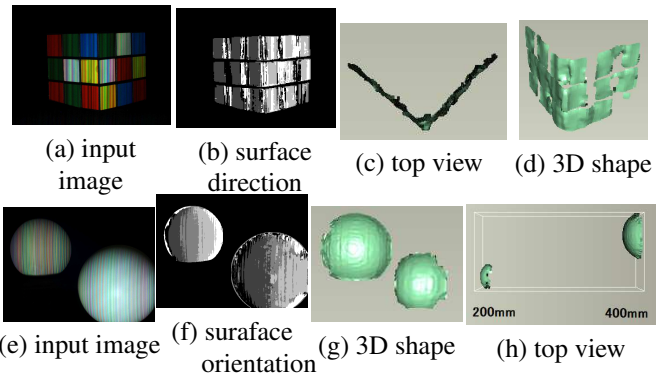(e) input image (f) suraface orientation (g) 3D shape (h) top view

Figure 12. Shape reconstruction result of slanted planes and curved surfaces.

To confirm the effectiveness of the slanted image synthesis technique for arbitrary rotation angles, we reconstruct a cube-shaped object and a sphere-shaped object. From the sampled images of the reference plane, virtually rotated reference images with rotation angles of -60 ,-30, 0, 30, and 60 degrees were synthesized. From the real and synthesized image set, the shapes of the objects were reconstructed. The results are shown in Fig. 12, including the estimated directions of the normal vectors. As the reconstructed shape and the visualized normal directions show, the positions of the points were accurately estimated, although the normal directions were just roughly estimated. For the cube-shaped object, we extracted 3D points and calculated the RMSE of the 3D points by fitting them to the dominant plane. The value was 2.9mm where the distance from the camera to the cube was about 300mm. The reason why the result is worse than previous experiments is that each face of cube consists of different color blocks with black border lines, and it is more challenging object than previous cases.

5.4. Arbitrary shape and wide depth range test

We estimated the depth of more generic objects. First, we measured the objects placed widely apart as shown in Fig.13(a) and (b). Four objects are placed about 150mm, 300mm, 450mm and 620mm from the lens, respectively. Fig.13 (a) shows the captured image with the projected pattern and (c) and (d) show the reconstruction results. We

can confirm that the shapes are correctly estimated at right position with small details. With this experiment, we need to capture a high dynamic range (HDR) image with different exposure time. An efficient HDR capturing system is desired for achieve real one-shot scan.

Finally, we applied our method to shapes with curved surfaces and non-uniform texture as shown in Fig.14(left column). Those objects are placed between 250mm to 450mm from the projector. Fig.14 middle column shows the reconstruction results with five pixel intervals for fast calculation and the right column shows all pixel reconstruction to show capability of dense reconstruction. Fig.14(j) shows zoom-up views of the dense reconstruction results of (a). We can confirm that the curved surface is restored accurately without any postprocess. In the results, we also observe some parts are missing in 3D shapes. This is because the texture presents dark areas where no pattern is observed.

5.5. Shape reconstruction using DOE projector

We also tested shape reconstruction using DOE based system as shown in Fig. 15(a). A target object is placed at 150mm from the camera. Fig.15(b) shows the captured image and (c) shows the reconstruction result. In the experiment, since we use regular pattern for first DOE and just two different regular patterns for second DOE, uniqueness of the pattern and possible depth range is limited, however, we can confirm that the shape is correctly restored with our prototype system. Construction of more unique patterns will be investigated in the future.

6. Conclusion

In this paper, we propose a one-shot shape reconstruction method using a light field projector which is not a central projection model. The projector is constructed by using a combination of a special projected pattern and a coded aperture, which preserves high frequency information while maintaining a wide depth range. We also propose an image-based stereo matching technique, which achieves robust reconstruction despite the severe distortion which inevitably occurs with actual optics. Because of the heavy computational requirements of the basic image-based technique, a hierarchical matching as well as ANN search are introduced. By using our technique, arbitrary objects with complicated texture are reconstructed for wide depth range with high accuracy. In the future, joint optimization for designing the projected pattern and coded aperture will be investigated.
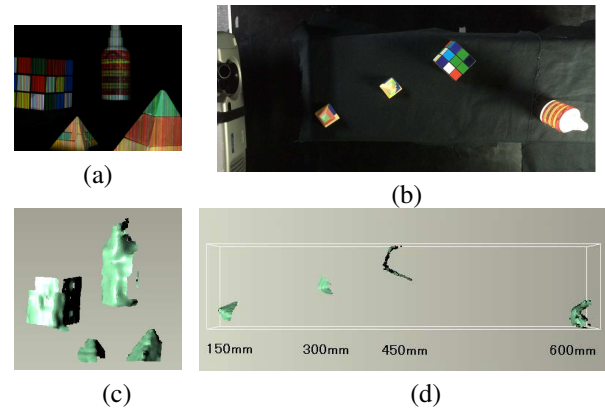
Acknowledgment

(a)

(b)

(c)

(d)

Figure 13. Wide depth range shape reconstruction result. (a) Input image, (b) top view, and (c, d) reconstruction results.

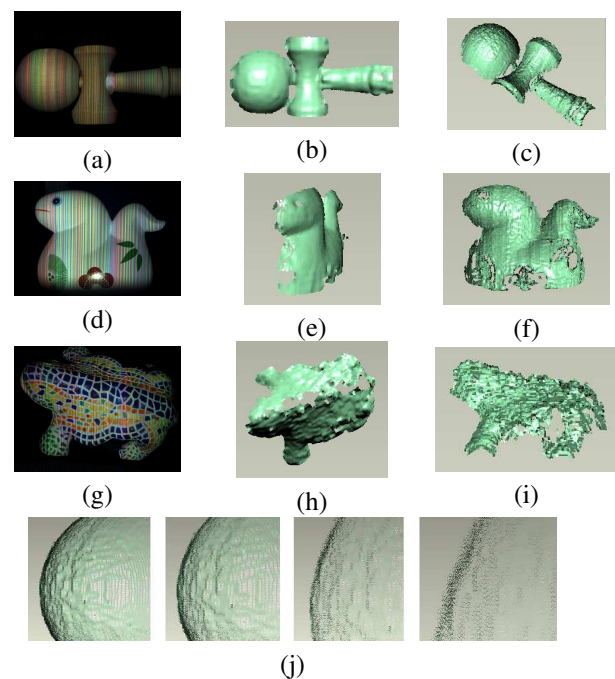(a)　(b)　(c)

(d)　(e)　(f)

(g)　(h)　(i)

(j)

Figure 14. Arbitrary shape with texture object reconstruction. Left column (a,d,g) shows target object, middle (b,e,h) shows sparse reconstruction and right column (c,f,i) shows dense reconstruction results. Bottom row (j) shows zoomed up views, from left to right, of the high resolution reconstruction to show the density.
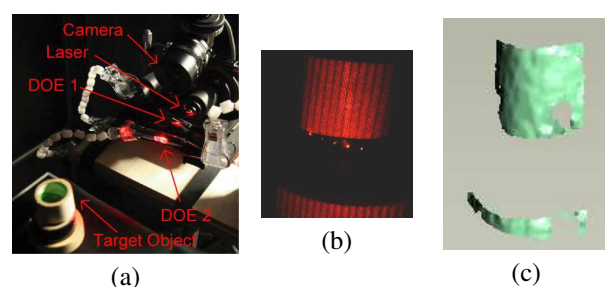

(a)　(b)　(c)

Figure 15. DOE system experimental result. (a) The system configuration, (b) target object illuminated by DOE, and (c) reconstruction results.

References

[1] Lytro redefines photography with light field cameras, June 2011. http://www.lytro.com. 1

[2] S. Achar and S. G. Narasimhan. Multi focus structured light for recovering scene shape and global illumination. In *European Conference on Computer Vision*, pages 205–219. Springer, 2014. 2

[3] P. Felzenszwalb and D. Huttenlocher. Efficient belief propagation for early vision. *International Journal of Computer Vision*, 70:41–54, 2006. 6

[4] B. Girod and S. Scherock. Depth from defocus of structured light. In *1989 Advances in Intelligent Robotics Systems Conference*, pages 209–215. International Society for Optics and Photonics, 1990. 2

[5] M. Grosse, G. Wetzstein, A. Grundhöfer, and O. Bimber. Coded aperture projection. *ACM Transaction on Graphics*, 29(3):1–12, 2010. 2

[6] M. Gupta and S. K. Nayar. Micro Phase Shifting. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, Jun 2012. 2

[7] M. Hirsch, G. Wetzstein, and R. Raskar. A compressive light field projection system. *ACM Transaction on Graphics*, 33(4):58, 2014. 2

[8] M. Hitoshi, K. Hiroshi, and F. Ryo. Depth from projector's defocus based on multiple focus pattern projection. *IPSJ Transactions on Computer Vision and Applications (CVA)*, 6:88–92, jul 2014. 2

[9] J. Jurik, A. Jones, M. Bolas, and P. Debevec. Prototyping a light field display involving direct observation of a video projector array. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society Conference on*, pages 15–20. IEEE, 2011. 1, 2

[10] H. Kawasaki, Y. Horita, H. Masuyama, S. Ono, M. Kimura, and Y. Takane. Optimized aperture for estimating depth from projector's defocus. In *International Conference on 3D Vision (3DV)*, pages 135–142, 2013. 2, 6

[11] H. Kawasaki, Y. Horita, H. Morinaga, Y. Matugano, S. Ono, M. Kimura, and Y. Takane. Structured light with coded aperture for wide range 3D measurement. In *IEEE Conference on Image Processing (ICIP)*, pages 2777–2780, 2012. 2

[12] A. Levin, R. Fergus, F. Durand, and W. T. Freeman. Image and depth from a conventional camera with a coded aperture. In *ACM SIGGRAPH 2007 papers*, SIGGRAPH '07, New York, NY, USA, 2007. ACM. 4

[13] Microsoft. Xbox 360 Kinect, 2010. http://www.xbox.com/en-US/kinect. 1, 2

[14] M. Muja and D. G. Lowe. Fast approximate nearest neighbors with automatic algorithm configuration. 2:331–340, 2009. 6

[15] K. Nagano, A. Jones, J. Liu, J. Busch, X. Yu, M. Bolas, and P. Debevec. An autostereoscopic projector array optimized for 3d facial display. In *ACM SIGGRAPH 2013 Emerging Technologies*, page 3. ACM, 2013. 1, 2

[16] S. Nayar, M. Watanabe, and M. Noguchi. Real-Time Focus Range Sensor. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(12):1186–1198, Dec 1996. 2

[17] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan. Light field photography with a hand-held plenoptic camera. *Computer Science Technical Report CSTR*, 2(11):1–11, 2005. 1

[18] R. Sagawa, H. Kawasaki, R. Furukawa, and S. Kiyota. Dense one-shot 3D reconstruction by detecting continuous regions with parallel line projection. In *Proc. 13th IEEE International Conference on Conputer Vison(ICCV 2011)*, pages 1911–1918, 2011. 1

[19] J. Salvi, J. Pages, and J. Batlle. Pattern codification strategies in structured light systems. *Pattern Recognition*, 37(4):827–849, 4 2004. 2

[20] P. Sturm and S. Ramalingam. A generic concept for camera calibration. In *European Conference on Computer Vision*, pages 1–13. Springer, 2004. 1

[21] Y. Takeda, S. Hiura, and K. Sato. Fusing depth from defocus and stereo with coded apertures. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2013. 4

[22] S. Thirthala and M. Pollefeys. Multi-view geometry of 1d radial cameras and its application to omnidirectional camera calibration. In *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, volume 2, pages 1539–1546. IEEE, 2005. 1

[23] A. Veeraraghavan, R. Raskar, A. Agrawal, A. Mohan, and J. Tumblin. Dappled photography: Mask enhanced cameras for heterodyned light fields and coded aperture refocusing. *ACM Transaction on Graphics*, 26(3):1–12, July 2007. 2, 4

[24] B. Wilburn, N. Joshi, V. Vaish, E.-V. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy. High performance imaging using large camera arrays. *ACM Transaction on Graphics*, 24(3):765–776, 2005. 1

[25] L. Zhang and S. Nayar. Projection defocus analysis for scene capture and image display. In *ACM Transaction on Graphics*, volume 25, pages 907–915, 2006. 2

[26] C. Zhou, S. Lin, and S. K. Nayar. Coded Aperture Pairs for Depth from Defocus. In *IEEE International Conference on Computer Vision (ICCV)*, Oct 2009. 4

[27] C. Zhou and S. Nayar. What are good apertures for defocus deblurring? In *Computational Photography (ICCP), 2009 IEEE International Conference on*, pages 1–8. IEEE, 2009. 4