16th International Conference on Machine Vision Applications (MVA)
National Olympics Memorial Youth Center, Tokyo, Japan, May 27-31, 2019.

Single-wavelength and multi-parallel dotted- and solid-lines for dense and robust active 3D reconstruction

Genki Nagamatsu
Kyushu University
Fukuoka, Japan

Ryo Furukawa
Hiroshima City University
Hiroshima, Japan

Ryusuke Sagawa
AIST
Tsukuba, Japan

Hiroshi Kawasaki
Kyushu University
Fukuoka, Japan

Abstract

A dense one-shot scanning technique that is robust to subsurface scattering is proposed. In this technique, a novel pattern, consisting of multiple parallel dotted lines and solid lines, that are aligned alternately, is proposed. To project such a pattern efficiently, a single-wavelength laser-based pattern projector is developed. To detect patterns robustly from captured images, a black and white camera attached with a narrow-band-path filter is used in conjunction with our novel deep learning based algorithm, which is based on a convolutional neural network (CNN). Because the detected lines must be identified for shape reconstruction, we apply a gap-coding technique, which is originally based on a grid-line pattern, to the dot pattern. To this end, we introduce a virtual grid-line structure, which is generated from the dot pattern. Additionally, we propose a calibration algorithm specialized for our system, where the pattern is static and shared with the shape reconstruction algorithm, i.e., correspondence problem remains. For a solution, gap-coding is further applied to find correspondences under epipolar constraints. The experimental results of scanning real objects are presented to demonstrate the effectiveness of our calibration and reconstruction techniques.

1 Introduction

Many approach are available for reconstructing three-dimensional (3D) shapes of object. Among them, structured light methods have been used for practical applications because of their simplicity, stability, and high precision. Recently, spatial encoding techniques, requiring only single images, have attracted considerable attention. One limitation of spatial encoding methods is if positional information is encoded into a small region, patterns tend to be complicated and are degraded easily by environmental conditions, such as noise, specularity, and blur. To avoid such limitations, techniques based on geometric constraints rather than local decoding have been proposed [1]. However, because these techniques are based on detected patterns, their results are affected by the degradation of pattern detection. Recently, grid patterns robust to severe degradation, such as subsurface scattering, have been proposed [2], where local information is embedded into gaps between lines. However, because it requires line detection for decoding, accuracy decrease if the lines are scattered. To solve such problems, we propose a one-shot scanning method that employs dot and line patterns, instead of using a grid pattern. To efficiently detect dots and lines from captured images, a learning-based algorithm, specifically convolutional neural network (CNN), is proposed. To apply CNN to one-shot scanning, we construct two types of CNNs, one for line and dot detections and the other for code detection. Both outputs of CNNs from a single captured image are integrated to generate a virtual grid-graph with gap information between lines. By using the grid-graph, each line is identified using gap information and 3D shapes are reconstructed using a light sectioning method. Additionally, we propose a calibration method for our system, where the remainder of the correspondence problem is solved effectively by adding one another CNN for detecting contour of the calibration object, which is a sphere in our method. The contributions of our approach are as follows. (1) A dot-line pattern that is robust to severe subsurface scattering is proposed. (2) A CNN-based technique for detecting and decoding dot-line patterns is proposed. (3) An automatic calibration technique for the system is proposed.

2 Related Works

For three dimensional (3D) reconstruction, structured light is one of the most practical techniques. There are two major approaches to encoding positional information into patterns in structured light systems, such as temporal and spatial encoding. Given that temporal encoding requires multiple images, it is unsuitable for capturing moving objects [3]. Spatial encoding requires only a single image and it can capture fast-moving objects. Recently several commercial systems based on this approach have been made available [4]. One practical issue with one-shot scanning is because the codes depend directly on spatial pattern distribution, reconstruction accuracy is affected severely by degradation of the captured pattern. To avoid this limitation, techniques based on geometric constraints rather than decoding have been proposed [1, 5]. However, because such techniques depend on pattern detection, the results are affected by pattern quality. Recently, solutions for typical degradation, caused by subsurface scattering, have been pro-

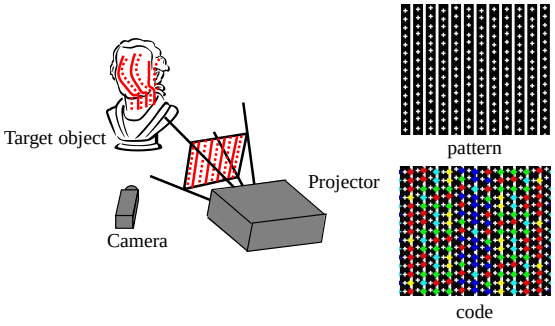


Figure 1. (Left) Scanning system, (top right) projection of patterns onto target object, (bottom right) and embedded code words.

posed [2, 6]. However, because these techniques still requires the intersection of lines for decoding, pattern resolution decreases to maintain robust detection of intersections in the captured images. In the present paper, we propose a new pattern that is robust to such degradation in conjunction with a CNN based algorithm, which is used widely in computer vision applications. As the CNN, U-Net [7], a FCNN architecture for generating a pixel-wise labeled image, is modified to fit our purpose.

3 Overview

3.1 System setup

As Fig. 1 shows, the proposed 3D measurement system consists of a camera and a DOE projector. The camera and the projector are set in almost parallel to each other and are assumed to be calibrated. The projector pattern is fixed and does not change; therefore, no synchronization is required. The geometric pattern is projected from the projector onto the objects, and the result is captured by the camera. In our system, to achieve convenient scanning without using PC systems, all processes from image capture to 3D shape reconstruction are performed using a JetsonTx2 [8].

3.2 Dot-line pattern

Under large variations of texture, strong subsurface scattering, and inter-reflection or specularity, it is difficult to extract a few typical types of information about the structured light, such as color or high-frequency information, and sometimes, such pieces of information are lost completely. To avoid loss of important detailed information, we adopted a pattern with a single color and independent structure, *i.e.*, sparse dots and straight lines with uniform intervals. Usually, with sparse dots, it is difficult to encode information using a wide baseline stereo setup with a large encoding window, because the pattern is distorted heavily

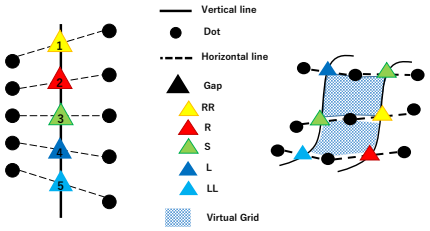


Figure 2. Geometric relationship among dot, detecting gap code, code class indexes (numbers in gap square) (Left), and virtual grid (Right).

under such conditions. To alleviate the problem, Furukawa *et al.* proposed a technique using only lines, where information is embedded as "*Gap*" of line segments [2]. One remaining problem is it requires the intersection of two lines, which is easily blurred our and degraded by subsurface scattering effects. In the proposed technique, we take an intermediate solution, such as a pattern consisting of both dots and lines with gap information. The key point is horizontal line segments are substituted by dots, even though they are not connected directly. Fig. 1 shows that the actual dot-line pattern consisting of dots and line segments.In the pattern, dots are configured to create sparsely dotted lines and are located between the solid lines. To each pair of dots aligned horizontally and adjacently across a solid line, codes are assigned by modulating the positions of the dots in the vertical direction. As Fig. 2 shows, the code classes are either S/L/LL/R/RR. The gaps are located at the intersection of a vertical line and the line connecting between dots. Moreover, the DOE projector has a limit in terms of the number of branches because noise is generated when the limit is exceeded. In the proposed pattern comprising dots instead of lines, a greater number of codes can be embedded compared to that in the conventional gird pattern. Furthermore, while lines cannot uniquely determine the position due to distortion, codes and dots can be corresponded because dots can be located uniquely on the captured image in the proposed pattern. Due to Increasing the number of codes and corresponding dots, the proposed pattern can be reconstructed more densely than the conventional gird pattern.

3.3 Algorithm

The proposed method consists of two stages: a pattern decoding stage and a 3D reconstruction stage as shown in Fig. 3. In the pattern decoding stage, the captured image is first input into a CNN for vertical line and dot detection. Simultaneously, the image is input into a CNN for pixel-wise classification of local feature codes embedded into the pattern. Then, by integrating the two results, a virtual grid-graph consisting of virtual grid, is generated (Sec. 4.3).

After generating the virtual grid-graph, by using the gird-graph as the input, 3D shapes are recovered in

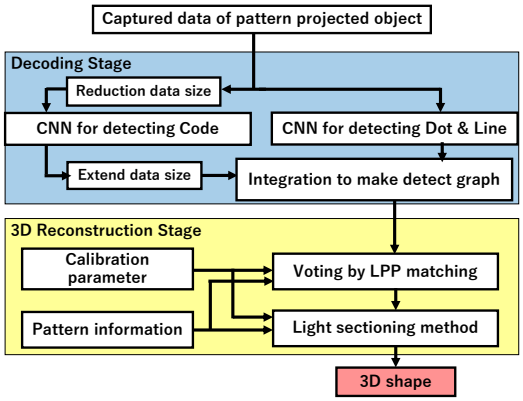


Figure 3. Overview of proposed algorithm: CNN-based decoding and 3D reconstruction for one-shot scan. Note that we must use two CNNs for vertical line and dot detections, and another CNN for code and coarse gap-position detections.

the 3D reconstruction stage. In the 3D reconstruction stage, each line is identified using the voting method and 3D shapes are recovered using a light sectioning method (Sec. 4.4).

4 Robust pattern detection using CNN

In this paper, we use U-Nets [7] to extract pattern structures and gap code information. The outputs of U-Net reflect both fine and coarse features of input images. By applying a U-Net to a captured image, N-dimensional feature maps of the same size as the input image can be obtained. In the resulting N-dimensional feature maps, where each pixel is an N-D vector, maximum element for each N-D vector is extracted to construct a N-labels classified image.

4.1 Detection of lines and dots

In the proposed method, lines and dots are detected simultaneously by using the U-Net; we call it "Structure U-Net" hereinafter. In this study, we configure a single network for detecting two primitives of lines and dots of a given pattern. We choose this approach because we expect that the kernel sizes required to detect dots and lines are almost the same because both are both the basic structural elements of the same pattern. Notably, calculation time can be halved by using a single network to detect two features.

The output data of the Structure U-Net consists of five feature maps. Note that projected lines are observed as curves on 3D scene; therefore curve detection is required. For detecting curve position, curves are represented by the left-side and right-side labels of curves, as proposed in [6]. Including the no-curve region, three maps are used for line detection. The other two maps are used for detecting dots and non-dot regions. Fig. 4(b) shows the output of Structure U-Net with five labels.

The positions of the detected vertical line are calculated using the following method. First, a segmentation image is created by setting IDs for the maximum values of the three output maps for each pixel. Then, we calculate the horizontal sub-pixels of the vertical lines as the intersection of the right and the left sides of the lines by using the value of each region.

Similarly, a segmentation image of the dot regions is created by taking the maximums of two of the feature maps from the output. Then, the positions of the dots are calculated as the center of gravity of the dot regions. Fig. 4(c) shows the result of dots detection.

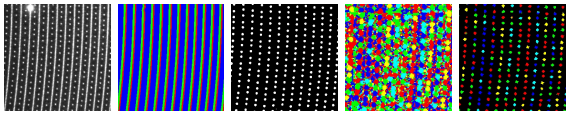
4.2 Detection of gap codes

In the proposed method, gap codes are estimated directly by applying U-Net to the image signal, as opposed to being estimated from the geometric relationships between the results of line and dot detection. Because the direct method is not affected by the quality of dot and line detection results, gap code detection can be more stable than the geometric method, an important advantage of our method.

The architecture of U-Net for detecting gap-codes, which we call "Code U-Net" hereinafter, is the same as that of the Structure U-Net, except for the kernel size of the convolution layers and the number of output maps. To accelerate the calculation while reducing memory requirement, the size of the input image is reduced to as appropriately by using bilinear interpolation. The output of Code U-Net consists of six feature maps. After extending all feature maps to the original input size, two segmentation images are generated as shown in Fig. 4(d) and (e). One is an image with five labels, where each label corresponds to each gap code, and the other is an image with six labels including "background" label with the aforementioned five labels. The image with five labels is used to decide gap-codes for each gap, whereas the image with six labels is used to estimate the distance from the nearest gap along the vertical line, which is used to refine wrong detections. The detail will be explained later.

4.3 Virtual grid-graph generation

The decoded gap consists of two pieces of information: position of the gap and feature code of the gap. In the original gap-coding technique, gap positions are given as the intersection of detected vertical and horizontal lines. In the present work, we form a virtual straight line by connecting the detected dots toward the horizontal direction. Because the projector and the camera are set up in the front parallel configuration, we can detect the virtual line easily by scanning pixels along the horizontal line. Then, gap positions



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

Figure 4. (a) captured image, (b) line detection results, where "line background" regions are colored in blue, "line left side" in green, "line right side" in red, "dot background" in black and "dot" in white, (c) dot detection results, (d) code estimation results with five labels, and (e) image with six labels, where each colors denotes a code class, same as Fig. 2 and the "background" label in case of the image with six labels.

can be retrieved as intersection points between vertical and virtual horizontal lines. Thereafter, the feature codes of the obtained gaps can be retrieved by simply referring to the pixel values at the positions of the gaps of the segmented code image with five labels. A virtual grid-graph is generated by connecting the gaps, where the gaps have four connections: Up, Down, Left, and Right. Because there may be missing gaps, if the absolute distance along the vertical direction between gaps is larger than the threshold value, an appropriate number of gaps is created between those gaps. Then, by following the stored connections of the dots, the horizontal connections (Left and Right connections) of the gap are determined.

4.4 Line identification by voting

After generating the virtual grid-graph, 3D shapes are recovered using the light sectioning method. To apply the light sectioning method, each line should be identified. In the proposed method, identification is decided by the voting approach proposed in [2], where information about connectivity in the detect graph and epipolar constraints are used with a voting scheme to increase robustness. In the proposed method, we extended the voting method by using a new similarity-matching scheme as follows. By allocating the indexes to each feature code as shown in Fig.2, the similarity-matching score is calculated as the average of the absolute difference in the indexes of the detect gaps in the detect graph and those of the correct gaps in the projected pattern. If the matching score is smaller than the threshold value, the matching gaps in the graph are voted. The U-Net for detecting codes cannot label all pixels perfectly, but, the code obtained by failure to detect is close to the correct code. Also, If a corresponded gap in a captured image is connected to dots, connected dots can be obtained the correspondence. Thus, by using the matching score in the voting scheme, we can efficiently obtain correspondence points. Once the correspondence points are retrieved, 3D shapes can be reconstructed using the light sectioning method.

5 Automatic calibration with sphere

Because we use the DOE projector, the projected pattern is static, and thus, the projector's intrinsic parameter and the relative positions and orientations between devices should be calibrated by using knowledge about the pattern for 3D reconstruction. However, it is difficult to use the pattern because we cannot use the epipolar constraint for this case, meaning finding correspondences becomes a 2D search and cannot be solved by using pattern information alone.

As a solution, we use special markers on pattern and a calibration tool with a known 3D shape, which is a sphere herein. In the calibration process, positional information about the contour of the sphere is detected by another CNN network to eliminate scaling ambiguity. By using the gap-code with the virtual grid-graph as well as special markers, calibration can be conducted in a fully automatic manner. Note that, none of the processes are manual, which is another strength of our method.

5.1 Detection of sphere contour by U-Net

In the proposed method, we use U-Net for detecting the contour of a sphere. The output data of this U-Net consists of three feature maps, similar to the output of Structure U-Net. The three labels represent regions inside and outside the contour and other pixels. By assigning the ID to the maximum of three values, a segmentation image is generated, as shown in Fig. 5 (middle). We apply the subpixel estimation technique, which is as same as the technique for line detection, to acquire precise contour positions, as shown in Fig. 5 (right). This information is used efficiently to decide the intrinsic parameter of the projector.

5.2 Acquisition of one-to-one correspondence

By using the positions and gap-codes of the virtual grid-graph, the possible line IDs of each line are narrowed. To eliminate the remaining ambiguity identifying a correct ID for each line, additional information is employed, for example, the zero-dimensional-light of the DOE projector is assigned to the pattern as a special marker and used. This special marker can be detected easily by finding the maximum intensity of the detected dots. Once the marker is found, the line ID on the marker is decided. Then, the IDs are propagated to adjacent lines to acquire all correspondences of the virtual grid-graph. The sphere is moved within the are where the special marker is projected, a few images are captured, and the above-described processing is executed to obtain the corresponding point relationship. Finally, bundle adjustment is applied to estimate the calibration parameters.

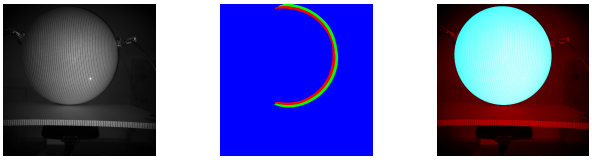


Figure 5. Automatic calibration: (left) calibration image, (middle) detection of contours, and (right) detected sphere region.

Table 1. Accuracies of scanned results (RMSEs [mm]) obtained using a previous method [2] and proposed method.

Table	Mannequin	Sponge	Shoes
Previous [2]	1.9	4.2	3.7
Proposed	1.7	3.3	3.2

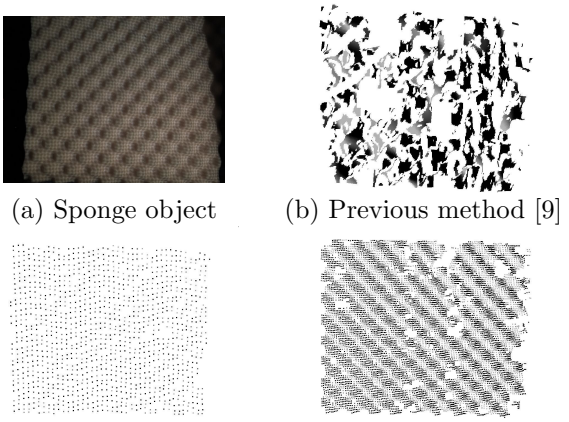
6 Experiment

6.1 Evaluation of shape reconstruction method

To confirm the effectiveness of the proposed method, we first apply the conventional method [9], which is strongly affected by the subsurface scattering as shown in Fig. 6(a). Because previous method [9] uses local geometric features, if detecting embedded features is failed due to the disturbance, it is difficult to reconstruct 3D shapes and only collapsed shapes are recovered as shown in Fig. 6(b). We also applied our method and another previous work [2] to scan the same object. Because the method [2] also uses global geometric features as ours, it is reported to be more resistant to subsurface scattering than previous methods [9]. Result is shown in Fig. 6(c) and result of our method is shown in Fig. 6(d). To make fair condition for the experiment, both projection patterns contain the same number of projection points, *i.e.*the same number of branches of DOE projector. Since horizontal lines do not contribute to shape reconstruction, our result becomes much denser than previous methods [9]; this is our another strength of the method. Fig. 7 shows the scanned results of other objects, such as mannequin head and pair of shoes. We compared the results with the ground truth 3D shapes obtained by graycode projection, and the RMSEs relative to the ground truth shapes are given in Table 1. Table 1 indicates that our technique is more accurate than [2]. In [2], because the intersections of two lines are used as the corresponding points, the ambiguity of line detection reduces accuracy.

6.2 Evaluation of calibration method

To validate the proposed calibration method, we reconstructed the 3D shapes of a board and sphere by using the parameters estimated with the proposed calibration method that uses a ball, and by using the parameters estimated with the calibration method that uses a cube shape with checker patterns. The accuracies are estimated by measuring RMSEs relative to the ground truth shapes. The ground truth shapes are measured by gray-code projection with calibration using the cube shape. Table 2 shows the results. The RMSE values obtained using the proposed method are approximately on par or comparable with those obtained using calibration with the cube. This shows the validity of the proposed calibration method.



(a) Sponge object (b) Previous method [9]

(c) Previous method [2] (d) Our method

Figure 6. Reconstruction result on strong subsurface scattering object with various methods.

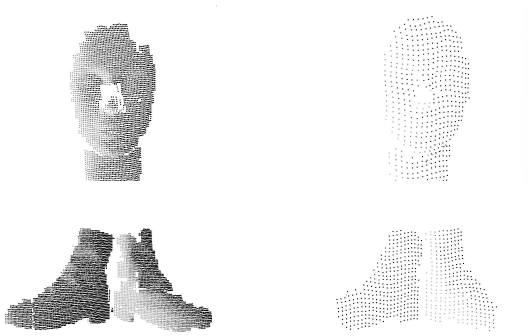


Figure 7. Reconstruction results: (left column) obtained using our dot-line technique, (right column) those obtained using the grid-gap technique [2]. We can confirm that our technique yields considerably denser results than the previous technique using the same number of points for projected pattern.

6.3 Reconstruction results with DOE projector

We created an actual system using a DOE laser projector with infrared wavelength and a narrow bandpath filter. The reconstruction results are shown in Fig. 8 and Fig. 9. The proposed method can reconstruct objects with various materials including a ceramic bottle, highly specular board and mannequin made of soft vinyl.

Table 2. Shape accuracies (RMSEs [mm]) with proposed calibration method and calibration using a cubic object with checker patterns.

Table	Board	Sphere	Mannequin
Checker cube	0.60	1.2	1.8
Proposed	0.63	3.1	3.2

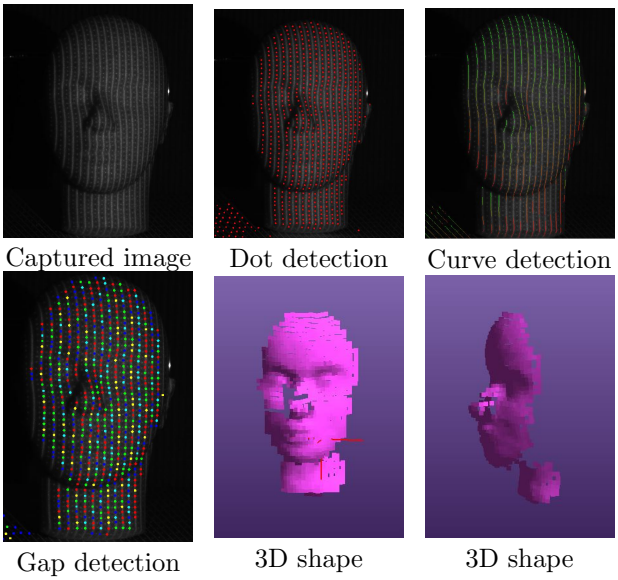


Figure 8. Infrared DOE projector results: Top row: Source image and dot and curve detection results. Bottom row: gap detection and reconstructed shape.

7 Conclusion

In this paper, we proposed a dense one-shot scanning technique, that is robust to subsurface scattering, as well as an original pattern consisting of lines and dots. Moreover, we proposed a calibration method, that was developed specifically our system. Two networks are constructed for dot and line detection and for feature-code detection. By integrating these pieces of detected information, a virtual grid-graph that consists of virtual grids with gap codes is generated. Moreover, we also constructed another network for detecting the contour of a sphere for automatic calibration process. Evaluations are conducted to show the effectiveness of our proposed method both qualitatively and quantitatively. In the future, we plan to increase the calculation speed of the proposed algorithm by using multi processing techniques.

Acknowledgment

This work was supported by JSPS/KAKENHI 16H02849, 16KK0151, 18H04119, 18K19824f and MSRA CORE.

References

[1] H. Kawasaki, R. Furukawa, , R. Sagawa, and Y. Yagi, "Dynamic scene shape reconstruction using a single structured light pattern," in "CVPR," (2008), pp. 1–8.

[2] R. Furukawa, H. Morinaga, Y. Sanomura, S. Tanaka, S. Yoshida, and H. Kawasaki, "Shape acquisition and registration for 3D endoscope based on grid pattern projection," in "The 14th," , vol. Part VI (2016), vol. Part VI, pp. 399–415.

[3] Y. Taguchi, A. Agrawal, and O. Tuzel, "Motion-aware structured light using spatio-temporal decodable patterns," Computer Vision–ECCV 2012 pp. 832–845 (2012).

[4] Microsoft, "Xbox 360 Kinect," (2010). Http://www.xbox.com/en-US/kinect.

[5] A. O. Ulusoy, F. Calakli, and G. Taubin, "One-shot scanning using de bruijn spaced grids," in "The 7th," (2009), pp. 1786–1792.

[6] R. Furukawa, D. Miyazaki, M. Baba, S. Hiura, and H. Kawasaki, "Robust structured light system against subsurface scattering effects achieved by cnn-based pattern detection and decoding algorithm," in "ECCV Workshop 3D reconstruction in the wild," (2018).

[7] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in "MICCAI," (Springer, 2015), pp. 234–241.

[8] NVIDIA, "Jetsontx2 developer kit," (2017).

[9] R. Sagawa, K. Sakashita, N. Kasuya, H. Kawasaki, R. Furukawa, and Y. Yagi, "Grid-based active stereo with single-colored wave pattern for dense one-shot 3D scan," in "3DIMPVT," (2012), pp. 363–370.

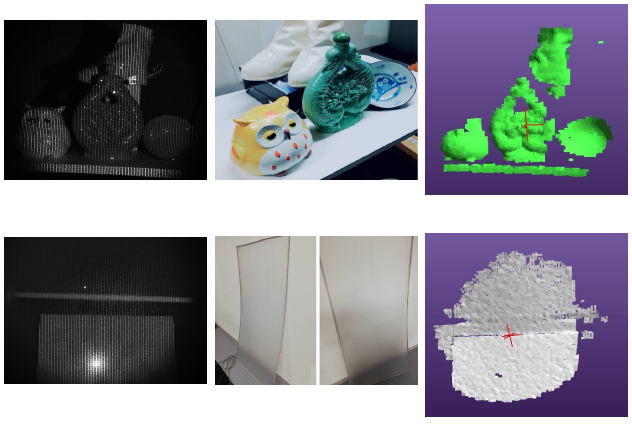


Figure 9. Infrared DOE projector reconstruction results. Top: objects with various materials. Bottom: objects with subsurface scattering.