

Robust structured light system against subsurface scattering effects achieved by CNN-based pattern detection and decoding algorithm

Ryo Furukawa¹ Daisuke Miyazaki¹ Masashi Baba¹, Shinsaku Hiura¹
Hiroshi Kawasaki²

¹Hiroshima City University, Japan, ²Kyushu University, Japan,

Abstract. To reconstruct 3D shapes of real objects, a structured-light technique has been commonly used especially for practical purposes, such as inspection, industrial modeling, medical diagnosis, etc, because of simplicity, stability and high precision. Among them, oneshot scanning technique, which requires only single image for reconstruction, becomes important for the purpose of capturing moving objects. One open problem of oneshot scanning technique is its instability, when captured pattern is degraded by some reasons, such as strong specular, subsurface scattering, inter-reflection and so on. One of important targets for oneshot scan is live animal, which includes human body or tissue of organ, and has subsurface scattering. In this paper, we propose a learning-based approach to solve pattern degradation caused by subsurface scattering for oneshot scan. Since patterns are significantly blurred by subsurface scattering, robust decoding technique is required, which is effectively achieved by separating the decoding process into two parts, such as pattern detection and ID recognition in our technique; both parts are implemented by CNN. To efficiently achieve robust pattern detection, we convert a line detection into segmentation problem. For robust ID recognition, we segment all the region into each ID using U-Net. In the experiments, it is shown that our technique is robust against strong subsurface scattering compared to state of the art technique.

1 Introduction

In order to reconstruct 3D shapes of real objects, mainly two approaches exist, one is camera only algorithm and the other is camera and active-lighting based algorithm. Although there are several important advantages on camera only algorithm, *i.e.*, systems can be compact and low energy consumption, critical limitations exists, such as textureless objects cannot be recovered. Because of the limitation, it is difficult to apply the camera only algorithm to practical purposes, such as inspection, industrial modeling or medical diagnosis. To the contrary, camera and active-lighting algorithm do not have such limitation and wide varieties of products have been developed. Among them, structured-light

technique is most popular because of its simplicity, stability and high precision. Along a long history of structured light, problems on scanning static objects with smooth and Lambertian surfaces have been mostly solved. One important remaining problem is a reconstruction of moving objects and it has been intensively researched recently. Most popular solution is based on static pattern projection with single image capture, which is called oneshot scan [12, 16].

Generally, to retrieve correspondences between the captured image and the original illuminated patterns of a projector for oneshot scan systems, features and codes are extracted from the captured image. For this purpose, common feature detection methods such as edge/corner detection are the first choice. Since projected patterns for recent oneshot scan techniques usually consist of simple structures, such as grids or dots [12, 26, 16], using the prior knowledge of such structure is used to improve the detection accuracy. After feature extraction, unique correspondences between the projected pattern and the captured image are retrieved by using spatial distribution of those features. To achieve efficient and stable matching between the patterns and the captured images, many existing oneshot methods embed a distinctive “codes” into local area of pattern. These codes are detected from the captured image (*i.e.*, decoding) and used for matching. If these codes are detected with smaller error rate than the assumption of the matching algorithm, the matching process succeeds. One practical issue of oneshot scan is that since the codes are embedded as spatial pattern distribution, reconstruction accuracy is severely affected by degradation of the captured pattern; note that such degradation is frequently caused by common physical phenomena, such as specularly, strong subsurface scattering or inter-reflection.

In this paper, we propose a learning-based approach to solve subsurface scattering effect, which is typical on live animals including human body and tissue or organ. Since patterns are significantly blurred out by subsurface scattering effects, robust decoding technique is required, which is efficiently achieved by separating the decoding process into two parts, such as pattern detection and ID recognition part. Both parts are implemented by CNN to avoid analyzing complicated phenomena of subsurface scattering and related effects. In the paper, we further propose a robust line detection algorithm by converting the problem into a segmentation problem, where lines are detected as a boundary of two segments; note that CNN works surprisingly well on segmentation and our technique can take full advantage of it. We also propose a robust ID recognition technique which is achieved by segmenting all the region into each ID using U-Net.

In the experiments, it is shown that our technique is robust to strong subsurface scattering effects compared to state of the art technique. The advantages of our approach are as follows: (1) Novel CNN-based technique for detecting and decoding projected patterns, where grid-structures of patterns are detected by two line-detection CNNs, and ID recognition CNN, is proposed. (2) Line detection problem is solved by segmentation algorithm. (3) The CNNs are trained by examples with strong subsurface scattering properties and the stability of the method against blurring of the projected patterns are confirmed.

2 Related work

There are two major shape recovery techniques using active light, such as photometric stereo and structured light. Photometric stereo recovers the surface normal of each pixel using multiple images captured by a camera while changing the light source direction [10, 9]. Although photometric stereo can recover surface normals, they need usually more than 40 images for stable reconstruction and cannot be used for oneshot scan [27, 11]. The structured-light technique has been used for practical applications [22, 27, 17]. There are two primary approaches to encode positional information into patterns, such as temporal and spatial encoding. Because temporal encoding requires multiple images, it is not suitable for capturing moving objects [23, 25]. Spatial encoding requires only a single image and is possible to capture fast-moving objects [16, 12, 15, 26, 20] and recently draw a wide attention. One severe problem for spatial encoding method is that they encode positional information into small regions, patterns tend to be complicated and easily degraded by environmental conditions, such as noise, specularly, blur, etc. To avoid such limitations, some techniques are based on geometric constraints rather than decoding [13, 21, 18, 14, 26], but not considered for strong degradation like subsurface scattering.. There are several techniques for compensation of such degradation [6, 7, 5], however, they assume capturing multiple images with projecting multiple patterns, thus cannot be applied to oneshot scan. Recently, solution for subsurface scattering objects is proposed [4], but it requires a specifically designed pattern and more general technique for wide variations of patterns is strongly demanded.

Recently, CNNs have become common tools for vision applications. Image segmentation is one of the most successful examples of such applications [8, 3]. U-Net [19] is an architecture of FCNN (Fully convolutional neural network), which can receive an image and produces a pixel-wise labeled image. It has contracting paths (signal flow where image resolution becomes coarse) to capture large-scale context information and symmetric expanding path (signal flow where low-resolution features are merged into high-resolution features) to realize precise localization. It is known to outperform previous FCNN architectures such as sliding window convolutional networks [2] in segmentation tasks of medical images. Song *et al.* proposed to detect code information of structured light using a CNN[24]. We not only use CNNs for classifying codes of the structured light, but also use them for detecting structures of the pattern such as lines that form grid-like structures.

3 Overview

3.1 System configuration

The proposed 3D measurement system consists of a camera and a projector as shown in Fig. 1(a). The camera and the projector are assumed to be calibrated (*i.e.*, the intrinsic parameters of the devices and their relative positions and orientations are known). The projector pattern is fixed and does not change, so

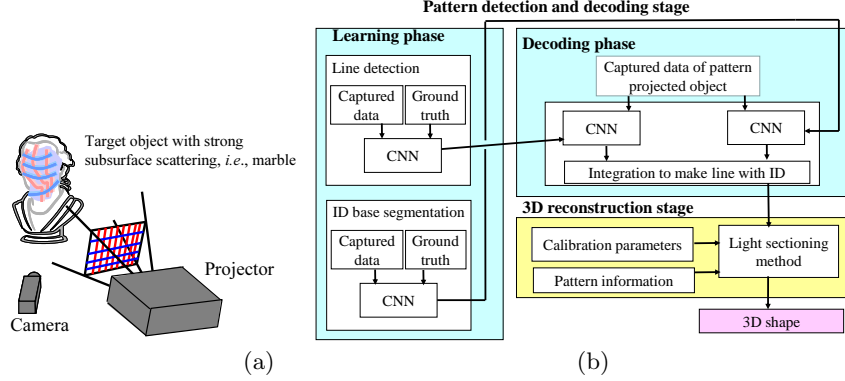


Fig. 1. Projector-camera scanning system and algorithm overview: (a) Scanning system: patterns are projected onto the objects including strong subsurface scattering. Geometric features of patterns, such as lines or intersections, are detected and used for reconstruction. (b) Algorithm overview of CNN-based decoding and 3D reconstruction for oneshot scan. Note that we have two CNNs for vertical and horizontal line detections, and another CNN for decoding IDs of grid points.

no synchronization is required. Some geometric patterns are projected from the projector and captured by the camera. In the work of spatial encoding method, the projected pattern is extracted from the captured image and the accuracy of such pattern detection is the key for accuracy and stability of 3D reconstruction. In the method, pixel-wise pattern information is efficiently decoded by CNN-based technique.

3.2 Algorithm

Our method consists of two stages: pattern decoding stage and 3D reconstruction stage as shown in Fig. 1(b).

The pattern decoding stage can also be divided into two phases, such as a learning phase and a decoding phase. In the pattern learning phase, first, actual patterns are projected onto the strong subsurface scattering objects and captured by a camera. Then, correct lines and code IDs are manually given as for the ground truth. It is a tough task even for humans, thus, learning data augmentations such as image translations or rotations are used to decrease the burden. Then, parameters and kernels of U-net [19] are estimated for lines and IDs independently using deep learning framework so that cost functions are minimized. The cost function is basically a difference between an output of U-net and the ground truth.

In the decoding phase, the captured image is first applied to CNNs for vertical and horizontal line detections. At the same time, the image is also applied to CNN for region-wise classification of local feature codes embedded into the

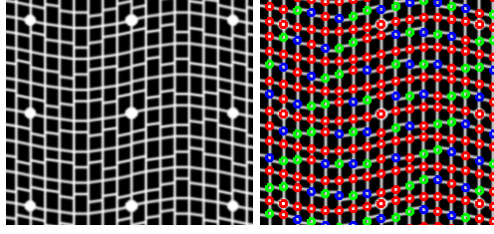


Fig. 2. The projected pattern(top), and embedded codewords of S colored in red, L in blue, and R in green (bottom). S means edges of the left and the right sides have the same height, L means the left side is higher, and R means the right is higher.

pattern. Then, both results are combined to produce final output, *i.e.*, detected lines with estimated local codes in the pattern.

Finally, by using the image with detected lines with pattern ID as the input, 3D shapes are recovered in the 3D reconstruction stage. Since a single local code is not sufficient for unique decision of correspondences, information of connectivity and the epipolar constraints are used with a voting scheme to increase robustness, similarly as [4]. Once correspondences are retrieved, 3D shapes are reconstructed by light sectioning method.

4 CNN-based feature detection and decode for active stereo

In this paper, we use “grid pattern with gapped codes” described in [4], which is claimed to be robust to defocus on projection. The pattern is shown in Fig. 2. A major feature of this pattern is a grid-like structure and discrete codes given to each grid point. The grid-like structure is composed of vertical and horizontal line segments. In the pattern, a discrete feature (gap code) is attached to each of the grid point represented by the level gap between the left and right edges of the grid point. The classes of the code are either of S / L / R as shown in different color in Fig. 2(right).

We believe such a grid-like structure is suited for CNN-based detection framework. Because pattern structures are repetitive, common features of different points can be easily extracted. Moreover, the grid-structure itself can be a large help for detecting local features such as lines, even in disturbances such as noise or blurring.

4.1 Detection of grid structure

In this paper, we propose to extract grid-structure and gap-code information using U-Nets[19], which is a kind of FCNN (fully convolutional neural network). The structure of the U-Net is shown in Fig. 3. The numbers in the figure represents dimensions of the feature maps. For example, the 1-D image (intensity

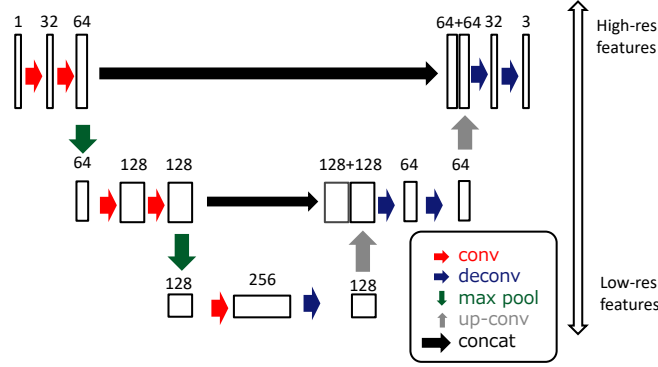


Fig. 3. Structure of U-Nets.

image) of the input is converted to a 64-D feature map by applying 2-steps of convolutions. Then, the spatial size of the feature map is sub-sampled to become a lower resolution ($1/2$ for both x and y-axis) feature map by max pooling. This coarse resolution feature map is later up-sampled by up-convolution and concatenated with the high-resolution feature maps. The information flow in the feature map looks as if “U” character in Fig. 3, thus, it is called “U-Net”. Because of this network structure, both fine and coarse resolution features are accounted for in the outputs of U-Nets.

U-Nets are originally used for pixel-wise labeling or segmentation of images. Applying a U-Net to an image, it finally produces a feature map of the same size. In the resulting N-dimensional feature maps, each pixel is an N-D vector. By taking the index of the maximum element for each N-D vector, image of N-labels is obtained.

The training process of a U-Net for detecting vertical lines is as follows. First, image samples of the pattern-illuminated scene is collected. Then, the vertical line locations for the image samples are designated manually as curves of 1-dot widths. The 1-dot width curves are too sparse and narrow to be directly used as regions of teacher data. Thus, regions with 5-dot width of left and right side of the thin curves are extracted, and labeled as 1 and 2, respectively, as shown in Fig. 4. The rest of the pixels are labeled as 0. These 3 labeled images are used as teacher data. Then, a U-Net is trained to produce such labeled regions using the loss function of the softmax entropy between the 3-labeled teacher data and the 3-D feature map produced by the trained U-Net.

By applying the trained U-Net to the image, we can get the 3-labeled image, where left and right side of the vertical curves are labeled as 1 and 2, respectively. Thus, by extracting the 2 horizontally-adjacent pixels where the left is 1 and the right is 2, and connecting those pixels vertically, vertical curve detection is achieved.

The horizontal curve detection is achieved similarly. However, the horizontal edges may be disconnected due to the gaps at the grid points. Even in those

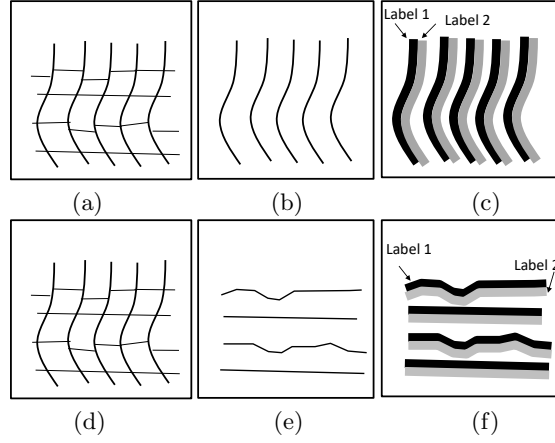


Fig. 4. Teacher data for vertical line detection (a-c), and horizontal line detection (d-f). (a)(d): Sample image of the projected pattern. (b)(e) Vertical lines that are manually annotated. (c)(f) Labeled regions used as training data. For horizontal line detection, discontinuities at the grid points are intentionally connected in the teacher data.

cases, teacher data is provided as continuous curves that go through the center point of the gaps as shown in Fig. 4(e). By training a U-Net using such teacher data, we can expect results where horizontal curves are detected as continuous at grid points, even if they are actually disconnected by gap codes.

An advantage of using U-Net for line detection of the grid structure is stability to disturbances such as blur, noise, or specularities. Fig. 5 shows an example to show the stability of the line detection ability using a surface of a squid. In the examples, strong noise is added to the image, however, the line-detection result does not degrade significantly. We think that the stability comes from that the CNN uses features extracted by image convolution, which is known to be stable to small noises. In addition, we think that the U-Net uses global information such as grid-like structures for detecting local features of line, because the lines that are almost completely wiped out by the noise and saturation are also detected.

On the contrary, we have found that the line detection of U-Net can be disturbed by scale changes that are not trained. In Fig. 6, we used U-Net that are trained for patterns with about 20x20 (pixel) grid size, and we can see that the result of U-Net was disturbed by scale changes (40x40 grid size) of input image for this example. This problem can be solved by adding training data with proper sizes, for example, if the grid sizes of the input images may be as large as 40x40, then training data with this size should be included. This can be achieved by data augmentation.

Another advantage of using U-Net for grid detection is that the horizontal edges that are actually disconnected by the gaps are intentionally detected as continuous curves by providing such training data (Fig. 7). Such a task is not easy for conventional line detection algorithms. Thanks to the continuously-

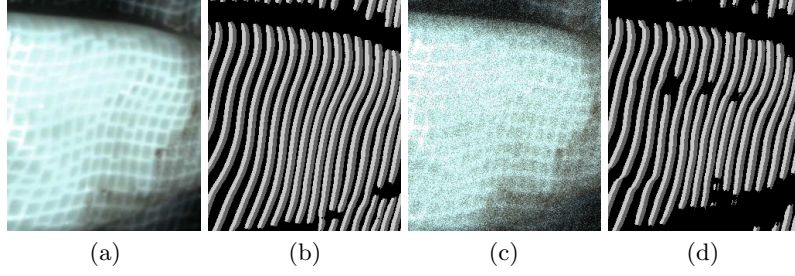


Fig. 5. Stability to noisy image. (a) Sample image(squid surface). (b) CNN (vertical line detection) result of (a). (c) Noised image of (a). (d) CNN (vertical line detection) result of (c).

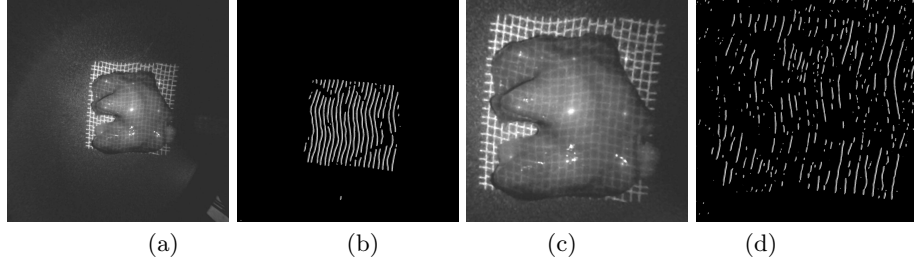


Fig. 6. Scale mismatch of the input image. (a) Input image with matched scale (about 20×20 grid size). (b) CNN (vertical line detection) result of (a). (c) Input image with mismatched scale (about 40×40 grid size). (d) CNN (vertical line detection) result of (c).

detected horizontal curves, analysis of grid-structure becomes much simpler than the previous work[4].

The max pooling and up convolution of the U-Net provide feature maps for different resolutions. For line detection process, we use 4 different resolutions. In the coarsest resolutions, the size of a “pixel” feature map is 8×8 pixels of the original image. Thus, the convolution in this resolution uses information of about 24×24 pixel patches, which is larger than typical grid size that is about 20×20 of the original image. Thus, the U-Net is considered to use information of the grid structures for local line detections.

4.2 Detection of pattern codes

In the proposed method, identification of gap codes is processed by directly applying U-Net to the image signal, not from the line detection results. Thus, the gap code estimation does not depend on line segment detection, which is advantageous for stable detection of gap codes. Note that such a direct method is not easy to implement by conventional image processing.

The training data generation is shown in Fig. 8. In the training process, the white background pixels of Fig. 8(c) are treated as “don’t care” regions.

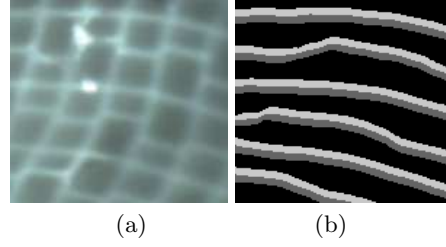


Fig. 7. Detection of discontinuous lines with gaps as continuous lines. (a) Input image (squid surface). (b) CNN (horizontal line detection) result of (a).

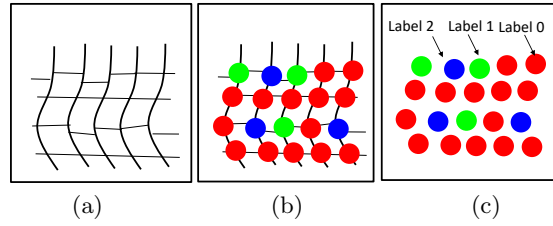


Fig. 8. Teacher data for code detection. (a): Sample image of the projected pattern. (b) Codes that are manually annotated. (c) Labeled regions used as training data. Background pixels are treated as “don’t care” data for the loss function.

The advantage of directly detecting the pattern code is that the stability of the code detection. Since, in the previous work [4], identification of gap codes have been achieved by using results of line detection, failure of line detection or failure of grid-structure analysis consequently leads to code-detection failures. The proposed method is free from such problems of sequential processing.

5 Experiment

5.1 Evaluation using subsurface scattering objects

To examine the ability of our technique, we actually scan multiple objects, which also exhibit subsurface scattering effects, by using the system as shown in Fig. 9. The 3D points on the projected lines are calculated by triangulation. Since the lines are sparse points, we filled space between the lines using interpolation based on RBFs (radial-basis functions) [1].

For the training of U-Nets, we have captured images of several materials with subsurface scattering (*e.g.*, bio-tissues, squids, or candles) while projecting the grid pattern of Fig. 2. Then the vertical and horizontal lines are annotated by human hands, and teacher data samples shown in Fig. 4 are generated from the data. Teacher data of code detection (Fig. 8) are also generated from human annotation. The numbers of the annotated images were 42, 40, and 42, respectively for the vertical lines, horizontal lines, and codes detection. The U-Nets are



Fig. 9. Actual objects capturing setup. A camera and a projector are precisely calibrated in advance.

trained with image patches with size of 80 extracted from the training data set, and with batch size of 50. The number of updates of weights of U-Nets were 2000 for each of the three nets. In this training process, we augment the training data by adding noise and scaling the intensity, since the intensity of the illuminated patterns may change significantly.

The tested objects are squid, bottled milk and wax materials. Fig. 10 shows the grid and code detection results for squid using our technique as well as previous technique [4]. The code error were improved 15% by applying our technique from [4]. Fig. 11 shows all the reconstruction results, where objects are multiple times scanned in order to increase their density. From the results, it is clearly shown that reconstruction quality and density are significantly improved from previous technique [4]. We have compared the results with ground truth 3D shapes obtained by Gray-code projection, and evaluated the RMSE of the shapes. . RMSE value is improved from 1.70mm to 1.27mm for milk and 1.78mm to 1.44mm for squid, respectively. We also scan the same objects with Kinect to verify the subsurface scattering effect. As shown in Fig. 12, shapes cannot be recovered because of strong subsurface scattering effect, such effect are clearly shown in infra-red images (it increases speckle noise to intervene correct decoding process).

5.2 Bio-tissue scanning

Next, to demonstrate practicability of the proposed method, we measured an actual organ tissue using endoscopic camera system. Bio-tissues such as organ tissues are generally problematic material, because they have strong subsurface scattering, but important for medical diagnosis and have strong demand.

Fig. 13 shows the results of grid detection and code detection. Note that, since the background regions are trained as “Don’t care”, the background pixels are labeled arbitrarily in (d). Also, to qualitatively evaluate correctness of (e), compare the color arrangements of (e) with (f).

For quantitative evaluation of the accuracy of the code detection, we manually counted the number of erroneous code detection in Fig. 13(e), and the error rate was 1.6%. It is shown that, despite the effects of subsurface scattering, which causes strong blurring on the projected lines or bright regions between the lines, the grid structures and the codes in the projected pattern are stably detected.

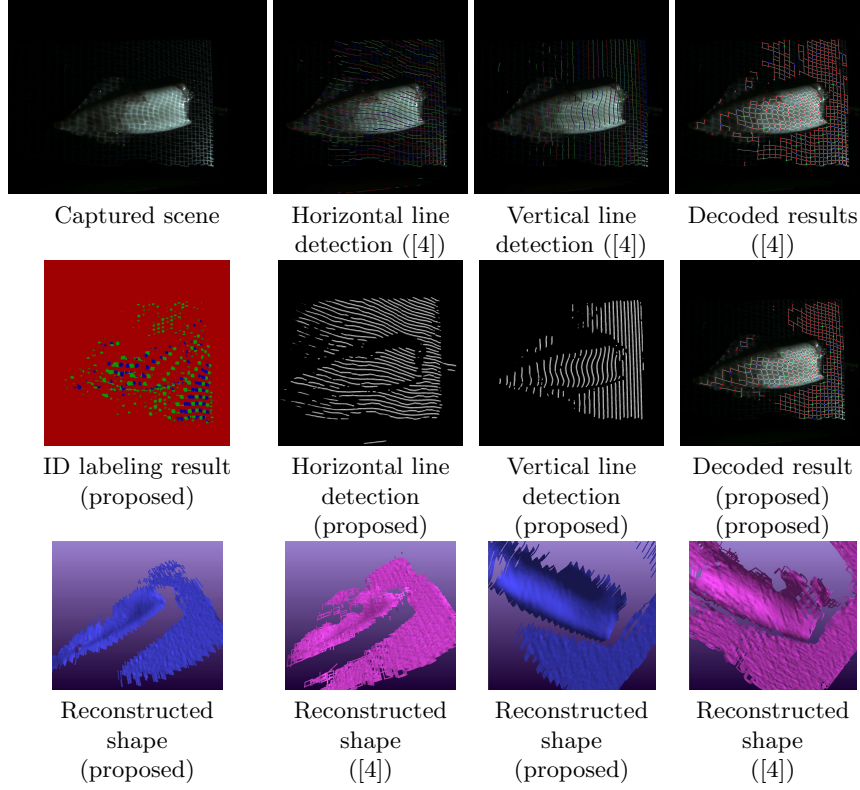


Fig. 10. Grid and code detection results for squid: Top row: Source images and line detection and decoding results of [4]. Middle row: ID segmentation, line detection and decoding results of our method. Bottom row: Shape reconstructed results. Since encoding pattern is as same as Fig. 13(f), it is confirmed that our decoding result is more correct than [4]. Further, from 3D reconstruction results, it is confirmed that our method achieves dense and smooth reconstruction without any smoothing algorithm, whereas previous method [4] creates many holes and bumps.

Fig. 14 (a-d) shows the results of the proposed algorithm. For comparison, Fig. 13 (e-g) shows grid and code detection results and the 3D reconstruction result with line detection algorithm of [4]. The reconstructed area of the proposed method is far wider than the result of [4].

6 Conclusion

This paper proposed a CNN-based grid pattern detection algorithm for active stereo to solve pattern degradation problem caused by subsurface scattering. Two independent networks are constructed and trained for both line detection and code based segmentation purposes, respectively. They are integrated to retrieve robust and accurate line detection results with pattern IDs. With our

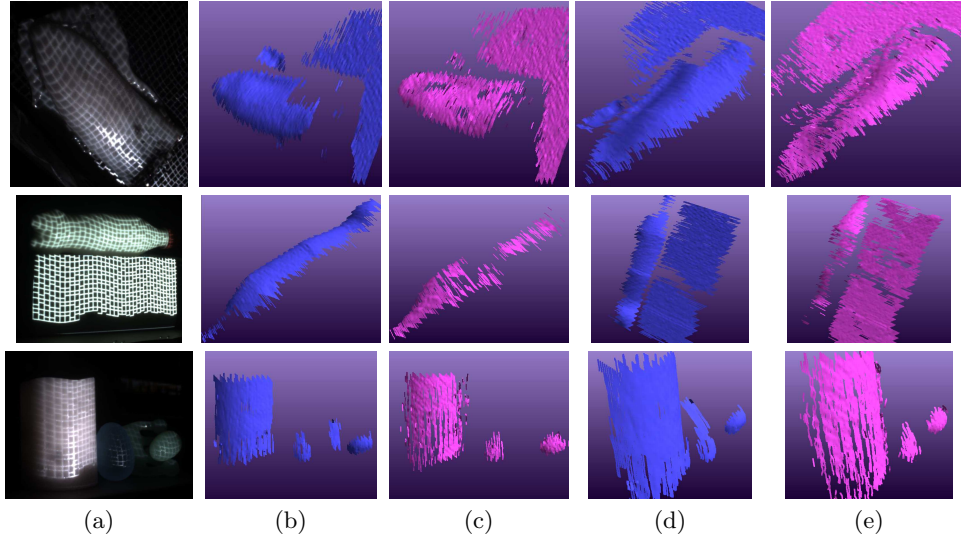


Fig. 11. Reconstruction results compared to previous technique [4]. Top row: squid, middle row: bottled milk, and bottom row: waxed objects. Left to right: (a) scanning scene, (b) and (d) shape reconstructed by our method, and (c) and (e) shape reconstructed by [4]. It is confirmed that our method achieves dense and smooth reconstruction, whereas previous method [4] creates many holes and bumps.

experiments using several target objects with strong subsurface scattering and specular effects, the proposed method shows stable detection of the grid structure and codes that are embedded into the grid points. In addition, 3D shapes of strong subsurface scattering objects are successfully reconstructed, which is only scarcely reconstructed even with the previous technique which is designed to robust to blurring effects. In the future, scale dependencies would be addressed by analyzing the training data sets.

Acknowledgment

This work was supported by JSPS/KAKENHI 16H02849, 16KK0151, 18H04119, 18K19824, and MSRA CORE14.

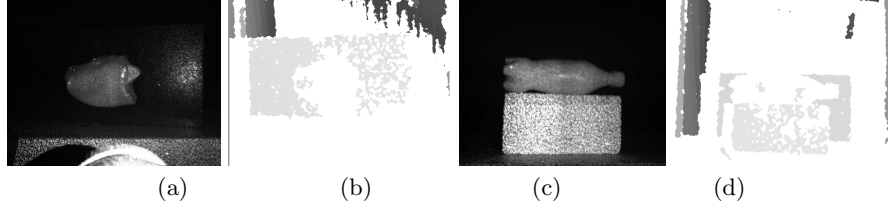


Fig. 12. Objects with strong subsurface scattering effects scanned by Kinect. (a): Infra-red image of squid. (b) Depth image of scene (a), where it is shown that the depth of the squid cannot be retrieved. (c): Infra-red image of bottled milk. (d) Depth image of scene (c), where it is shown that the depth of the bottled milk cannot be retrieved.

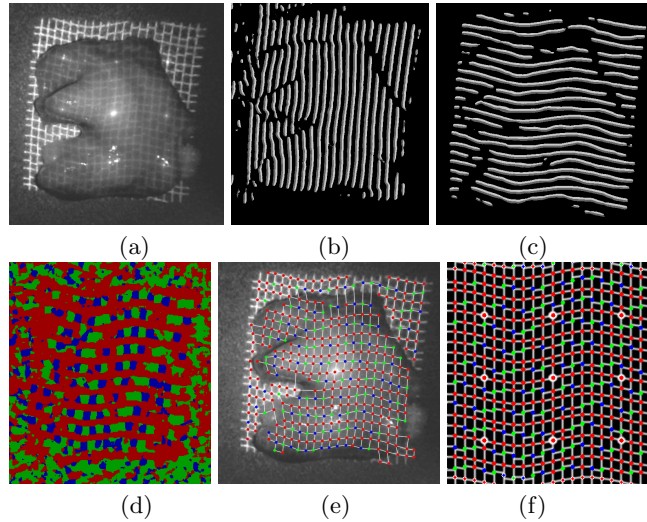


Fig. 13. Grid and code detection results for measuring a piece of organ tissue: (a): Source image. (b) Output label image for vertical line detection. (c) Output label image for horizontal line detection. (d) Output label image for code detection (Note that, since the background regions are trained as “Don’t care”, the background pixels are labeled arbitrary). (e) Extracted grid-structures and codes for grid points from (b), (c) and (d). (f) Original gap codes with same coloring with (e). To evaluate correctness of (e), compare the color arrangements of (e) with (f).

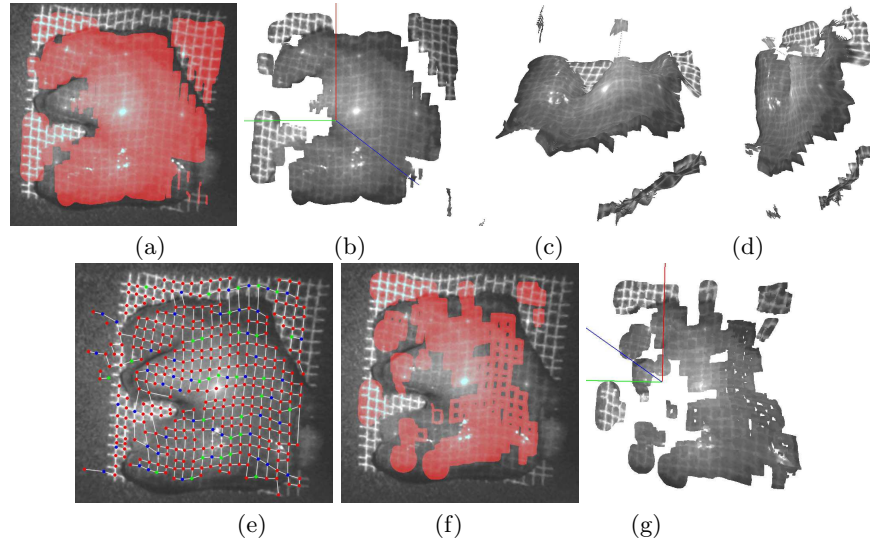


Fig. 14. 3D reconstruction of Fig. 13. (a-d) The results using the proposed grid and code detection algorithm. (e-g) The results using a conventional line detection algorithms of [4]. (a) Reconstructed regions. (b-d) Reconstructed 3D shapes. (e) Extracted grid-structures and codes for grid points based with a conventional method [4]. (f) Reconstructed regions with [4]. (g) Reconstructed shape with [4].

References

1. J. C. Carr, R. K. Beatson, J. B. Cherrie, T. J. Mitchell, W. R. Fright, B. C. McCallum, and T. R. Evans. Reconstruction and representation of 3d objects with radial basis functions. In *Proceedings of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '01, pages 67–76, New York, NY, USA, 2001. ACM.
2. D. Ciresan, A. Giusti, L. M. Gambardella, and J. Schmidhuber. Deep neural networks segment neuronal membranes in electron microscopy images. In *Advances in neural information processing systems*, pages 2843–2851, 2012.
3. J. Dai, K. He, and J. Sun. Convolutional feature masking for joint object and stuff segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3992–4000, 2015.
4. 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 ECCV*, volume Part VI, pages 399–415, 2016.
5. M. Gupta, A. Agrawal, A. Veeraraghavan, and S. G. Narasimhan. A practical approach to 3d scanning in the presence of interreflections, subsurface scattering and defocus. *International Journal of Computer Vision*, 102:33–55, 03 2013.
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.
7. M. Gupta, Q. Yin, and S. K. Nayar. Structured light in sunlight. In *The IEEE International Conference on Computer Vision (ICCV)*, December 2013.
8. B. Hariharan, P. Arbeláez, R. Girshick, and J. Malik. Simultaneous detection and segmentation. In *European Conference on Computer Vision*, pages 297–312. Springer, 2014.
9. B. K. P. Horn. Shape from shading. chapter Obtaining Shape from Shading Information, pages 123–171. MIT Press, Cambridge, MA, USA, 1989.
10. K. Ikeuchi. Determining surface orientations of specular surfaces by using the photometric stereo method. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (6):661–669, 1981.
11. C. Inoshita, Y. Mukaigawa, Y. Matsushita, and Y. Yagi. Shape from single scattering for translucent objects. In *ECCV*, pages 371–384, 10 2012.
12. H. Kawasaki, R. Furukawa, , R. Sagawa, and Y. Yagi. Dynamic scene shape reconstruction using a single structured light pattern. In *CVPR*, pages 1–8, June 23–28 2008.
13. H. Kawasaki, S. Ono, Y. Horita, Y. Shiba, R. Furukawa, and S. Hiura. Active one-shot scan for wide depth range using a light field projector based on coded aperture. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3568–3576, 2015.
14. T. P. Koninckx and L. Van Gool. Real-time range acquisition by adaptive structured light. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(3):432–445, 2006.
15. Mesa Imaging AG. SwissRanger SR-4000, 2011. <http://www.swissranger.ch/index.php>.
16. Microsoft. Xbox 360 Kinect, 2010. <http://www.xbox.com/en-US/kinect>.
17. M. O’Toole, S. Achar, S. G. Narasimhan, and K. N. Kutulakos. Homogeneous codes for energy-efficient illumination and imaging. *ACM Trans. Graph.*, 34(4):35:1–35:13, July 2015.
18. M. Proesmans and L. Van Gool. One-shot 3d-shape and texture acquisition of facial data. In *Audio-and Video-based Biometric Person Authentication*, pages 411–418. Springer, 1997.

19. O. Ronneberger, P. Fischer, and T. Brox. U-net: Convolutional networks for biomedical image segmentation. In *MICCAI*, pages 234–241. Springer, 2015.
20. S. Ryan Fanello, C. Rhemann, V. Tankovich, A. Kowdle, S. Orts Escolano, D. Kim, and S. Izadi. Hyperdepth: Learning depth from structured light without matching. In *CVPR*, June 2016.
21. R. Sagawa, Y. Ota, Y. Yagi, R. Furukawa, N. Asada, and H. Kawasaki. Dense 3D reconstruction method using a single pattern for fast moving object. In *ICCV*, pages 1779–1786, 2009.
22. J. Salvi, J. Pages, and J. Batlle. Pattern codification strategies in structured light systems. *Pattern Recognition*, 37(4):827–849, 4 2004.
23. K. Sato and S. Inokuchi. Three-dimensional surface measurement by space encoding range imaging. *Journal of Robotic Systems*, 2:27–39, 1985.
24. L. Song, S. Tang, and Z. Song. A robust structured light pattern decoding method for single-shot 3d reconstruction. In *2017 IEEE International Conference on Real-time Computing and Robotics (RCAR)*, pages 668–672, July 2017.
25. Y. Taguchi, A. Agrawal, and O. Tuzel. Motion-aware structured light using spatio-temporal decodable patterns. *Computer Vision–ECCV 2012*, pages 832–845, 2012.
26. A. Ulusoy, F. Calakli, and G. Taubin. One-shot scanning using de bruijn spaced grids. In *Proc. The 2009 IEEE International Workshop on 3-D Digital Imaging and Modeling*, 2009.
27. J. Wang, A. C. Sankaranarayanan, M. Gupta, and S. G. Narasimhan. Dual structured light 3d using a 1d sensor. In *Computer Vision - ECCV 2016 - 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part VI*, pages 383–398, 2016.