Sensors and Materials, Vol. 35, No. 4 (2023) 1327–1337
MYU Tokyo

S & M 3249

Assessment of All-in-focus Image Quality in Shape-from-focus Technique

Shinya Onogi* and Yoshikazu Nakajima

Department of Biomedical Information, Tokyo Medical and Dental University,
2-3-10 Kanda-Surugadai, Chiyoda-ku, Tokyo 101-0062, Japan

(Received October 31, 2022; accepted February 22, 2023)

Keywords: shape from focus, all-in-focus image, image processing, 3D measurement

Shape from focus (SFF) is a 3D shape measurement technique using a monocular camera. SFF also provides an all-in-focus image from multiple defocus images. In SFF, the pixelwise focused camera position and pixel data are determined from the degree of focus (focus measure). We previously reported that the optimal window size in the focus measure computation for accurate 3D shape measurement depends on the local texture frequency of objects. In this study, we validate the performance of the all-in-focus image reconstruction. We performed two tests: image quality validation using human tissue slides and a feasibility test using animal tissue. First, the image quality was validated using the similarity between reference images and reconstructed images of five human tissue slides. Second, the surface texture of an animal tissue was reconstructed as a feasibility test. The experimental results show that the similarity of the reconstructed images was significantly high, and the precise texture of the tissue 3D surface could be clearly observed. From the experiments, we demonstrate the reconstruction performance of the all-in-focus image by SFF both quantitatively and qualitatively.

1. Introduction

The 3D measurement of an object is still an interesting topic in sensing and measurement. Accurate 3D shape measurement is practically performed for the validation and inspection of objects in both industrial and scientific fields, for example, for comparison between CAD models and 3D scans and for the 3D visualization of tissues and cell microstructures. In medical imaging, 2D or 3D shapes of tissues or organs are obtained by computed tomography (CT), magnetic resonance imaging (MRI), ultrasonography, and so on. These modalities can provide the shape and inner structure based on the respective imaging principles but not the surface texture. Moreover, optical imaging can obtain the texture but not the 3D shape. Therefore, a measurement technology that can simultaneously measure both the shape and the texture has a potential to improve the accuracy of diagnosis including detection by deep neural networks. Moreover, a textured 3D shape can contribute to surgical navigation. Texture information can improve nonrigid registration accuracy.(1)

*Corresponding author: e-mail: shinya.onogi.1979@ieee.org
https://doi.org/10.18494/SAM4219

ISSN 0914-4935
https://myukk.org/

Many methods of 3D shape measurement have been proposed, including laser scanning[2,3] and optical image processing. Although laser scanning is popular and accurate, both a laser scanner and a camera are required, and the scanner itself does not provide the texture of the object. Three-dimensional shape measurement by stereoscopy using two or more cameras is a representative optical image processing method and is employed in many cases.[4,5] However, the ambiguity of stereo matching sometimes results in incorrect measurements, and matching points must be visible at both cameras. Shape estimation from the illumination difference, known as shape from shading, is another measurement method with one camera.[6,7] Although it works well on smooth and textureless surfaces, it is very difficult to distinguish the illumination from the brightness of the texture itself. Hence, its range of applicable conditions is limited. Shape from focus (SFF) and shape from defocus are also passive methods for obtaining the 3D shape of an object by estimating the depth of each pixel from images taken with different focus settings.[8–16] The best focus setting at a pixel is directly related to its depth. SFF works well on textured surfaces and it is suitable for natural objects including human tissues because it provides both the 3D shape and the texture simultaneously. Previously, we proposed and developed a prototype endoscope that can control the image sensor position to change the depth of focus, and we validated its reconstruction accuracy using planar and cylindrical objects with a checker texture.[17]

In SFF, a focus measure, the degree of focus, is estimated from images of different focus levels. Many focus measures have been proposed, including the Tenengrad focus measure (TEN),[18] modified Laplacian (ML),[9] sum of ML (SML), Laplacian of Gaussian (LoG), and gray level variance (GLV).[19,20] The relative performance of different focus measures has been reported to strongly depend on imaging conditions and the size of the operator's window.[21] To compute a focus measure, different fixed window sizes, ranging from 3 × 3 to 7 × 7 or more, have been used. Malik and Choi reported a study on illumination effects and window sizes.[22] They showed that increasing the window size reduces the depth resolution due to oversmoothing effects. On the other hand, small windows increase the sensitivity to noise.[23,24] Pertuz *et al.* suggested that the optimum window size is a trade-off between spatial resolution and robustness.[21] As a more robust approach, an adaptive window size determined by the median absolute deviation (MAD) has been reported.[24] The window size is enlarged until the MAD is larger than a threshold. However, the threshold should be determined for an object by an object. We previously reported that the optimal window size can be determined from the local frequency of the object's texture.[25] Note that the above studies focused on the accuracy of 3D shape measurement and not on the quality of the all-in-focus image.

From the above survey, SFF would be useful for medical purpose. As focused on reconstructed all-in-focus images, one of the expected applications is microscopic imaging, which is used for pathological diagnosis and microsurgery. Microscopy generally has a very low depth of field due to its high magnification, and partial blur is inevitable for some biological specimens with 3D shapes. SFF can be useful to remove the blur, and the quality of the reconstructed all-in-focus image is important in the applications. In this study, we validate the quality of the reconstructed all-in-focus image from a focal stack both quantitatively and qualitatively.

2. Materials and Methods

2.1 SFF

Figure 1 shows the principle of imaging under focused and defocused conditions. Let P be a point on the surface of an object. Its focused image is obtained at Q on the image plane. Then, the distance U between the lens and point P and the distance V between the lens and the image plane have the following relationship, where f is the focal length:

$$\frac{1}{U}+\frac{1}{V}=\frac{1}{f}. \tag{1}$$

A defocused image is obtained at Q'. The blurring effect is generally modeled by the point spread function (PSF) as

$$h(x,y)=\frac{1}{2\pi\sigma^2}\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right). \tag{2}$$

Image formation involves the convolution of the actual image $I(x, y)$ and the PSF:

$$I'(x,y)=h(x,y)*I(x,y). \tag{3}$$

To find the focused position of each pixel, multiple frames are captured with various focusing parameters by translating the camera position, the image sensor position, or the lens. Focus measures expressing the degree of focus are calculated for each pixel in the captured frames. The focusing parameter giving the maximum focus measure is estimated as the focused position at the pixel. Therefore, the depth of each pixel can be obtained. The focus measure f_m at a pixel is obtained as:

$$f_m(x,y)=LOG(x,y)*I'(x,y), \tag{4}$$

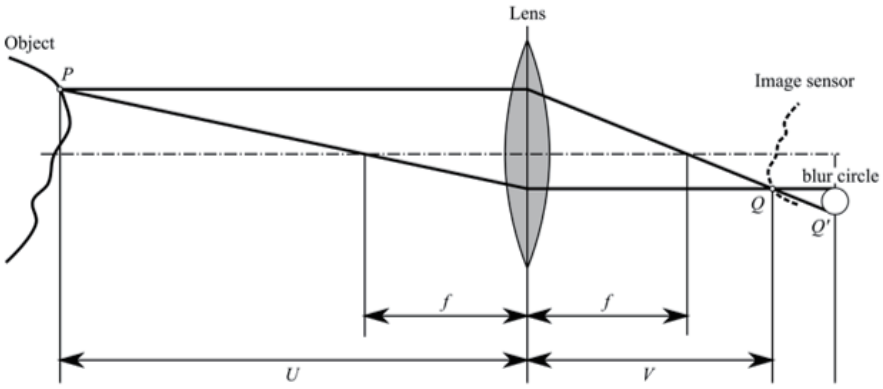


Fig. 1. Principle of depth from focus.

where *LOG* is a LoG filter acting as a band-pass filter. Let $I(s, x, y)$ denote multiple defocused images at camera position s and $f_m(s, x, y)$ denote the corresponding focus measures. Then, the camera position that provides the maximum focus measure is given for each pixel as

$$s(x,y) = \arg\max_{s} f_m(s,x,y), \tag{5}$$

and the depth of point (x, y) is directly obtained from the camera position using Eq. (1). Finally, the focused pixel value for each pixel is given by

$$I^*(x,y) = I(s(x,y),x,y), \tag{6}$$

where I^* is the reconstructed all-in-focus image.

2.2 Experimental setup

Figure 2 shows the experimental setup used in this study. A monocular camera (IK-4KH, Canon Global) with a resolution of 3840 × 2160 pixels was mounted on a microscope (SZ61, Olympus). The microscope was fixed with a linear actuator (OSMS20-85, Sigma Koki) to control its height.

We validated the reconstructed image quality using human tissue slides and the reconstruction feasibility of an object with a 3D shape using *ex vivo* animal tissues. For quality validation, five

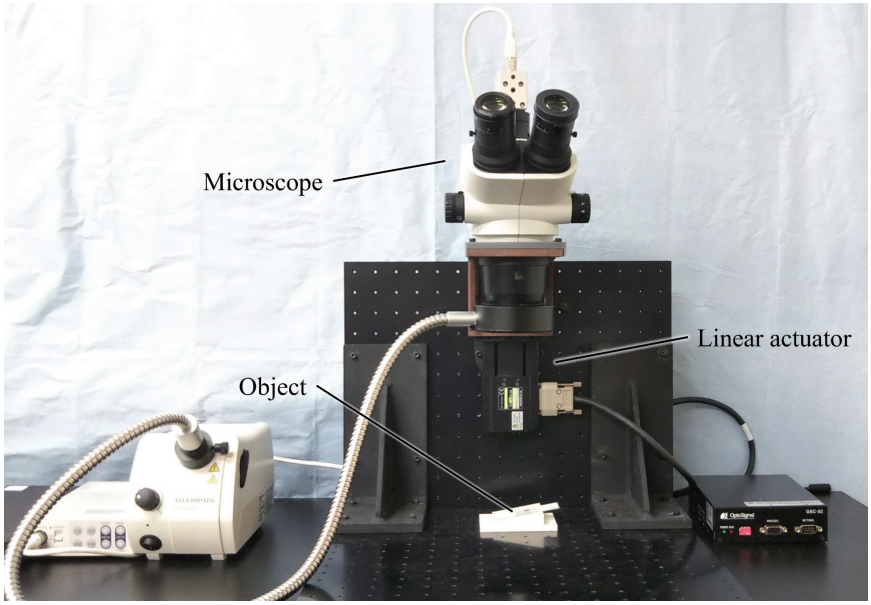


Fig. 2. (Color online) Setup for image acquisition. The tissue slide was placed on the stage with a slope for defocused image acquisition or without a slope for reference image acquisition. The microscope was electrically moved from the lower position to the upper position by the linear actuator.

slides of human tissues, namely, stomach, liver, duodenum, thyroid, and areolar connective tissues (Normal Human Histology, Basic Set, 3B Scientific), were used. First, images of the slides were acquired from perpendicular to the focused position as reference images. Next, the objects were placed at a tilt angle of 30°, as shown in Fig. 3; then, the camera was moved from a too-near defocused position to a too-far defocused position at 1 μm intervals, and the camera position and image were recorded at each position. The focus measure at each pixel was computed; then, all-in-focus images were reconstructed. The defocused and reconstructed images were registered to the reference images using an image registration technique based on the AKAZE algorithm,[26] which is implemented in the OpenCV library. Then, the similarity between the reconstructed and reference images was calculated using the zero-mean normalized cross-correlation (ZNCC).

For the feasibility test, the fresh pig intestine shown in Fig. 4 was used. The specimen was opened, and its inner surface was photographed in the same way as in the above experiment. The all-in-focus image of the specimen was also reconstructed.

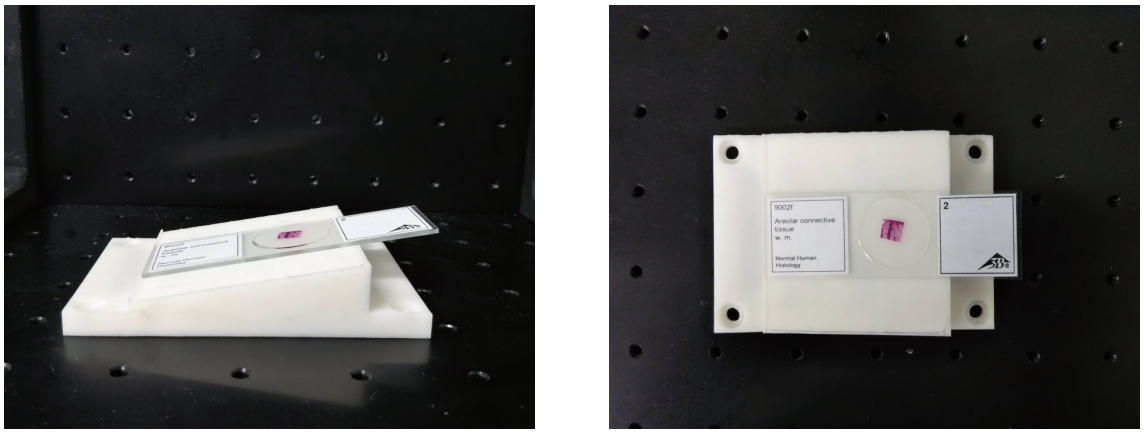


Fig. 3. (Color online) Tissue slide on stage with slope of 30°.

Fig. 4. (Color online) Pig intestine used for *ex vivo* experiment.

3. Validation of All-in-focus Image Quality

Figure 5 shows the reconstructed depth maps and all-in-focus images of the slides. A window size of 5 × 5 was used. The slope shape was accurately obtained in each sample. Some errors

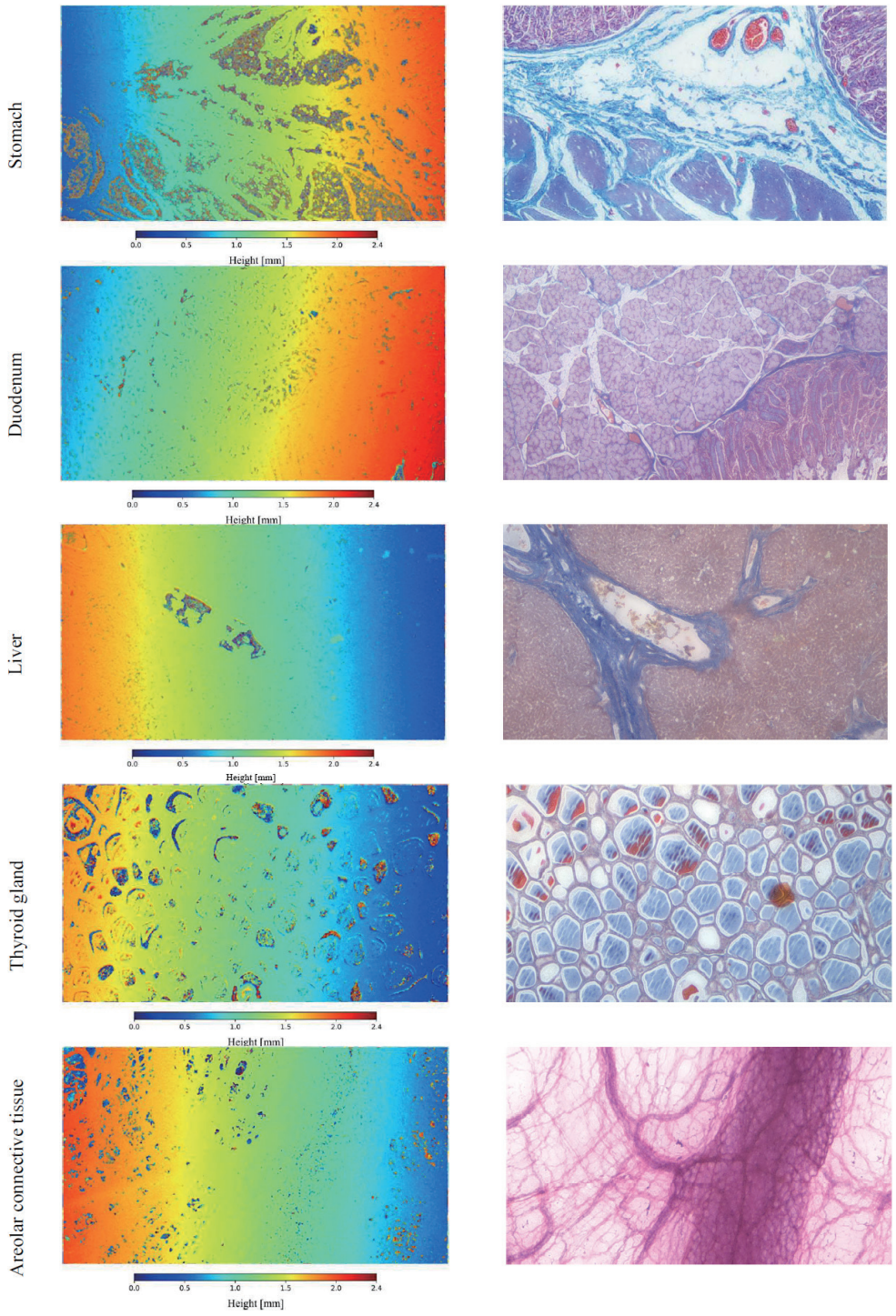


Fig. 5. (Color online) (a) Depth maps and (b) all-in-focus images of five human tissues. The slope was accurately obtained, and some errors were observed in the poor texture region (white area).

were observed in the poor texture region (white area). However, all-in-focus images were well reconstructed because the white area is not changed in focus or defocus condition. Figure 6 shows the defocused, reconstructed, and reference images. The defocused images that were focused at the center were selected. The reconstructed all-in-focus images were compared with the corresponding focused reference images. Table 1 shows the results of the experiment. The similarities of the reconstructed images were significantly higher than those of the defocused images. Finally, the reconstruction results of the *ex vivo* tissue are shown in Fig. 7. A window size of 21 × 21 and a 1280 × 960 region were considered. Figures 7(a) and 7(b) show the raw and all-in-focus images, respectively. Figure 7(c) shows the absolute difference between them with contrast enhancement. Figure 7(d) shows the depth map. The results suggest that the all-in-focus image could be reconstructed in the whole area and that the precise texture could be observed.

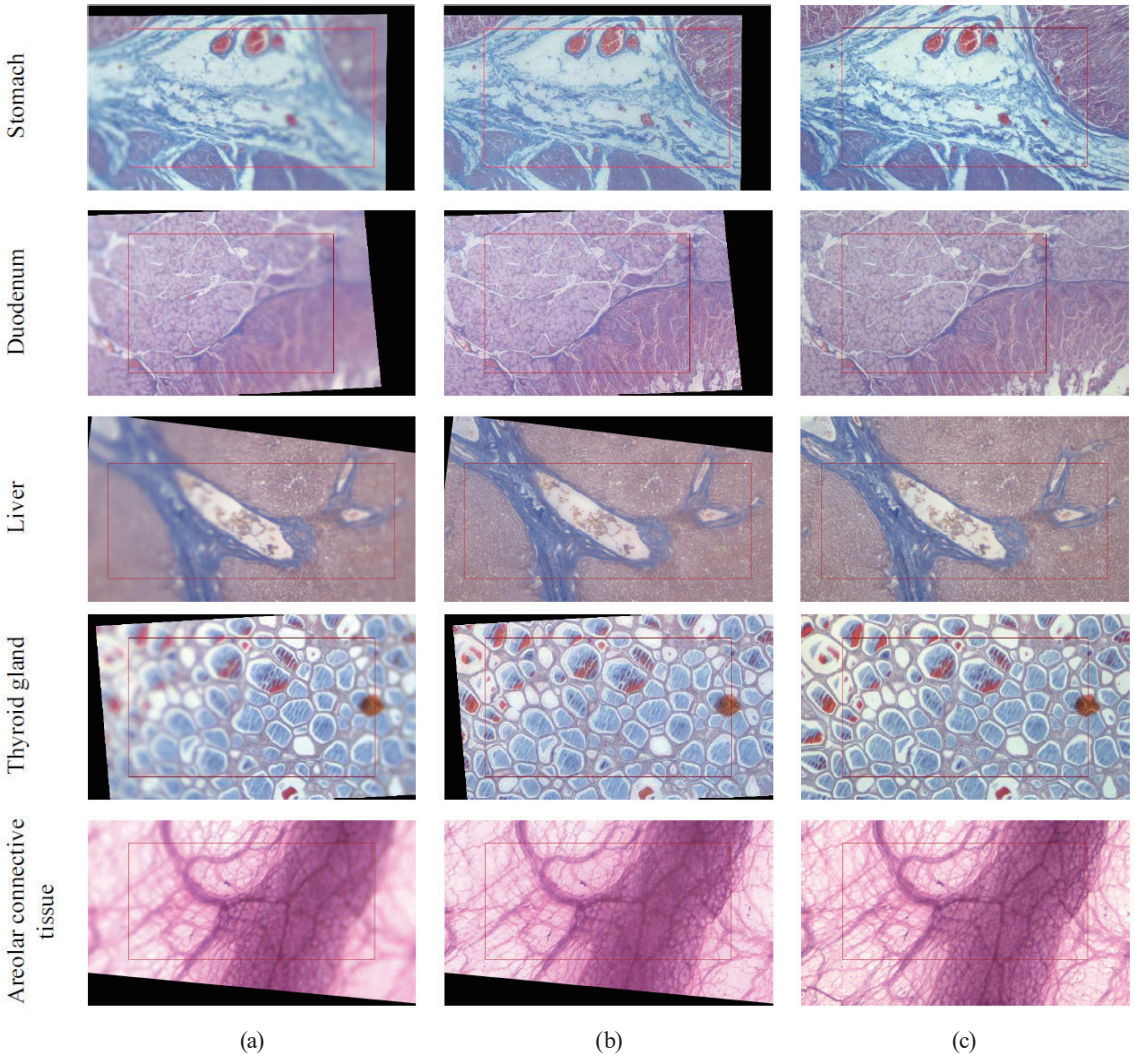
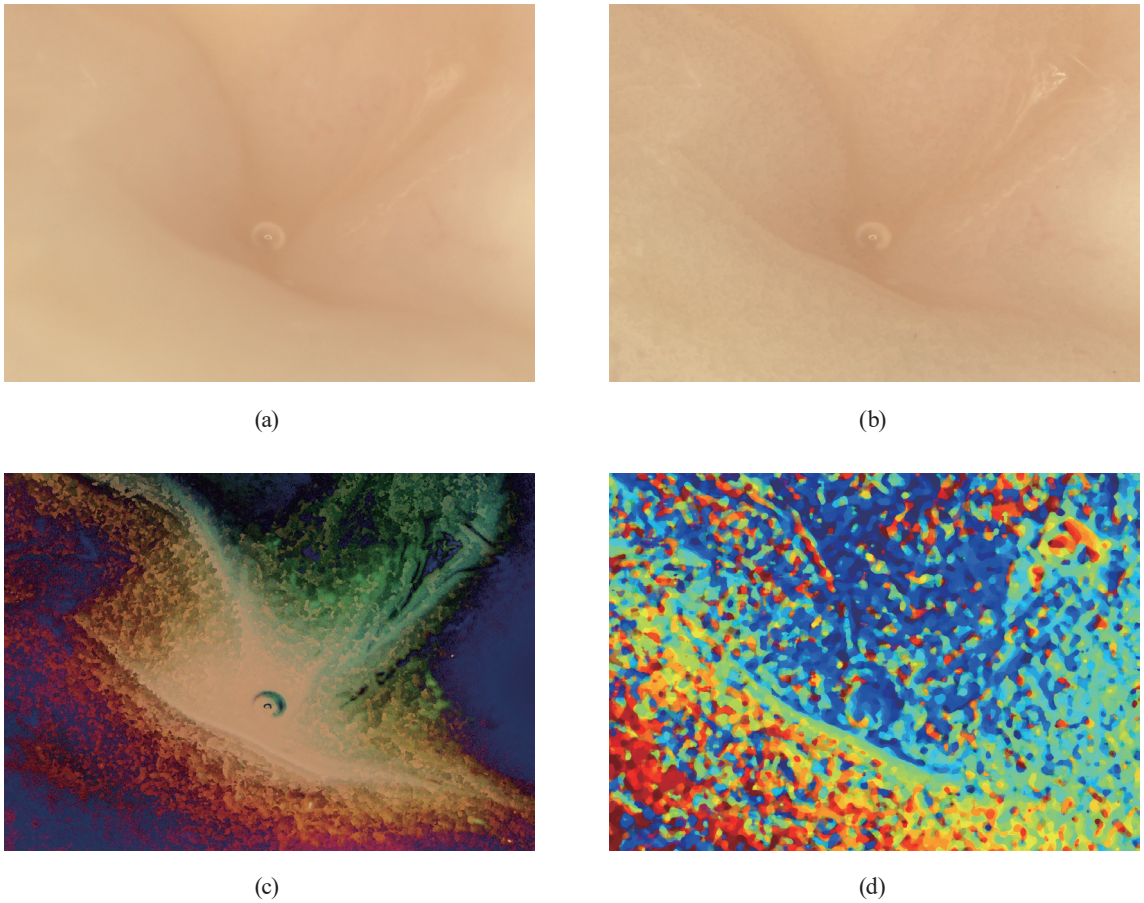


Fig. 6. (Color online) (a) Defocused, (b) reconstructed, and (c) reference images of five human tissues. The defocused and reconstructed images were registered to the corresponding reference images. The rectangles express the region used for the computation of correlation coefficients.

Table 1
Similarities of the defocused and reconstructed images.

Tissue	Defocused image	Reconstructed image
Stomach	0.91	0.94
Duodenum	0.87	0.96
Liver	0.91	0.96
Thyroid gland	0.89	0.95
Areolar connective tissue	0.97	0.97
Overall*	0.91 ± 0.04	0.96 ± 0.01

*$p < 0.05$



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

Fig. 7. (Color online) Results of all-in-focus reconstruction of pig intestine: (a) defocused image, (b) all-in-focus image, (c) absolute difference image with contract enhancement, and (d) depth map.

4. Discussion

From the results, the similarities were significantly increased in the all-in-focus images by using the SFF technique. The small difference between the reconstructed and reference images was caused by the difference in image resolution because the slides for reconstruction were

photographed at an angle. To obtain the ground truth image directly, a lens with the deep depth of field is needed; however, other camera parameters are also changed and the image is not suitable for pixelwise validation. Therefore, we believe that the validation method is practical.

The image quality was validated using flat objects without 3D shapes. The possibility of all-in-focus image reconstruction for a 3D object was demonstrated by the test using the tissue surface; however, the quantitative performance for an object with a curved surface is not yet clear. Theoretically, a smooth curved surface does not affect the reconstruction accuracy; however, occlusions and shade cause a decrease in accuracy. Thus, some improvement of the focus measure computation would be needed for more realistic objects.

The SFF technique requires a rich texture for focus measure computation and cannot determine the focused position for textureless objects. Some compensation methods such as shape measurement from shading and pattern projection have been reported.[27] A combined approach is expected to be reasonable for shape and texture measurements. Another technical problem is a change in magnification during focusing. In the present implementation, each pixel of all acquired images is assumed to be at the same point in an object without the consideration of magnification. The compensation of magnification shift[28] would be required to further improve the reconstruction accuracy.

5. Conclusions

This study validated the quality of all-in-focus images by using the SFF technique. Validation results showed that the reconstructed all-in-focus images have significantly higher similarities to the in-focus images than the raw photographed images. We confirmed that the SFF technique can provide accurate all-in-focus images as well as the 3D shape.

Acknowledgments

This work was partly supported by JSPS KAKENHI Grant Number 16H03191.

References

1 Y. Nakajima, Y. Sohma, and J. Jiang: Sens. Mater. **32** (2020) 935. https://doi.org/10.18494/SAM.2020.2630
2 K. N. Kutulakos and E. Steger: Int. J. Comput. Vis. **76** (2008) 13. https://doi.org/10.1007/s11263-007-0049-9
3 S. K. Nayar and M. Gupta: Proc. 2012 IEEE Int. Conf. Comput. Photogra. (ICCP) (2012) 1–11. https://doi.org/10.1109/ICCPhot.2012.6215216
4 S. T. Barnard and W. B. Thompson: IEEE Trans. Pattern Anal. Mach. Intell. **PAMI-2** (1980) 333. https://doi.org/10.1109/TPAMI.1980.4767032
5 H. Hirschmuller and D. Scharstein: Proc. 2007 IEEE Conf. Comput. Vis. Pattern Recognit. (2007) 1–8. https://doi.org/10.1109/CVPR.2007.383248
6 R. T. Frankot and R. Chellappa: IEEE Trans. Pattern Anal. Mach. Intell. **10** (1988) 439. https://doi.org/10.1109/34.3909
7 J. D. Durou, M. Falcone, and M. Sagona: Comput. Vision Image Underst. **109** (2008) 22. https://doi.org/10.1016/j.cviu.2007.09.003
8 A. P. Pentland: IEEE Trans. Pattern Anal. Mach. Intell. **9** (1987) 523. https://doi.org/10.1109/TPAMI.1987.4767940
9 S. K. Nayer and Y. Nakagawa: IEEE Trans. Pattern Anal. Mach. Intell. **16** (1994) 824. https://doi.ieeecomputersociety.org/10.1109/34.308479

10 M. Subbarao and N. Gurumoorthy: Proc. Comput. Vis. Pattern Recognit. (1988) 498. https://doi.org/10.1109/CVPR.1988.196281
11 M. Subbarao and G. Surya: Int. J. Comput. Vis. **13** (1994) 271. https://doi.org/10.1007/BF02028349
12 M. Subbarao and T. Choi: IEEE Trans. Pattern Anal. Mach. Intell. **17** (1995) 266. https://doi.org/10.1109/34.368191
13 M. Subbarao and J.-K. Tyan: IEEE Trans. Pattern Anal. Mach. Intell. **20** (1998) 864. https://doi.org/10.1109/34.709612
14 J. Ens and P. Lawrence: IEEE Trans. Pattern Anal. Mach. Intell. **15** (1993) 97. https://doi.org/10.1109/34.192482
15 P. Favaro and S. Soatto: IEEE Trans. Pattern Anal. Mach. Intell. **27** (2005) 406. https://doi.org/10.1109/TPAMI.2005.43
16 E. Wong: Proc. Int. Conf. Acoust. Speech Signal Process. **3** (2006) 1184. https://doi.org/10.1109/ICASSP.2006.1660871
17 T. Takeshita, M. Kim, and Y. Nakajima: Int. J. Comput. Assist. Radiol. Surg. **8** (2012) 451. https://doi.org/10.1007/s11548-012-0794-2
18 A. Thelen, S. Frey, S. Hirsch, and P. Hering: IEEE Trans. Image Process. **18** (2008) 151. https://doi.org/10.1109/tip.2008.2007049
19 F. Groen, I. Young, and G. Ligthart: Cytometry **6** (1985) 81. https://doi.org/10.1002/cyto.990060202
20 T. Yeo, S. Ong, S. Jayasooriah, and R. Sinniah: Imag. Vis. Comput. **11** (1993) 629. https://doi.org/10.1016/0262-8856(93)90059-P
21 S. Pertuz, D. Puig, and M. A. Garcia: Pattern Recognit. **46** (2013) 1415. https://doi.org/10.1016/j.patcog.2012.11.011
22 A. S. Malik and T. S. Choi: Pattern Recognit. **40** (2007) 154. https://doi.org/10.1016/j.patcog.2012.11.011
23 J. Marshall, C. Burbeck, D. Ariely, J. Rolland, and K. Martin: J. Opt. Soc. Am. A **13** (1996) 681. https://doi.org/10.1364/JOSAA.13.000681
24 I. Lee, M. T. Mahmood, and T. S. Choi: Opt. Laser Technol. **45** (2013) 21. https://doi.org/10.1016/j.optlastec.2012.08.003
25 S. Onogi, T. Kawase, T. Sugino and Y. Nakajima: Electronics **10** (2021) 1870. https://doi.org/10.3390/electronics10161870
26 P. F. Alcantarilla, J. Nuevo, and A. Bartoli: Proc. British Machine Vision Conf. (2013) 13.1–13.11. http://dx.doi.org/10.5244/C.27.13
27 O. Ghita, P. F. Whelan, and J. Mallon: J. Electron. Imag. **14** (2005) 023021. https://doi.org/10.1117/1.1900743
28 S. Pertuz, D. Puig, and M. A. Garcia: Proc. 2010 20th Int. Conf. Pattern Recognit. (2010) 802–805. https://doi.org/10.1109/ICPR.2010.202

About the Authors

Shinya Onogi received his B.S. and M.S. degrees from Hokkaido University, Japan, in 2002 and 2004, respectively, and his Ph.D. degree from The University of Tokyo, Japan, in 2007. From 2007 to 2010, he was a project researcher at The University of Tokyo. From 2010 to 2011, he was a postdoctoral fellow at Johns Hopkins University, USA. From 2011 to 2014, he was an assistant professor at Tokyo University of Agriculture and Technology, Japan. From 2014 to 2018, he was an associate professor at Kyushu University, Japan. From 2018 to 2020, he was an associate professor at Fukuoka Institute of Technology, Japan. Since 2020, he has been an associate professor at Tokyo Medical and Dental University. His research interests are in biomedical engineering, mechatronics, and system engineering. (shinya.onogi.1979@ieee.org)

Yoshikazu Nakajima received his Ph.D. degree from Osaka University, Japan, in 1997. He worked for the head office and Information Technology R&D Center of Mitsubishi Electric Corporation from 1997 to 2000, then for the Division of Functional Diagnostic Imaging,

Graduate School of Medicine, Osaka University, where he was an assistant professor from 2001 to 2005. He was then an associate professor of the Department of Bioengineering and Intelligent Modeling Laboratory of the University of Tokyo from 2005 to 2017. Currently, he is a professor of the Institute of Biomaterials and Bioengineering, Tokyo Medical and Dental University (TMDU). (nakajima.bmi@tmd.ac.jp)