2022 44th Annual International Conference of
the IEEE Engineering in Medicine & Biology Society (EMBC)
Scottish Event Campus, Glasgow, UK, July 11-15, 2022

2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC) | 978-1-7281-2782-8/22/$31.00 ©2022 IEEE | DOI: 10.1109/EMBC48229.2022.9871060

3D endoscope system with AR display superimposing dense and wide-angle-of-view 3D points obtained by using micro pattern projector

Michihiro Mikamo[1], Ryo Furukawa[1], Shiro Oka[2], Takahiro Kotachi[2], Yuki Okamoto[2],
Shinji Tanaka[2], Ryusuke Sagawa[3], Hiroshi Kawasaki[4]

Abstract— **In recent years, augmented reality (AR) technologies have been widespread for supporting various kinds of tasks, by superimposing useful information on the users' view of the real environments. In endoscopic diagnosis, AR systems can be helpful as an aid in presenting information to endoscopists who have their hands full. In this paper, we propose a system that can superimpose shapes, which are reconstructed from an endoscope image, onto the field of view. The feature of the proposed system is that it reconstructs 3D shapes from the images captured by the endoscope and superimposes them onto the real views. As a result, the superimposed view allows the doctor to keep operating the endoscope while observing the patient's internal body with additional information. The proposed system is composed of the reconstruction module and the display module. The reconstruction module is for acquiring 3D shapes based on an active stereo method. In particular, we propose a novel projection pattern that can reconstruct wide areas of the endoscopic view. The display module shows the 3D shape obtained by the reconstructed module, superimposing on the field of view. In the experiments, we show that it is possible to perform a wide range of dense 3D reconstructions using the new projection patterns. In addition, we confirmed the usefulness of the AR system by interviewing medical doctors.**

I. INTRODUCTION

Endoscopic examination and treatment is an indispensable part of modern healthcare. For endoscopic systems, 3D information can be beneficial for understanding the size of tumor, providing information of endoscope position, and aiding operations. For those purposes, researches of acquiring 3D information from endoscopic systems have been widely researched.

On the other hand, how to show endoscope information including 3D shape to endoscopists is another challenging task, since the endoscopists needs to pay attention to several things simultaneously during the operation, such as watching the image from endoscope, controlling endoscope by hand, observing patients condition, measuring tumor size by special device, etc. If it is possible to keep watching endoscopic image during those operations, it greatly helps to improve the efficiency of entire operation.

In this paper, we propose a system with functions of 3D measurement and augmented reality (AR) display. The system is capable of measure 3D shapes of the target captured by the endoscopic camera, and showing the shape by superimposing it on real scenes.

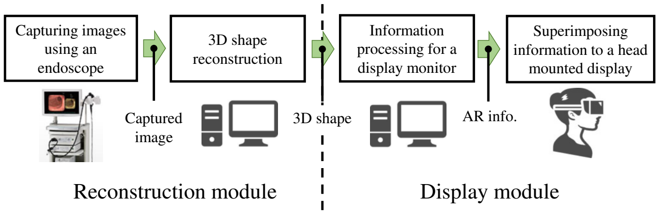


Fig. 1. System overview

For 3D measurement part, we propose active stereo system using a static pattern projector with wide angle of view, up to 90° in angle of projection. This is achieved by our new pattern for the projector, which is manufactured using a new optical element (diffractive optical element: DOE) that splits the laser to make the projection pattern with wide field of view. Note that new calibration and shape reconstruction techniques are also proposed to cope with the new pattern, which is our important contribution of the paper. By using the new DOE and reconstruction technique, we can obtain pixel-wise depth information with wide area in the endoscopic camera view.

For AR display part, we constructed a system which can display the measured 3D shapes to the endoscopist. The 3D shape is displayed as a texture-colored point cloud superimposed onto the real scene.

As a result, our proposed AR system becomes highly flexible and informative; the doctors can keep operation even if they need to conduct other tasks. In the experiments, we show that our proposed system successfully reconstruct 3D scene of wide angle of view using fisheye lens of endoscope for both phantom real stomach of pig.

II. RELATED WORKS

Structured light technique has been widely used for practical applications for 3D scanning purposes [1]. For endoscope systems, since the endoscope head always moves during operation, one-shot scanning techniques are suitable and were actually applied to endoscopic systems [2]–[5]. Nagakura *et al.* or Stoyanov *et al.* build an endoscopic stereo system by adding a static pattern projector to endoscopic or laparoscopic cameras [6], [7]. Furukawa *et al.* proposed to use a micro-sized projector that can be inserted through an

---

*This work was supported by 20H00611, 18K19824, 18H04119, 16KK0151, and Nedo 3.0 in Japan

[1] Hiroshima City University, JAPAN
[2] Hiroshima University Hospital, JAPAN
[3] Natl. Inst. of Advanced Industrial Science and Technology, JAPAN
[4] Kyushu University, JAPAN

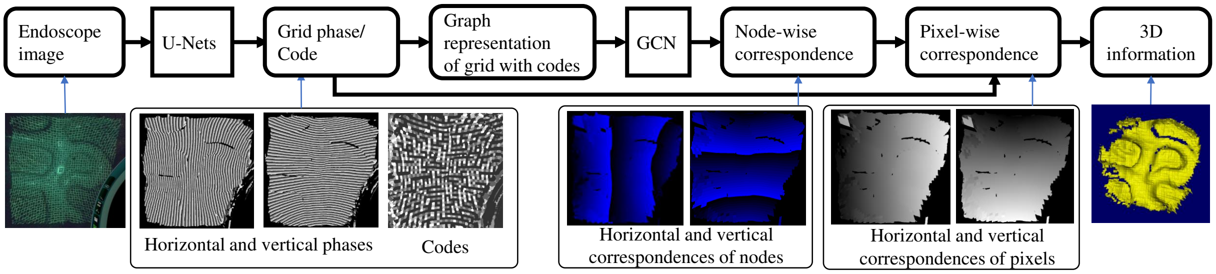


Fig. 2. Overview of the reconstruction process

instrument channel of flexible endoscopes [8]. One severe limitation of existing one-shot scan approaches is that the field of view (FOV) or projected pattern has been narrow because of limitation of optical elements and uniqueness of local pattern. To overcome the limitation, we developed a new DOE pattern which has two times larger FOV then previous one in this paper.

In terms of efficient search of correspondences between stereo pairs, deep-learning-based approaches has been proposed [9], [10]. In this paper, since structured-light pattern represented as a grid graph is used for our endoscope systems, graph convolutional networks (GCN) [11], which was proposed for aggregating node-wise features of graphs, is extended [12].

In terms of efficient way to present wide variety of information to doctors, many surgical support systems using AR displays have been developed [13]. By using the AR techniques, information obtained before and during examinations can be displayed on the visual fields. The preoperative information includes the shape model of a lesion obtained from MR images, the planned operation path of the surgical device, the size of the lesion measured before examination and so on. Gustavo *et al.* developed a method that stably positions the model on the tumor [14]. A robotic-assisted surgery shows the path of instruments [15]. On the other hand, the information obtained during examination includes heart rate, blood pressure, images of the surgical field and so on. A method shows the probability that the interested area is tumor or not, through endoscopic images [16]. Besides, by showing the shape of the neighboring organs by AR, doctors can avoid the risk of extra injures [17].

III. SYSTEM OVERVIEW

The proposed system is composed of two modules. That is, the 3D shape reconstruction module and AR display module (Fig.1). The 3D shape reconstruction module computes the shapes from captured endoscopic images. The AR display module integrates the information to the view of medical doctors with a head-mounted display (HMD) superimposed onto the surrounding environment shown by video pass-through.

The flow of the 3D reconstruction algorithm is similar to [18] and is shown in Fig.2. In this process, grid-structured pattern is projected to target surface and captured with the endoscopic camera. Then, pixel-wise grid structure in the image, and pattern codes are detected using U-Nets. The information is represented as a graph and processed with GCN to predict node-wise correspondences. The node-wise correspondences are densified to pixel-wise using pixel-wise grid information. Then, pixel-wise 3D reconstruction for the endoscopic camera images can be achieved.

The reconstruction results are then, sent to the AR display module. The 3D information is stored in the depth image and displayed as point clouds by the displaying system. The input images are used to add color to the point clouds.

Fig.5 shows the overview of the operation scene. The display system consists of a head-mounted display with pose sensing functions, and a desktop computer for rendering. The view of the medical doctor is shared to the operator by showing it on the monitor. We used a head-mounted display of HTC VIVE PRO [19] in our experiment. As for the development of AR application, Unity 2018 [20] with SR Works SDK [21] was used for the platform.

The proposed system has the following features.

- The areas of interest are reconstructed in real time from the images obtained from an endoscope.
- The reconstructed shape is superimposed on the surrounding environment by video pass-through and displayed on the HMD. This reduces the change of visual fields of medical doctor's and allows him/her to view the patient while operating the device.
- While operating the device, the visual field can be shared, which can be used as educational material.

IV. 3D RECONSTRUCTION MODULE

A. PROJECTED PATTERN

As shown in Fig.3(a), the 3D endoscope system acquire depth information with active stereo system which is composed of the endoscopic camera and a small-scaled pattern projector that can be inserted through the instrument channel (Fig.3) similar to Furukawa *et al.* [18].

The pattern is a 47x47 grid, and the grid points have five types of features represented by horizontal and vertical gaps at the grid points. In this paper, we call these features the code of the grid points. Fig.3 (b) and (c) show the projection pattern, and the codes at the grid points of the pattern, respectively. The five different colors in the figures correspond to the five different codes of the grid points.

Although the proposed system is similar to Furukawa *et al.* [18], the proposed pattern projector has been improved

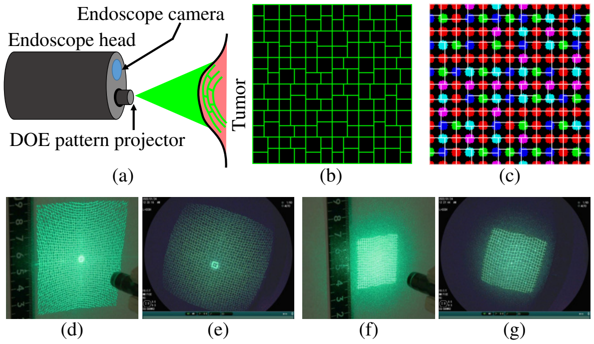


Fig. 3. The proposed system and pattern. (a) The system configuration. (b) projection pattern. (c) The corresponding node position and five kinds of features. (d) The proposed pattern with scale. (e) The proposed pattern captured with endoscopic camera. (f) The pattern of previous method [18]. (g) The pattern of [18] captured with endoscopic camera with the same condition with (e). As shown in (e) and (g), the pattern of the proposed method covers about 77 percent by the radius (about 60 percent in area) in the endoscopic camera image, whereas the pattern of [18] covers about 50 percent by the radius (about 25 percent in area). The angle of projection of the proposed method and the pattern density are 90.0° and 0.477 nodes/mm^2, respectively. Those of the previous pattern are 51.7° and 0.447 nodes/mm^2.

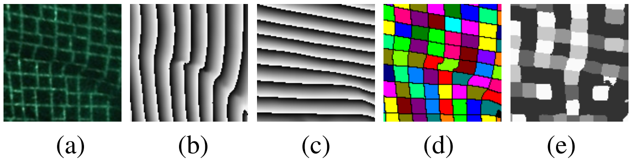


Fig. 4. Phase detection, grid detection by segmentation, and code detection: (a) An image patch of a pattern-projected surface captured by an endoscopic camera. (b)(c) Vertical and horizontal phase detection results of (a) by U-Nets trained with CG images. (d) Grid-detection results by segmenting (b) and (c). (e) Code-detection results of (a).

in two ways. First, compared to [18], the projector of the proposed system has wider angle of projection. Thus, the maximum area of the depth measurement in the endoscopic camera largely increases. Second, the projected pattern has a new codification strategy. In [18], each grid points had codes from three symbols using line gaps for only vertical direction. In the pattern of the proposed system, each grid points had codes from five symbols by using gaps for both vertical and horizontal directions. By utilizing the increased number of the symbols, we can achieve stable correspondence estimation, even with wider and increased correspondence candidates. Moreover, the proposed pattern has less directional dependency which is important because the projector can rotate in the instrument channel of the projector.

As shown in Fig.3(d) the pattern projector of the proposed system is distorted with pincushion distortion. For 3D reconstruction, intrinsic parameters of the projector (*e.g.*, the focal length), and the distortion parameters should be calibrated. For the purpose, we project the pattern onto a surface and minimized the errors between the predicted projection of the grid points on the surface and the observed grid points with respect to the intrinsic parameters, distortion parameters and the pose parameters of the projector.

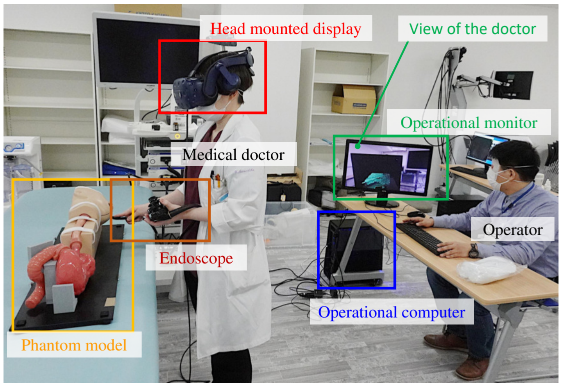


Fig. 5. Overview of the operation room

B. DENSE 3D RECONSTRUCTION

The projector of the proposed method is inserted into the instrument channel of the endoscope, and the pattern is projected onto the target surface. The repetition of the grid pattern is detected as pixel-wise phase values in the captured image. Here, phases are cyclic values (*e.g.*, $[0, 2\pi)$ in radians) that represent the vertical and horizontal repetition of the grid patterns. The detection of phase values are done by U-Nets trained by CG-synthesized images similarly to [18]. Examples of the detected phase values are shown in Fig.2(a)-(c).

Since the phase values represent repetition of the grid structure, they can be used to segment image regions into grids. To do this, images are segmented by a fixed phase values (*e.g.*, phase value of 0) of vertical and horizontal phases. Then, the segmented regions represent grid unit as shown in Fig.4(d). Then, the grid-regions are connected by horizontal and vertical connections to form a graph representing the grid.

In parallel with the phase detection, grid-codes are also detected. The codes are five classes of features assigned to the grid points. An example of the detected code information is shown in Fig.4(e). The code is then, sampled for each grid points detected as regions, and assigned to nodes of the graph representing the grid.

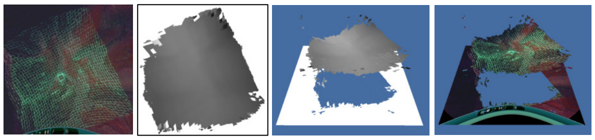
The proposed system uses a graph convolutional network (GCN) for correspondence inference. GCN can be thought of as an extension of CNN to graphs. Our system detects the grid and code information in the endoscopic image, and represent it as a graph with node-wise code attributes. The graph is then processed by a GCN to produce node-wise correspondence prediction. The GCN is trained using dataset generated from CG-synthesized images similarly with [18].

Using both node-wise correspondences and the pixel-wise phase values global pixel-wise mapping between the endoscope image and the original pattern can be obtained. As a result, we can obtain a pixel-wise 3D depth.

V. AR DISPLAY MODULE

A. SUPERIMPOSING INFORMATION ON REAL ENVIRONMENTS USING VIDEO PASS-THROUGH

Fig.5 shows an endoscopist using the proposed system. The endoscopist performs the examination while wearing the



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

Fig. 6. Colorization of a 3D shape. (a) Endoscope image. (b) The depth image obtained by the 3D reconstruction, (c) displaying the depth image with points, and (d) Points colorized by (a).

HMD. Behind the doctor, an engineer switches the images and adjusts the screen to be used. The surrounding doctors can also see the doctor's field of view during the operation by watching the monitor.

Reconstructed shapes from endoscope images are superimposed to the view of the doctor. In addition, the endoscope is superimposed on the patient, so the doctor can confirm the position of the endoscope easily.

B. COLORIZATION OF A 3D SHAPE BY TEXTURES

To colorize the point cloud, we use the depth images, the captured images and the internal camera parameters. The depth image represents the 3D points of the point clouds projected on the camera screen. The points at (X,Y,Z) in 3D space can be projected to (u,v) by the following equations.

$$u = \frac{f}{\delta_u}\frac{X}{Z} + c_u, \quad (1)$$

$$v = \frac{f}{\delta_v}\frac{Y}{Z} + c_v, \quad (2)$$

where, f is the focal length, δ_u, δ_v are the sizes of pixels, c_u, c_v are the center of the camera screen. They are included in the internal camera parameter.

By the equations, we can align Z on the camera screen. The captured images and depth images share the screen, therefore, we can colorize each point using the position on the screen. Fig.6 shows the images on the process of the colorization. Fig.6 (a) is an endoscopic image. Fig.6 (c) is the point representation of the depth image (b). By using the corresponding positions between Fig.6 (a) and (b), the points are colorized as Fig.6 (d).

VI. EXPERIMENTS

A. DENSE 3D RECONSTRUCTION

We measured objects such as medical phantoms using the proposed method, and compared the reconstruction results with a measurement result captured with a previous method [18]. The results are shown in Fig.7. In the figure, it is confirmed that our method can reconstruct wider area than the previous method. The reconstructed points of our method (a.3), (b.3) are 420,614 and 373,846 while that of the previous method (f.3) is 205,438. This significant improvement comes from the wider projection area (2 times larger than previous one) of the proposed method. Results Fig.7(a.3) and (e.3) were fit to the ground-truth shapes with an interactive closest point (ICP) algorithm [22] for evaluation.

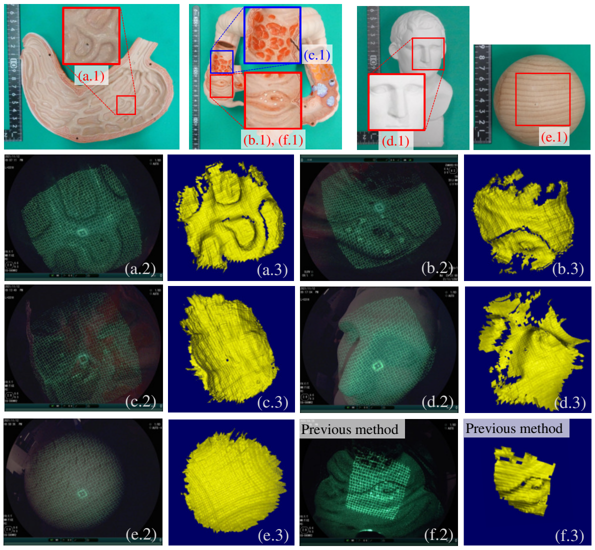
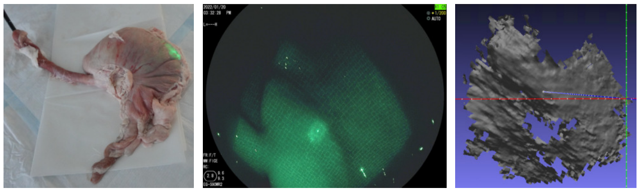


Fig. 7. The reconstruction results. The letters a–f in the legends show target regions, and the numbers 1,2 and 3 represent the appearances, input images, and the reconstruction results, respectively. The RMSE between the phantom model (a.1) and (a.3) is 1.53 [mm] and that between the actual sphere (e.1) and (e.3) is 1.04 [mm].



(a) (b) (c)

Fig. 8. The reconstruction result of surfaces of a pig's stomach. (a) An outside view of the stomach. (b) An image inside the stomach captured by the endoscope. (c) The reconstruction result of (b).

Also, we measured a pig's stomach as shown in Fig.8. It is also confirmed that the proposed method can reconstruct 3D shape for real tissues inside the stomach. In the result, there were reconstruction failure at regions around the bright spot at the center of the projected pattern and the peripheral regions. Reconstructions of such irregular reflection regions are our future works.

B. SYSTEM REVIEW BY MEDICAL DOCTORS

We implemented the AR system using a head-mounted display (HMD). To achieve AR, a camera attached to the HMD captures the surrounding environment and transfers it to the system to achieve video-pass-through. Fig.9 shows the screen that users will see. The reconstructed shape is shown on the virtual display superimposed on the surrounding environment , which is placed on the left side, whereas the virtual endoscope model is superimposed on the phantom of the human body. The endoscope model is overlaid on the phantom during the doctor manipulates the endoscope in real.

By using the system, we interviewed two doctors about their opinions about the system on the following two points;

1) Whether it is good that the displayed content follows to the movement of the visual field (Yes / No)
2) How do you feel while using the device?

Opinions regarding 1)

The display content should accompany the movement of the

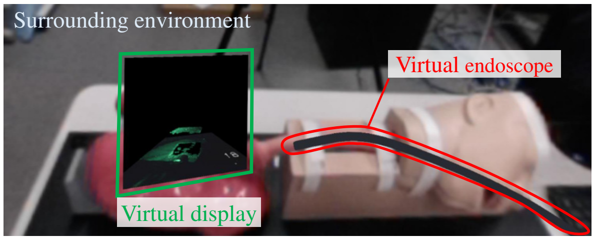


Fig. 9. The screen displayed on HMD by video pass-through

visual field. This is because that make the doctors keep eyes on the patient's condition.

Opinions regarding 2)

The doctors felt that the superimposed display was useful because it allowed them to perform the examination while simultaneously diagnosing the superimposed information and the patient. An improvement is needed in the point that there exists a difference in the distance between the image displayed on the device and the real world. It took some time to get used to it. It seemed good to be able to display more contents, such as heart rate, blood pressure, tumor size, etc.

Based on these opinions, it is confirmed that our AR display system for 3D endoscopic system is useful for real operation. In the future, it should be considered to increase the type of displaying contents and use optical-see-through devices brings to the improvement of the usability. We believe that by incorporating these opinions, the system becomes more useful.

VII. CONCLUSIONS

We proposed a 3D endoscope system with capability of showing the 3D information as augmented reality (AR) displayed on HMD. Users can measure pixel-wise depth information from the images captured with the endoscopic camera by projecting a newly-designed, wide-angle-of-projections patterns to the target surface. The 3D information is superimposed on the real operation environments shown with video pass-through on a HMD, so that the additional 3D information does not interfere the endoscopists' operations or views. We have confirmed that the pixel-wise, wide-angle-of-view 3D depth information can be measured with the proposed system. In the experiments, we constructed the real AR display system which superimpose the 3D information on the scene and got the opinions from the medical doctors that such AR system is useful and promising.

REFERENCES

[1] K. Ikeuchi, Y. Matsushita, R. Sagawa, D. Miyazaki, Y. Mukaigawa, R. Furukawa, and H. Kawasaki, *Active Lighting and Its Application for Computer Vision –40 Years of History of Active Lighting Techniques–*. Springer, 2020.

[2] C. Schmalz, F. Forster, A. Schick, and E. Angelopoulou, "An endoscopic 3d scanner based on structured light," *Medical image analysis*, vol. 16, no. 5, pp. 1063–1072, 2012.

[3] X. Maurice, C. Albitar, C. Doignon, and M. de Mathelin, "A structured light-based laparoscope with real-time organs' surface reconstruction for minimally invasive surgery," in *2012 Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, 2012, pp. 5769–5772.

[4] R. Furukawa, M. Mizomori, S. Hiura, S. Oka, S. Tanaka, and H. Kawasaki, "Wide-area shape reconstruction by 3d endoscopic system based on cnn decoding, shape registration and fusion," in *OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis*. Springer, 2018, pp. 139–150.

[5] J. Geurten, W. Xia, U. Jayarathne, T. M. Peters, and E. C. Chen, "Endoscopic laser surface scanner for minimally invasive abdominal surgeries," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2018, pp. 143–150.

[6] T. Nagakura, T. Michida, M. Hirao, K. Kawahara, and K. Yamada, "The study of three-dimensional measurement from an endoscopic images with stereo matching method," in *2006 World Automation Congress*. IEEE, 2006, pp. 1–4.

[7] D. Stoyanov, M. V. Scarzanella, P. Pratt, and G.-Z. Yang, "Real-time stereo reconstruction in robotically assisted minimally invasive surgery," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2010, pp. 275–282.

[8] R. Furukawa, R. Masutani, D. Miyazaki, M. Baba, S. Hiura, M. Visentini-Scarzanella, H. Morinaga, H. Kawasaki, and R. Sagawa, "2-dof auto-calibration for a 3d endoscope system based on active stereo," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2015, pp. 7937–7941.

[9] S. Zagoruyko and N. Komodakis, "Learning to compare image patches via convolutional neural networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 4353–4361.

[10] J. Žbontar and Y. LeCun, "Stereo matching by training a convolutional neural network to compare image patches," *The journal of machine learning research*, vol. 17, no. 1, pp. 2287–2318, 2016.

[11] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Advances in neural information processing systems*, 2016, pp. 3844–3852.

[12] R. Furukawa, S. Oka, T. Kotachi, Y. Okamoto, S. Tanaka, R. Sagawa, and H. Kawasaki, "Fully auto-calibrated active-stereo-based 3d endoscopic system using correspondence estimation with graph convolutional network," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 10 2020, pp. 4357–4360.

[13] L. Qian, J. Y. Wu, S. P. DiMaio, N. Navab, and P. Kazanzides, "A review of augmented reality in robotic-assisted surgery," *IEEE Transactions on Medical Robotics and Bionics*, vol. 2, no. 1, pp. 1–16, 2020.

[14] G. A. Puerto-Souza, J. A. Cadeddu, and G.-L. Mariottini, "Toward long-term and accurate augmented-reality for monocular endoscopic videos," *IEEE Transactions on Biomedical Engineering*, vol. 61, no. 10, pp. 2609–2620, 2014.

[15] M. Linxweiler, L. Pillong, D. Kopanja, J. P. Kühn, S. Wagenpfeil, J. C. Radosa, J. Wang, L. G. T. Morris, B. Al Kadah, F. Bochen, S. Körner, and B. Schick, "Augmented reality-enhanced navigation in endoscopic sinus surgery: A prospective, randomized, controlled clinical trial," *Laryngoscope Investigative Otolaryngology*, vol. 5, no. 4, pp. 621–629, 2020.

[16] N. Mahmud, J. Cohen, K. Tsourides, and T. M. Berzin, "Computer vision and augmented reality in gastrointestinal endoscopy," *Gastroenterology Report*, vol. 3, no. 3, pp. 179–184, 2015.

[17] D. Lee, H. J. Kong, D. Kim, J. W. Yi, Y. J. Chai, K. E. Lee, and H. C. Kim, "Preliminary study on application of augmented reality visualization in robotic thyroid surgery," *Annals of Surgical Treatment and Research*, vol. 95, no. 6, pp. 297–302, 2018.

[18] M. Mikamo, R. Furukawa, S. Oka, T. Kotachi, Y. Okamoto, S. Tanaka, R. Sagawa, and H. Kawasaki, "Active stereo method for 3d endoscopes using deep-layer gcn and graph representation with proximity information," in *2021 43rd Annual International Conference of the IEEE Engineering in Medicine Biology Society (EMBC)*, 2021, pp. 7551–7555.

[19] "VIVE pro," https://www.vive.com/us/product/vive-pro/.

[20] "Unity," https://unity.com/our-company.

[21] "SRworks SDK," https://developer-express.vive.com/resources/vive-sense/srworks-sdk/.

[22] P. J. Besel and N. D. McKay, "A method for registration of 3-d shapes," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 14, no. 2, pp. 239–256, 1992.