

# Simultaneous shape and camera-projector parameter estimation for 3D endoscopic system using CNN-based grid-oneshot scan

Ryo Furukawa<sup>1</sup> ✉, Genki Nagamatsu<sup>2</sup>, Shiro Oka<sup>3</sup>, Takahiro Kotachi<sup>4</sup>, Yuki Okamoto<sup>4</sup>, Shinji Tanaka<sup>4</sup>, Hiroshi Kawasaki<sup>2</sup>

<sup>1</sup>Graduate School of Information Sciences, Hiroshima City University, Hiroshima, Japan

<sup>2</sup>Graduate School and Faculty of Information Science and Electrical Engineering, Kyushu University, Fukuoka, Japan

<sup>3</sup>Department of Gastroenterology & Metabolism, Hiroshima University Hospital, Hiroshima, Japan

<sup>4</sup>Department of Endoscopy, Hiroshima University Hospital, Hiroshima, Japan

✉ E-mail: ryo-f@hiroshima-cu.ac.jp

Published in Healthcare Technology Letters; Received on 19th September 2019; Accepted on 2nd October 2019

For effective in situ endoscopic diagnosis and treatment, measurement of polyp sizes is important. For this purpose, 3D endoscopic systems have been researched. Among such systems, an active stereo technique, which projects a special pattern wherein each feature is coded, is a promising approach because of simplicity and high precision. However, previous works of this approach have problems. First, the quality of 3D reconstruction depended on the stabilities of feature extraction from the images captured by the endoscope camera. Second, due to the limited pattern projection area, the reconstructed region was relatively small. In this Letter, the authors propose a learning-based technique using convolutional neural networks to solve the first problem and an extended bundle adjustment technique, which integrates multiple shapes into a consistent single shape, to address the second. The effectiveness of the proposed techniques compared to previous techniques was evaluated experimentally.

**1. Introduction:** Endoscopic diagnosis and treatment of digestive tract tumours have become popular because such procedures facilitate early-stage diagnosis and minimally invasive surgery. Currently, accurate tumour-size measurement is an open problem for practical endoscopic systems to be effective. To improve measurement accuracy, three-dimensional (3D) endoscope systems have been investigated. Such systems employ various techniques such as shape from shading (SFS) [1], binocular stereo vision [2], or a single-line laser scanner attached to the head of the scope [3]. Among them, systems based on active stereo techniques that use ultra-small-sized pattern projectors that can be inserted through the instrument channel of common endoscopes, have been proposed because of their simple configurations and potential high accuracies [4–6].

A limitation of the system based on an active stereo technique proposed by Furukawa *et al.* [4] is that it easily fails to recover shapes due to strong subsurface scattering and specular effects which are common in internal tissues. To address the pattern detection problems caused by complicated surface reflectance, we propose a convolutional neural network (CNN)-based deep learning approach using U-Nets. With our approach, two independent networks, for line detection and ID-based segmentation, are constructed and trained. Then, a light sectioning method, which uses detected lines and decoded IDs is employed to reconstruct 3D shapes.

To address the problems associated with the limited reconstructed region, we propose an extended bundle adjustment (BA) technique that integrates multiple shapes into a consistent single shape by estimating 3D shapes and the calibration parameters simultaneously. Note that, if shapes are distorted due to calibration errors, they cannot be registered correctly using previous techniques. In this Letter, we address distortions caused by erroneous or unfixed relative poses between the pattern projector and the endoscopic camera, and demonstrate that our proposed technique can reduce such distortions. The proposed techniques are evaluated experimentally and compared to previous techniques [4, 7] using *ex vivo* human tumour samples, a liver phantom, and a real pig's stomach.

The primary contributions of this study can be summarised as follows. (i) 3D endoscope system that can measure a wide area

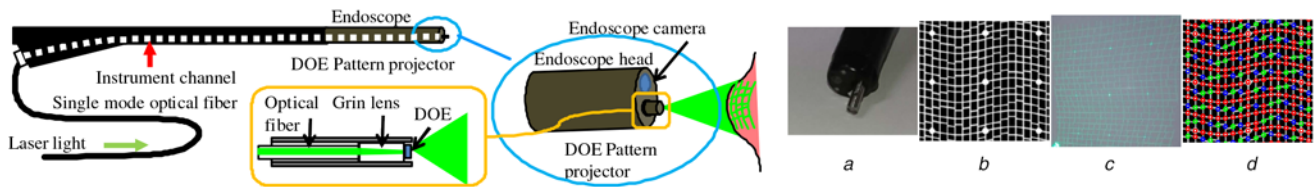
by scanning target surfaces and fusing multiple frames is realised and evaluated using a 120 frame image sequence. The system comprises a commercially available endoscope and a small-sized pattern projector that is inserted into the instrument channel. (ii) A shape fusion algorithm that does not assume fixed relative pose parameters between the camera and the projector is implemented. Note that previous studies [4, 7] do not account for distortions of depth images caused by erroneous or unfixed relative poses. (iii) To support the proposed 3D endoscope system, a CNN-based grid and code detection algorithm is implemented and evaluated experimentally. The CNN-based algorithm was shown to be more stable than a previously proposed method [4], and we demonstrated that all images in the test sequence could be reconstructed using the algorithm.

## 2. Method

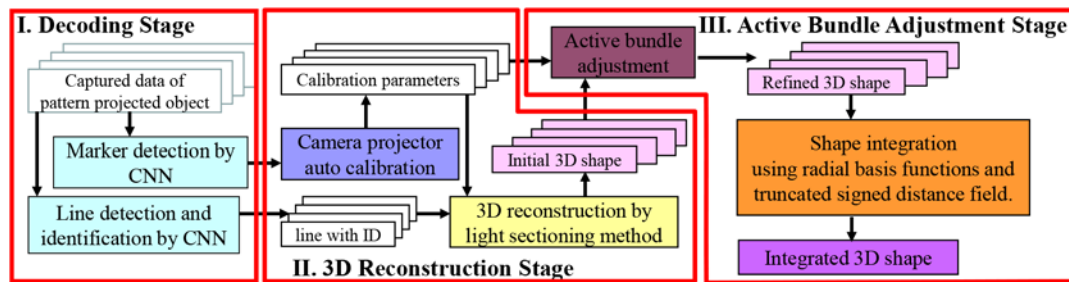
**2.1. System configuration:** For this study, a projector-camera system was constructed by inserting a fibre-shaped, micro pattern projector into the instrument channel of a standard endoscope. We used a Fujifilm EG-590WR endoscope and a pattern projector with a diffractive optical element (DOE) to generate structured-light illumination. The pattern projector can be inserted into the endoscope's instrument channel and patterns are projected from the projector to surfaces in front of the head of the endoscope, as shown in Fig. 1. As shown in Fig. 1, we used a grid pattern that is robust against subsurface scattering [8]. All vertical edges are connected; horizontal edges have small gaps, representing code letters *S*, *L* and *R*, as shown in Fig. 1*d*, where red dots mean that the right and the left edges of the grid point have the same height (code letter *S*) blue means the left side is higher (code letter *L*), and green means the right is higher (code *R*).

**2.2. Algorithm overview:** Our objective is to reconstruct 3D shapes from endoscopic images. The algorithm consists of a decoding stage, a 3D reconstruction stage, and an active BA stage, as shown in Fig. 2.

Two types of CNNs are used for the decoding stage and they should be trained in advance. To train the CNNs, actual patterns



**Fig. 1** (left) System configuration  
a DOE pattern projector,  
b Projected pattern,  
c Real pattern, and,  
d Image of the code letters represented by a colour



**Fig. 2** Overview of CNN-based decoding, 3D reconstruction and shape integration of multiple scans for endoscope systems. Shape integration involves frame-wise depth interpolation using radial basis functions [9] and shape fusion of multiple depth images using truncated signed distance fields [7]

are projected onto strong subsurface scattering objects and captured by a camera. Then, ground-truth lines and code IDs are assigned manually. Then, a U-Net [10] is trained to detect the lines and codes using the manually-assigned training data. By applying the U-Net to endoscopic images, grid structures and code letters are extracted from the projected pattern.

In the 3D shape reconstruction stage, we use an active stereo method for frame-wise 3D reconstruction. A single code letter is insufficient to determine the correspondences between the pattern projector and the endoscopic camera; therefore, similar to a previous study [4], connectivity information and epipolar constraints are used with a voting scheme to increase robustness. Once correspondences are determined, 3D shapes are reconstructed using a light sectioning method.

From the above processes, frame-wise depth images are obtained. Since the pattern projector is not fixed to the endoscope, the reconstruction process is executed using auto-calibrated extrinsic parameters. In general, auto-calibration precision is not high; thus, typically, the reconstructed 3D shape for each frame is distorted. Such distorted shapes cannot be registered correctly using a simple alignment algorithm, such as iterative closest point (ICP).

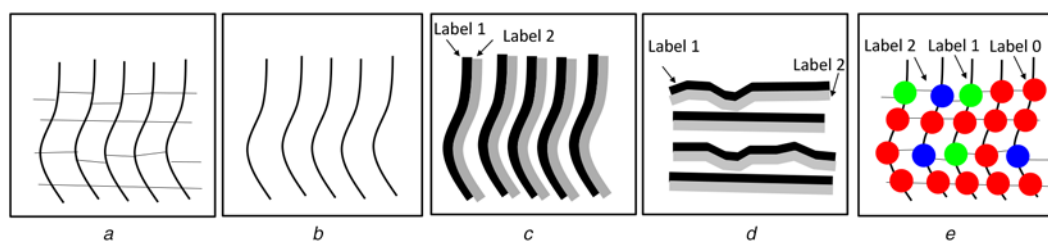
To address this problem, we propose an extended BA technique, where the extrinsic and intrinsic projector and camera parameters are

estimated such that inconsistencies between multiple frames are minimised. Note that a naive BA for multi-frame camera images cannot be applied in a simple manner to this case, because sequentially captured active stereo frames do not have the information required to retrieve correspondences between different frames. Thus, simultaneously estimating correspondences between frames and applying BA is crucial for our proposed technique.

**2.3. CNN-based feature detection and decoding for active stereo:** As described previously, the projected pattern is a grid structure with code letters (*S/L/R*) associated with the grid points.

We extract the grid-structure and gap-code information using U-Nets [10]. U-Nets can use global image structural context information to detect local features. The projected pattern has global grid structures; therefore, we expect U-Nets to use global grid structure information to detect local line features.

Training a U-Net to detect vertical lines proceeds as follows: First, sample images of the pattern-illuminated scene are collected. Then, vertical line locations for the samples are assigned manually as one-dot width curves. The one-dot width curves (Fig. 3b) are too sparse and narrow to be used directly as training label regions. Therefore, regions of five-dot width are constructed on the left and right sides of the true curves and labelled 1 and 2, as shown



**Fig. 3** Training data for U-Nets  
a Captured image,  
b Manually annotated vertical lines,  
c Labels for training vertical-line detection,  
d Labels for horizontal-line detection, and,  
e Labels for code detection

in Figs. 3c and d. By applying the trained U-Net to the endoscope images, we can obtain 3-labelled images, where the left and right side of the vertical curves are labelled 1 and 2. Here, curves are detected by extracting the borders between regions of label 1 and label 2. Note that horizontal curves are detected in a similar manner.

Gridpoint code letters are detected by applying another U-Net directly to the image rather than from the line detection results. The training data is annotated as shown in Fig. 3e. In the training process, unlabelled white background pixels are considered ‘don’t care’ regions.

Stable line and code detection are key issues for 3D reconstruction for medical applications since stabilities of measurement processes are highly important for practical use in the medical scenes. By using the labels shown in Figs. 3c–e, we can use U-Nets which have been proven to be a powerful tool for semantic segmentation for biological images. Detecting the pattern code letters directly from the input image also contributes to stable code detection. This is in contrast with the previous study [4], where gap codes are identified using line detection results and failures of line detection generally lead to code-detection failures. The proposed method, on the contrary, is free from such sequential processing problems.

**2.4. Active bundle adjustment:** As described previously, the pattern projector is not fixed to the endoscope and the projector pose parameters may vary frame by frame. Thus, projector pose parameters must be auto-calibrated for each frame. However, estimated pose parameters may include errors that cause shape distortions, and, due to such distortions, shapes from multiple frames cannot be registered using a simple ICP algorithm.

In addition, a naive BA for multi-frame data is not applicable because a sequence of active stereo data only includes correspondences between the projector and the camera within each frame, not correspondences between different frames.

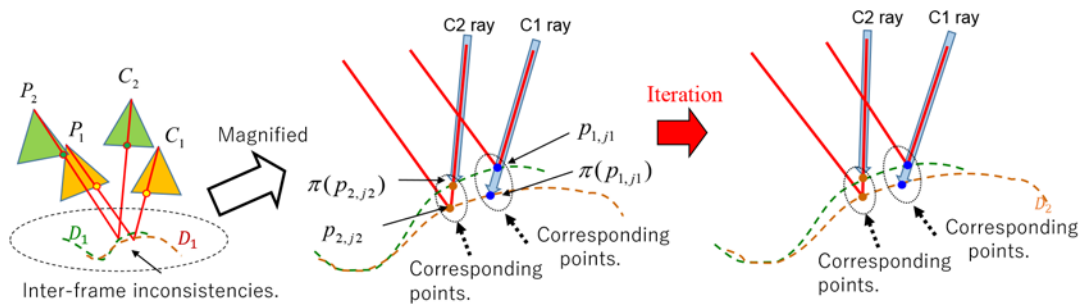
Consequently, we propose an active BA algorithm, where an ICP-like inter-frame correspondence search and BA are applied alternatively. The steps of the active BA algorithm are as follows:

*Step 1:* Initial parameters of the relative pose between the projector and the camera as well as the positions of the shapes of the frames are given as input.

*Step 2:* The 3D shape for each frame is reconstructed from the current pose information and the given correspondences between the projector and the camera for each frame.

*Step 3:* The corresponding points between different frames are sampled using a proximity relation between the frame surfaces. Note that this is similar to the process used to retrieve the corresponding point pairs in an ICP algorithm.

*Step 4:* Using the inter-frame correspondences obtained in Step 3, corresponding pairs of 3D points with associated 2D projection information onto the projectors and the cameras are generated. This step will be explained later in more detail.



**Fig. 4** Active BA correspondence finding process and cost between points. An ICP-like correspondence search between frames and BA-like optimisation of camera and projector parameters for all frames is applied iteratively until convergence

*Step 5:* The locations of the 3D points with 2D projection information obtained in Step 4 are optimised with the intrinsic and extrinsic parameters of the projectors and the cameras. In this step, both the reprojection errors of the 3D points and the distances between the corresponding points between different frames are minimised with respect to pose parameters and intrinsic parameters of the projectors and the cameras. Using the solutions of the minimisation, dense 3D shapes and the relative position between frames are updated.

*Step 6:* Repeat Steps 2–5 until convergence.

Step 4 is illustrated in Fig. 4. For the frame  $k$ , denote the camera and projector as  $C_k$  and  $P_k$ , respectively. By processing the image of frame  $k$  (captured with  $C_k$ ), the 2D correspondences between  $C_k$  and  $P_k$  are obtained. Let the  $j$ th pair of correspondences be a pair  $u_{k,j}^c$  of  $C_k$  and  $u_{k,j}^p$  of  $P_k$ . The 3D point obtained by the triangulation of  $u_{k,j}^c$  of  $C_k$  and  $u_{k,j}^p$  of  $P_k$  is denoted  $p_{k,j}$ . Let frame  $l$  be another frame; further, all correspondence pairs of frame  $l$  are also reconstructed. If the reconstructed points  $p_{l,j}$  are sparse, they should be interpolated, and the dense depth image  $D_l$  with the view  $C_l$  is obtained. Further,  $p_{k,j}$  is projected onto  $D_l$  using the pose and intrinsic parameter of  $C_l$ . If the projected pixel is a valid 3D point, we can define this point in the corresponding point of  $p_{k,j}$  in frame  $l$ . Let this corresponding point be  $\pi_l(p_{k,j})$ . The 2D projection of  $\pi_l(p_{k,j})$  can be calculated using camera  $C_l$  and projector  $P_l$ . Let these 2D points be  $v_{k,j,l}^c$  and  $v_{k,j,l}^p$ , respectively.  $p_{k,j}$  and  $\pi_l(p_{k,j})$  are corresponding points between different frames. In general, these corresponding points are different, but they are neighbouring points of frames  $k$  and  $j$ .

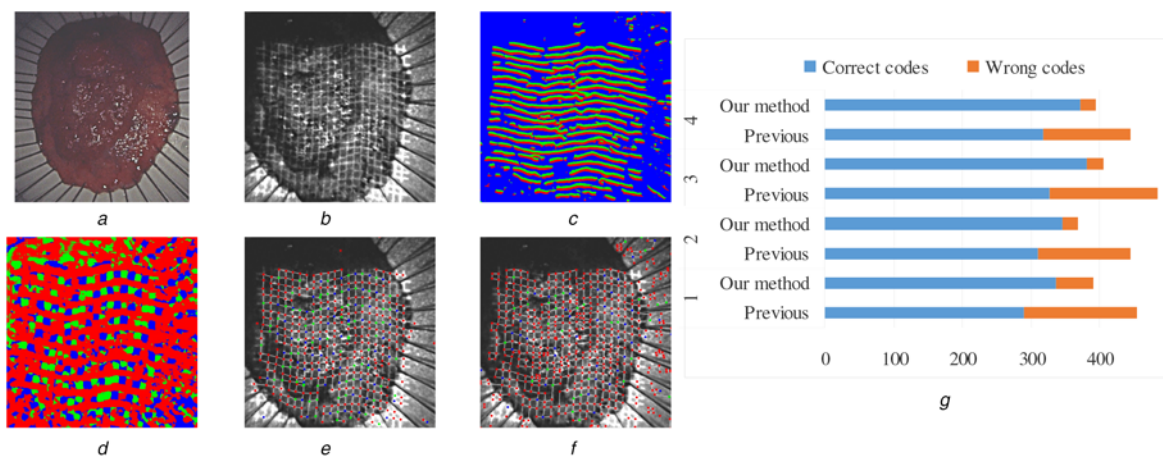
In our algorithm, we calculate BA-style reprojection errors of points  $p_{k,j}$  and  $\pi_l(p_{k,j})$ , respectively, within each frame ( $k$  and  $l$ ), and the distance errors between the corresponding points. Then, the total cost to be minimised is the weighted sum of reprojection errors of all points  $p_{k,j}$  and distance errors of all pairs of  $p_{k,j}$  and  $\pi_l(p_{k,j})$ .

The cost function to be minimised is expressed as follows:

$$L(I, E, P) = \sum_k \sum_j \left\{ \text{reproj}(p_{k,j}; I_{C_k}, E_{C_k}) + \text{reproj}(\pi_l(p_{k,j}); I_{P_k}, E_{P_k}) \right\} + w_c |p_{k,j} - \pi_l(p_{k,j})|^2 + w_b \{S(E) - \text{Const}\}^2 \quad (1)$$

where  $I_{C_k}$  and  $E_{C_k}$  are intrinsic and extrinsic parameters of camera  $C_k$ ,  $I_{P_k}$  and  $E_{P_k}$  are intrinsic and extrinsic parameters of projector  $P_k$ ,  $\text{reproj}()$  denotes BA-style reprojection errors,  $I$  is the set of intrinsic parameters  $I_{C_k}$  and  $I_{P_k}$ ,  $E$  is the set of extrinsic parameters  $E_{C_k}$  and  $E_{P_k}$ , and  $P$  is the set of  $p_{k,j}$  and  $\pi_l(p_{k,j})$ .  $L(I, E, P)$  is minimised with respect to  $I, E$  and  $P$ .  $S(E)$  is a scale function that determines the scale of the scene, and  $\text{Const}$  is a constant value. We use the sum of distances for randomly sampled devices (projectors or cameras) for  $S(E)$ . Note that, in the experiments described in Section 3,  $\text{Const}$  is set to the initial value of  $S(E)$  such that the scale





**Fig. 5** Grid and code-detection results for a cancer specimen  
*a* Appearance,  
*b* Captured image,  
*c* U-Net output for horizontal-line detection,  
*d* U-Net code detection output,  
*e* Extracted grid structures and codes using the proposed method,  
*f* Extracted grid structures and codes using a previous method [4],  
*g* Number of correct and incorrect detections of gap codes for four polyp specimens

of the processed model does not vary in the optimisation process.  $w_c$  and  $w_b$  are weights for cost terms. In real applications, such as the endoscopic systems considered in this study, the intrinsic parameters of the camera for all frames  $k$  are the same. Thus, we use a common intrinsic for all  $k$ .

The active BA can be considered a variation of the ICP algorithm, where ICP is used to estimate only a rigid transformation between frames. However, in this study, the proposed algorithm estimates the projector-camera relative pose, which significantly affects the shapes of the frames.

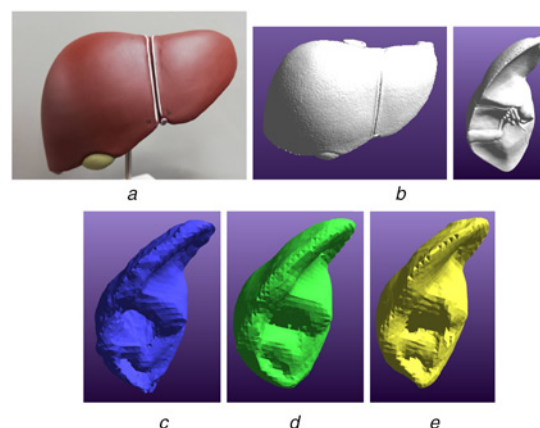
### 3. Experiment

**3.1. Evaluation of CNN-based robust grid/code detection:** To evaluate the proposed CNN-based pattern-feature extraction for endoscope images, we measured resected cancer specimens. The dataset used to train the CNN was obtained from 47 actual endoscopic images where the pattern projected by the DOE pattern projector was shown in Fig. 1*b*. The training dataset was annotated, as shown in Fig. 3. From this set, sub-image patches ( $120 \times 120$ ) were extracted and trained for 4000 iterations using Adam. For each iteration, 20 sub-images with rotation augmentation ( $-30^\circ$  to  $30^\circ$ ) and scaling (0.5–2) are input as a batch.

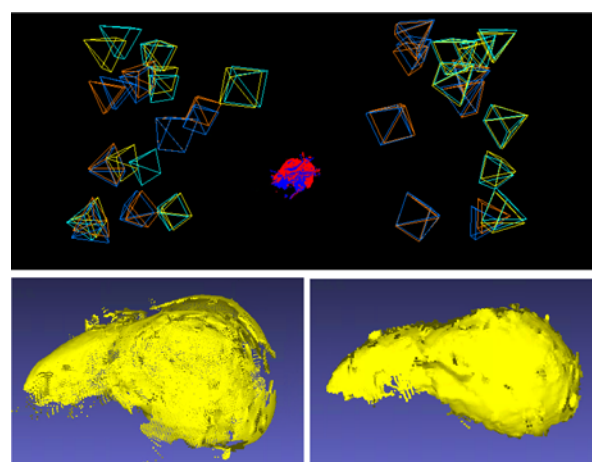
The appearance of a specimen, the image captured by the 3D endoscope with the projected pattern, and the U-Net outputs for horizontal-line detection and code-letter detection are shown in Figs. 5*a–d*, respectively. The grid structures and codes extracted from the U-Net result are shown in Fig. 5*e* and the grid extracted using a previous technique [4] is shown in Fig. 5*f*. By comparing the code colours between Figs. 5*e* and *f*, we can confirm that, in terms of code extraction, the proposed method returns better results than the previous technique.

We applied the proposed and the previous techniques to other specimens. The extraction results are summarised in Fig. 5*g*. The graph represents the numbers of grid points where gap codes are correctly or erroneously estimated. As can be seen, the total number of points detected using the previous technique [4] is larger than the number detected using the proposed technique. However, a large number of error points are included in the total number of detected points; therefore, the total number of correctly detected points is greater for the proposed technique for all the cases.

Compared to the conventional approach [4], the proposed approach is more stable due to the *contextual* line detection ability



**Fig. 6** Captured shapes of a liver phantom using a CCD-based system  
*a* Captured image for active stereo,  
*b* Ground truth shape captured using grey-code projection,  
*c–e* KinectFusion results for one, two, and three rotations around the target object, respectively. Note that since there is no explicit loop closure mechanism in KinectFusion, a large inconsistency remains



**Fig. 7** (Top) Camera positions (yellow/cyan), projector positions (orange/azure), and reconstructed points (red/blue) before/after BA. (Bottom left) 3D points before BA and (bottom right) 3D points after BA

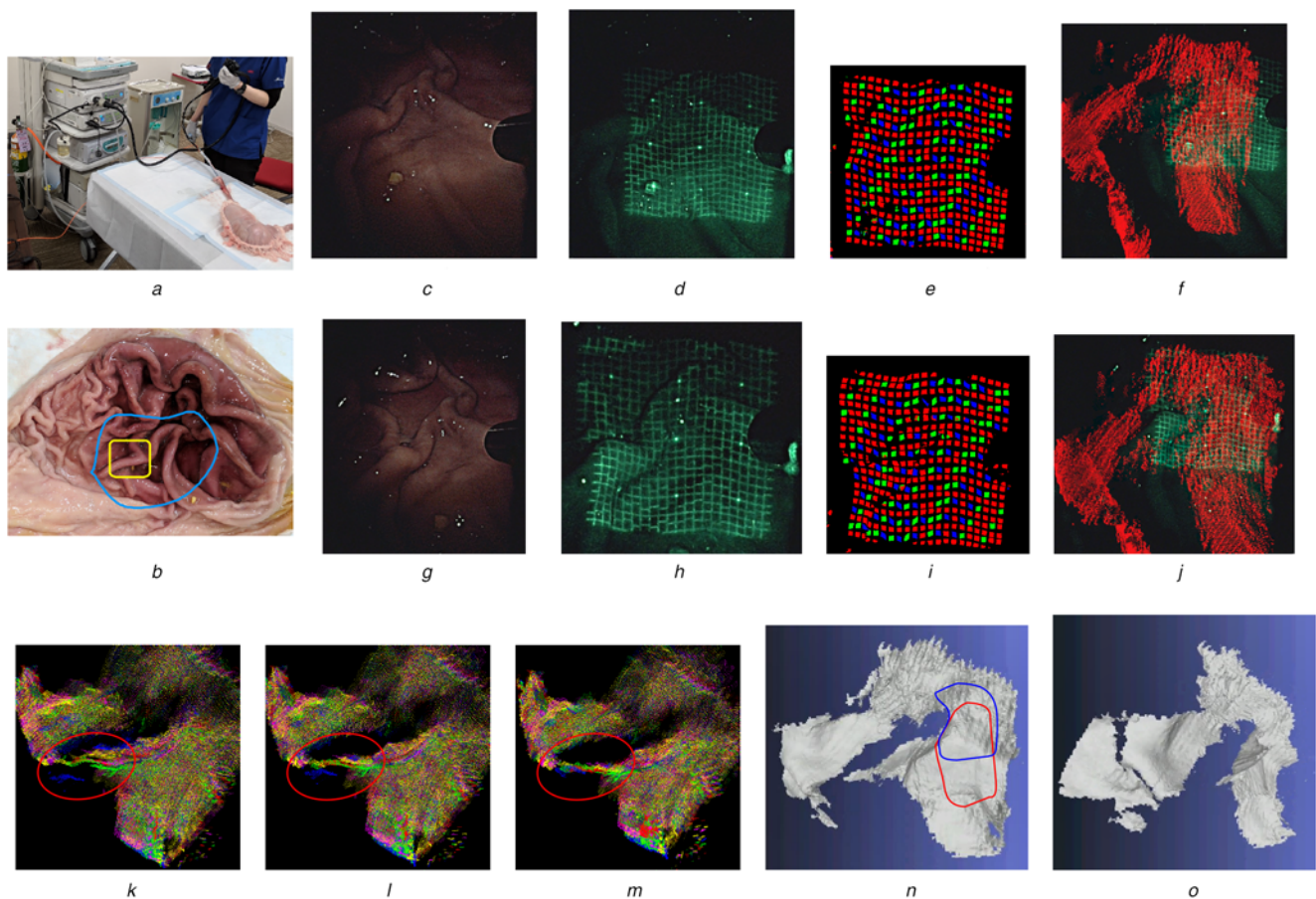
of a CNN. In other words, continuous lines are detected taking the grid-like structures around the local area into account, even if the local pixel intensity signals are extremely weak. In addition, the ability to detect codes directly from input images contributes to the stability of the proposed approach. Even if a detection failure occurs, such as a continuous line detected as separate lines, such

failures do not affect code detection, whereas, in the conventional approach [4], code detection is affected by line detection.

3.2. Evaluation of active BA using a CCD-based system: In an experiment to evaluate active BA, initially, we captured a single object (a liver phantom) multiple times (15 sets) using an experimental projector-camera system, equipped with a CCD camera. The liver phantom is approximately the same size as a human liver. We used a CCD-camera system instead of an endoscopic system because we can capture a precise shape data that can be used as a ground truth shape simply by projecting grey-code patterns [11] using an off-the-shelf video projector. Obtaining effective ground truth shapes is difficult with endoscopic cameras due to their low resolution and inherent fisheye distortions. In the evaluation, a grid pattern is projected onto the object, a liver phantom (Fig. 6a), and the images are

**Table 1** RMSE [mm] between integrated shape and ground truth obtained by KinectFusion and proposed method

Initial	Proposed	KinectFusion (single rotation) Fig. 6c	KinectFusion (two rotations) Fig. 6d	KinectFusion (three rotations) Fig. 6e
8.37	2.49	4.11	3.90	3.13



**Fig. 8** 3D reconstruction of bio-tissue inside a pig stomach. (a) Experimental environment. (b) Pig-stomach cut open after the experiment. The yellow rectangle is the area scanned by a single frame; blue region is the integrated shape region. (c)–(f) Images captured under normal and structured light conditions, extracted grids and codes, and merged shape after active BA superimposed on (c). (g)–(j) Images processed in the same manner for another frame. (k)–(m) Shapes coloured by frames before and after active BA. (n) and (o) Merged shape after active BA. In (n), the regions outlined in red and blue are the regions captured in (c) and (g), respectively

- a Environment of the experiments
- b Stomach cut open,
- c Captured image with normal light,
- d Captured image with structured light,
- e Extracted grids and codes,
- f Merged shape superimposed on (d),
- g Captured image with normal light,
- h Captured image with structured light,
- i Extracted grids and codes,
- j Merged shape superimposed on (h),
- k Shapes before active BA,
- l Shapes after 100 iterations,
- m Shapes after 210 iterations (final),
- n Merged shape,
- o Merged shape with another angle

captured by a CCD camera. The relative positions of the projector and the camera differ slightly for each frame to simulate the endoscope system. Under these conditions, auto-calibration and 3D reconstruction processes are performed for each frame.

The ICP alignment results for the initial shapes and the active BA results based on these data are shown in Fig. 7. The alignment results for the initial shape differ considerably between frames; however, the differences are reduced significantly by active BA. We also scan the same object using KinectFusion [7] with the camera rotating around the target 1, 2 and 3 times.

The results are shown in Figs. 6c–e, respectively. KinectFusion does not have a loop closure mechanism; consequently, inconsistent shapes are reconstructed. To evaluate the accuracy of the results, we use a commercial scanner to capture the same object as the ground truth, as shown in Fig. 6b. Then, we apply ICP to calculate RMSEs (Table 1). The results indicate that, due to simultaneous optimisation of entire shapes and parameters, the proposed technique can recover consistent shapes more accurately than KinectFusion.

3.3. Scan using a pig's stomach: To confirm the strength of the proposed method for actual operations, we scanned a pig's stomach (Fig. 8a). As mentioned previously, the projector and camera cannot be rigidly attached; therefore, the relative position of the devices changes continuously during operation. We applied the proposed auto-calibration technique to retrieve the initial pose and shape for each frame, and then applied active BA. Figs. 8c and d are images captured under normal and structured light conditions, respectively, and Fig. 8e is the extracted grids and codes. The merged shape after active BA is superimposed on Fig. 8f. Figs. 8g–i are images processed in the same manner for another frame. Figs. 8k–m are shapes coloured by frames before active BA, after active BA (100 iterations), and after active BA (210 iterations), respectively. Note that the gaps between shapes are minimised as active BA iterations proceed as shown by the red ellipses in the images. Figs. 8n and o are merged shapes after active BA. In Fig. 8n, the regions outlined in red and blue are the regions captured in Figs. 8c and g, respectively. From Figs. 8d, j, n and o, we could confirm that a single, consistent merged shape was obtained.

The readers may refer to the supplementary movie for more visual results.

**4. Conclusion:** We have proposed a CNN-based grid pattern detection algorithm for active stereo to solve pattern degradation problem caused by subsurface scattering and specularities. In this study, two independent networks, U-Nets, were constructed and trained for line detection and code-based segmentation purposes, respectively. These networks are integrated to retrieve robust and accurate line detection results with pattern IDs. In addition, to retrieve an integrated shape from multiple scans for an endoscope system, we propose an extended BA to estimate the shapes and parameters of a camera and a projector. Experiments conducted using several real cancer tissues with strong subsurface scattering

and specular effects demonstrated that the proposed method shows stable detection of the grid structure and codes embedded in the grid points. In addition, 3D shapes of large regions are successfully reconstructed from small reconstructed regions in each scan.

In future, to evaluate the effectiveness of the proposed system, we intend to conduct *in-vivo* experiments for test and real diagnosis purposes. In addition, we also need to evaluate the applicability of the proposed system to organs other than stomachs, such as colons.

**5. Acknowledgment:** This work is supported by JSPS/KAKENHI 16H02849, 16KK0151, 18H04119, 18K19824, and MSRA CORE14.

## 6 References

- [1] Visentini Scarzanella M., Stoyanov D., Yang G.Z.: 'Metric depth recovery from monocular images using shape-from-shading and specularities'. 19th IEEE Int. Conf. on Image Processing, Orlando, USA, 2012, pp. 25–28
- [2] Stoyanov D., Visentini Scarzanella M., Pratt P., *ET AL.*: 'Real-time stereo reconstruction in robotically assisted minimally invasive surgery'. Medical Image Computing and Computer-Assisted Intervention, Beijing, China, 2010, pp. 275–282
- [3] Grasa O.G., Bernal E., Casado S., *ET AL.*: 'Visual slam for handheld monocular endoscope', *IEEE Trans. Med. Imaging*, 2014, **33**, (1), pp. 135–146
- [4] Furukawa R., Morinaga H., Sanomura Y., *ET AL.*: 'Shape acquisition and registration for 3D endoscope based on grid pattern projection'. The 14th ECCV. Vol. Part VI, Amsterdam, the Netherlands, 2016, pp. 399–415, [http://dx.doi.org/10.1007/978-3-319-46466-4\\_24](http://dx.doi.org/10.1007/978-3-319-46466-4_24)
- [5] Lin J., Clancy N.T., Stoyanov D., *ET AL.*: 'Tissue surface reconstruction aided by local normal information using a self-calibrated endoscopic structured light system'. Int. Conf. on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 2015, pp. 405–412
- [6] Geurten J., Xia W., Jayarathne U., : 'Endoscopic laser surface scanner for minimally invasive abdominal surgeries', Frangi A.F., Schnabel J. A., Davatzikos C., *ET AL.*: (Eds): 'Medical image computing and computer assisted intervention – MICCAI 2018' (Springer International Publishing, Cham, 2018), pp. 143–150
- [7] Newcombe R.A., Davison A.J., Izadi S., *ET AL.*: 'KinectFusion: real-time dense surface mapping and tracking'. IEEE Int. Symp. on Mixed and Augmented Reality, Basel, Switzerland, 2011, pp. 127–136, Available at: [http://ieeexplore.ieee.org/xpls/abs\\_all.jsp?arnumber=6162880](http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6162880)
- [8] Furukawa R., Sanomura Y., Tanaka S., *ET AL.*: '3D endoscope system using DOE projector'. The 2016 38th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society (EMBC), Orlando, USA, 2016, pp. 2091–2094
- [9] Carr J.C., Fright W.R., Beatson R.K.: 'Surface interpolation with radial basis functions for medical imaging', *IEEE Trans. Med. Imaging*, 1997, **16**, (1), pp. 96–107
- [10] Ronneberger O., Fischer P., Brox T.: 'U-net: convolutional networks for biomedical image segmentation'. Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 2015, pp. 234–241
- [11] Inokuchi S., Sato K., Matsuda F.: 'Range imaging system for 3-D object recognition'. ICPR, Montreal, Canada, 1984, pp. 806–808