

# Shape acquisition and registration for 3D endoscope based on grid pattern projection

Ryo Furukawa<sup>1</sup>, Hiroki Morinaga<sup>2</sup>, Yoji Sanomura<sup>3</sup>, Shinji Tanaka<sup>3</sup>,  
Shigeto Yoshida<sup>4</sup>, and Hiroshi Kawasaki<sup>2</sup>

<sup>1</sup>Hiroshima City University, Japan,  
ryo-f@hiroshima-cu.ac.jp

<sup>2</sup>Kagoshima University, Japan,  
{sc110080, kawasaki}@ibe.kagoshima-u.ac.jp

<sup>3</sup>Hiroshima University Hospital, Japan,  
{sanomura, colon}@hiroshima-cu.ac.jp

<sup>4</sup>Hiroshima General Hospital of West Japan Railway Company, Japan,  
yoshida7@hiroshima-cu.ac.jp

**Abstract.** For effective endoscopic diagnosis and treatment, size measurement and shape characterization of lesions, such as tumors, is important. For this purpose, 3D endoscopic systems based on active stereo to measure the shape and size of living tissue have recently been proposed. In those works, a large problem is the degree of reconstruction instability due to image blurring caused by the strong subsurface scattering common to internal tissue. To reduce this instability problem, using a coarse pattern for structured light is an option, however it reduces the resolution of the acquired shape information. In this paper, we tackle these shortcomings by developing a new micro pattern laser projector to be inserted in the scope tool channel. There are hardware and software contributions in the paper. First, the new projector uses a Diffractive Optical Element (DOE) instead of a single lens which we proposed to solve the off-focus blur. Second, we propose a new line-based grid pattern with gap coding to counter the subsurface scattering effect. The proposed pattern is a coarse grid pattern so that the grid features are not blurred out by the subsurface scattering. Third, to increase shape resolution of line-based grid pattern, we propose to use a multiple shape data registration technique for the grid-structured shapes, which are acquired sequentially by small motions, is proposed. Quantitative experiments are conducted to show the effectiveness of the method followed by a demonstration using real endoscopic system.

## 1 Introduction

Endoscopic diagnosis and treatment on digestive tracts has become increasingly accepted methods. For example, in the diagnosis of early-stage gastric tumors, the size of the tumor is one of the most important factors for the choice of treatment. However, this is currently evaluated either by manipulation with forceps or by visual assessment alone, both of which are time consuming and

possible human errors occurring. For this reason, an easy to deploy, accurate tumor size estimation technique is necessary for endoscopic diagnosis systems.

Recently, several papers have proposed 3D endoscope systems to measure the shapes and sizes of living tissue [1–9]. Among them, we have adopted active stereo systems for developing 3D endoscopic systems, because of stability, accuracy and cost effectiveness [5, 7]. These systems use micro-sized pattern projectors with the endoscope cameras, and have successfully reconstructed several *ex vivo* human tumor samples. To implement the micro-sized pattern projector, a micro chip of the pattern and a lens are used to project a focused pattern image on the target surface. One significant limitation of such system is that a lens-based projection method can project a clear pattern only within a narrow depth range. This is because the pattern projector is based on a single lens, and thus, off-focus blurring as well as aberrations in the periphery of the field of view inevitably occur. Another critical problem in an active stereo system for endoscope, is the strong subsurface scattering effect, which is common for internal tissue. This not only blurs the projected pattern, making it more difficult to detect, but it also diminishes its brightness. One more important problem, is the sparse reconstruction of the system. Since the human body moves dynamically, we are required to scan the target object within a short period of time; such a system is also known as a oneshot scan [10–12]. Usually, the resolution of a oneshot scanning system is low because a certain area of the pattern is required to embed the projector’s positional information. Consequently, resolution of oneshot active stereo system for endoscope becomes low.

In this paper, we propose a three approach solution for the aforementioned problems. The first, is using a special optical device called Diffractive Optical Element (DOE). The DOE is a device that can project sharp patterns regardless of its depth, it solves the narrow depth of field problem. Furthermore, the DOE light efficiency is usually more than 90%, which helps to maximize the pattern detection accuracy by preserving its visibility. The second, is a novel line-based grid pattern, processing gap coding. Since high frequency information is easily lost in the presence of strong scattering effects, low frequency patterns are generally more robust. However, a low frequency pattern is not suitable for encoding rich information. We address this issue by intentionally adding gaps between adjacent lines in the grid pattern, this creates implicit unambiguous higher-level label structures that can be easily detected under strong scattering effects. The third contribution is, using a multiple-shape alignment algorithm producing a grid-like shape, which then is reconstructed using the line-based grid pattern. With this method, multiple sparse shapes of the grid patterns are effectively aligned using grid information and merged into a finer shape.

By using our DOE micro pattern projector with a line-based grid pattern, we can achieve an efficient and accurate reconstruction of tissue in metric 3D with a wide depth of field using an ordinary endoscopic system. In the experiments, we show the effectiveness of our technique with several tests using a projector-camera system, and demonstrate the reconstruction of an *ex vivo* tumor sample imaged at several distances from the camera using the endoscopic system.

## 2 Related work

For a 3D reconstruction method using endoscopes, techniques using Shape from Shading (SfS) [13–16] have been proposed. However, the SfS techniques often have stringent assumptions concerning the types of images that can be processed, such as the known or uniform diffuse reflection rates of the target surface, also the precise size measurement is generally difficult. 3D endoscopes based on binocular stereo [17, 18] are actively being researched at the present. For the binocular stereo algorithm, which is a typical passive stereo technique, correspondence problems are often very difficult, especially on textureless surfaces. Visual SLAM has also been applied on endoscope images [6]. The 3D reconstruction is only up-to-scale. Thus, it cannot be directly applied for measuring real sizes of 3D tissues. As an example of active stereo applications in endoscopy, in [1] a single-line laser scanner attached to the scope head was used to measure tissue shape, however the scope head had to be actuated in a direction parallel to the target, which limited the practical applicability of the technique. Some other vision techniques using special cameras being applied to endoscopes such as the Shape from Polarization (SfP) [2] which uses an endoscope with a rotating polarizing filter on light source, and ToF sensors [4]. Considering laparoscope systems, computer guided surgery operations have been actively researched, such as Penne *et al.* [8] or Kunert *et al.* [9]. However, the 3D system used in a laparoscope is not necessarily small and therefore cannot be used to endoscope. Recently, Furukawa *et al.* proposed a structured light system for endoscope [5, 7], which allows users to update a common endoscope system without any reconfiguration. However, there are several problems with the system and our technique can provide practical solutions to them.

The Structured light based 3D scanning systems have been studied for a number of years and these techniques are well summarized by Salvi *et al.* [19]. Based on the analysis, the techniques can be largely categorized into two method, the temporal method and the spatial encoding method. Multiple patterns are required for a temporal encoding method [20], whereas just a single static pattern is required for a spatial encoding method. Because of such differences, one of the most important advantages of a spatial encoding method is that it can capture a moving object or, in other words, the sensor can be moved during the scan. Another important benefit is a potential for compact implementation. Note that those two advantages are applicable for endoscopic systems. Based on such advantages, spatial encoding techniques have been intensively studied [21, 22, 11, 23]. To increase their stability and accuracy, most techniques use color information, however, it is difficult to use multiple colors in endoscopic systems because of space limitations. Koninckx *et al.* proposed a single color method using parallel lines [24], but it has several limitations in practical usage. Sagawa *et al.* proposed a single color method using a wave shaped pattern to encode additional information into the phase of the wave [25]. Kinect is another successful implementation using a random dot pattern [12]. However, those patterns are not considered to be useful under strong subsurface scattering environment, and thus, not suitable for endoscopic systems.

Our third contribution is based on a rigid registration algorithm. Rigid registration algorithms estimate translation and rotation of an object from two point sets, with the ICP algorithm [26] and its extension to multiple point sets [27] being the two best-known approaches. Since then, improved techniques have been intensively researched on realtime registration [28], large scale simultaneous registration [29, 30] and color compensated registration [31]. However, since they all assume a large overlap of dense shapes, they generally cannot be used whenever the shape is sparse such as a grid based reconstruction [10, 11, 23]. Recently, an ICP for the sparse point set was proposed [32], however, since the technique is still based on the correspondences of closest points, lines in the same direction are inevitably pulled together, and thus, all the grid based shapes are bundled into a single grid liked shape. Banno *et al.* proposed a method to align the multiple 3D curves which are reconstructed by the light sectioning method into single consistent shape [33], however, they assumed a base shape with holes captured in advance with 3D curves that are aligned with it to fill the holes. Therefore, the technique cannot be applied to the data which consists of independent curves only. Another approach to achieving robust registration of multiple shapes is based on 3D features extracted from input shapes [34–38]. However, stable 3D features are usually extracted only from dense 3D points and cannot be applied to grid based shapes, whose points are sparse and unevenly distributed.

### 3 DOE projector for endoscopy

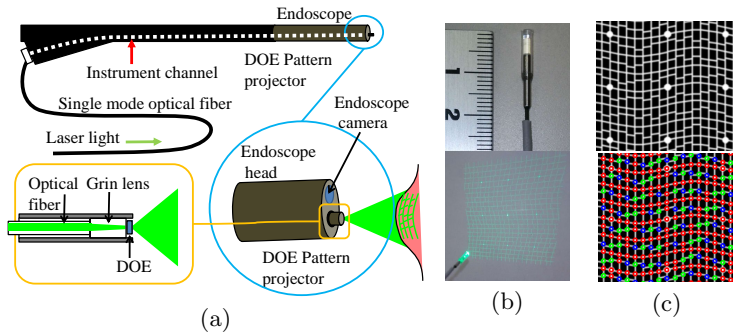
#### 3.1 System configuration

A projector-camera system is constructed by installing a micro pattern projector on a standard endoscope as shown in Fig. 1(a). For our system, we used a FujiFilm VP-4450HD system coupled with a EG-590WR scope. The DOE-based pattern projector is inserted in the endoscope through the instrument channel, the projector protrudes slightly from the endoscope head and emits structured light. The light source of the projector is a green laser module with a wavelength of 517nm. The laser light is transmitted through a single-mode optical fiber to the head of the DOE projector. In the head, the light is collimated by a grin lens, and go through the DOE. The DOE generates the pattern through diffraction of the laser light.

#### 3.2 DOE PROJECTOR FOR ENDOSCOPY

In the previous work [7], a lens with a mask pattern is used for the pattern projection. From our experience, such an optical system has a generally narrow depth of field, such as approximately an 8mm depth for a working distance of 40mm. Another problem with such optical systems is brightness efficiency, which is important since light exposure for the cameras on the endoscopes are low. To solve both problems, we have created a micro-pattern projector consisting of a DOE, a grin lens with single-mode optical fiber and a laser light source as shown





**Fig. 1.** System configuration: (a) system components, (b) DOE micro projector, (c) 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.

in Fig. 1(b). Since the DOE can project a fine, complex pattern at a greater depth range without requiring lenses, the energy loss is less than 5%. The actual specifications of the micro pattern projector are as follows. To lead the micro DOE projector into the head of the endoscope through the instrument channel, its dimensions were needed to be  $2.8\text{mm}$  in diameter and  $12\text{mm}$  in length. The working distance, valid depth range and area for the pattern projector are of  $30\text{mm}$ ,  $(-10\text{mm to } +40\text{mm})$  and  $30\text{mm}^2$ , respectively.

### 3.3 DESIGN OF PROJECTED PATTERN

**Avoidance of subsurface scattering** Since the reflectance conditions inside the body are very different from the ordinary environments for which active scanners are built, we then tailored an original pattern design specifically for the intra-operative environment. One significant casual effect under endoscopic environment is a strong subsurface scattering on the surface of internal organs. In the previous work [7], a grid pattern that consists of waved lines was used. We have found that, under the conditions of strong subsurface scattering, some of the important information of the waved grid patterns, such as the wave curvature, is difficult to be extracted or sometimes is lost completely. To avoid losing important detailed information, we considered a pattern with a larger, low-frequency structure. Existing patterns of this kind include sparse dots or a straight line-based pattern with a wide interval. However, sparse dots are difficult to decode with wide baselines and large windows, because the pattern is heavily distorted under such conditions. On the other hand, a simple line-based pattern cannot encode distinctive information efficiently [10]. Instead, we propose a line-based pattern with large intervals with a new encoding technique which is robust against the scattering effect.

Our proposed pattern consists of line segments only as shown in Fig. 1(c). The vertical lines of the pattern are all connected and straight, whereas the

horizontal segments are designed in such a way to leave a small variable vertical gap between adjacent horizontal segments and their intersections with the same vertical line. With this configuration, a higher-level ternary code emerges from the design with the following three codewords:  $S$  (the end-points of both sides have the same height),  $L$  (the end-point of the left side is higher), and  $R$  (the end-point of the right side is higher). In our actual implementation, we assign all  $S$  code for every other horizontal line, because it increasing the robustness of the line detection process. Therefore, the final codes of the pattern of Fig. 1(c)(top) are shown by color in Fig. 1(c)(bottom) and by graph representation in Fig. 3 (left).

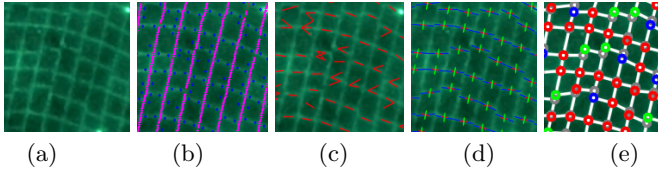
**Eliminating the singular rotation angle of pattern** As shown in Fig.1(a), since a DOE pattern projector cannot be fixed to the head of the endoscope, a rotation angle of the pattern may have some freedom, such as  $\pm 30$ degrees. If the rotation angle is near 0 degrees, the epiloral lines, which are drawn on the pattern image of Fig. 1(c), nearly coincide with the horizontal lines and the number of candidate points on the pattern for the intersection on the captured image increases; such a condition increases the ambiguity and results in low reconstruction accuracy. To mitigate the instability at such a singular rotation angle, each set of horizontal line segments in the same column is slightly inclined with a specific degree; *e.g.*, according to a piecewise long wavelength sinusoid as shown in Fig. 1(c).

## 4 3D reconstruction

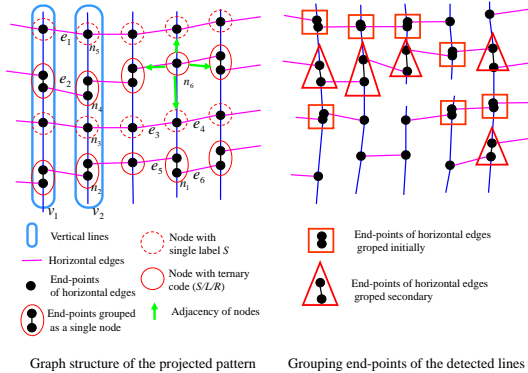
### 4.1 Detection of line patterns

The source image is first geometrically corrected on fisheye lens distortion. Noises in the image are suppressed using Gaussian filters or median filters simultaneously. Fig. 2 (a) shows an example of an input image. Then, the vertical lines on the captured image are detected because vertical lines projected onto the objects are still connected if the surface of the objects is smooth; remember that the vertical lines are straight whereas the horizontal lines are frequently disconnected small segment. Fig. 2 (b) shows detected vertical lines.

Next, the horizontal segments connecting the vertical lines are measured. In this stage, intensities of both sides of the detected vertical lines are traced and the peak values on each side are measured. The position of these peaks are the candidate end-points of the horizontal segments as shown in Fig 2 (b) as blue dots. From all line segments that connect the candidate end-points, those within a predefined range of lengths are then selected as initial candidates for the horizontal edges as shown in Fig 2 (c) as red lines. Then, to correct for small positional errors of the initial edge candidates, every pixel on the edge is moved to the local peak position along the vertical line from the original pixel. Finally, all the distances between the corrected edge candidates are calculated and if there is a pair of edge candidates that are close to each other, the edge



**Fig. 2.** Example of grid graph detection: (a) source image (a partial region), (b) identified vertical line (violet dots) and candidate end-points (blue dots), (c) initial candidates of the horizontal edges, (d) identified horizontal edges (blue line segments), and (e) detected grid graph with gap codes (colors represent same codes with Fig. 1(c)).



**Fig. 3.** Grid graph construction: (left) structure of the projected pattern, and (right) grouping end-points of the identified lines.

candidate with the smaller average intensity is removed. The final determined horizontal edges are shown in Fig. 2 (d).

## 4.2 Constructing Grid Graphs

From the identified line patterns, a grid-graph structure is constructed. The grid graph structure of the proposed pattern is shown in Fig. 3 (left). Note that the vertical lines are first identified (e.g.,  $v_1$  and  $v_2$  of the figure), and the horizontal edges are then measured between them (e.g.,  $e_1$  and  $e_2$ ). Thus, pairs of horizontal edges from both sides of a vertical line may be classified as ‘continuous’ or ‘non-continuous’. Here, ‘continuous’ horizontal edges do not mean only that they are geometrically continuous as  $e_3$  and  $e_4$ . Edges  $e_5$  and  $e_6$  are also considered continuous because their end-points are regarded as a single node  $n_1$ . Thus, classifying the continuity of horizontal segments is an important process.

As shown in Fig. 3 (left), nodes have ternary  $S/L/R$  gap codes. For example, on the vertical line  $v_2$ , nodes have  $S$  codes at  $n_3$  and  $n_5$ ,  $L$  codes at  $n_2$  and  $R$  codes at  $n_4$ . Taking this property into account, we first label the nodes from continuous horizontal edges with the label  $S$  by selecting those satisfying the small distance threshold between end points of consecutive horizontal segments

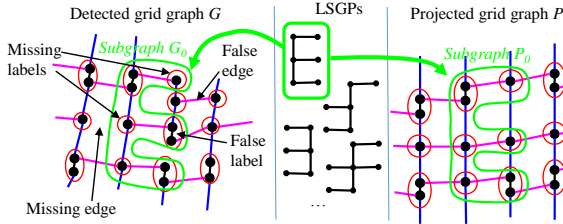


Fig. 4. Matching the detected grid graph and the projected pattern using LSGPs.

of the identified lines (end-point groups shown by rectangles in Fig. 3 (right)) Then, the remaining edge pairs with a larger vertical distance between end-points are selected, and if the horizontal lines above or below have continuous nodes, then the pair is labeled as belonging to the same horizontal piecewise sinusoid and then joined together as a single node (end-point groups shown by triangles in Fig. 3 (right)).

Each node is connected by vertical or horizontal edges with its up, down, left, or right adjacent nodes, such as  $n_6$ . Some horizontal edges might have a missing edge because of mis-identification. In this case, the node will only have either a left or a right edge, which may be matched later by looking at other connectivity in the grid graph. Fig. 2 (e). shows an example of the identified vertical and horizontal patterns with estimated gap codes.

### 4.3 Finding Correspondences using sub-graph patterns

Let the grid-graph detected be  $G$ , and let the pattern of the grid-graph in Fig. 1(c) be  $P$ . Note that graph  $G$  may lack some edges, or have undesired false edges, missing labels, or false labels of  $S/L/R$  as shown in Fig. 4. To match  $G$  to  $P$  allowing for topological errors, we exploit the notion of local sub-graph patterns (LSGPs). We define the LSGP to be a sub-graph of a grid-graph, which can be used as a template for matching the common local topologies of  $G$  to  $P$  as shown in Fig. 4. Given a dictionary of LSGPs,  $G$  may be matched to  $P$  robustly for missing or false edges. By providing multiple LSGPs and trying to match  $G$  to  $P$  using each of them, matching flexibility can be realized. In our implementation, an LSGP is represented by a path that traces all of its edges. To merge all the matching results of LSGPs, voting is used.

The matching algorithm is as follows. From a node  $n$  of  $G$ , the path of an LSGP is traced, checking for the existence of missing edges with the sub-graph denoted as  $G_0$ . Then, the a corresponding sub-graph  $P_0$  of  $P$  is searched, under the condition that the topology of  $P_0$  matches that of  $G_0$ , all the nodes of  $P_0$  fulfill epipolar constraints with the corresponding nodes of  $G_0$ , and the  $S/L/R$  labels of nodes of  $P_0$  and the corresponding nodes  $G_0$  should match to at least some pre-defined agreement ratio. Since our proposed pattern structure has a low number of candidate nodes fulfilling the epipolar constraints (at most 10

depending on the point), the dictionary search can be performed efficiently and with a low degree of ambiguity.

If a  $P_0$  with the above condition is found, all the nodes of  $P_0$  are voted on as candidate matches for the nodes of  $G_0$ . The above process is repeated for all the nodes of  $G$  with all the pre-defined LSGPs. After each iteration finishes, each node of  $G$  is checked if it fulfills the predefined thresholds of minimum number and minimum percentage of votes. If the thresholds are fulfilled, it is matched with the corresponding node of  $P$  with the maximum votes.

Once the correspondence of the captured image to the pattern is obtained, the points on the vertical and horizontal lines are reconstructed in 3D using a light-sectioning method.

#### 4.4 Taking consensus of vertical and horizontal line positions

In a real system, the calibration of the camera and the projector includes errors. With the existence of calibration errors, vertical and horizontal lines that are reconstructed by a light sectioning method generally do not intersect in 3D space (*i.e.*, they result in skewed positions). The inconsistencies between vertical and horizontal lines are not desirable for obtaining a consistent shape of the target surface.

The direct cause of a inconsistency of the vertical and horizontal lines at an intersection is the displacement of an intersection point from the corresponding epipolar line. Thus, to solve this problem, we propose a deformation of the local segments of both of the detected lines around the intersection in 2D image space so that the intersection is moved strictly onto the epipolar line. This process can be done on each of the intersections after the identified grid-graph is mapped to the corresponding projected grid pattern. Fig. 5 shows the approach.

This deformation of the lines is done locally, so that the correction of a grid-graph node does not interfere with the adjacent grid-graph nodes. One more policy about the deformation is to move the points of the lines only in the direction that is vertical to the epipolar line. This deformation can be realized by, first calculating the displacement vectors at each intersection that is vertical to the epipolar lines and move the intersections onto the epipolar lines, and then shift each of the points on the lines with the weighted means of the displacement vectors at the adjacent two nodes in both of the line directions at the point. The weights can be decided using the distances from the two nodes.

#### 4.5 Registration of reconstructed grid patterns

Once the correspondence from the captured image to the pattern is obtained, the points on the vertical and horizontal lines are reconstructed as 3D curves using a light-sectioning method. Since the line intervals between parallel lines are wide enough to avoid mis-detection by the subsurface scattering effect, the shapes can be only coarsely reconstructed. To increase the density of the sparse grid shaped 3D points, one solution is to capture the object multiple times by moving the sensor, and then, align and integrate them.

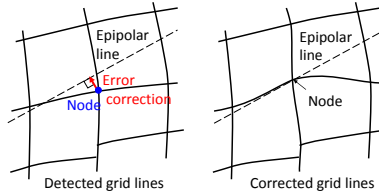


Fig. 5. Correction of position of grid nodes.

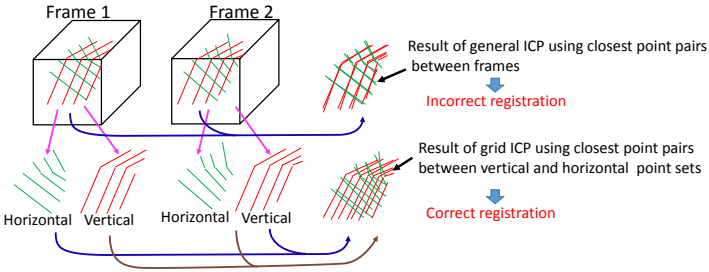
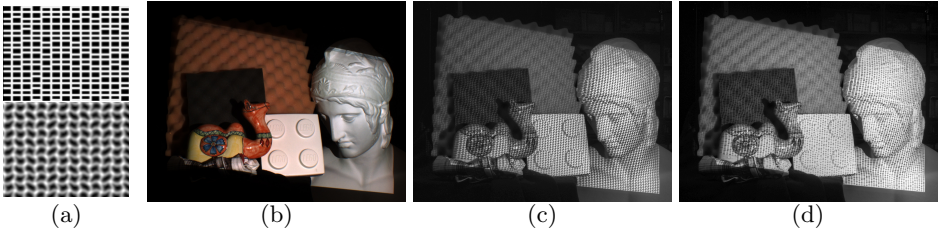


Fig. 6. Process of grid ICP

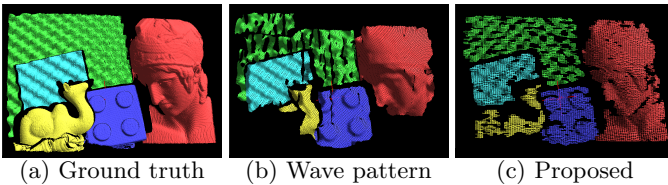
The ICP algorithm is the most used solution to conduct shape alignment between 3D shapes of a static object. The algorithm consists of two steps such as 1) searching for the closest point  $q_i$  of the scene object from point  $p_i$ , which belongs to the target object, and 2) estimate a rigid transformation  $R, t$  by minimizing  $\sum_i \|p_i - (R q_i + t)\|^2$ . Final parameters of  $R, t$  are obtained by iterating the two steps until convergence.

However, such a naive ICP algorithm does not work properly on sparse grid shapes, because the closest points from vertical/horizontal lines of the scene object are usually found on the line in same direction as the target shape, note that such incorrect corresponding points are pulled together to minimize the differences to configure an incorrect wrong shape. Noteworthy, if multiple shapes are captured with small translational motions, grid lines tends to be bundled together.

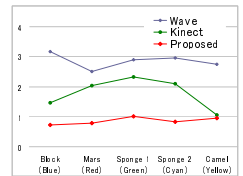
In this paper, we propose a new ICP algorithm to solve this problem. Fig. 6 shows the process of our algorithm. We first divide the grid shape into two sets of lines depending on the line directions, *i.e.*, the vertical set and the horizontal set. Then, the closest point  $q_i^v$  in the vertical line set from the point  $p_i^h$  in the horizontal line set is searched. Similarly, the closest point  $q_i^h$  in the horizontal line set from the point  $p_i^v$  which belongs to the vertical line set is found. Finally, rigid transformation parameters  $R, t$  are estimated by minimizing  $\sum_i \|p_i^h - (R q_i^v + t)\|^2 + \sum_j \|p_j^v - (R q_j^h + t)\|^2$ . Final results are obtained by iterating the aforementioned steps until convergence. Within this scenario, grid lines of the final shapes are evenly distributed realizing dense reconstruction of the object surface.



**Fig. 7.** Captured images: (a) projected patterns (top is proposed and bottom is wave pattern), (b) measurement scene, (c) a scene projected the wave pattern, and (d) a scene projected the proposed pattern.



**Fig. 8.** Reconstruction result



**Fig. 9.** RMSE (mm) of wave pattern and proposed method.

## 5 Experiments

### 5.1 3D reconstruction based on gap-coded grid pattern

To confirm the effectiveness of our gap-coded grid pattern, which is robust to objects with strong subsurface scattering and complicated textures, we prepare the scene with various materials. In addition, to compare our technique with the existing state-of-the-art technique, we reconstruct the same scene using Kinect1 and wave a pattern [39, 7], which are also a single color and oneshot scanning techniques. Projected patterns are shown in Fig. 7(a). For the ground truth, we also capture the same scene with a time coded technique, *i.e.*, gray code [20]. We use a video projector and a CCD camera for the experiment and the actual captured scene is shown in Fig. 7(b)-(d). Fig. 8 shows the reconstructed shapes and the results are summarized in Fig. 9. For the evaluation, we divided the scene with each object as shown in different colors in Fig. 8 and then calculate the value for each segment. From the results, we can confirm that our technique successfully reconstructed the shape with higher density and accuracy than the previous technique, especially on the strong subsurface scattering object (sponge) and the complicated texture object (camel figurine).

### 5.2 Evaluation of grid based ICP

Next, we evaluated the grid-based ICP technique by using 15 frames as an input, which are captured with a slight movement of the device. We also used a common

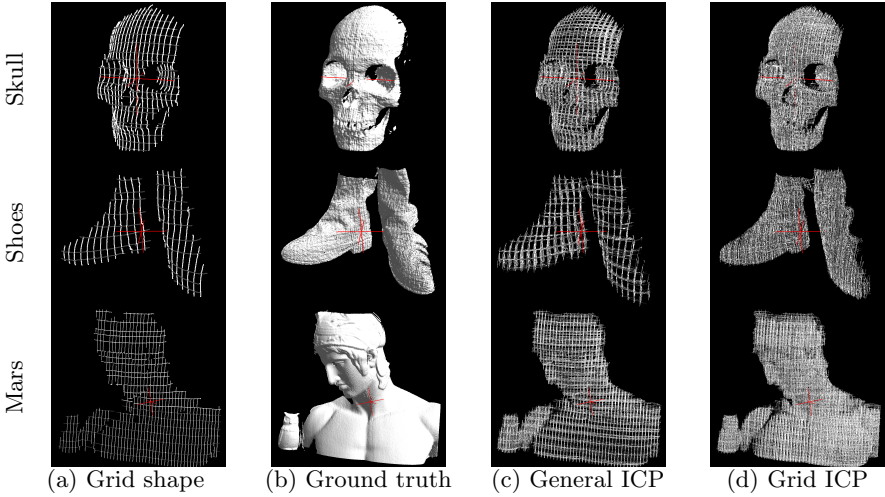


Fig. 10. Registration result

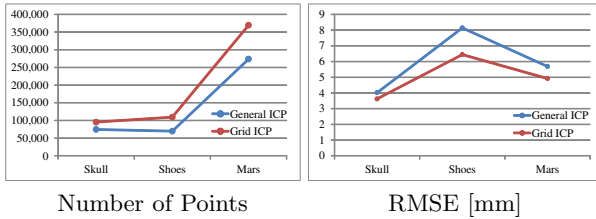


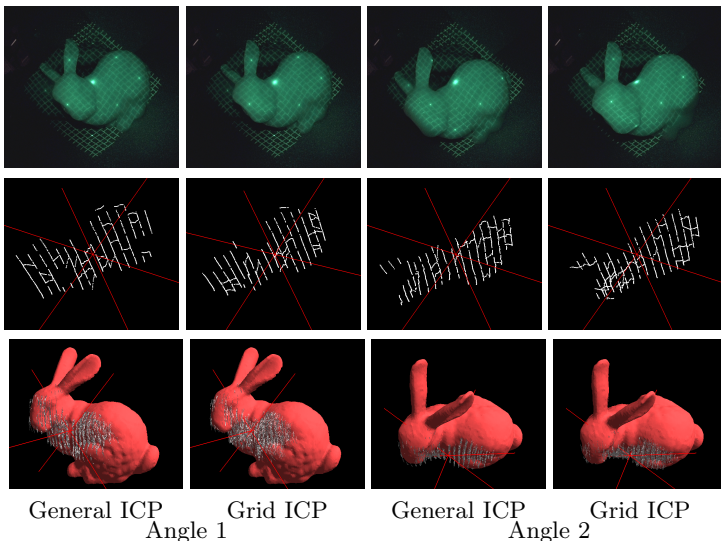
Fig. 11. Registration comparison result

ICP for comparison. Fig. 10(a) is an example of a reconstructed shape from the first frame. Fig. 10(b) is the 3D shape reconstructed with a time-encoded technique. Registration result with the grid-based ICP and a common ICP are shown in Fig. 10(c)(d) and all the results are summarized in table 11. In the table, the *Number of Points* are calculated by projecting all the points onto the image plane of the camera and counting their pixels. From the figure and the graph, we can confirm that integrated 3D points with our grid-based ICP are more evenly distributed than that with the general ICP. The RMSE is also improved by 22% compared to a common ICP.

### 5.3 3D reconstruction using endoscope images

We first evaluated the proposed 3D endoscopic system by measuring a 3D object with a known shape. The target is an output of a 3D printer using the 3D shape model of Stanford bunny object. We used this object because its ground-truth data is available. It is first scanned with the 3D endoscope by moving the object as shown in Fig. 12(top row), and each of the frames are reconstructed as shown



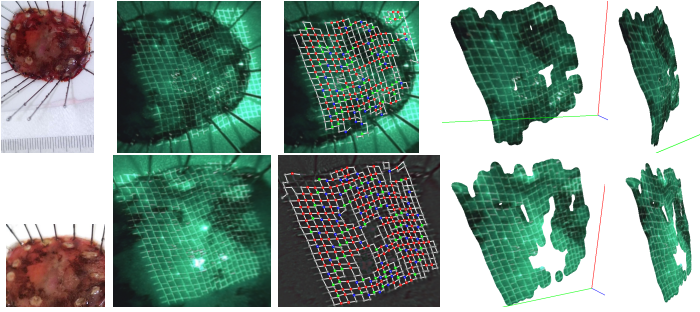


**Fig. 12.** Registration of shapes measured by the 3D endoscope. (Top row) Input images, (Middle row) reconstructed 3D points, and (Bottom row) registration results.

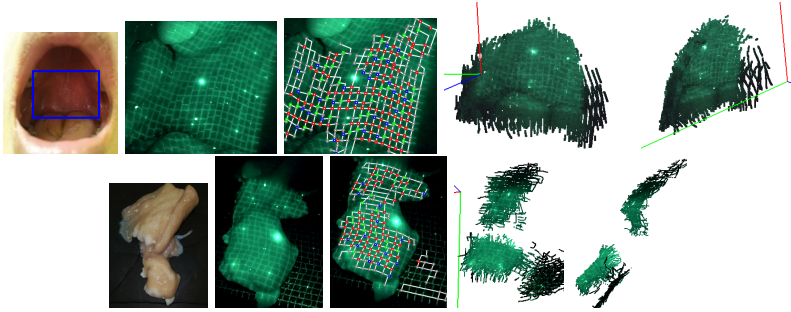
in Fig. 12(middle row). Finally, those reconstructed frames are registered with both a general ICP algorithm and our grid ICP algorithm, and compared with the ground-truth shape data. The registration results that are fit to the true shape data is shown in Fig. 12(bottom row). The results show that, in the result shape of the general ICP, the grid lines of multiple frames were pulled together and aggregated, whereas in the result of our grid ICP, the grid lines were uniformly distributed. The RMSEs between the registered 3D shapes and the ground-truth were  $1.54\text{ mm}$  for the general ICP, and  $1.20\text{ mm}$  for our grid ICP.

To evaluate the system in realistic conditions, a biological specimen extracted from a human stomach in an endoscopy operation is measured, as shown in Fig. 13. We captured the tissue image, from different distances, about  $25\text{mm}$  and  $15\text{mm}$ . Note that, since the projector is based on DOE, the projected patterns from both images, at different distances are sharp enough. Using our proposed algorithm, both images could be nearly reconstructed, except for the regions that were affected by bright specular highlights. Our future work will be to avoid the effects of these specular highlights. Other than these points, the correspondence between the detected grid points and the projected pattern were estimated accurately, and the shape of the specimen was successfully reconstructed.

We also measured the inside of the human mouth (palate), and a piece of intestines of a cattle, with the endoscopic system and the result were registered with the proposed grid ICP, as shown in Fig. 14. From the figures, we can confirm that shapes are densely reconstructed with our gap-based DOE pattern projector, and grid based ICP technique.



**Fig. 13.** 3D reconstruction of real tissues from human stomach measured with distance of about  $25mm$  (top row) and about  $15mm$  (bottom row). (From left column to right) The appearance of the sample, the captured image, the identified grid graph with gap codes, and the reconstructed shapes (the front and the side).



**Fig. 14.** 3D reconstruction of (top row) the inside of the human mouth (palate), and (bottom row) a piece of intestines of a cattle.

## 6 Conclusion

We proposed a 3D endoscopic system based on an active stereo, where the pattern projector consists of a DOE that generates a special line based grid pattern. By using a DOE which is free from the blur effect, sharp patterns are projected at wide depth ranges while keeping a strong intensity (usually more than 90% efficiency). In addition, by using a line based grid pattern, the severely blurred pattern caused by the subsurface scattering effect, is also robustly detected and decoded. We also propose a new reconstruction algorithm for the gap coding implemented on the pattern. Since shapes reconstructed from grid pattern are usually sparse, an ICP algorithm specialized for grid patterns is proposed. The potential of the technique was verified by intensive experiment using projector-camera systems and demonstrated by reconstructing the shape of some bio tissues such as the surface of a human palate, or a biological specimen extracted from the stomach of a human subject at various distances from the endoscope. Our future work is to test the system with real diagnosis.

## References

1. Grasa, O., Bernal, E., Casado, S., Gil, I., Montiel, J.: Visual slam for handheld monocular endoscope. *Medical Imaging, IEEE Transactions on* **33**(1) (Jan 2014) 135–146
2. Liao, H.: 3d endoscopic image reconstruction using shape from polarization and central projection model compensation. *International Journal of Computer Assisted Radiology and Surgery* **5**(1) (2010) S281
3. Aoki, H., Furukawa, R., Aoyama, M., Hiura, S., Asada, N., Sagawa, R., , Kawasaki, H., Tanaka, S., Yoshida, S., , Sanomura, Y.: Proposal on 3D endoscope by using grid-based active stereo. In: *The 35th Annual International Conference of IEEE EMBC*. (2013)
4. Köhler, T., Haase, S., Bauer, S., Wasza, J., Kilgus, T., Maier-Hein, L., Feußner, H., Hornegger, J.: Tof meets rgb: novel multi-sensor super-resolution for hybrid 3-d endoscopy. In: *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2013, Springer* (2013) 139–146
5. Furukawa, R., Aoyama, M., Hiura, S., Aoki, H., Kominami, Y., Sanomura, Y., Yoshida, S., Tanaka, S., Sagawa, R., Kawasaki, H.: Calibration of a 3d endoscopic system based on active stereo method for shape measurement of biological tissues and specimen. In: *EMBC*. (2014) 4991–4994
6. Grasa, O.G., Bernal, E., Casado, S., Gil, I., Montiel, J.: Visual slam for handheld monocular endoscope. *Medical Imaging, IEEE Transactions on* **33**(1) (2014) 135–146
7. Furukawa, R., Masutani, R., Miyazaki, D., Baba, M., Hiura, S., Visentini-Scarzanella, M., Morinaga, H., Kawasaki, H., Sagawa, R.: 2-dof auto-calibration for a 3d endoscope system based on active stereo. In: *Engineering in Medicine and Biology Society (EMBC), 2015 37th Annual International Conference of the IEEE*. (Aug 2015) 7937–7941
8. Penne, J., Schaller, C., Engelbrecht, R., Maier-Hein, L., Schmauss, B., Meinzer, H.P., Hornegger, J.: Laparoscopic quantitative 3d endoscopy for image guided surgery. In: *Bildverarbeitung für die Medizin, Citeseer* (2010) 16–20
9. Kunert, W., Storz, P., Kirschniak, A.: For 3d laparoscopy: a step toward advanced surgical navigation: how to get maximum benefit from 3d vision. *Surgical endoscopy* **27**(2) (2013) 696–699
10. Kawasaki, H., Furukawa, R., , Sagawa, R., Yagi, Y.: Dynamic scene shape reconstruction using a single structured light pattern. In: *CVPR*. (June 23-28 2008) 1–8
11. Sagawa, R., Ota, Y., Yagi, Y., Furukawa, R., Asada, N., Kawasaki, H.: Dense 3D reconstruction method using a single pattern for fast moving object. In: *ICCV*. (2009)
12. Microsoft: Xbox 360 Kinect (2010) <http://www.xbox.com/en-US/kinect>.
13. Deguchi, K., Sasano, T., Arai, H., Yoshikawa, Y.: Shape reconstruction from an endoscope image by shape from shading technique for a point light source at the projection. *CVIU* **66**(2) (1997) 119–131
14. Visentini-Scarzanella, M., Stoyanov, D., Yang, G.: Metric depth recovery from monocular images using shape-from-shading and specularities. In: *ICIP, Orlando, USA* (2012) 25–28
15. Wu, C., Narasimhan, S., Jaramaz, B.: A multi-image shape-from-shading framework for near-lighting perspective endoscopes. *IJCV* **86** (2010) 211–228

16. Ciuti, G., Visentini-Scarzanella, M., Dore, A., Menciassi, A., Dario, P., Yang, G.Z.: Intra-operative monocular 3d reconstruction for image-guided navigation in active locomotion capsule endoscopy. In: *BioRob*. (2012) 768–774
17. Nagakura, T., Michida, T., Hirao, M., Kawahara, K., Yamada, K.: The study of three-dimensional measurement from an endoscopic images with stereo matching method. In: *Automation Congress, 2006. WAC '06. World*. (July 2006) 1–4
18. Stoyanov, D., Visentini-Scarzanella, M., Pratt, P., Yang, G.: Real-time stereo reconstruction in robotically assisted minimally invasive surgery. In: *MICCAI*. (2010) 275–282
19. Salvi, J., Batlle, J., Mouaddib, E.M.: A robust-coded pattern projection for dynamic 3D scene measurement. *Pattern Recognition* **19**(11) (1998) 1055–1065
20. Sato, K., Inokuchi, S.: Range-imaging system utilizing nematic liquid crystal mask. In: *Proc. Int. Conf. on Computer Vision*. (1987) 657–661
21. Je, C., Lee, S.W., Park, R.H.: High-contrast color-stripe pattern for rapid structured-light range imaging. In: *ECCV*. Volume 1. (2004) 95–107
22. Zhang, L., Curless, B., Seitz, S.: Rapid shape acquisition using color structured light and multi-pass dynamic programming. In: *3DPVT*. (2002) 24–36
23. Sagawa, R., Kawasaki, H., Furukawa, R., Kiyota, S.: Dense one-shot 3D reconstruction by detecting continuous regions with parallel line projection. In: *ICCV*. (2011) 1911–1918
24. Koninckx, T., Gool, L.V.: Real-time range acquisition by adaptive structured light. *IEEE Transaction Pattern Analysis Machine Intelligence* **28**(3) (2006) 432–445
25. Sagawa, R., Sakashita, K., Kasuya, N., Kawasaki, H., Furukawa, R., Yagi, Y.: Grid-based active stereo with single-colored wave pattern for dense one-shot 3d scan. In: *3D Imaging, Modeling, Processing, Visualization and Transmission (3DIMPVT), 2012 Second International Conference on*, IEEE (2012) 363–370
26. J.Besl, P., D.McKay, N.: A method for registration of 3-d shapes. *IEEE Trans. Pattern Analysis and Machine Intelligence* **14**(2) (1992) 239–256
27. NEUGEBAUER, P.: Geometrical cloning of 3d objects via simultaneous registration of multiple range image. *Proc. Int'l Conf. Shape Modeling and Applications* (1997) 130–139
28. Hall-Holt, O., Rusinkiewicz, S.: Stripe boundary codes for real-time structured-light range scanning of moving objects. In: *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*. Volume 2., IEEE (2001) 359–366
29. Oishi, T., Sagawa, R., Nakazawa, A., Kurazume, R., Ikeuchi, K.: Parallel alignment of a large number of range images. In: *IEEE Conf. 3DIM2003*. (2003) 195–202
30. Oishi, T., Kurazume, R., Nakazawa, A., Ikeuchi, K.: Fast simultaneous alignment of multiple range images using index images. In: *3-D Digital Imaging and Modeling, 2005. 3DIM 2005. Fifth International Conference on*, IEEE (2005) 476–483
31. Johnson, A.E., Kang, S.B.: Registration and integration of textured 3d data. *Image and vision computing* **17**(2) (1999) 135–147
32. Bouaziz, S., Tagliasacchi, A., Pauly, M.: Sparse iterative closest point. *Computer Graphics Forum (Symposium on Geometry Processing)* **32**(5) (2013) 1–11
33. Banno, A., Masuda, T., Oishi, T., Ikeuchi, K.: Flying laser range sensor for large-scale site-modeling and its applications in bayon digital archival project. *Int. J. Comput. Vision* **78**(2-3) (July 2008) 207–222
34. Li, H., Hartley, R.: The 3d-3d registration problem revisited. In: *Proc. Int'l Conf. Computer Vision*. (2007) 1–8
35. Yang, J., Li, H., Jia, Y.: Go-icp: Solving 3d registration efficiently and globally optimally. In: *IEEE Int'l Conf. Computer Vision*. (2013) 1457–1464

36. Wang, R., Choi, J., Medioni, G.: 3d modeling from wide baseline range scans using contour coherence. The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2014) 4018–4025
37. Rusu, R., Blodow, N., Beetz, M.: Fast point feature histograms (FPFH) for 3d registration. In: Int'l Conf. Robotics and Automation. (May 2009) 3212–3217
38. Rusu, R., Blodow, N., Marton, Z., Beetz, M.: Aligning point cloud views using persistent feature histograms. In: Int'l Conf. Intelligent Robots and Systems. (2008) 3384–3391
39. Sagawa, R., Sakashita, K., Kasuya, N., Kawasaki, H., Furukawa, R., Yagi, Y.: Grid-based active stereo with single-colored wave pattern for dense one-shot 3D scan. In: 3DIMPVT. (2012) 363–370