



# Auto-calibration Method for Active 3D Endoscope System Using Silhouette of Pattern Projector

Ryo Furukawa<sup>1</sup>(✉), Masahito Naito<sup>1</sup>, Daisuke Miyazaki<sup>1</sup>, Masahi Baba<sup>1</sup>, Shinsaku Hiura<sup>1</sup>, Yoji Sanomura<sup>2</sup>, Shinji Tanaka<sup>2</sup>, and Hiroshi Kawasaki<sup>3</sup>

<sup>1</sup> Hiroshima City University, Hiroshima, Japan  
{ryo-f,miyazaki,baba,hiura}@hiroshima-cu.ac.jp,  
naito@ime.infohiroshima-cu.ac.jp

<sup>2</sup> Hiroshima University Hospital, Hiroshima, Japan  
{sanomura,colon}@hiroshima-cu.ac.jp

<sup>3</sup> Kyushu University, Fukuoka, Japan  
kawasaki@ait.kyushu-u.ac.jp

**Abstract.** In this paper, we develop an active stereo system for endoscope which requires auto-calibration, because a micro pattern projector is inserted through the instrument channel during an operation and cannot be fixed to the endoscope. For solution, a new auto-calibration technique with full 6-DOF estimation of an active stereo system without any extra devices nor extra pattern projections is proposed. In the technique, the pattern projector itself is simultaneously captured with a target scene by an endoscope camera and the silhouette of the pattern projector is used to conduct 2D-3D matching by using the knowledge of the shape of the projector. In addition, the markers which is included in the projection pattern are extracted and the distances from the closest epipolar lines are calculated as for the cost function. To enhance the robustness of the reconstruction, we also propose a simple high dynamic range (HDR) imaging system for an endoscope by alternating the input power of the pattern projector ON and OFF to blink the pattern so that exposure time will vary with beat frequency, realizing a virtual multi-exposure camera. By applying our auto calibration technique with HDR imaging system, we achieved a robust and accurate reconstruction of tissue in metric 3D under practical operation of the endoscopic system, such as reconstruction of the inside of a real stomach of a pig.

## 1 Introduction

In this paper, we develop an active stereo system for endoscope which requires auto-calibration, because a micro pattern projector is inserted through the instrument channel during an operation and cannot be fixed to the endoscope.

Auto-calibration is a basic, but yet an important research topic for computer vision. Recently, auto-calibration is applied to active stereo systems and it has been researched for two decades. In general, calibration process requires stable

and sufficient number of correspondences, however, this requirement is hard to achieve for active stereo systems before calibration, because correspondences are only acquired after calibration; this is a typical chicken and egg problem. To solve the problem, additional devices, such as a spherical calibration tool with known size [1] or extra pattern projections, such as gray code are used [2]. Since it is not possible to install such extra devices nor extra patterns to endoscopes, it makes auto-calibration more difficult than ordinary case.

In this paper, we propose a new auto-calibration technique for an active stereo system with full 6-DOF estimation without any extra devices nor extra pattern projections. In our technique, a pattern projector itself is simultaneously captured by an endoscope camera as well as a target scene; note that simultaneous capture of the projector and the scene is not difficult for the endoscope camera because the field of view of the camera is usually extremely wide. In actual process, we extract the silhouette of the pattern projector and conduct 2D-3D matching by using the knowledge of the shape of the projector. In addition, we extract markers which are included in the pattern and calculate the distances from the epipolar lines as for the cost function; note that we just impose 9 markers into the pattern, which makes it possible to specify the markers in the captured images.

In the paper, we also propose a simple high dynamic range (HDR) imaging system for an endoscope. Usually HDR images are synthesized by using multiple-exposure images, however, it is impossible to precisely control the exposure of the camera frame by frame for commonly available endoscope systems. Therefore, we alternate the input power of the pattern projector ON and OFF to blink the pattern with certain frequency so that exposure time will be varied with beat frequency, realizing a virtual multi-exposure camera.

By applying our auto calibration technique to the HDR images which are efficiently captured by our proposed system, we can achieve a robust and accurate reconstruction of tissue in metric 3D under practical operation of the endoscope system. In the experiments, we show the effectiveness of our technique with several tests using the real system, and demonstrate the successful reconstruction of the inside of a real stomach of a pig.

## 2 Related Work

3D endoscopes based on binocular stereo are actively being researched at the present [3,4]. For the binocular stereo algorithm, which is a typical passive stereo technique, correspondence retrieval is essentially difficult, especially on textureless surfaces. To cope with textureless surfaces, techniques using Shape from Shading (SfS) have been proposed [5], however, the 3D reconstruction is only up-to-scale and it cannot be directly applied for measuring real sizes of 3D tissues.

An active stereo technique is a simple solution for the aforementioned problems. Using color-coding for the projected pattern is one solution [6,7]. However, accurate classification of illumination colors on textured tissues may need careful

adjustment. Some other vision techniques using special cameras being applied to endoscopes such as ToF sensors are proposed [8,9]. However, the resolution of ToF sensor is inevitably low [8] or the size is larger than those of RGB sensors and only applicable to laparoscope systems [9]. Recently, Furukawa *et al.* proposed a structured light system for endoscope [1,10,11], which allows users to update a common endoscope system without any reconfiguration.

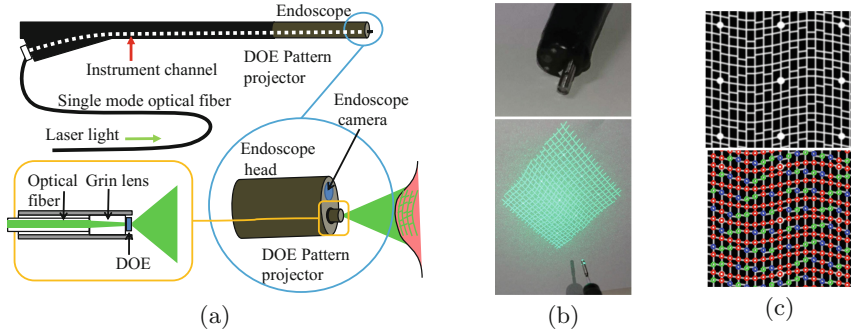
In terms of auto-calibration, there is a long history for binocular stereo and many papers have been published so far and they are summarized in [12]. On the other hand, there is a few techniques for active stereo systems, especially for structured light systems [1,2,13–15]. The structured light system can be categorized into three, such as light sectioning method, temporally encoding method and spatially encoding method. An auto-calibration for light sectioning method estimates three unknown parameters for each laser sheet by using intersections of multiple lasers [13,14]. In terms of an auto-calibration for temporally encoding method, since the method usually assumes video projector, which can project an arbitrary pattern such as gray code, dense and accurate correspondences are easily retrieved and a calibration is efficiently achieved [2,15]. Unlike the temporally encoding method, spatially encoding method, which is used in our endoscope system, basically cannot change the pattern and has essential difficulty on retrieving correspondences, and thus, only few technique is proposed [1]. In the method [1], since the authors attached the projector to the head of the endoscope and only 2-DOF remains, they can use a special calibration tool with known size to estimate the remaining 4-DOF in advance. In our technique, we estimate full 6-DOF without any extra devices nor projections, and similar techniques has never been proposed yet with our best knowledge.

About HDR image synthesis, usually multiple-exposure images are assumed [16]. However, it is difficult to capture images with different exposures using video cameras. There are several techniques which achieve HDR synthesis and tone-mapping for video [17,18], however active lighting conditions are not considered. If a lighting condition can be precisely controlled with camera synchronization, multi-exposure images are easily retrieved, however, the system becomes complicated. In our system, instead of adopting such a complicated system, we simply alternate the input power to blink the pattern ON-and-OFF with no synchronization mechanism, and such an approach has never been published yet.

### 3 DOE-Based Laser Pattern Projector for Endoscopy

#### 3.1 System Configuration

A projector-camera system is constructed by installing a micro pattern projector on a standard endoscope system 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 laser 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



**Fig. 1.** System configuration: (a) System components. (b) DOE micro projector inserted through the instrument channel of an endoscope. (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. (Color figure online)

a wavelength of 517 nm. The laser light is transmitted through a single-mode optical fiber to the head of the DOE projector. The DOE generates the pattern through diffraction of the laser light.

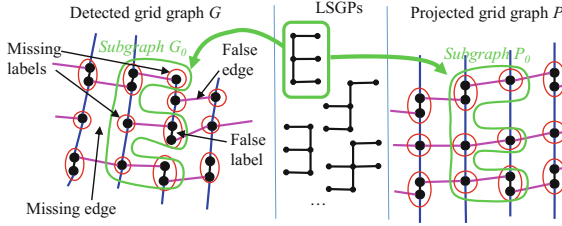
Our system is based on active stereo method proposed by Furukawa *et al.* [11], in which a gap-based grid pattern is used for avoiding effect of subsurface scattering that is harmful for 3D reconstruction. Here, we describe the pattern and 3D reconnection method briefly.

The projected pattern consists of only line segments as shown in Fig. 1(c) (top). The vertical lines of pattern are all connected and straight, whereas the horizontal segments are designed in 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 left side is higher). The codes of the pattern of Fig. 1(c) (top) are shown by color in Fig. 1(c) (bottom).

### 3.2 3D Reconstruction

The source image is first geometrically corrected on the fish-eye lens distortion. Noises of the image are suppressed using Gaussian filters or median filters at the same time. The projected vertical and horizontal lines are detected in the undistorted image using the line detection algorithm from Sagawa *et al.* [19]. This method can detect projected parallel lines whose approximate directions are known, ignoring intersecting non-vertical lines, based on loopy belief propagation.

From the detected line patterns, grid-graph structure is constructed by detecting intersections between the horizontal and vertical lines. Then, each



**Fig. 2.** Matching the detected grid graph and the projected pattern using LSGPs.

node is connected with its up, down, left, or right adjacent nodes by vertical or horizontal edges. Some horizontal edges might have a missing edge because of misdetection. In this case, the node will only have either a left or a right edge, which may be later matched by looking at other connectivity of the grid graph. Figure 11(f) shows examples of the detected vertical and horizontal patterns with estimated gap codes.

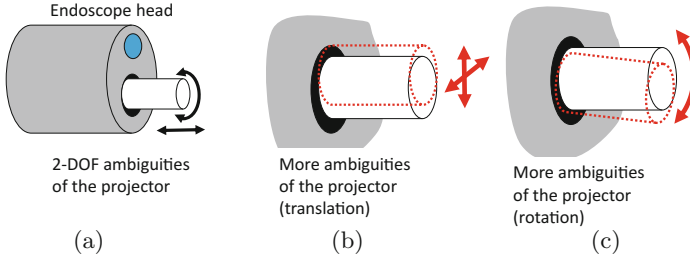
Let the detected grid-graph be  $G$  and let the grid-graph of the pattern in Fig. 1(c) be  $P$ . Note that graph  $G$  may lack some edges or have undesired false edges, missing labels or false labels of  $S/L/R$  as shown in the left part of Fig. 2. To match  $G$  and  $P$  allowing topological errors, we exploit the notion of local sub-graph patterns (LSGPs). We define an LSGP to be a sub-graph of a grid-graph used as a template for matching common local topologies of  $G$  and  $P$ . In Fig. 2, the left part shows  $G$ , the right part shows  $P$ , and the middle part shows LSGPs. Given a dictionary of LSGPs,  $G$  may be matched to  $P$  robustly to missing or false edges. By providing multiple LSGPs and trying to match  $G$  and  $P$  using each of them, flexible matching 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 scheme is used.

Once the correspondences 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 Auto-calibration of the Projector Position

In this system, the target surface, which is projected by pattern projector, is captured by the endoscope camera. Since the head of the projector is not tightly fixed to the endoscope, the relative position between the projector head and the endoscope camera varies during endoscopic operations, such as bending the head. Since, for active stereo techniques, the position of the projector is an important parameter for 3D reconstruction, such unstable condition is problematic for robust and accurate shape measurement.

Furukawa *et al.* [1] modeled the relative position by 2-DOF rigid transformation, where projector translates along or rotates around the axis of the instrument channel (Fig. 3(a)). This 2-DOF model could be applied to our system if the pattern projector's outer diameter perfectly fitted to the inner diameter



**Fig. 3.** Ambiguities of the projector. (a) 2-DOF ambiguity proposed in Furukawa *et al.* [1]. (b), (c) More freedom that is not modeled in Furukawa *et al.* [1] caused by margin between the projector and the instrument channel.

of the instrument channel. However, there should be some margin between the projector and the channel in order for the projector to be inserted during the endoscopic operations. Thus, in real situations, the projector have more freedom to move beyond the 2-DOF model within the margin (Fig. 3(b), (c)).

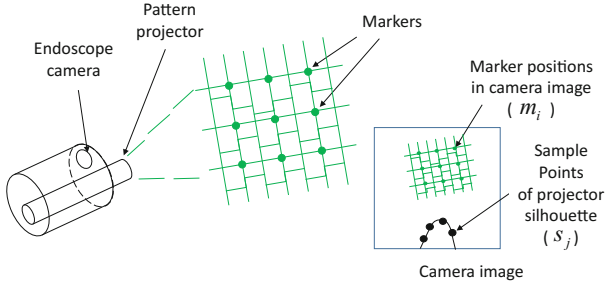
Another limitation of the work of Furukawa *et al.* [1] is that they estimate the projector's position by detecting a marker drawn on the projector from the endoscope image. In real situations where endoscope image is captured in dark environments, markers drawn on the projector are difficult to be detected from the captured image.

In the proposed system, we use silhouette of the projector and the markers embedded in the grid pattern projected onto the target surface. The silhouette of the projector can be observed from the captured image, even if there are not illumination except for the projected pattern. The markers in the grid pattern can be also detected from the same image (see Fig. 11(d) for an example, where the projector silhouette can be observed at the bottom of the image).

The actual process is as follows: From the input image captured for measurement, markers in the grid patterns ( $m_i$ ) are detected. Also, several points in the projector's silhouette ( $s_j$ ) are also sampled (Fig. 4). The auto-calibration is processed as an optimization of 6-DOF rigid transformation parameters that represent projector's position, while using the 2-DOF freedom described in [1] as 'soft' constraints. To achieve this, we divide the estimated 6 parameters into 2 sets of parameters: one set is for 2-DOF freedom described in [1] and the other is for the rest 4 parameters. We regard the 2-DOF parameters (the former set) as freely changing parameters, since they represent the motion of the pattern projector that rotates around and translates along the axis of the instrument channel, while we suppress the rest 4 parameters (the latter set) since they are deviation from the 2-DOF freedom of [1]. Because of this 'soft' 2-DOF constraint, the estimated projector position does not have scale ambiguity.

The optimized cost function is defined as follows:

1. The cost function takes 6 parameters  $p_1, p_2, q_1, q_2, q_3, q_4$ , representing the 3D position of the projector (rotation  $\mathbf{R}$  and translation  $\mathbf{t}$ ) relative to the endoscope camera, where  $p_1$  and  $p_2$  are 2-DOF parameters described in [1], and  $q_1, \dots, q_4$  represents the rest of the 6-DOF rigid transformation.



**Fig. 4.** Input points for auto-calibration of the projector.

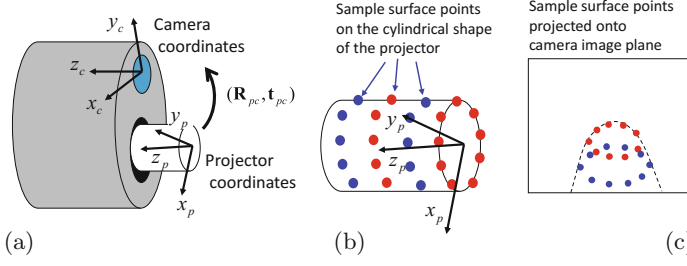
2. For the markers  $m_i$ , the corresponding epipolar line is calculated, and the distance between  $m_i$  and the epipolar line is calculated as  $g_i(p_1, p_2, q_1, q_2, q_3, q_4)$ .
3. The virtual silhouette of the projector is rendered as a cylinder moved by the rigid transformation  $\mathbf{R}$  and  $\mathbf{t}$ . From each  $s_j$ , the minimum distance from  $s_j$  to contours of the rendered silhouette is calculated as  $h_j(p_1, p_2, q_1, q_2, q_3, q_4)$ .
4.  $\sum_i \{g_i(p_1, p_2, q_1, q_2, q_3, q_4)\}^2 + w_1 \sum_j \{h_j(p_1, p_2, q_1, q_2, q_3, q_4)\}^2 + w_2 \sum_{k=1}^4 (q_k)^2$  is calculated as the cost value, where  $w_1$  is weight of the cost of silhouette fitting and  $w_2$  is weight for the ‘soft’ constraints of the 2-DOF freedom [1], which suppress the parameters  $q_1, \dots, q_4$ .

The cost function is optimized with respect to  $p_1, p_2, q_1, q_2, q_3$  and  $q_4$ . In current implementation, selection of the marker position and sampling points on the silhouette contour are conducted manually for each frame in image sequences and auto-calibration should be conducted for each frame. Further automation for point selection will be our future work.

#### 4.1 Details of Implementation

Generally, projectors and cameras can be described in the same model (*i.e.*, pinhole-camera model). The standard coordinates of 3D camera calibration is the camera coordinates, which is  $(x_c, y_c, z_c)$  in Fig. 5(a), where the origin is the optical center of the endoscope camera, and the  $z$ -axis goes through both the optical center and the principal point of the image plane. The projector coordinates can be also modeled similarly, as shown as  $(x_p, y_p, z_p)$  in Fig. 5(a). The relative position between the projector and the camera is described as the rigid transformation  $(\mathbf{R}_{pc}, \mathbf{t}_{pc})$  between these two coordinates.

In the work of Furukawa *et al.* [1], this rigid transformation is defined by 2-DOF rigid transformation, which is a composition of rotation around the  $z$ -axis of the projector coordinates, and translation in parallel with the same axis. Let the two parameters be  $p_1$  and  $p_2$ , and the rest 4-DOF of rigid transformation be  $q_1, \dots, q_4$ , then the transformation from the projector-coordinates to the camera coordinates be described as



**Fig. 5.** (a) The camera/projector coordinates and rigid transformation  $(\mathbf{R}_{pc}, \mathbf{t}_{pc})$ . (b) Surface points on the cylinder-like projector shape. (c) Sample surface points projected onto the image plane of the camera.

$$\begin{pmatrix} x_c \\ y_c \\ z_c \end{pmatrix} = \mathbf{R}_x(q_1) \mathbf{R}_y(q_2) \mathbf{R}_z(p_1) \begin{pmatrix} x_p \\ y_p \\ z_p \end{pmatrix} + \begin{pmatrix} q_3 \\ q_4 \\ p_2 \end{pmatrix} = \mathbf{R}_{pc} \begin{pmatrix} x_p \\ y_p \\ z_p \end{pmatrix} + \mathbf{t}_{pc} \quad (1)$$

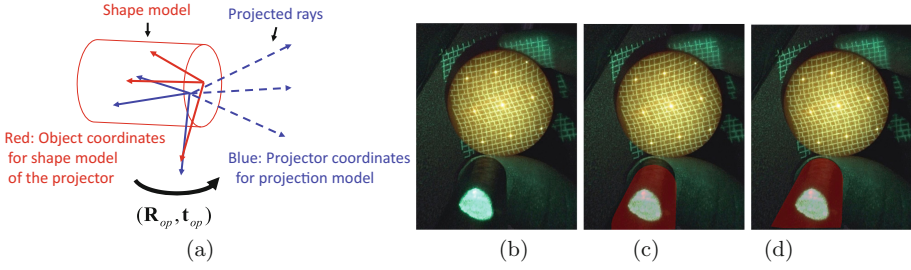
where  $\mathbf{R}_x$ ,  $\mathbf{R}_y$  and  $\mathbf{R}_z$  rotations around  $x$ ,  $y$ , and  $z$ -axis, respectively.

In calculation of the cost function, the silhouette of the cylinder-shaped projector is rendered using transformation (1). We project the 3D points onto the image plane and render a 2D convex hull of them (Fig. 5(b), (c)). In the cost function, the distances from the contours of the rendered virtual silhouette to sampled points on the imaged silhouette ( $s_j$ ) should be calculated. To estimate these distances, we obtain Euclidean distance transformation from the convex hull using the OpenCV library and look-up the pixels of the distance transformation at the sampled points. Approximately, the point set on the surface of the projector can be modeled as a cylinder surface whose axis is the same as  $z$ -axis of the projector coordinates, and the cylindrical bottom is at the origin of the projector coordinates. However, the projector coordinates are defined from the geometry of optical projection, where the cylinder is the physical shape of the projector. Thus, the precise relative positional relationships between them are unknown, and the object coordinates of the cylinder shape and the projector coordinates have a small deviation as shown in Fig. 6(a). This deviation can be described as a rigid transformation  $(\mathbf{R}_{op}, \mathbf{t}_{op})$  and can be calibrated in the following steps.

First, the pattern is projected onto a sphere with a known size, and image with the projected patterns and the projector silhouette is captured. Then, from the projected patterns, the relative position of the projector coordinates is estimated using the calibration method described in [10]. Then, the deviation between the object coordinates and the projector coordinates is estimated by fitting the virtual silhouette of the projector to the imaged silhouette points with respect to  $(\mathbf{R}_{op}, \mathbf{t}_{op})$  using the similar method as the auto-calibration except that the epipolar constraints are not used.

The effect of calibration of  $(\mathbf{R}_{op}, \mathbf{t}_{op})$  is shown in Figs. 6(b), (c) and (d). Figure 6(b) is not the image used for calibration of  $(\mathbf{R}_{op}, \mathbf{t}_{op})$ , so that we can





**Fig. 6.** (a) Projector coordinates and object coordinates of the projector shape rigid transformation  $(\mathbf{R}_{op}, \mathbf{t}_{op})$  between them. (b) An image with calibration object. (c) Projector shape overlayed onto (b) without calibration of  $(\mathbf{R}_{op}, \mathbf{t}_{op})$ . (d) Projector shape overlayed onto (b) with calibration of  $(\mathbf{R}_{op}, \mathbf{t}_{op})$ .

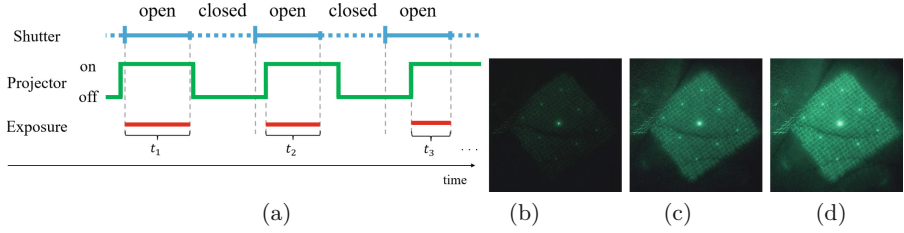
validate the estimated  $(\mathbf{R}_{op}, \mathbf{t}_{op})$ . In Fig. 6(c), the virtual shape of the projector is not overlayed correctly onto the real image of the projector. In Fig. 6(d), the error between these shapes is drastically reduced.

## 5 HDR Synthesis Using Asynchronous Blinking Pattern

To synthesize HDR image, usually multiple-exposure images are required. However, it is not possible to capture such images with commonly available endoscopic systems. To solve the issue, we control the light source instead of the camera, *i.e.*, blinking the pattern. Note that we just switch the pattern ON-and-OFF periodically without synchronization mechanism.

Using just two levels of intensity for the projector are fine for synthesizing HDR images because we utilize ‘beat’ between the frequencies of the camera capturing and the projector illumination. Suppose  $n$  Hz for the camera and  $m$  Hz for the projector, then, beat frequency becomes  $n - m$  Hz ( $n > m$ ). With ON-and-OFF switching signal, it makes half of exposure time, *i.e.*,  $1/(2m)$  sec. for the projector, whereas we cannot control the shutter speed of the camera. Suppose the camera shutter speed to be  $\alpha/n$  sec ( $\alpha < 1$ ), then the exposure time varies between  $\min(0, 1/(2m) + \alpha/n - 1/m)$  to  $\max(1/(2m), \alpha/n)$  as shown in Fig. 7. In our experiment, we set  $n = 30$  and  $m = 26$  and then the exposure time varies with 4 Hz and we can synthesize HDR using the 8 frames. Then, tone mapping is applied to the HDR images to make 8 bit images, which allows to use conventional image processing tools.

To synthesize HDR images, exposure times are supposed to be known, however, since the camera and the projector are not synchronized, it cannot be retrieved with our system. For solution, we estimate the exposure time only from captured image set. In our implementation, we simply average the intensity of the pattern excluding outliers with simple thresholding technique for each frame and use the ratio of the average as for the exposure time.



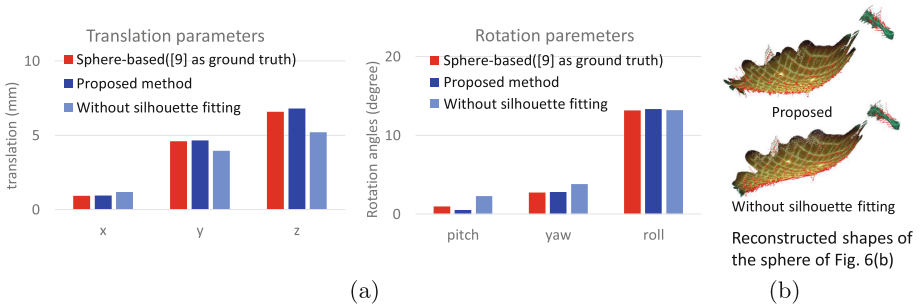
**Fig. 7.** Acquiring multi-exposure images: (a) Timings of exposure time and pattern projection time ( $t_i$  represents an exposure time). (b)–(d): Multi-exposure images. One cycle of multi-exposure images includes about 8 images  $I_1, I_2, \dots, I_8$ .  $I_1, I_3$  and  $I_5$  are shown here.

## 6 Experiments

### 6.1 Auto-calibration of the Projector Position

To confirm the accuracy of the proposed auto-calibration, we compare the results of auto-calibration and calibration based on known objects. As already explained, Fig. 6(b) shows an image with a sphere that can be used as a calibration object. Generally, calibration by using a known-shaped object such as sphere-based calibration [10] is supposed to be more accurate than auto-calibration which cannot use known-shaped objects. Thus, we compared the auto-calibration results with this data, assuming that [10] is the ground truth.

In the experiment, the image of Fig. 6(b) is calibrated by sphere-based calibration and auto-calibration. To show effectiveness of using silhouette of the pattern projector, the auto-calibration was tested with and without silhouette fitting cost function (*i.e.*, weight  $w_1$  in the cost function was set to zero in the case without silhouette fitting). Figure 8(a) shows the results, in which 6 parameters of the translation and rotation are compared. The proposed method was more accurate than the method without silhouette fitting and par with calibration method based on a known-shaped object [10]. In Fig. 8(b), we can observe



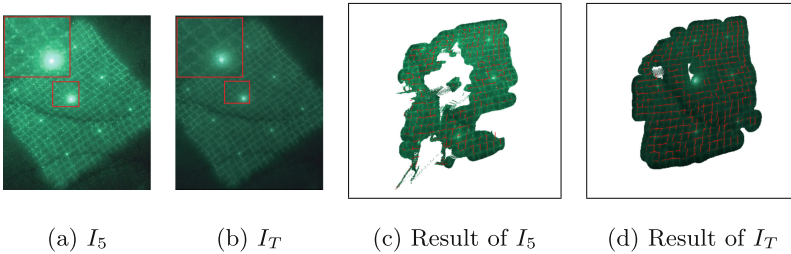
**Fig. 8.** Comparison of the proposed method with sphere-based calibration [10] as ground truth and an auto-calibration without silhouette fitting of the pattern projector.

that the reconstructed sphere by the method without silhouette fitting was partially distorted, whereas the shape generated by the proposed method did not have such distortion.

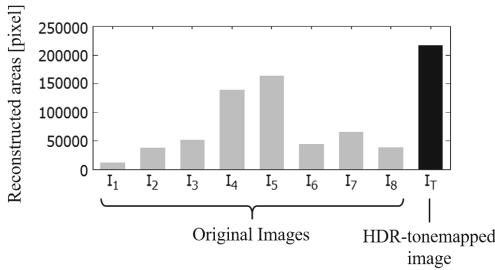
### 6.2 Improvement Using HDR Image for 3D Reconstruction

To show effectiveness of the HDR image generation, we tested our algorithm using a human hand as the target object. We first captured images of the target surface which is projected by a blinking laser pattern projector. Although pattern is just illuminated bright and dark repeatedly, we could obtain an image sequence with different exposures. We have extracted 8 images of one multi-exposure cycle  $I_1, I_2, \dots, I_8$ .  $I_1, I_3$ , and  $I_5$  are shown in Fig. 7.

Then, HDR image is created from the sequence, and then, tone-mapped for 3D reconstructed algorithm. The HDR image is shown in Fig. 9(b). The 3D reconstructed results with/without HDR algorithm are shown in Fig. 9(c) and (d). The numbers of reconstructed points for each frame and HDR image are shown in Fig. 10, proving that the area which was successfully reconstructed from the HDR image was larger than the results of any of the original input images. In Fig. 9, we can see that the regions around the brightest center marker



**Fig. 9.** Comparison between the original and the tone-mapped HDR images. (a)  $I_5$  from Fig. 7, which was most successfully reconstructed in images  $I_1$  to  $I_8$ . (b) The tone-mapped HDR image generated from  $I_1$  to  $I_8$ . (c) 3D reconstruction result of (a). (d) 3D reconstruction result of (b).

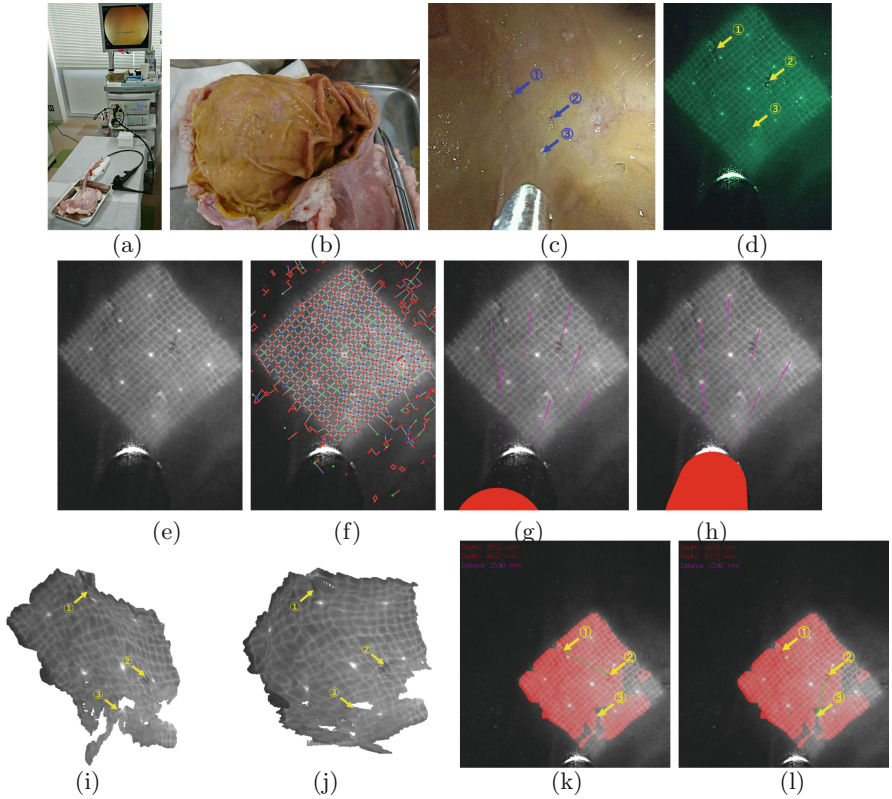


**Fig. 10.** Comparison of 3D reconstructed areas of original and tone-mapped HDR images.

were reconstructed in the result of  $I_T$  (HDR image), whereas, in the result of  $I_5$  (note that  $I_5$  was the most successfully reconstructed image from Fig. 10), the same regions were not reconstructed. In Fig. 9, we can also see that the noises in the source images are reduced in the tone-mapped HDR image, because multiple images are merged so that independent noises are suppressed.

### 6.3 3D Reconstruction Inside a Stomach of a Pig

To evaluate the system in more realistic conditions, we captured shapes inside a stomach of a pig, which is often used for evaluation purpose and a practice of an endoscopist. To evaluate the scales captured by the 3D endoscope, we first curved



**Fig. 11.** 3D reconstruction of bio-tissue inside a pig stomach with markers. (a) The environment of the experiment. (b) The pig-stomach cut open after experiment session. (c) The appearance inside the stomach with marker positions. (d) The captured image with the pattern projected. (e) The HDR enhanced image. (f) The detected grid graph. (g), (h) Before and after the auto-calibration of the projector. The rendered projector positing is the read cylinder and the epipolar lines are pink line segments. (g) is before the auto-calibration and (h) is after the auto-calibration. (i), (j) Reconstructed 3D shape rendered from two different view points. (k), (l) Distance measurements between the two markers. Red regions are reconstructed areas. (Color figure online)

**Table 1.** Estimated distances between two markers on the surface inside a pig stomach

Marker IDs	Ground truth	Ground truth (expanded)	Our result
1 and 2	24.6 mm	29.4 mm	25.9 mm
2 and 3	14.2 mm	15.1 mm	13.9 mm

several markers on the surface of pig’s stomach, then, reconstruct 3D shape of the entire surface. The distances between the two markers are estimated and compared to the ground truth, which is obtained by measuring the real distances between the markers after the measurement process; we cut and opened the stomach. Since the stomach was inflated while the endoscopy diagnosis, the ground truth distances that are actually measured were considered to be smaller than the estimated distances. To compensate such error, we also measure the ground truth distance while expanding the stomach surface manually.

Figure 11 shows the experimental situation and the results. Comparison between estimated results and ground truth are shown in Table 1. The precision was about 5.0% and 2.1% from the unexpanded ground truth. Considering the difference of measurement situation, we could conclude that the measurement was sufficiently accurate. In Fig. 11(g), (h), we also show the result of auto-calibration. In Fig. 11(g), which shows situation before auto-calibration, the rendered silhouette of the projector is different from the captured silhouette of Fig. 11(e). After auto-calibration, the projector position fits to the captured image, and the epipolar lines lie on the marker position as shown in Fig. 11(h).

## 7 Conclusion

We proposed a 3D endoscopic system based on an active stereo, where the micro laser pattern projector is inserted through an instrumental channel. Since there is a margin between the projector and the channel and the head of endoscope dynamically moves during an actual operation, the relative position of a camera and a projector is not fixed with respect to each other. For 3D reconstruction the relative position should be known, we propose an auto-calibration technique using the silhouette of the pattern projector. In addition, since the laser projector has a strong light intensity and dynamic range of the camera is not enough, we propose a new HDR image synthesis technique using a blinking modulation applied to the projector. The ability of the techniques were confirmed by intensive experiments using real endoscopic systems and demonstrated by reconstructing the 3D shape of the inside surface of a pig’s stomach. Our future work is to construct the realtime system and use it to actual diagnosis and operations.

**Acknowledgment.** This work was supported in part by JSPS KAKENHI Grant No. 15H02779, 16H02849, MIC SCOPE 171507010 and MSR CORE12.

## References

1. 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: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), pp. 7937–7941, August 2015
2. Furukawa, R., Kawasaki, H.: Uncalibrated multiple image stereo system with arbitrarily movable camera and projector for wide range scanning. In: IEEE Conference on 3DIM, pp. 302–309 (2005)
3. 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: World Automation Congress, WAC 2006, pp. 1–4, July 2006
4. Stoyanov, D., Scarzanella, M.V., Pratt, P., Yang, G.-Z.: Real-time stereo reconstruction in robotically assisted minimally invasive surgery. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6361, pp. 275–282. Springer, Heidelberg (2010). [https://doi.org/10.1007/978-3-642-15705-9\\_34](https://doi.org/10.1007/978-3-642-15705-9_34)
5. Visentini-Scarzanella, M., Stoyanov, D., Yang, G.: Metric depth recovery from monocular images using shape-from-shading and specularities. In: ICIP, Orlando, USA, pp. 25–28 (2012)
6. Schmalz, C., Forster, F., Schick, A., Angelopoulou, E.: An endoscopic 3D scanner based on structured light. *Med. Image Anal.* **16**(5), 1063–1072 (2012)
7. Lin, J., Clancy, N.T., Elson, D.S.: An endoscopic structured light system using multispectral detection. *Int. J. Comput. Assist. Radiol. Surg.* **10**(12), 1941–1950 (2015)
8. 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: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013. LNCS, vol. 8149, pp. 139–146. Springer, Heidelberg (2013). [https://doi.org/10.1007/978-3-642-40811-3\\_18](https://doi.org/10.1007/978-3-642-40811-3_18)
9. 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, pp. 16–20. Citeseer (2010)
10. 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, pp. 4991–4994 (2014)
11. Furukawa, R., Sanomura, Y., Tanaka, S., Yoshida, S., Sagawa, R., Visentini-Scarzanella, M., Kawasaki, H.: 3D endoscope system using doe projector. In: The 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2016) (2016)
12. Forsyth, D., Ponce, J.: *Computer Vision: A Modern Approach*, 2nd edn. Pearson Education Inc., London (2011)
13. Furukawa, R., Kawasaki, H.: Self-calibration of multiple laser planes for 3D scene reconstruction. In: 3DPVT, pp. 200–207 (2006)
14. Furukawa, R., Kawasaki, H.: Laser range scanner based on self-calibration techniques using coplanarities and metric constraints. *Comput. Vis. Image Underst.* **113**(11), 1118–1129 (2009)

15. Yamazaki, S., Mochimaru, M., Kanade, T.: Simultaneous self-calibration of a projector and a camera using structured light. In: 2011 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 60–67. IEEE (2011)
16. Debevec, P.E., Malik, J.: Recovering high dynamic range radiance maps from photographs. In: SIGGRAPH 2008, pp. 1–10. ACM, New York (2008)
17. Kalantari, N.K., Shechtman, E., Barnes, C., Darabi, S., Goldman, D.B., Sen, P.: Patch-based high dynamic range video. *ACM Trans. Graph.* **32**(6), 202–1 (2013)
18. Eilertsen, G., Mantiuk, R., Unger, J.: A comparative review of tone-mapping algorithms for high dynamic range video. In: *Computer Graphics Forum*, vol. 36, pp. 565–592. Wiley Online Library (2017)
19. 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)