

# Optimized depth fusion of multiple measurements by 3D endoscope based on active stereo technique\*

Ryo Furukawa<sup>1</sup>, Michihiro Mikam<sup>2</sup>, Ryusuke Sagawa<sup>3</sup>, Yuki Okamoto<sup>4</sup>, Shiro Oka<sup>4</sup>, Shinji Tanaka<sup>4</sup>, and Hiroshi Kawasaki<sup>5</sup>

<sup>1</sup> Kindai University, Japan

<sup>2</sup> Hiroshima City University, Japan

<sup>3</sup> AIST, Japan

<sup>4</sup> Hiroshima University, Japan

<sup>5</sup> Kyushu University, Japan

**Abstract.** 3D measurement of inside human body becomes important and several techniques have been proposed. Among them, ultra-small pattern projector with off-the-shelf endoscope is promising because of stability, accuracy and cost-effectiveness. One drawback of the system is that since 3D shape is reconstructed for each frame while endoscope head is moving, they are not registered. In addition, shapes are distorted independently, since the position of the projector cannot be fixed to the head. In this paper, we propose a novel bundle adjust (BA) algorithm to solve both distortion and shape registration problems for a 3D endoscope system based on pattern projection. By using the proposed method, multiple measurements that potentially have scale inconsistencies and shape distortions are registered, with simultaneous correction of inter-frame inconsistencies. The effectiveness of the technique is proved for measurements of concave shapes made of clay, phantom models, and a real stomach of a pig.

**Keywords:** 3D endoscope · structured light · registration.

## 1 Introduction

3D measurement systems for the inside of the human body using endoscope have drawn wide attention. However, it is often difficult to obtain sufficient texture information on the surface of internal organs, which results in unstable reconstruction for passive stereo algorithms including SfM and SLAM. In contrast, a 3D endoscope using active stereo algorithm, in which a pattern is projected from an ultra-small fiber projector that passes through the instrument channel of a common endoscope is proposed [7, 9]. In those systems, it is difficult to fix the relative positions between the camera and the pattern projector scale inconsistencies and shape distortions between different frames inevitably occur.

---

\* Supported by JSPS Kakenhi JP20H00611, JP18H04119, JP21H01457 and NEDO JPNP20006

To deal with such problems in passive-stereo systems, multi-frame global optimizations are often used for correcting scaling mismatches and shape distortions, and also for widening the measurement areas, *i.e.*, SfM or SLAM using only a camera. In general, such systems are often based on optimization techniques of multi-frame information called bundle adjustment (BA), where 3D feature points and the camera positions are optimized from 2D observations of the 3D points by minimizing a loss function that represents the sum of reprojection errors, *i.e.*, the errors between the 2D observations and the projections of the 3D points onto the camera images.

Since projectors are often modeled by pinhole camera, one may consider that it is possible to optimize the cameras and projectors parameters by BA for passive stereo, however, it is not true. The reasons for this are as follows. (1) In the active stereo system, textureless environment is assumed, and thus, explicit correspondences between different frames cannot be obtained, whereas the correspondence between the projector and the camera is obtained with high density only in single frame, (2) From the above fact, it can be concluded that the frame number of 2D points corresponding to a 3D point is only two, which is in contrast to the passive stereo systems.

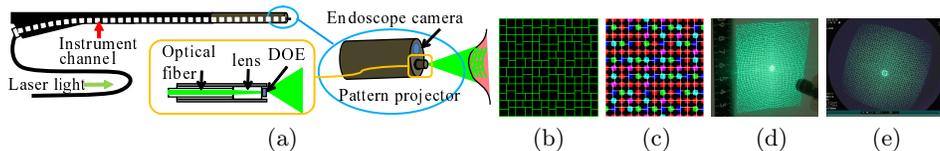
In this paper, to achieve a multi-frame global optimization for an active stereo system where the relative position between the projector and camera is not fixed, we define a new loss function and propose a method to perform the global optimization using the defined loss function by an inverse rendering technique. This system consists of the following contributions. (1) a new loss function that directly models pattern projection in stereo active system, and (2) multi-frame global optimization by minimizing the above loss function by using inverse rendering technique, are proposed.

As far as we know, this work is the first study that discusses how multi-view active stereo is different from multi-view passive stereo, and models the active stereo method using projection mapping for global optimization purposes. In the following sections, these contributions are explained and finally, we demonstrate the effectiveness of our method by experiments.

## 2 Related Works

Mathematically, pattern projectors can be represented using the same model as cameras. For estimating the projector and the camera parameters, a typical approach is projecting patterns onto a camera calibration object such as a grid-printed or white plane [13, 19, 18, 1, 5]. These methods are for pre-calibration, where the projector and the camera can be fixed, and the system can be calibrated before the measurement process.

With a sufficient number of correspondences between a projector and a camera, single-frame self-calibration is possible. For example, Furukawa *et al.* proposed a static grid pattern modulated by gaps for endoscopic shape measurement, and performed auto-calibration [9].



**Fig. 1.** (a) The system configuration. (b-e): The projected patterns.

For application for medical problems, many researchers have worked for multi-frame 3D reconstruction for endoscopy images. For example, 3D reconstruction based on passive camera, *e.g.*, SLAM or SfM, for endoscopic images have been researched in medical image analysis, such as Mahmoud *et al.* [15], Chen *et al.* [2], and Leonard *et al.* [12]. Recently, non-rigid SLAM have been proposed, such as Song *et al.* [17], Lamarca *et al.* [11], and Zhou *et al.* [20]. These methods need 3D feature points, thus needs textures. However, medical images and endoscopic images generally have less detectable image features than natural images, Thus, these methods becomes often unstable for medical purposes.

For 3D registration for medical purposes, ICP algorithm has been used [3, 16]. In this paper, our target is not only registration of multiple 3D scenes, but simultaneously correcting inter-frame inconsistencies by taking the observation model into account. For such a purpose, Furukawa *et al.* proposed a modification of bundle adjustment for passive stereo [8]. Their method does not directly model dynamics of active stereo observations and has often problems in convergence. In contract, our technique is faster and stable. Note that since active stereo techniques for laparoscopes and endoscopes have been widely studied [10, 14, 8], the multi-frame optimization method proposed in this paper can be applied to general active stereo systems and useful.

The main challenge of this paper is achieving multi-frame optimization for structured-light-based endoscopic systems. Prior work of such a problem is Furukawa *et al.* [8]. In their approach, ICP-like correspondence search and optimization is alternatively applied. Since the ICP-like correspondence search does not guarantee decrease of the optimized loss function, their approach often does not optimize robustly. In this paper, we newly introduce a differentiable, loss function that can be globally minimized.

### 3 System configuration and single-frame correspondence estimation

In this paper, an active stereo system consisting of a projector and a camera is constructed by inserting a fiber-shaped, micro pattern projector into the instrument channel of a standard endoscope as shown in Fig.1(a). It is based on a similar method proposed by Furukawa *et al.* [6]. We used a Fujifilm EG-590WR endoscope and a pattern projector with a diffractive optical element (DOE) to generate structured-light illumination (Fig.1(b)). We used a grid pattern, where all vertical and horizontal edges have small gaps, representing code symbols  $S$ ,  $L$

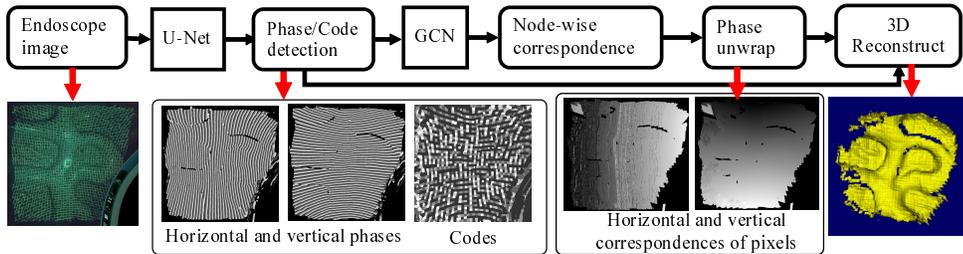


Fig. 2. Overview of the reconstruction process

and  $R$  [9] (Fig.1(c)), Actual pattern projected onto the object surface are shown in Fig.1(d) and (e).

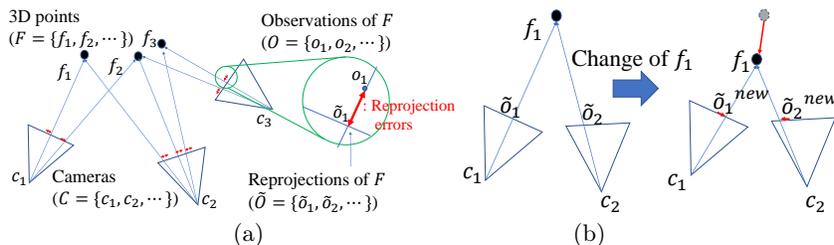
The flow of the 3D reconstruction algorithm is shown in Fig.2. In this process, the grid pattern is projected onto the target surface and captured by the endoscopic camera. Then, pixel-wise phase information as well as pattern code for each node are detected by using U-Nets. The grid structure and codes are represented by a graph and graph convolutional network (GCN) is applied to predict node-wise correspondences to unwrap the phase. Finally, pixel-wise 3D reconstruction are achieved by light sectioning method using the unwrapped phase. For more details of the phase and pattern code detection of UNets, please refer to Furukawa *et al.* [6].

Note that the multi-frame optimization described in the following sections does not depend on the above pattern and decoding algorithm. As long as dense 2D correspondence mapping between the projector and the camera can be obtained, any patterns and algorithms can be used.

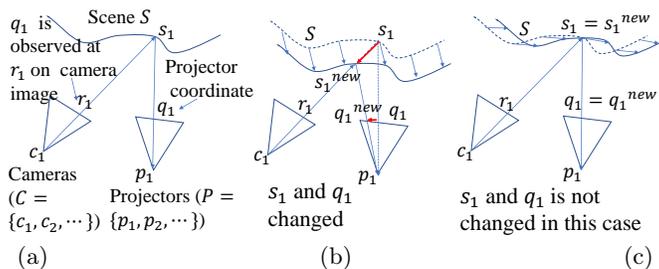
## 4 Observation model of active stereo

In the case of **passive stereo**, multi-frame optimization can be done by bundle adjustment (BA) with reprojection error minimization, as shown in Fig.3. The input for the algorithm is a set of 2D positions  $O = \{o_1, o_2, \dots\}$  where 3D points are observed by the camera. The parameters to be optimized in this algorithm are the set of 3D feature points  $F = \{f_1, f_2, \dots\}$ , the camera intrinsic and extrinsic parameters  $C = \{c_1, c_2, \dots\}$ . The loss function  $L(F, C)$  for optimization is a sum of reprojection errors of the 3D points, which are the difference between the observed positions  $O$ , and the 'predicted' positions  $\tilde{O} = \{\tilde{o}_1, \tilde{o}_2, \dots\}$  calculated from  $F$  and  $C$ , as shown in Fig.3(a). In BA, a 3D point, for example,  $f_1$ , is often observed in multiple frames. The fact that a 3D point is observed from multiple cameras is a strong constraint on the positional relationship between those cameras, and greatly contributes to the stability of the BA algorithm. Note that BA is a direct modeling of a set of actually existing 3D feature points  $F$  and their observations  $O$ .

In case of **active stereo**, the observed values are represented as a mapping from 2D points on the camera image ( $r_1$  of Fig.4(a)) to 2D points on the pro-



**Fig. 3.** Observation model for **passive stereo** (normal bundle bundle adjustment): (a) reprojection errors, (b) changes of feature points.



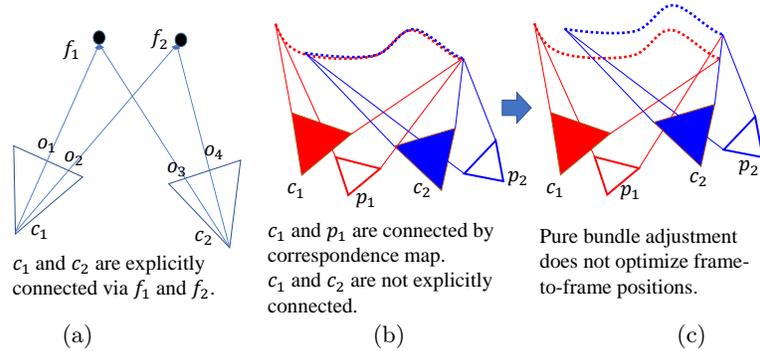
**Fig. 4.** Observation model of active stereo systems: (a) observation model, (b), (c) changes of position of scene  $S$

jected pattern ( $q_1$ ). The static geometric relationship of this case 'seems' similar to passive stereo, but the physical phenomenon is different. In active stereo, the ray from  $q_1$  hits the surface of the 3D scene  $S$  at  $s_1$  and  $S_1$  is observed by the camera at  $r_1$ . as shown in Fig.4(a). Let the parameters of the projectors be  $P = \{p_1, p_2, \dots\}$ . The observed values in active stereo are the intersection of the ray from  $P$  and  $S$  being observed by  $C$ .

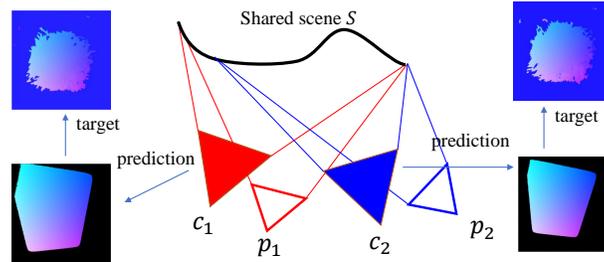
Note that  $S$  is not a set of sparse 3D points, but the continuous 3D surfaces in the scene. In passive stereo model, changes of 3D feature points  $F$  can be defined independently of cameras  $C$ , as shown in Fig.3(b). In active stereo model, the motion of scene  $S$  affects the observation, but the real effect is much different from the passive-stereo case. For example, if scene  $S$  moves as Fig.3(b), the observation at the ray  $r_1$  moves as in the figure, whereas, if  $S$  moves as Fig.3(c), observation at  $r_1$  may be unchanged. An extreme case is a plane-shaped  $S$  moving in parallel to the plane, where observations of active stereo is totally unchanged.

These interactions between  $S, C$  and  $P$  are largely different from the case of passive stereo, This makes a significant impact on optimization algorithms that uses the gradient of loss function with respect to the the camera or projector parameters.

In addition to the difference in the observation model, there is another difference in inter-frame constraints. In passive stereo, a point  $F_1$  of sparse 3D feature points  $F$  is observed across multiple frames. In contrast, in active stereo, there are no such distinctive feature points observed in multiple frames. Thus, a pure bundle adjustment model does not constraint positional relationships between



**Fig. 5.** Inter-frame constraints: (a) passive-stereo cases, where inter-frame correspondences exists, (b) active-stereo cases, where there are no explicit inter-frame correspondences, (c) pure bundle adjustment does not constrain inter-frame alignment for active-stereo cases.



**Fig. 6.** Inter-frame optimization via scene  $S$ : the prediction is projector coordinates (Red: $x$ , Green: $y$ ) predicted for  $C$ ,  $P$ , and  $S$

different frames (Fig.5). This is a negative aspect of the active stereo where it works under textureless scenes.

## 5 Multi-frame optimization for active stereo

In this section, we define the observation error model for active stereo. As shown in Fig.4, the 2D coordinates of the camera and the projector are related via the scene surface  $S$ . For virtual, CG scene, the same information can be rendered as a projection mapping.

Note that it can be easily rendered by using a *pixel shader* of GPU, where the *pixel shader* receives a 3D point on surface  $S$ , and projects it onto a projector. This image represents a mapping from the camera 2D coordinates to the 2D coordinates of the projected pattern, and is a function of  $S$ ,  $C$ ,  $P$ .

Let the images obtained by the projection mapping be denoted as  $R(S, C, P)$ . In active stereo,  $R(S, C, P)$  is the same as the 2D correspondence map from the camera to the projector, which is obtained by analyzing a pattern-projected camera image. Thus, for optimized geometrical solution,  $R(S, C, P)$  should mimic the 2D correspondence map.

By using  $R(S, C, P)$  for optimization, multi-frame optimization of active stereo system can be achieved as shown in Fig.6. Note that the rendered and optimized images are not projected images themselves, but the projector coordinates, which has much better properties for optimization (please refer to a supplemental document).

The procedure is as follows.

- Step 1:** Analyze the pattern projected image for each frame obtain a dense mapping from camera to projector. Let  $M_i$  be an image of this mapping, where  $i$  is the frame number.
- Step2:** For each frame  $M_i$ , perform self-calibration to estimate the position between the projector and the camera, and calculate the 3D shape  $\sigma_i$ .
- Step3:** Roughly align  $\sigma_i$  ( $i = 1, 2, \dots, N$ ) and retrieve the initial camera and projector positions  $C_0$  and  $P_0$ , respectively. This can be done manually or by using ICP algorithm.
- Step 4:** By using the initial values of  $C_0$  and  $P_0$ , calculate a rough mesh shape  $C_0$  close to  $\sigma_i$  ( $i = 1, 2, \dots, N$ ).
- Step 5:** By using camera position  $C$  (initial value  $C_0$ ), projector position  $P$  (initial value  $P_0$ ), and scene shape  $S$  (initial value  $S_0$ ), calculate  $R(S, C, P)$ . Let the  $i$ -th frame function be  $R_i(S, C, P)$ , set  $L(S, C, p) = \sum_i \log(|R_i(S, C, P) - M_i|^2 + 1)$ . as the loss function ( $\log(x+1)$  is Cauchy loss for robustifying the least square loss), and optimize  $S, C$ , and  $P$  as follows:  $\tilde{S}, \tilde{C}, \tilde{P} = \underset{S, C, P}{\operatorname{argmin}} L(S, C, P)$

Note that the dynamic properties shown in Fig.4 is directly modeled in  $R(S, C, P)$ . Thus, the optimization of  $L(S, C, p)$  proceeds effectively using gradient-based optimization such as Adam. The inter-frame positions are also optimized with this method, because all the views share a single scene model  $S$ .

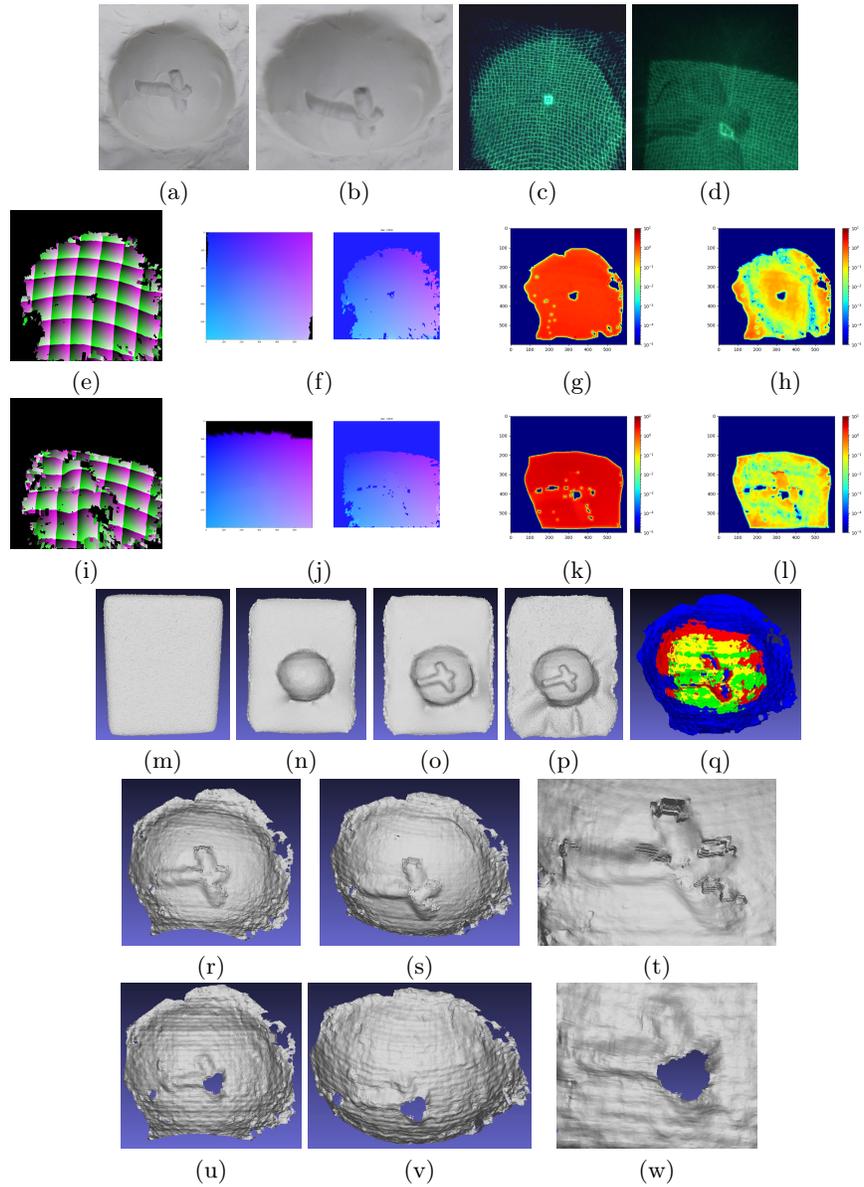
## 6 Experiments

We have implemented a global optimization system using inverse rendering in Pytorch3D (ver.0.6.1). In the rendering system of Pytorch3D, the positions of the projectors are sent to the pixel shaders, and 2D coordinates of the projected patterns are drawn as color values.

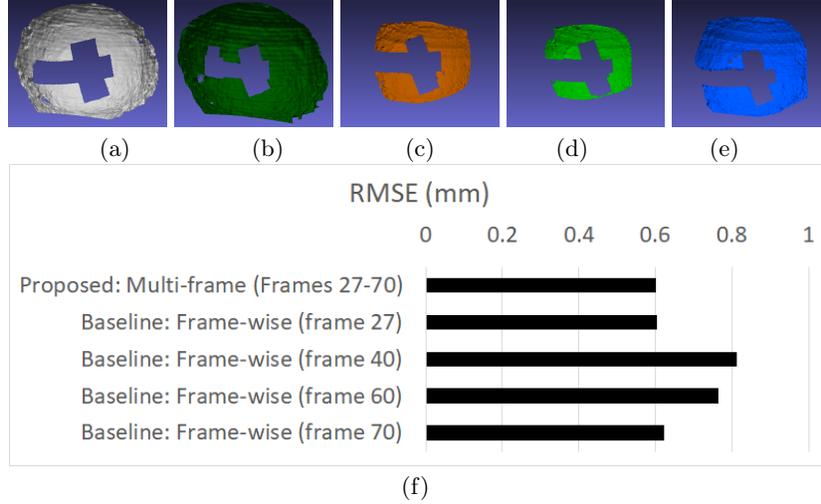
After images  $R_i(S, C, P)$  of all the viewpoints are rendered, the loss function is calculated from  $M_i$  and  $R_i(S, C, P)$ . The gradients of the loss function with respect to  $S, C$  and  $P$  are calculated by back propagation, and the loss is minimized by Adam algorithm.

### 6.1 Multiple shape integration using test shapes

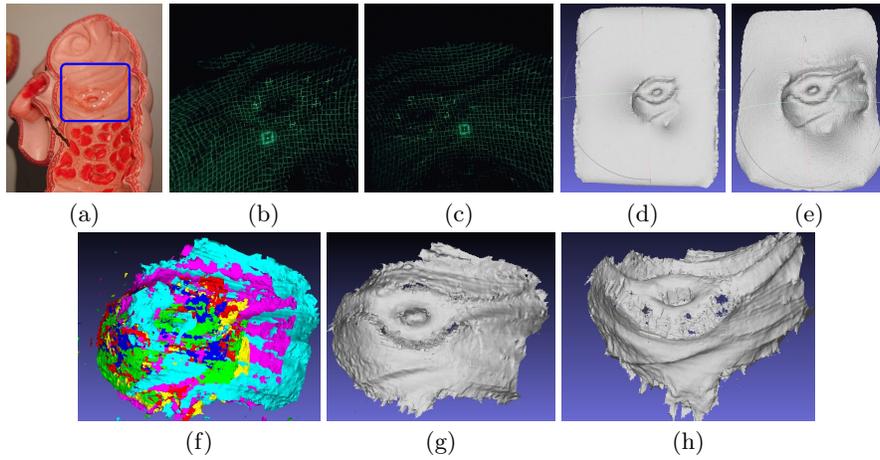
To show effectiveness of the proposed method, we measured three test shapes. Since the endoscope and the projector have wide field of view, the system is appropriate for measuring concave shapes. Thus, we provided concave test objects, which are a concave sphere made of clay, a colon phantom model, and a stomach



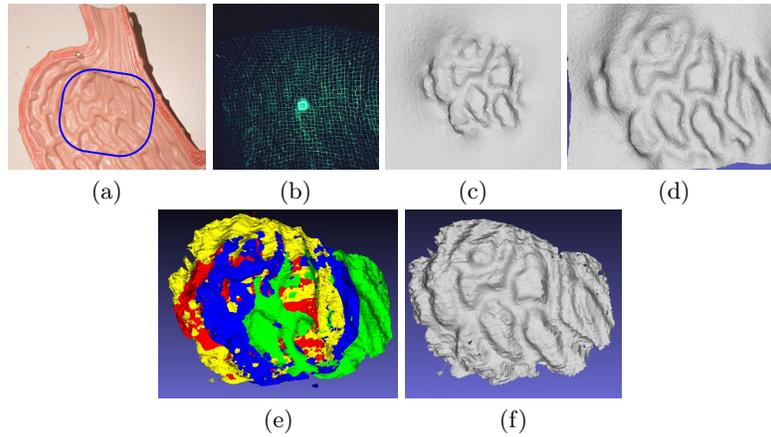
**Fig. 7.** Measuring a concave sphere with shape features with 4 frames. (a)(b) Appearance of the concave spherical sample. (c)(d) Captured images (frame 1 and frame 3) with pattern projection. (e)(i) Projector coordinates estimated from images (c) and (d). (f)(j) Samples of rendered image and target values, where pixel colors represent projector coordinates. (g)(k) Initial residuals. (h)(l) Residuals after 500 iterations. (m),(n),(o),(p) Shape  $S$  after 0, 10, 100, 600 iterations. (q) Estimated aligned shapes colored by frames (the number frames is four). (r)(s)(t) Final results merged with volumetric method [4] using optimized parameters w.r.t. 4 frames. (r)(s)(t) Results merged with with volumetric method [4] using a single frame.



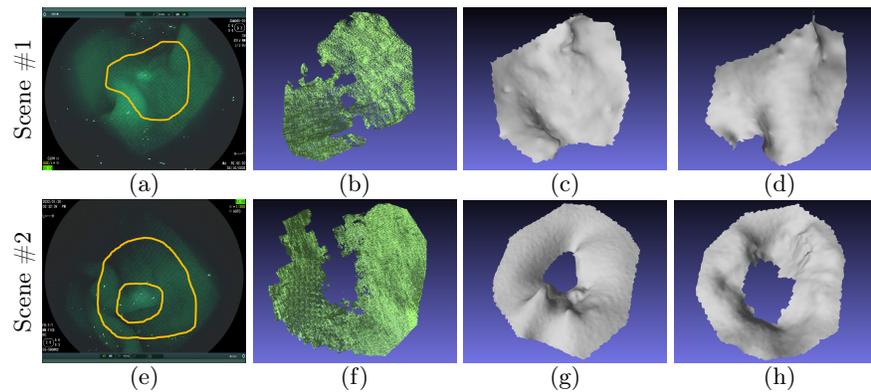
**Fig. 8.** Quantitative evaluation measurement of a concave sphere. (a-e) Regions fit to spheres for (a) the multi-frame optimized shape (proposed) and (b-e) frame-wise calibrated shapes for frames 27, 40, 60, 70, respectively (baselines). (f) RMSEs from the ground-truth shapes (spheres).



**Fig. 9.** Measuring a colon phantom with 6 frames . (a) Appearance and captured region (blue rectangle). (b)(c) Captured images. (d) Shape  $S$  estimated from one frame. (e)  $S$  estimated from 6 frames. (f) Estimated shapes colored by frames. (g)(h) Final results merged with volumetric method [4] using 6 frames.



**Fig. 10.** Measuring a stomach phantom with 4 frames . (a) Appearance and captured region (blue rectangle). (b) Captured image. (c) Shape  $S$  estimated from one frame. (d)  $S$  estimated from 4 frames. (e) Estimated shapes colored by frames. (f) Final results merged with volumetric method [4] using 4 frames.



**Fig. 11.** Real data: pig's stomach. (a)(e) reconstructed areas, (b)(f) reconstructed points using a single frame, (c)(g) initial shape before the optimization, and (d)(h) integrated shape after the optimization.

phantom model. The test shapes were continuously measured by our 3D endoscope system, and the shapes and projector-camera parameters were optimized. The results are shown in Fig.7, Fig.9 and Fig.10.

**Qualitative evaluation with a concave sphere** The measurement of a concave sphere made of clay is shown in Fig.7, where (a) and (b) shows the appearances, (c) and (d) shows the captured images. Four frames were used for this sample. Image (d) is captured with close-up view than image (c). Images (e) and (i) show the projector coordinates obtained by analyzing images (c) and (d). These  $xy$ -maps are used as target images for rendering, which are the right images of (f) and (j). The left images of (f) and (j) are rendered coordinates, and the differences between the right and left images of (f) and (j) are minimized.

The residuals between them are shown in (g),(h),(k) and (l), where (g) and (k) are the initial residual, and (h) and (l) are the residual after optimization. Note that the residuals are log scale, thus, values of (h) and (l) are smaller than (g) and (k) by orders of  $10^2$  to  $10^3$ . Images (m)-(p) show the optimization process of shape  $S$ . Fig.7(q) shows the shapes colored by frames after optimization. Images (r),(s) and (t) are a merged shape using all the frames. Images (u), (v) and (w) are a merged shape using only a single-frame optimization. By comparing (r)-(t) with (u)-(w), we can see that the multi-shape merging effectively improved the quality of the shape.

**Quantitative evaluation with a concave sphere** As a quantitative evaluation of the effectiveness of the proposed method, we compared the multi-frame-optimized shape (*i.e.*, the proposed method) shown in Fig.7, and shapes reconstructed by frame-wise auto-calibration (*i.e.*, the baseline method that does not use multi-frame information) as shown in Fig.8.

To evaluate the errors of the reconstructed shapes, we first removed the outlier regions that are distant from the sphere shapes (Fig.8(a-e)). Then, we fit the shapes to sphere-shaped point clouds (the ground-truth shapes). To deal with scale ambiguity, we generated ground-truth point clouds with multiple radii around 20mm interleaved by 0.05mm. Then, we fit each of the evaluated shapes to ground-truth shapes for each radius using an ICP method. Of all the radii, fit with the smallest RMSE was adopted to calculate the real error.

The RMSE of the proposed approach was shown to be the most accurate. Generally, the larger the measurement area, the lower the RMSE tended to be (Fig.8(a-e) for the area differences). The reason may be that auto-calibration became more accurate if the correspondences of wider regions were used as inputs. The frame 27 of Fig.8(b) was the same frame as Fig.7(u-w). Note that, although the global shape accuracy of shape of frame 27 was similar to the multi-frame optimized shape, the small details or shape resolution of the frame-27 result were much worse than the result of the proposed method, as shown in Fig.7(r-t) and (u-w).

**Evaluations with phantom models** The measurement of a colon phantom model is shown in Fig.9, where 6 frames were used. Image (a) is the appearance and the scanned region. Images (b) and (c) are captured images. Images (d) is the optimized mesh  $S$  with a single frame, and (e) is  $S$  optimized with 6 frames. Image (f) shows shapes with frame-wise colors, and (g) and (h) show the merged shape reconstructed with [4].

The measurement of a stomach phantom model is shown in Fig.10, where 4 frames were used. Image (c) is the optimized mesh  $S$  with a single frame, and (d) is  $S$  optimized with 4 frames. Image (f) shows the merged shape reconstructed with 4 frames.

## 6.2 Shape integration of a real stomach of a pig

Next, the inside of the real stomach of pig was continuously measured using the 3D endoscope system and optimized. The results are shown in Fig.11. Six frames were integrated for each scene. Comparing Fig.11(b) and (d) of scene #1, it is confirmed that the generated holes in (b) are correctly restored in (c) and (d). In addition, in the results of scene #2, it is confirmed that high-frequency shapes are restored by our method.

## 7 Conclusion

In the paper, a novel bundle adjust (BA) algorithm for endoscope with pattern projector is proposed. By analyzing geometrical differences between a feature point of a scene and a feature point created by pattern projector on the scene, optimal cost function for BA is derived, which can achieve fast and robust convergence on finding relative position and orientation between a camera and a projector. By using the estimated parameters, all the 3D shapes reconstructed from sequential images of endoscope camera are readily integrated to make a large 3D scene. The effectiveness of the technique is confirmed by using phantom of human stomach and a real stomach of pig.

The future work of this paper is creating 3D multi-frame reconstruction in interactive time. Another direction of the future work is to use the 3D estimation for annotating depth maps for endoscopy image sequences, since lacking 3D-annotated endoscopy image database has been a large difficulty for supervised training of depth-estimation deep learning models.

## References

1. Audet, S., Okutomi, M.: A user-friendly method to geometrically calibrate projector-camera systems. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. pp. 47–54. IEEE (2009)
2. Chen, L., Tang, W., John, N.W., Wan, T.R., Zhang, J.J.: Slam-based dense surface reconstruction in monocular minimally invasive surgery and its application to augmented reality. *Computer methods and programs in biomedicine* **158**, 135–146 (2018)
3. Combès, B., Prima, S.: An efficient em-icp algorithm for symmetric consistent non-linear registration of point sets. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 594–601. Springer (2010)
4. Curless, B., Levoy, M.: A volumetric method for building complex models from range images. In: Proceedings of the 23rd annual conference on Computer graphics and interactive techniques. pp. 303–312 (1996)
5. Draréni, J., Roy, S., Sturm, P.: Geometric video projector auto-calibration. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. pp. 39–46. IEEE (2009)
6. Furukawa, R., Mikamo, M., Sagawa, R., Kawasaki, H.: Single-shot dense active stereo with pixel-wise phase estimation based on grid-structure using cnn and correspondence estimation using gen. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision (WACV). pp. 4001–4011 (January 2022)
7. Furukawa, R., Mizomori, M., Hiura, S., Oka, S., Tanaka, S., Kawasaki, H.: Wide-area shape reconstruction by 3d endoscopic system based on cnn decoding, shape registration and fusion. In: OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis, pp. 139–150. Springer (2018)
8. Furukawa, R., Nagamatsu, G., Oka, S., Kotachi, T., Okamoto, Y., Tanaka, S., Kawasaki, H.: Simultaneous shape and camera-projector parameter estimation for 3d endoscopic system using cnn-based grid-oneshot scan. *Healthcare Technology Letters* **6**(6), 249–254 (2019)
9. Furukawa, R., Oka, S., Kotachi, T., Okamoto, Y., Tanaka, S., Sagawa, R., Kawasaki, H.: Fully auto-calibrated active-stereo-based 3d endoscopic system using correspondence estimation with graph convolutional network. In: 2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC). pp. 4357–4360. IEEE (2020)
10. Geurten, J., Xia, W., Jayarathne, U., Peters, T.M., Chen, E.C.: Endoscopic laser surface scanner for minimally invasive abdominal surgeries. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 143–150. Springer (2018)
11. Lamarca, J., Parashar, S., Bartoli, A., Montiel, J.: Defslam: Tracking and mapping of deforming scenes from monocular sequences. *IEEE Transactions on robotics* **37**(1), 291–303 (2020)
12. Leonard, S., Sinha, A., Reiter, A., Ishii, M., Gallia, G.L., Taylor, R.H., Hager, G.D.: Evaluation and stability analysis of video-based navigation system for functional endoscopic sinus surgery on in vivo clinical data. *IEEE transactions on medical imaging* **37**(10), 2185–2195 (2018)
13. Liao, J., Cai, L.: A calibration method for uncoupling projector and camera of a structured light system. In: 2008 IEEE/ASME International Conference on Advanced Intelligent Mechatronics. pp. 770–774. IEEE (2008)

14. Lin, J., Clancy, N.T., Stoyanov, D., Elson, D.S.: Tissue surface reconstruction aided by local normal information using a self-calibrated endoscopic structured light system. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 405–412. Springer (2015)
15. Mahmoud, N., Collins, T., Hostettler, A., Soler, L., Doignon, C., Montiel, J.M.M.: Live tracking and dense reconstruction for handheld monocular endoscopy. *IEEE transactions on medical imaging* **38**(1), 79–89 (2018)
16. Sinko, M., Kamencay, P., Hudec, R., Benco, M.: 3d registration of the point cloud data using icp algorithm in medical image analysis. In: 2018 ELEKTRO. pp. 1–6. IEEE (2018)
17. Song, J., Wang, J., Zhao, L., Huang, S., Dissanayake, G.: Mis-slam: Real-time large-scale dense deformable slam system in minimal invasive surgery based on heterogeneous computing. *IEEE Robotics and Automation Letters* **3**(4), 4068–4075 (2018)
18. Wei, G., Liang, W., Zhan-Yi, H.: Flexible calibration of a portable structured light system through surface plane. *Acta Automatica Sinica* **34**(11), 1358–1362 (2008)
19. Yamauchi, K., Saito, H., Sato, Y.: Calibration of a structured light system by observing planar object from unknown viewpoints. In: Proceedings of the International Conference on Pattern Recognition (ICPR). pp. 1–4. IEEE (2008)
20. Zhou, H., Jayender, J.: Emdq-slam: Real-time high-resolution reconstruction of soft tissue surface from stereo laparoscopy videos. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 331–340. Springer (2021)