2-DOF auto-calibration for a 3D endoscope system based on active stereo

Ryo Furukawa[1], Ryunosuke Masutani[1], Daisuke Miyazaki[1], Masashi Baba[1], Shinsaku Hiura[1], Marco Visentini-Scarzanella[2], Hiroki Morinaga[2], Hiroshi Kawasaki[2], and Ryusuke Sagawa[3]

***Abstract*— For endoscopic medical treatment, measuring the size and shape of lesions, such as tumors, is important. We are developing a 3D endoscope system to measure the shape and size of living tissues based on active stereo. In previous works, our group attached a pattern projector outside the endoscope head. Since this increased the diameter of the endoscope, the burden and the risks of the patients would increase. In this paper, we set the pattern projector inside the instrument channel of the endoscope instead of mounting it outside, so that it can be deployed whenever required. This does not increase the size of the endoscope and facilitates the measuring process. However, since the projector is not physically fixed to the endoscope anymore prior to the operation, we propose an "auto-calibration" technique where extrinsic parameters are calibrated intra-operatively from a point marker on the projector observed simultaneously on the target surface. In the experiment, we show that the external parameters were successfully calibrated to obtain 3D reconstructions properly with the overall systems. The accuracy of the auto-calibration was validated by confirming that the epipolar constraints were kept, and a 3D reconstruction of a human tissue was demonstrated.**

I. INTRODUCTION

With recent developments in endoscopy technology, endoscopic diagnosis and treatment on digestive tracts has become increasingly widespread. However, specifically in the context of early-stage gastric tumors, treatments differ based on the size of tumors. This is currently evaluated either by manipulation with forceps or by visual assessment alone, which is time consuming and prone to human errors. For this reason, methods for objective, accurate tumor size estimation are important for endoscopic diagnosis systems.

Our ongoing work is the development of a 3D endoscope system to automatically measure the shape and size of living tissue based on active stereo. In our previous works[1], [2], the projector was attached externally to the endoscope. This has two main disadvantages: first, the size of the endoscope was increased because of the attached projector, which increased safety concerns and reduced maneuverability for endoscopists to stay within safe margins. The other is that the projector needs to be physically attached to the endoscope before the endoscope is inserted into the patient's esophagus.

Conversely, in our proposed system the projector is inserted through the instrument channel of the endoscope. Thus, the thickness of the endoscope remains the same while the projector can now be deployed like conventional tools through the instrument channel when appropriate.

However, since the projector now is not physically fixed to the endoscope anymore, it is not possible to calibrate its extrinsic parameters prior to the operation. Therefore, in this paper we propose an "auto-calibration" technique where the projector's extrinsic parameters are calibrated intra-operatively from a point marker on the projector observed on the tissue surface.

II. RELATED WORK

In [1], we estimated the parameters of the pattern projector by a plane-based calibration which used the surface of an LCD monitor. However, this procedure required a large number of images to be captured and did not estimate the focal length of the projector. In [2], spheres were used as pre-defined calibration objects with known dimensions, thus requiring only a few images for calibration and simplifying the process.

For 3D reconstruction method using endoscopes, techniques using Shape from Shading (SFS) method [3], [4], [5], [6] or using binocular stereo method [7], [8] have been proposed. However, these techniques often have stringent assumptions on the images that can be processed, or, in the case of binocular stereo, require specialized endoscopes.

As an example of active stereo applications in endoscopy, in [9] a single-line laser scanner attached to the head of the scope was used to measure tissue shape, however the scope head but the head of the endoscope needed to be actuated in a direction parallel to the target, which limited the practical applicability of the technique.

III. SYSTEM CONFIGURATION

A projector-camera system is constructed by attaching a micro pattern projector on a standard endoscope as shown in Fig. 1. For our system, we used the FujiFilm VP-4450HD system coupled with the EG-590WR scope. The pattern projector is inserted through the instrument channel of the endoscope and the lens that emits the structured light slightly protrudes from the endoscope head. The light source of the projector is a green laser module with a wavelength of 532nm. The laser light is transmitted through a plastic optical fiber (POF), and to the head of the pattern projector. A micro pattern chip on which the pattern image is printed is set at the tip of POF. The light that passes through the micro pattern chip also passes through the non-spherical lens placed in front of the chip, is then projected onto the target tissue.

[1]Faculty of information sciences, Hiroshima City Univ., Hiroshima, Japan
[2]Faculty of engineering, Kagoshima Univ. Kagoshima, Japan
[3]Natinal institute of advanced industrial science and technology, Tsukuba, Japan

978-1-4244-9270-1/15/$31.00 ©2015 IEEE

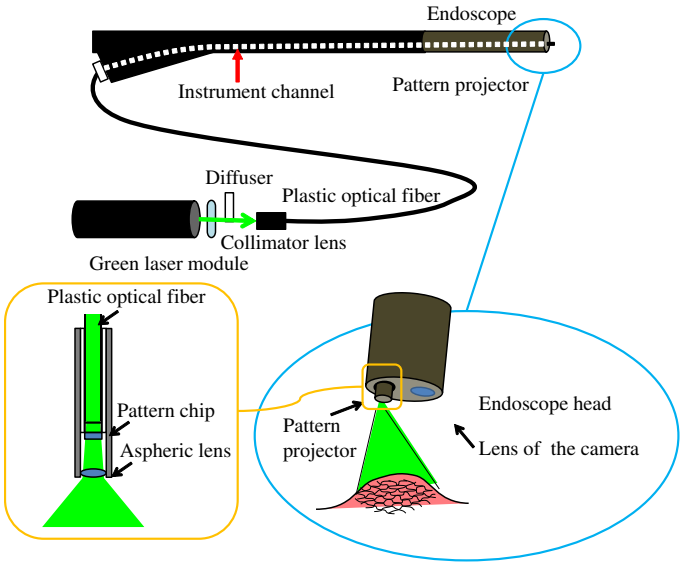


Fig. 1. The configuration of the system.

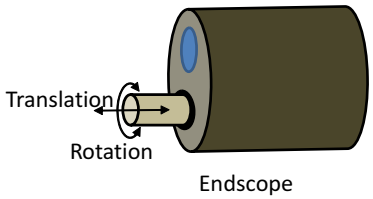


Fig. 2. The 2-DOF motion of the pattern projector.

The pattern is a grid of waved lines proposed by Sagawa *et al.* [10].

By observing the pattern projected onto the surface, the 3D shape can be reconstructed by triangulation. In this system, the pattern projector is not fixed to the endoscope, so that the pattern projector can be inserted freely during the operation.

In this configuration, the projector moves along the channel and rotates around the channel axis not only during its insertion, but also whenever the scope is bent during operation (Fig.2). Thus, the relative position between the projector and the endoscope camera is time-variant and cannot be estimated beforehand. To make this system practical for clinical use, auto-calibration, *i.e.*, the estimation of the extrinsic projector parameters without the need for custom calibration objects, is imperative.

One technique for auto-calibrating projector-camera systems consists in attaching markers to the projector. By observing the markers, the position of the projector can be estimated. By modeling the problem as the estimation of a translation along the instrument channel and rotation around its axis, the DOFs of the auto-calibration problem is 2 (one for the 1D translation and one for the 1D rotation). Thanks to this parametrization, a simple single-point marker is required, rather than the more involved solutions that would be required by the 6-DOF estimation of the full extrinsic matrix.

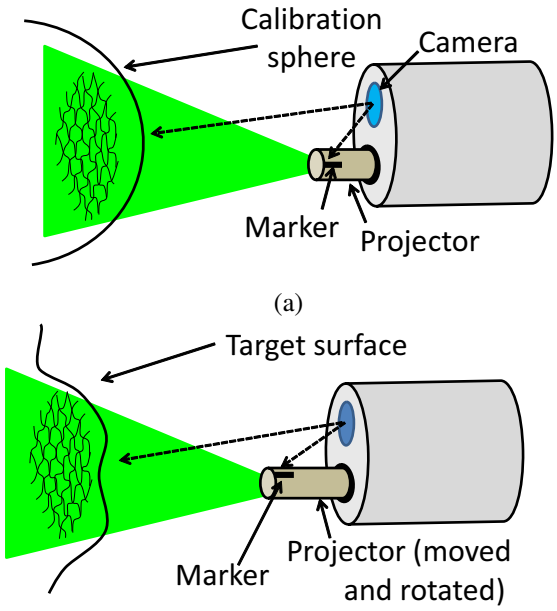


Fig. 3. The auto calibration approach: (a) base calibration using a sphere-shaped object, and (b) auto-calibration using marker observation.

IV. AUTO-CALIBRATION OF THE PROJECTOR-CAMERA SYSTEM

Our pattern projector projects a known static pattern into 3D space. Geometrical relationships between the projected points in 3D space and the original 2D points of the pattern image are related by the pinhole camera model. Therefore, parallels can be drawn between projector and traditional camera calibration procedures. In previous work [2], we proposed to obtain the intrinsic and extrinsic projection parameters using spherical objects, while in our proposed system, auto-calibration is performed using a point marker on the pattern projector.

Since our proposed auto-calibration method only estimates the 2 degrees of freedom relative to the displacement along the instrument channel and the rotation around its axis, we first pre-calibrate the extrinsic base parameters by using our previous calibration method[2] that uses a sphere as a calibration object (Fig. 3(a)). After this pre-operative calibration, the projector can move and rotate freely along the instrument channel. By decoupling the calibration procedure into a pre-operative and intra-operative phase, we restrict the intra-operative calibration ambiguity within a 2-DOF space that can be resolved by observation of a point marker.

Let us consider the projector's head, which is shaped like a cylinder whose axis coincides with the axis of the instrument channel. As the shape of this cylinder remains invariant with respect to the 2-DOF motion, we can estimate the 3D position of the marker by ray-tracing its detected position back from the camera (Fig. 3(b)) by considering the following discussion.

Let the rigid transformation from the camera coordinates

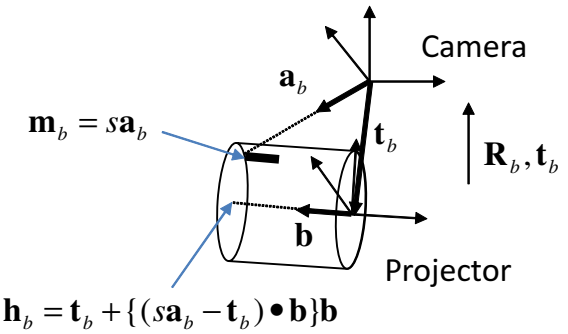


Fig. 4. The coordinates and variables for estimating the position of the marker in the projector.

to the projector coordinates in its pre-operative position be represented by a rotation matrix $\mathbf{R}_b$ and translation vector $\mathbf{t}_b$. $\mathbf{R}_b$ and $\mathbf{t}_b$ are calibrated pre-operatively.

Let the axis direction of the instrument channel be represented as $\mathbf{b}$. In this paper, $\mathbf{b}$ is assumed to coincide with the negative of the z axis of the projector coordinates, thus, $\mathbf{b} = \mathbf{R}_b\ (0\ 0\ -1)^t$. Let the radius of the cylindric shape of the projector be r, and we assume that optical center of the base projector position is on the axis of the instrument channel, which is equal to the origin of the base projector coordinates and can be calculated by $\mathbf{t}_b$. Then, the axis of the instrument channel is $\mathbf{t}_b + u\mathbf{b}$. Fig. 4 summarizes these relationships.

During the pre-operative calibration, we also observe the maker with the endoscope camera. From the observed marker in the camera image, we can calculate the direction of the marker from the camera. Let this direction be $\mathbf{a}_b$, and the distance from the camera to the maker be s. Then, the 3-D position $\mathbf{m}$ of the marker is

$$\mathbf{m}_b = s\mathbf{a}_b, \tag{1}$$

where s is an unknown variable.

Let the foot of the perpendicular to the channel axis passing through $\mathbf{m}_b$ be $\mathbf{h}_b = \mathbf{t}_b + u\mathbf{b}$. Since vector $\mathbf{m}_b - \mathbf{h}_b$ is perpendicular to $\mathbf{b}_b$, we get

$$\mathbf{h}_b = \mathbf{t}_b + \{(s\mathbf{a}_b - \mathbf{t}_b) \cdot \mathbf{b}_b\}\mathbf{b}_b. \tag{2}$$

Since the norm of vector $\mathbf{m}_b - \mathbf{h}_b$ is equal to r, we obtain

$$||\mathbf{m}_b - \mathbf{h}_b||^2 = r^2. \tag{3}$$

By expanding this equation with respect to the unknown variable s, we obtain

$$\begin{aligned} &As^2 + Bs + C = 0 \qquad (4)\\ A =&(\mathbf{a}_b \cdot \mathbf{a}_b) + (\mathbf{a}_b \cdot \mathbf{b})^2(\mathbf{b} \cdot \mathbf{b}) - 2(\mathbf{a}_b \cdot \mathbf{b})^2\\ B =& -2(\mathbf{a}_b \cdot \mathbf{b})(\mathbf{b} \cdot \mathbf{b})(\mathbf{t}_b \cdot \mathbf{b}) - 2(\mathbf{a}_b \cdot \mathbf{t}_b)\\ &+ 4(\mathbf{a}_b \cdot \mathbf{b})(\mathbf{t}_b \cdot \mathbf{b})\\ C =& -2(\mathbf{t}_b \cdot \mathbf{b}) + (\mathbf{t}_b \cdot \mathbf{t}_b) + (\mathbf{t}_b \cdot \mathbf{b})^2(\mathbf{b} \cdot \mathbf{b}) - r^2 \end{aligned}$$

Since A,B and C are known constants, s can be calculated in closed form. The two solutions correspond to the two intersections of the line of sight $s\mathbf{a}_b$ with $s > 0$, and the cylinder representing the pattern projector. Since the observed marker is visible, $\mathbf{m}_b$ corresponds the smaller solution of s and can therefore be disambiguated successfully.

After the pre-operative calibration is completed, the projector can be moved along the channel introducing a 2-DOF ambiguity. The ambiguity can be solved from the observed marker by considering the following procedure. Let the vectors corresponding to $\mathbf{a}_b$, $\mathbf{m}_b$, and $\mathbf{h}_b$ in the intra-operative marker position be $\mathbf{a}_a$, $\mathbf{m}_a$, and $\mathbf{h}_a$. Then, the rotation from the pre-operative to the intra-operative position is the signed angle between $\mathbf{m}_b - \mathbf{h}_b$ and $\mathbf{m}_a - \mathbf{h}_a$. Let this angle be $\theta_{a,b}$, and let the translation from pre-operative to intra-operative position be $\mathbf{h}_a - \mathbf{h}_b$.

Then, the auto-calibrated parameters of the pre-operative to intra-operative rigid transformation are:

$$\mathbf{t}_a = \mathbf{t}_b + (\mathbf{h}_a - \mathbf{h}_b) \tag{5}$$

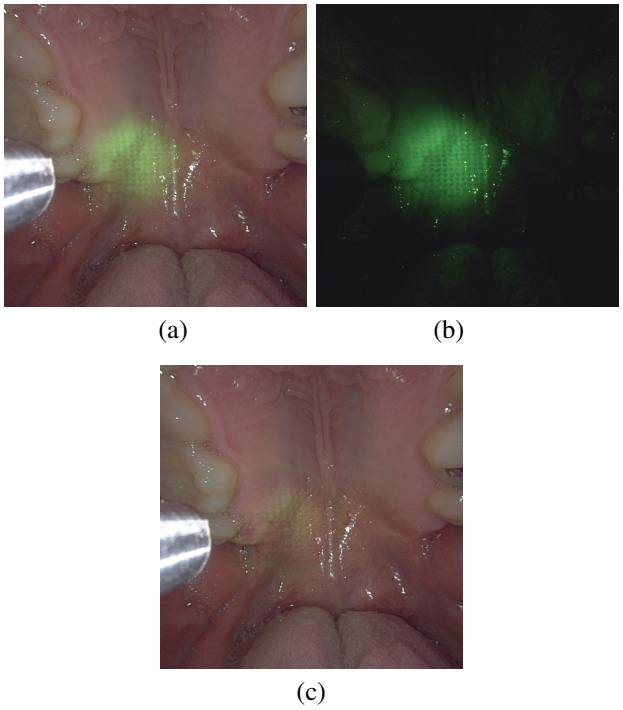
$$\mathbf{R}_a = \mathbf{R}_b\ \mathbf{R}_z(\theta_{a,b}) \tag{6}$$

where $\mathbf{R}_z(\theta_{a,b})$ is the rotation around the z-axis.

V. ACQUISITION OF COLOR TEXTURE AND RADIOMETRIC CALIBRATION

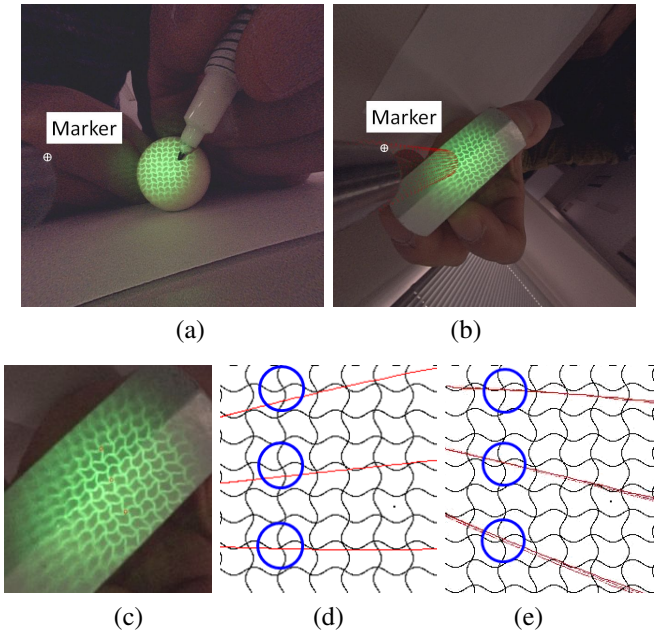
There is an inherent trade-off between the visibility and the reconstruction performance. On one hand, turning off the main endoscope light allows to have greater contrast to visualize the projected pattern, at the expense of visibility of the target tissue. Therefore, we capture the target surface both with and without the endoscope light, yielding images I_l and I_p respectively. The light is toggled on and off during capture, while the pattern projection is continuously kept on. From the two images acquired with the visible pattern, we also estimate the pattern-free images for visualization purposes. Being able to estimate the pattern-free images allows the endoscopists to better judge the extent of the target area without visual nuisances while keeping the procedural overhead low by not requiring to continuously turn the projector on and off. The acquired images are shown in Fig. 5.

First, the radiometric response function of the endoscopic camera is pre-operatively calibrated with a standard Macbeth chart. The radiometric response is calculating by fitting 2-parameter exponential curves of the form $y = ae^{bx}$ to the Macbeth values/observed intensity data entries for each color channel. The intensity of each incoming frame is then compensated according to the inverse of the estimated function. To estimate the pattern-free image, we notice that despite the narrow green wavelength of the projected pattern, overlap in spectral sensitivities between the camera's CCDs will result in non-zero images also in the red and blue channels, implying a white component in the dark pattern image. Therefore, at each nonzero pixel x of I_p we estimate the white component by considering the channel values as $m = \min(I_p^r, I_p^g, I_p^b)$. The green chromaticity is therefore $c = I_p^g - m$, which is removed per-pixel to yield the estimated tissue image $\hat{I}_t = I_l - c$. The brightness of the affected area is finally restored by calculating the mean value of the affected



(a) (b) (c)

Fig. 5. Color texture acquisition: (a) the image with the projected pattern and the texture, (b) the image for capturing the projected pattern, (c) the texture image composed from (a) and (b)



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

Fig. 6. Evaluation of auto-calibration: (a) an image for the pre-calibration, (b) an image for the auto-calibration (cylinder-like dots represents 2-DOF possible positions of the marker), (c) three points for epipolar tests (indicated as small red circles), (d) the epipolar lines of (c) drawn on the pattern image with the parameters of the base-calibration (blue circles are corresponding points of (c)), and (e) the epipolar lines of (c) calculated by the auto-calibrated parameters (three sets of lines calculated by operation of three subjects).

pixels and shifting it to match the mean of the unaffected pixels in the image.

VI. 3D RECONSTRUCTION

In the experiments described in this paper, the 3D reconstruction method is based on template stereo matching with patch-based normalization [11]. In this method, window matching under epipolar constraints is executed, where one image is the projected pattern image while the other is the observed camera image. The two images are rectified so that epipolar lines become horizontal, and intensity values in each window are first normalized to lie in the range $[0, 1]$. During the matching process, the pattern distortions in the camera image caused by the surface geometry are modeled by an affine transformation with 2 DOFs. The distortion and the disparities are simultaneously optimized for each patch, which results in a 3-DOF optimization for matching. This patch-based template matching cost is then regularized by a belief propagation method.

VII. EXPERIMENTS

A. Accuracy evaluation

To evaluate the accuracy of our proposed auto-calibration technique, we examined if the epipolar constraints are met for the auto-calibrated parameters. First, the pre-operative calibration of the base position was processed using a explicit calibration while simultaneously obtaining the marker position (Fig. 6(a)). Then, the head of the endoscope was bent, so that the pattern projector was moved and rotated within the instrument channel. We calibrated this position with our auto-calibration method from a captured image shown in Fig. 6(b). To avoid human bias in manually deciding the marker locations, the position of the projector has been estimated by three independent subjects, yielding three sets of the extrinsic parameters.

To evaluate the accuracy of the extrinsic parameters, we drew epipolar lines (Fig. 6(d) and (e)) of three points shown in (c). Fig. 6(d) shows that the epipolar lines drawn with the parameters of the "base position" calculated from Fig. 6(a). The epipolar lines do not pass through the points that are marked by the blue circles, which are corresponding points of (c). This means that the extrinsic parameters were not correct because the motion of the projector changed the extrinsic parameters.

Fig. 6(e) shows the epipolar lines drawn with three sets of auto-calibrated parameters calculated from marker positions decided by the tree subjects respectively. All sets of the lines pass near the points marked by the blue circles, which means that the results of the auto-calibration were correct, despite of the changes of the subjects.

B. 3D reconstruction of real tissues

To evaluate the system in realistic conditions, the membrane of the soft palate of a human was captured and its 3D shape was reconstructed using the auto-calibrated extrinsic parameters. During the image capture, the endoscopic lighting was multiplexed yielding two images with

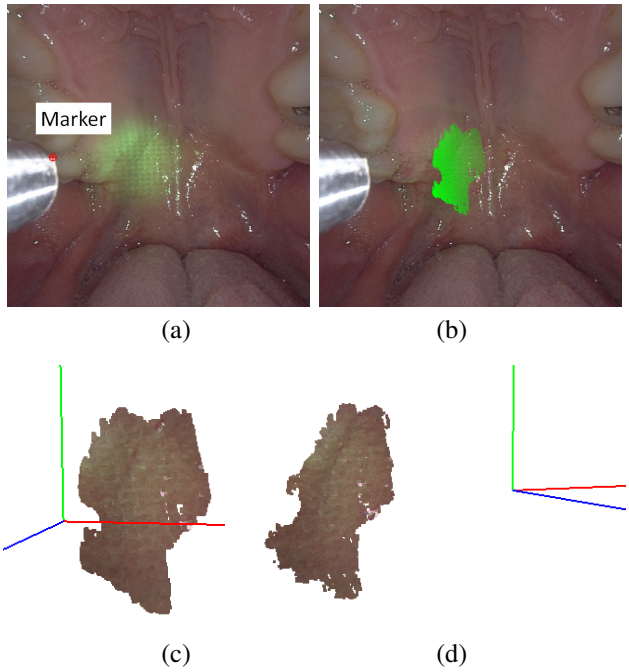


Fig. 7. Acquisition of the soft palate of a human: (a) the image for auto-calibration (the image of Fig. 5(a) with the marker position) , (b) the reconstructed area calculated from the image of Fig. 5(b), and (c)(d) the reconstructed 3D shape with texture and the camera coordinate axis.

and without main illumination, where the marker is observed from the bright image and the pattern is acquired from the dark image. We also estimate the underlying texture image without pattern projection from the dark and bright images using the method outlined in Sec. V. Fig. 7 shows the result. The 3D shape was successfully reconstructed. Note that, since the 3D reconstruction depends on epipolar constraints, the 3D reconstruction does not work properly without accurate calibration.

VIII. CONCLUSION

In this paper, we proposed a 3D endoscopic system based on pattern projection, where pattern projector is inserted trough the instrument channel of the endoscope. Since the projector is not fixed to the endoscope, an auto-calibration method for estimating the position of the projector intra-operatively is proposed. Using the proposed auto-calibration method, the extrinsic parameters of the projector-camera system were evaluated, and it was confirmed that the epipolar constraints were kept. The potential of the technique was demonstrated by reconstructing the geometry of the soft palate from a human subject. We also demonstrated the ability to estimate the underlying tissue texture free from the projected pattern overlay for visualization purposes, showing its efficacy.

REFERENCES

[1] H. Aoki, R. Furukawa, M. Aoyama, S. Hiura, N. Asada, R. Sagawa, , H. Kawasaki, S. Tanaka, S. Yoshida, and Y. Sanomura, "Proposal on 3-d endoscope by using grid-based active stereo," in *EMBC*, 2013, pp. 5694–5697.

[2] R. Furukawa, M. Aoyama, S. Hiura, H. Aoki, Y. Kominami, Y. Sanomura, S. Yoshida, S. Tanaka, R. Sagawa, and H. Kawasaki, "Calibration of a 3d endoscopic system based on active stereo method for shape measurement of biological tissues and specimen," in *EMBC*, 2014, pp. 4991–4994.

[3] K. Deguchi, T. Sasano, H. Arai, and Y. Yoshikawa, "Shape reconstruction from an endoscope image by shape from shading technique for a point light source at the projection," *CVIU*, vol. 66, no. 2, pp. 119–131, 1997.

[4] M. Visentini-Scarzanella, D. Stoyanov, and G.Z. Yang, "Metric depth recovery from monocular images using shape-from-shading and specularities," in *ICIP*, Orlando, USA, 2012, pp. 25 –28.

[5] C. Wu, S. Narasimhan, and B. Jaramaz, "A multi-image shape-from-shading framework fornear-lighting perspective endoscopes," *IJCV*, vol. 86, pp. 211–228, 2010.

[6] G. Ciuti, M. Visentini-Scarzanella, A. Dore, A. Menciassi, P. Dario, and Guang-Zhong Yang, "Intra-operative monocular 3d reconstruction for image-guided navigation in active locomotion capsule endoscopy," in *BioRob*, 2012, pp. 768 –774.

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

[8] D. Stoyanov, M. Visentini-Scarzanella, P. Pratt, and G.Z. Yang, "Real-time stereo reconstruction in robotically assisted minimally invasive surgery," in *MICCAI*, 2010, pp. 275–282.

[9] O.G. Grasa, E. Bernal, S. Casado, I. Gil, and J.M.M. Montiel, "Visual slam for handheld monocular endoscope," *IEEE Trans. on Medical Imaging*, vol. 33, no. 1, pp. 135–146, 2014.

[10] R. Sagawa, K. Sakashita, N. Kasuya, H. Kawasaki, R. Furukawa, and Y. Yagi, "Grid-based active stereo with single-colored wave pattern for dense one-shot 3D scan," in *3DIMPVT*, 2012, pp. 363–370.

[11] H. Kawasaki, H. Masuyama, R. Sagawa, and R. Furukawa, "Single colour one-shot scan using modified penrose tiling pattern," *Computer Vision, IET*, vol. 7, no. 5, pp. 293–301, 2013.