



Designing a New Endoscope for Panoramic-View with Focus-Area 3D-Vision in Minimally Invasive Surgery

Dinh Thai Kim^{1,3} · Ching-Hwa Cheng² · Don-Gey Liu^{1,2} · Kai Che Jack Liu⁴ · Wayne Shih Wei Huang⁴

Received: 16 April 2019 / Accepted: 27 November 2019 / Published online: 3 December 2019
© The Author(s) 2019

Abstract

Purpose The minimally invasive surgery (MIS) has shown advantages when compared to traditional surgery. However, there are two major challenges in the MIS technique: the limited field of view (FOV) and the lack of depth perception provided by the standard monocular endoscope. Therefore, in this study, we proposed a New Endoscope for Panoramic-View with Focus-Area 3D-Vision (3DMISPE) in order to provide surgeons with a broad view field and 3D images in the surgical area for real-time display.

Method The proposed system consisted of two endoscopic cameras fixed to each other. Compared to our previous study, the proposed algorithm for the stitching videos was novel. This proposed stitching algorithm was based on the stereo vision synthesis theory. Thus, this new method can support 3D reconstruction and image stitching at the same time. Moreover, our approach employed the same functions on reconstructing 3D surface images by calculating the overlap region's disparity and performing image stitching with the two-view images from both the cameras.

Results The experimental results demonstrated that the proposed method can combine two endoscope's FOV into one wider FOV. In addition, the part in the overlap region could also be synthesized for a 3D display to provide more information about depth and distance, with an error of about 1 mm. In the proposed system, the performance could achieve a frame rate of up to 11.3 fps on a single Intel i5-4590 CPU computer and 17.6 fps on a computer with an additional GTX1060 Nvidia GeForce GPU. Furthermore, the proposed stitching method in this study could be made 1.4 times after when compared to that in our previous report. Besides, our method also improved stitched image quality by significantly reducing the alignment errors or "ghosting" when compared to the SURF-based stitching method employed in our previous study.

Conclusion The proposed system can provide a more efficient way for the doctors with a broad area of view while still providing a 3D surface image in real-time applications. Our system give promises to improve existing limitations in laparoscopic surgery such as the limited FOV and the lack of depth perception.

Keywords Minimally invasive surgery (MIS) · Image stitching · 3D reconstruction · Real-time display

✉ Dinh Thai Kim
kdthai@ictu.edu.vn

Ching-Hwa Cheng
chengch@fcu.edu.tw

Don-Gey Liu
dgliu@fcu.edu.tw

¹ Ph. D. Program of Electrical and Communications Engineering, Feng Chia University, Taichung 40724, Taiwan, ROC

² Department of Electronic Engineering, Feng Chia University, Taichung 40724, Taiwan, ROC

³ University of Information and Communication Technology, Thai Nguyen, Vietnam

⁴ IRCAD Taiwan/AITS, Lukang, Taiwan, ROC

1 Introduction

Minimally invasive surgery (MIS) is gradually replacing traditional surgical methods because of its advantages such as causing less injury, preventing unsightly surgical scars, and resulting in faster recovery time. The difference between MIS and traditional surgery lies in their respective methods of observation and operation. Through the traditional methods, a doctor can look directly at the operating area, have a wide field of view (FOV) and viewing angles, and get tactile feedback during the operation. However, the incision through MIS is small enough for the instruments and endoscope to pass. Therefore, a doctor cannot look directly at the operating area and has to look at the image on the flat screen

transmitted from the endoscope. Being limited by the narrow view field of the endoscope makes it difficult for a doctor to see the full picture of the operation area. Furthermore, the display only provides 2D information, so it is also quite difficult to recognize the depth as well as the relative position of the organs from the surgical instruments. This increases surgical risks, making it challenging for less experienced doctors to perform a safe surgical operation.

Therefore, the two biggest challenges for MIS are narrow FOV and lack of depth information in the operating area. A number of commercially available stereo-endoscopic systems have been announced to provide surgeons with 3D images of the surgical area [1]. However, they are usually a concern in the hospitals because of their high cost.

Several studies have reported that image processing techniques overcome the limitations of laparoscopic surgery. As for the problem of the lack of depth information, there are some prominent improvements reported in [2–6]. These studies focus on reconstructing 3D images recorded from endoscopes. However, the narrow view field of the endoscope was not considered in these studies due to the system of the real-time requirements.

As regards the problem of the narrow-angle FOV, the earlier studies relied on the movement of an endoscope to create a static panorama picture that fill the operation area [7–11]. However, the position and shape of the internal organs as well as instruments, especially for abdominal MIS, frequently change during the operations. Therefore, this approach is not suitable for practical applications. Moreover, in our previous study, we proposed an MIS panoramic endoscope (MISPE) in order to provide the doctors with broad areas of view [12–14]. For this, we used two endoscopic cameras and a feature-based image stitching algorithm to create a panoramic dynamic image. However, this approach does not work well in MIS, which is often affected by smoke, vapor, viewpoints, and specular highlights. In this situation, the distribution of the features on the images would be ambiguous and the precision of the matching feature pairs would be degraded. Additionally, this approach requires expensive computation for extracting the features in an image; therefore, this feature-matching method is difficult to use in real-time applications.

Hence, in this study, we tried another approach to make a real-time system that can simultaneously reconstruct 3D images with the expanded FOV of an endoscope. We developed a new stitching method based on a method that originally calculated the disparity map. Our approach will make images stitching faster, more stable, and more accurate in MIS. In addition, this approach also provides additional 3D information that enables the physicians to be aware of the depth and distance of the operating environment.

The rest of the paper is organized as follows: Sect. 2 introduces the previous works on the problems of image stitching

and 3D reconstruction, and Sect. 3 describes our endoscope system and the proposed algorithm. Furthermore, Sect. 4 describes and discusses the experimental results, while Sect. 5 provides the conclusions along with directions for future research.

2 Previous Works

Several studies in the literature focus on solving the issues of a narrow FOV and the lack of depth perception, which persist in traditional MIS, separately. In this study, however, we tried to solve the two issues by one method. The separate technologies for stitching and 3D reconstruction have been reviewed below.

2.1 Image Stitching

In image processing, there is a method that allows multiple overlapping images to be merged into a single wide-field image, namely image stitching or image mosaicing. Comprehensive research can be found in [15]. This technique is also used in medicine such as slit lamp image mosaicing [7, 16, 17]. For instance, Cattin et al. [16] describe an approach for mosaic retina images based on Speeded-Up Robust Features (SURF) [18]. Moreover, Zanet et al. [7] proposed a method to improve slit lamp acquisitions by creating global mosaics of the retina when poor quality video frames were presented.

In MIS image engineering, there have been some studies to improve the calculation. For example, Behrens et al. [19] proposed a multithreaded image-mosaicing algorithm to perform the mosaicing of bladder images in real-time, while Yang et al. [20] proposed an approach based on scene-adaptive features for the mosaicing of placental vasculature images obtained during computer-assisted fetoscopic procedures.

Most of these studies adopted the features-based image registration to perform sequence-image mosaicing. The robustness of these methods depends on the availability of stable features. For instance, Hu et al. [21] proposed a robust technique for image registration by the Homographic Patch Feature Transform that they can detect features in gastroscopic image sequences with good robustness, precision, and uniformity.

Besides, there are also other approaches to implement image mosaicing. For example, Liu et al. [8] utilized a tracking device according to the images from a single-camera gastroscope with a dual-cubic projection method in order to simultaneously create both local and panoramic views. In [11], an image mosaicing scheme based on Simultaneous Localization and Mapping (SLAM) has been proposed for dynamic view expansion. Ali et al. [22] also proposed

a novel data term for motion estimation for robust bladder image mosaicing.

However, in all the studies described above, a panoramic image created by the movement of a monocular camera cannot easily reflect the tissue deformations or instrument motion outside of FOV. Therefore, this approach is difficult to apply in practical laparoscopic surgeries.

2.1.1 3D Reconstruction

In order to realize the depth information of the image, the 3D reconstruction technique was introduced. The essence of this technique is to map the available 2D image coordinates onto 3D world coordinates. In man-made environments, the 3D reconstruction using stereo images is a common approach for general problems [23, 24].

In the context of MIS, there are two approaches for this technique [25]. The first approach, used in traditional laparoscopy, is based on moving a monocular endoscope in order to reconstruct the 3D surface of the surgical area. Three methods are commonly used to obtain depth information: Structure from Motion (SfM) [3, 26], SLAM [27, 28], and Shape from Shading (SfS) [29]. However, a disadvantage for both SfM and SLAM is that the camera needs to move constantly in order to obtain 3D information. Moreover, the SfS method has the additional disadvantage that it is very sensitive to specular highlights; therefore, it is difficult to obtain accurate depth information.

The second approach is Robot-Assisted Surgery, which is based on the stereo endoscopes for getting the depth perception of the surgeon. The principle of this approach is matching pixels between the left and the right image in order to calculate the depth information through triangulation. Several studies are based on this approach in MIS. For example, Stoyanov et al. [30] presented a real-time stereo reconstruction in robotically assisted MIS. Bernhardt et al. [31] proposed a powerful approach for the dense matching between the two stereoscopic camera views so as to make a dense 3D reconstruction. Furthermore, a real-time dense GPU-enhanced surface reconstruction from stereo endoscopic images for intraoperative registration was also proposed in [32].

However, the 3D surface reconstruction of surgical endoscopic images is still an issue owing to certain challenges such as the abundance of texture-less areas, occlusions introduced by the surgical tools, specular highlights, smoke, and blood produced during the interventions [33]. Hence, few recent studies have focused on making the surface reconstruction more reliable, exact, and robust. For example, Penza et al. [4] introduced a novel method to enhance the dense surface reconstruction through disparity refinement that is based on simple linear iterative clustering (SLIC) super-pixels algorithm. Furthermore, Wang et al. [6]

proposed some advanced techniques aimed at reconstructing the 3D liver surface based on the stereo vision technique.

Besides, the use of stereo endoscopes is still not a common practice in traditional MIS, as thus far, their use is limited to robotic systems such as the da Vinci surgical system.

3 Materials and Methods

3.1 The Proposed Endoscope System (3DMISPE)

As is shown in Fig. 1, our device consisted of two cameras, a push-button, and a mechanical tube. The two cameras used here were the 2.0MP USB endoscope cameras, the specifications of which are shown in Table 1. The mechanical tube had a diameter of 13 mm. Figure 1a shows the primary state of our device, where the push-button had not been pushed down yet and the width of the gap between the two cameras was about 2 mm. In this state, our endoscope could be inserted into the patient's abdomen through a small hole about 15 mm in diameter. Figure 1b shows the working state of our device, where the push-button was pushed down and the distance between the two cameras was 15 mm. In the working state of our device, the two cameras were placed in parallel each other, and the geometric arrangement between them was as shown in Fig. 1c.

Our system consisted of two lenses connecting to a PC via two USB ports. Since our system was equipped with two lenses, it was convenient to reconstruct a 3D image as well as to expand the FOV of the endoscope. As shown in Fig. 2a, our system included two endoscopic cameras for capturing the input images of the surgical area. Then, the proposed algorithm performed image processing to simultaneously create the 3D image and stitched images.

3.2 Proposed Algorithm

The proposed algorithm consisted of four steps, as described in Fig. 2b. These input images were rectified to calculate the disparity. Then, the dense 3D reconstruction and image stitching process were performed from the rectified images and the disparity information. The processes will be described in detail in the subsections below.

3.2.1 Image Rectification

The two basic purposes of this step are as follows: the first was image correction, which was owing to the distortion of the lens. The second was the alignment of the two cameras into one viewing plane, which was so that the pixel rows between the cameras were exactly aligned with each other.

To achieve these objectives, we selected Bouguet's algorithm [34], which was available in the OpenCV with an

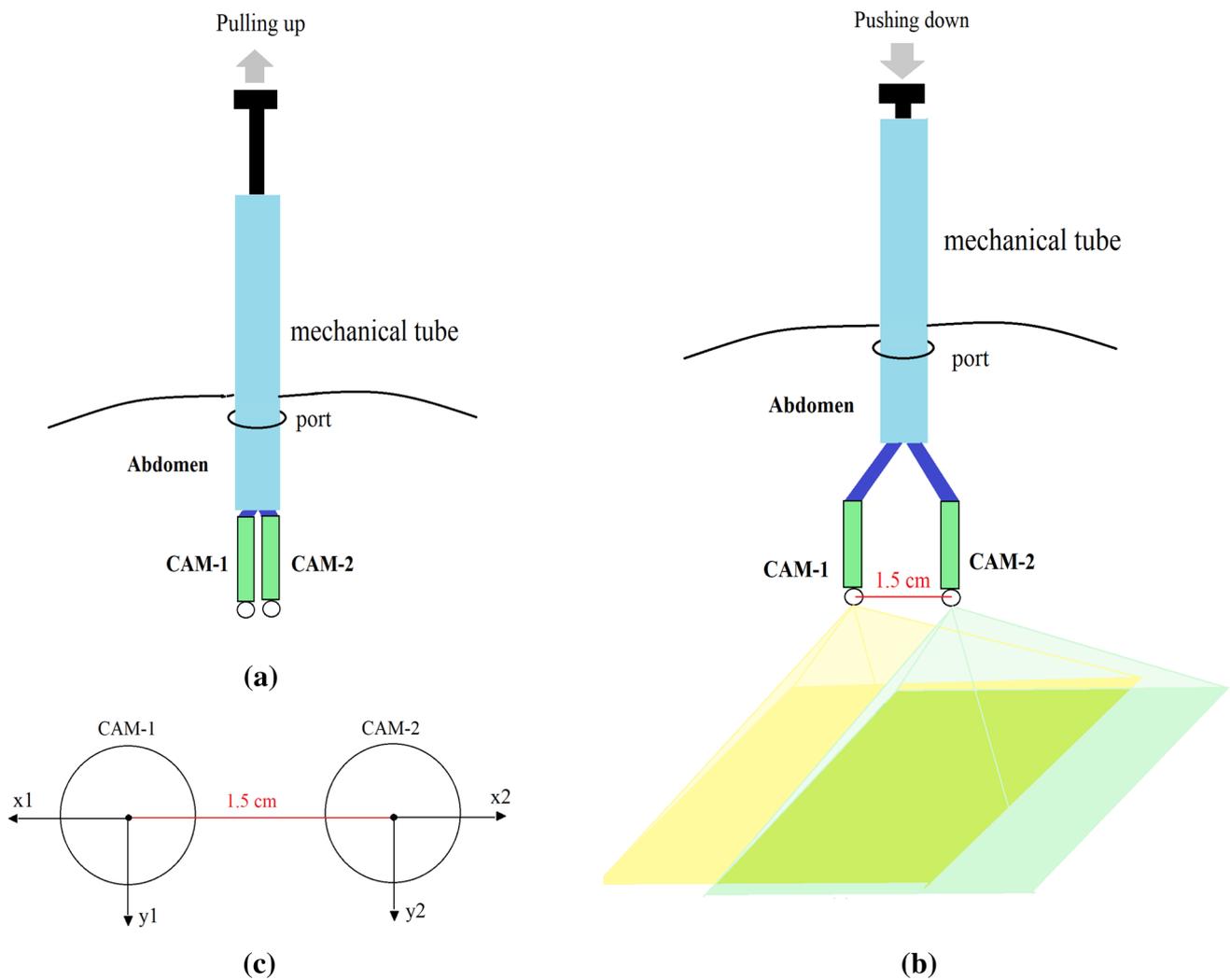


Fig. 1 The proposed endoscope system (3DMISPE) consisted of two cameras, a mechanical tube, and a push-button. The figure depicts **a** the primary state of the device, **b** the working state of the device, and **c** the geometric arrangement between the two cameras

Table 1 Technical specifications of the endoscopic cameras

USB endoscope camera, Shenzhen HuaSunTek Technology Inc.	
Camera diameter	5.5 mm
Resolutions	640×480, 1280×720 pixels at 30 fps
View angle	70°
Camera	2.0 Mp

assumption that the camera model was a pinhole model. First, we calibrated to obtain intrinsic and extrinsic parameters for each camera. In this study, we adopted Zhang’s method [35] for calibration. In order to obtain the precise calibration, we concurrently captured 20 images of a 14×11 chessboard, where the size of each square is 1.5×1.5 mm,

and placed it at a distance of 3 cm to 15 cm and different angles. Then, we employed Bouguet’s algorithm with the attempt to minimize the changes of reprojection while maximizing the common viewing area. As a result, the distorted and misaligned input images could be transformed into undistorted and rectified images and could make the corresponding pixels lie on the same horizontal line, i.e., the epipolar line.

3.2.2 Disparity Map

After the rectification process, the stereo correspondence, also known as stereo matching, was calculated. In this study, we selected the block matching (BM) algorithm, which was available in the OpenCV as the StereoBM module. This was because this algorithm is a fast and effective as well as similar to the one developed by Konolige [36]. It works by using

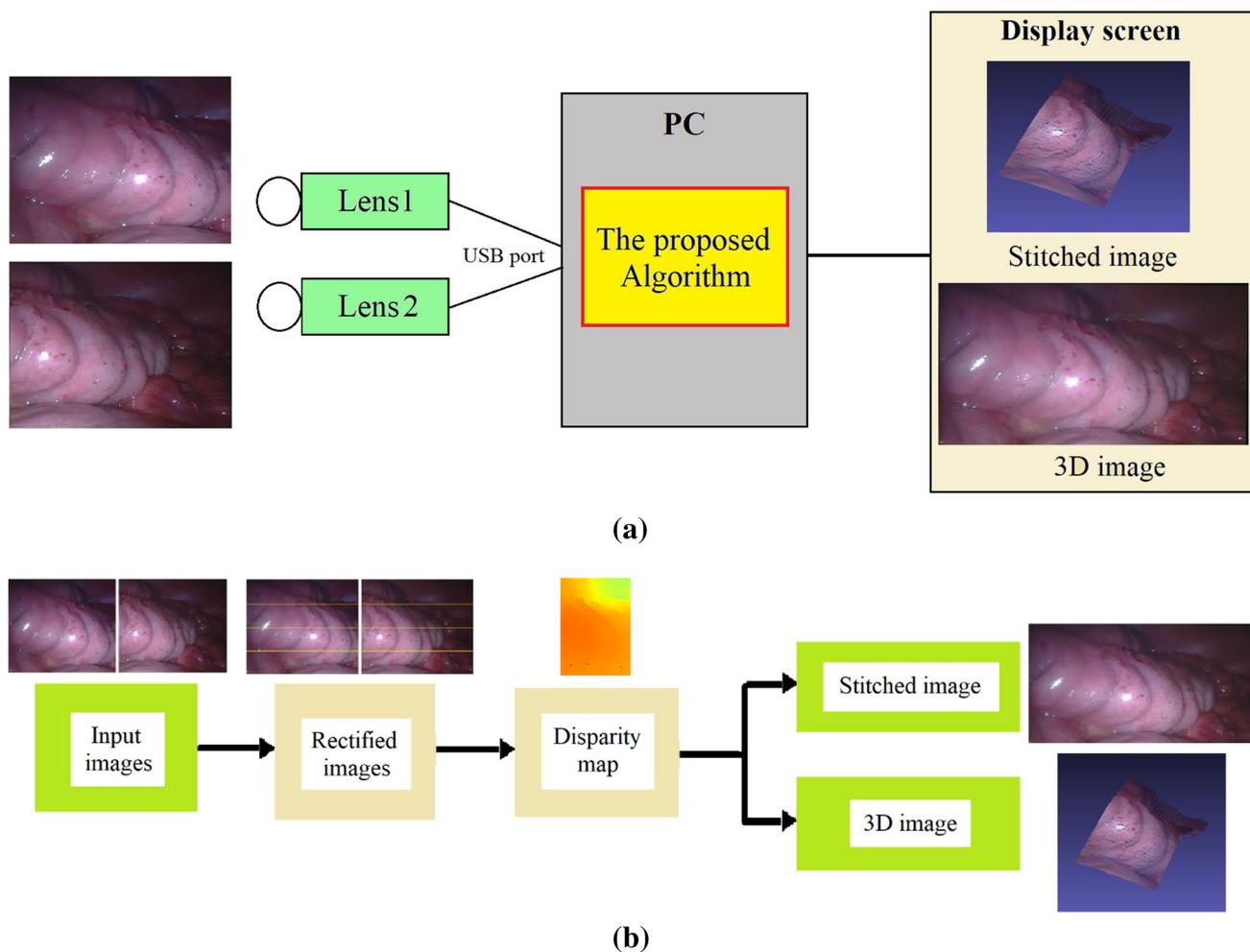


Fig. 2 The schematic diagram of our endoscope system and the proposed algorithm. **a** The schematic diagram of our endoscope system. The two images at the left side indicate the input images obtained from the two lenses. Through the USB ports on the PC in the center, two outputs can be derived by our algorithm. The two images at right

side indicate a window displaying a 3D image and another window showing an extended 2D view around the same area. **b** The proposed algorithm for rendering stitched image and 3D image. There are four steps: (1) Image rectification, (2) Disparity calculation, (3) 3D reconstruction and (4) Image stitching

a small “sum of absolute difference” (SAD) window to find the matching points between the left and right images. The disparity was calculated as the actual horizontal pixel difference. Here, the difference in the position of a point between the left and right images is called the disparity. This disparity map depicted all of the apparent differences in a pixel between the image pair.

However, the disparity map is computed by StereoBM, which usually contains invalid values (holes) is usually concentrated in uniform texture-less areas, half-occlusions, and regions near depth discontinuities. Therefore, we used the edge-aware filters as a post-filtering process in order to improve the quality of the disparity maps. This approach allows the aligning of the disparity map edges with those of the source image so as to propagate the disparity values from high- to low-confidence regions such as half-occlusions. The

two filters used in this study were: the Fast Global Smoothing [37] and the Fast Bilateral Solver [38], both of which were integrated into OpenCV as the DisparityWLSFilter (WLS) and FastBilateralSolverFilter (FBS) classes, respectively. These filters can enable a post-filtering process under the constraints of real-time computation by a CPU, where no additional GPU is needed.

The details regarding the implementation are depicted in Fig. 3. First, we computed two raw disparity maps using the StereoBM method. The first disparity map was obtained by taking the left image as the reference image (i.e., the left disparity map) and the second disparity map with the right image as the reference (i.e., the right disparity map). Then, we used the WLS in order to get a confidence map and refine the disparity map in half-occlusions and uniform areas (i.e., the WLS disparity map).

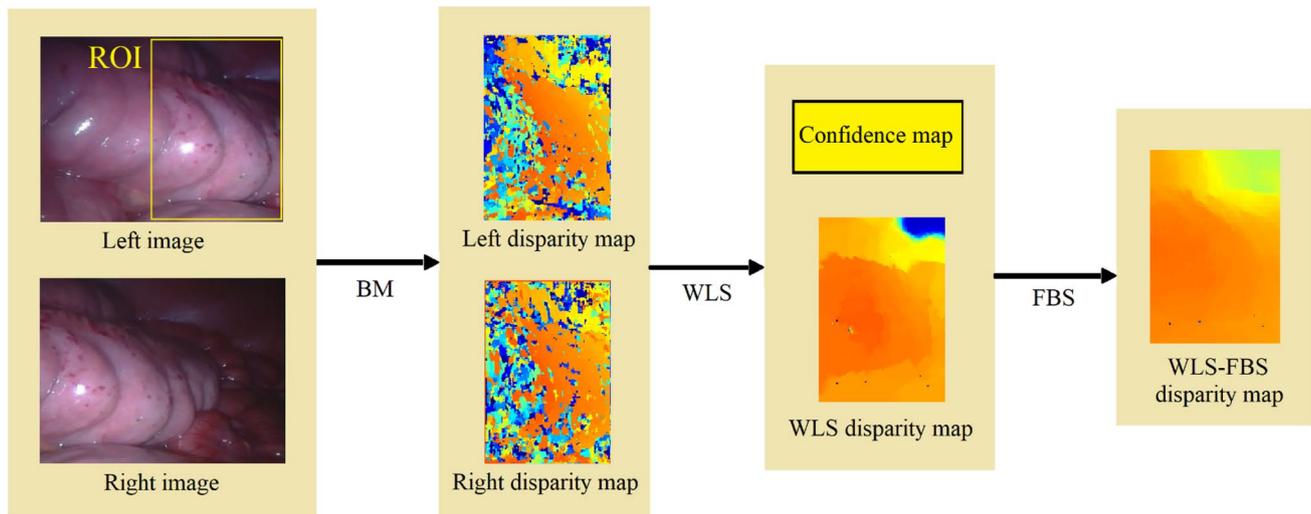


Fig. 3 Disparity map calculation algorithm includes three steps: (1) Compute two raw disparity map by StereoBM (BM), (2) Use WLS to get WLS disparity map and confidence map and (3) Use FBS to get WLS–FBS disparity map

Finally, we used the FBS with the confidence map and the ROI image, which served as a guide for filtering the WLS disparity map (i.e., the WLS–FBS disparity map).

3.2.2.1 3D Reconstruction Based on the camera’s geometry, the disparity value (d) can be converted into depth value (Z) by the following formula [34]:

$$Z = T \times f/d \quad (1)$$

Here, f is the focal length of the camera, and T is the baseline distance between the cameras.

In addition, the position of a point P in 3D-space can also be estimated by its coordinates in the left image and the disparity values (d) and calibrated camera parameters [34].

$$\begin{bmatrix} X \\ Y \\ Z \\ W \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & -c_x \\ 0 & 1 & 0 & -c_y \\ 0 & 0 & 0 & f \\ 0 & 0 & -1/T_x & (c_x - c'_x)/T_x \end{bmatrix} \times \begin{bmatrix} x \\ y \\ d \\ 1 \end{bmatrix} \quad (2)$$

Here, x and y as well as c_x and c_y are the corresponding pixel’s coordinates of the point P and the principal point in the left image, respectively. Moreover, c_x is the x -coordinate of the principal point in the right image, while T_x is the baseline length. The three-dimensional coordinates of the point P in the left camera coordinate system are X/W , Y/W , and Z/W .

Therefore, a 3D surface of the overlap region can be obtained from the original images and the disparity information. In order to visualize the 3D surface, we used the Viz-module in OpenCV for depicting the possible shape.

3.2.3 Image Stitching

To expand the endoscope’s FOV, we used the image stitching algorithm (or mosaicing) in order to combine two overlapped images into a larger picture. In this process, the image stitching algorithm consisted of two steps: image registration and image compositing [13].

The image registration consisted of matching the points in two overlap images in order to evaluate a homography matrix, which is a new registration method that is based on the stereo matching algorithm. We used two rectified-images instead of two input images for the stitching task. Furthermore, we came to the understanding that using all the matching pixels in the overlap region in order to evaluate the homography matrix was unnecessary and would have affected the computational speed.

Therefore, we defined a region of interest (ROI) as a region in the overlap of the left rectified-image where the pixel disparity was calculated, as shown in Fig. 4. Then, we divided this ROI by $m \times n$ grids. From the calculated results of the disparity, each peak of the grid (P) could find the matching point (Q) on the right rectified-image in the follows:

$$(x_Q, y_Q) = (x_P - \text{disparity}(P), y_P) \quad (3)$$

In this way, we have a set of ($m \times n$) correspondence point pairs of the two overlapped rectified-images. Moreover, because the homography matrix is a (3×3) matrix with 8 degrees of freedom (DoF), one needs at least four correspondence point pairs to determine the matrix. Hence,

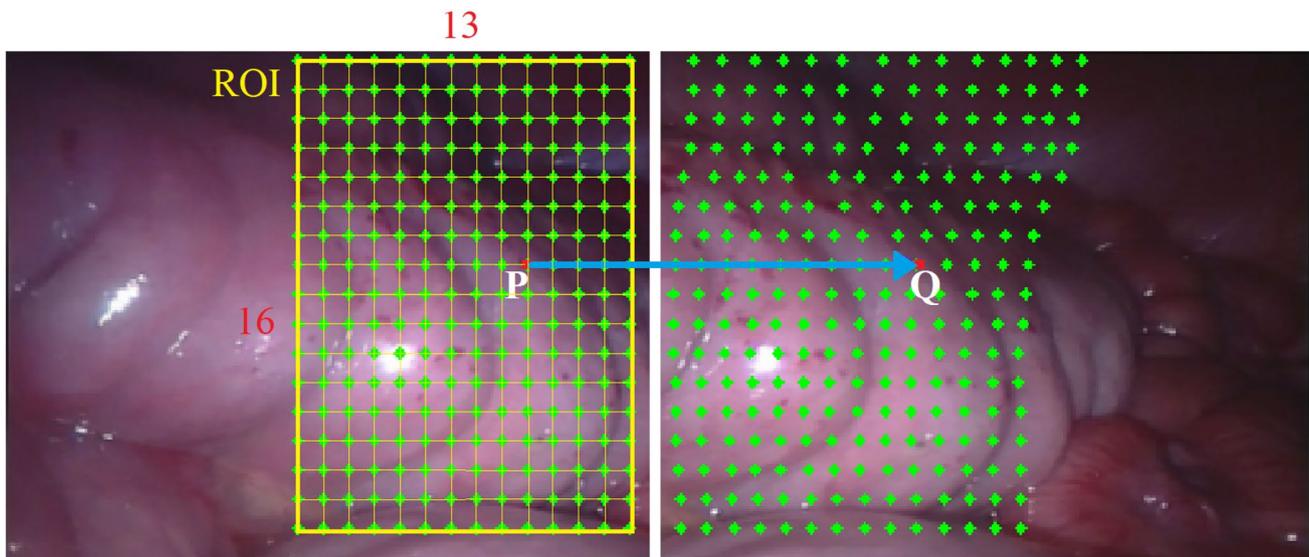


Fig. 4 Determine the matching point pairs based on its disparity value. ROI region (dark yellow) is the region used to compute disparity. Divide ROI by a 13×16 grid and each grid's peak (P) are used to find its matching point (Q) on the right rectified image

$(m \times n)$ was selected to ensure that the number of correspondence point pairs was not less than 4. In this study, the $(m \times n)$ we chose was (13×16) . As there were still some mismatched pairs due to invalid disparity values, we employed the Random Sample Consensus (RANSAC) algorithm [39] so as to remove the mismatch corresponding pairs. Then, the homography matrix was estimated based on the remaining set of corresponding pairs. As Fig. 4 shows, this approach ensured that a large number of matching pairs were evenly distributed in the ROI, making the stitching results more accurate and stable.

After completing the image registration, the image-compositing stage yielded wide-angle images. For this step, we did things as described in our previous study [13]. This means that we still used the graph-cut technique algorithm [40] to find an optimal seam for eliminating the appearance of “artifacts” or “ghosting”. Then, the multiband blending method [41] was adopted to effectively smoothen out the stitching results.

4 Results and Discussion

The experiments were performed using an Intel i5-4590 CPU @ 3.40 GHz with 16 GB of RAM and a GTX1060 Nvidia GPU on an Ubuntu 16.04 system. Furthermore, the program was implemented in C++ codes with OpenCV 4.1.1 and CUDA 10.0. The performance of this system could be accelerated by simultaneously including CPU and GPU operations.

In the following sections, the experimental results as well as the evaluation of the disparity and 3D reconstruction will

be presented. Subsequently, the evaluation of the effectiveness of the proposed stitching algorithm as compared to the feature-based stitching algorithm will be presented. Finally, the remaining limitations of this study will be discussed.

4.1 Experimental Results

To confirm the functions of our proposed system, we performed experiments for the in vivo animal and the phantom model trials.

For the experiments on the phantom model, we pushed the push-button downward in order to bring our endoscope to the working state. Then, our algorithm processed the images captured from the two cameras in order to create wide-range pictures and 3D images of the model frame by frame. Figure 5 demonstrates the two input images (a) with an overlap area (yellow), the proposed algorithm, which can generate (b) a 3D image, and (c) a stitched image.

Animal experiments were carried out in the IRCAD MIS research center of the Show Chwan Memorial Hospital, Taiwan. First, we made a small incision of about 1.5 cm so that in the primary state of the device, we could put our endoscope into the pig's abdomen. Next, we pushed the push-button downward in order to bring our device to the working state. At this state, our endoscope simultaneously captured the images inside the pig's abdomen. Finally, the proposed algorithm allowed for simultaneously creating 3D pictures and stitched images. Figure 6 shows an example of the in vivo animal experiment featuring (a) two input images with an overlap area (yellow), (b) 3D image of the overlap area, and (c) a stitched image.

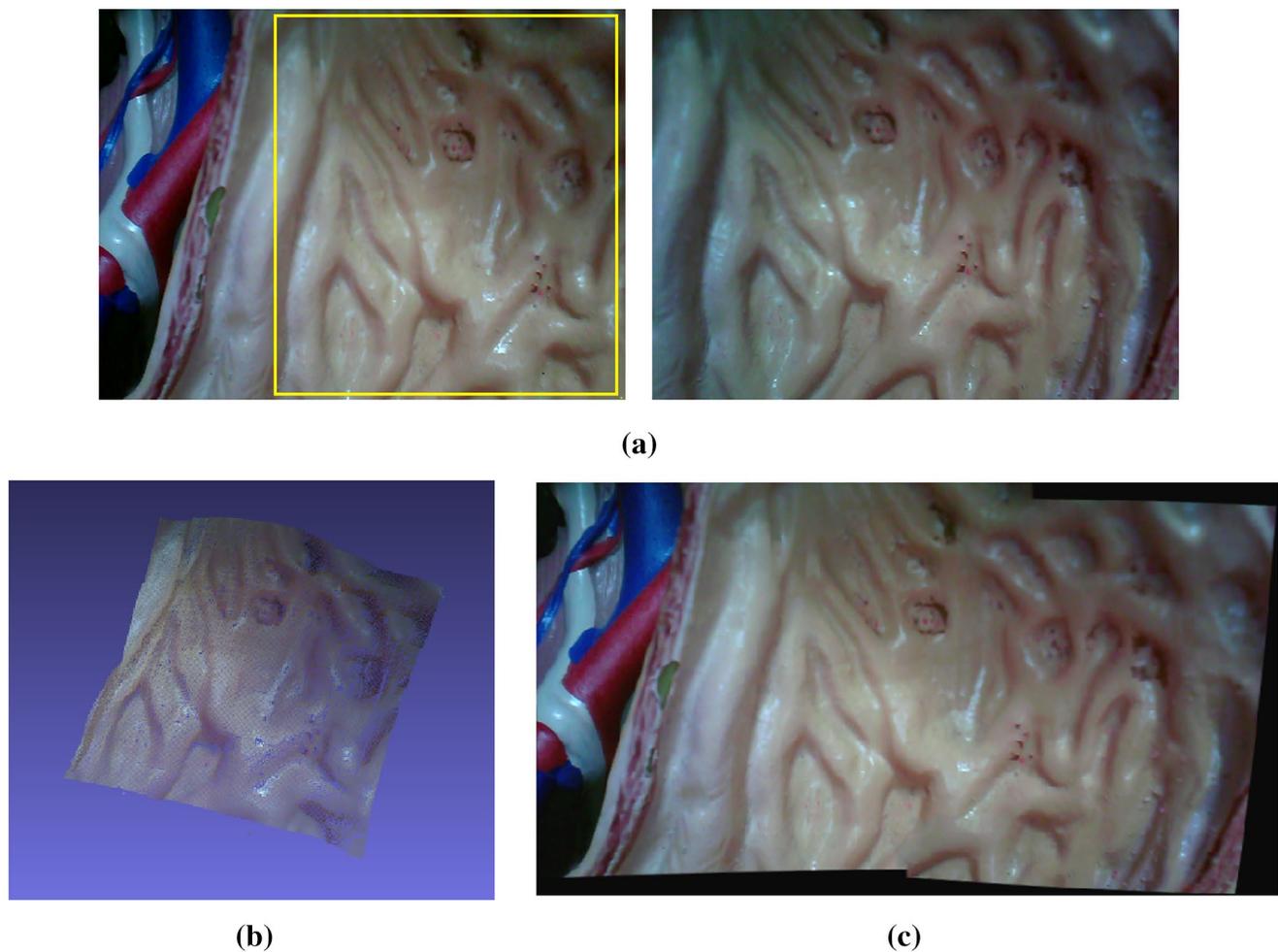


Fig. 5 Results in the phantom model experiment (sample 1): **a** two input images with an overlapped area (yellow), **b** 3D image and **c** Stitched image

Hence, these results confirmed that the proposed method can combine two images into one broader image and also reconstruct the 3D image of the overlap region. Moreover, we recorded the input videos for these two experiments. The detailed evaluations for both video datasets shown in Fig. 5 (sample-1) and Fig. 6 (sample-2) will be described in the sections below.

4.2 Evaluation of the Disparity Map and 3D Reconstruction

4.2.1 Qualitative Evaluation

In this study, we primarily focused on proposing a system that could simultaneously create a 3D image and a stitched image during MIS in real time. Hence, to ensure real-time requirements, we selected the stereoBM method to compute the disparity map. Then, the disparity map was filtered using WLS and FBS, which were integrated and optimized in

OpenCV as the disparity map post-filtering module. Unfortunately, the surgical video datasets, along with ground-truth information, were not available for evaluation. Therefore, we provided qualitative evaluations and compared the results of StereoBM with those of StereoBM after post-processing.

The main parameters for the OpenCV functions were used in the evaluation in the manner described in Table 2 below. In our program, these parameters could be adjusted on the control panel in order to get the disparity of the best quality. The disparity search range parameter (*numDisparities*) that was selected depended on the overlap width of the two cameras, while the remaining sub-parameters were selected by default.

The qualitative evaluation results of the disparity map and 3D reconstruction for both datasets have been presented in Figs. 7 and 8. It can be seen that the disparity map calculated by StereoBM had a lot of invalid values (holes), as represented by Figs. 7b and 8b. Although the WLS filled these invalid values, the WLS disparity map's depth discontinued

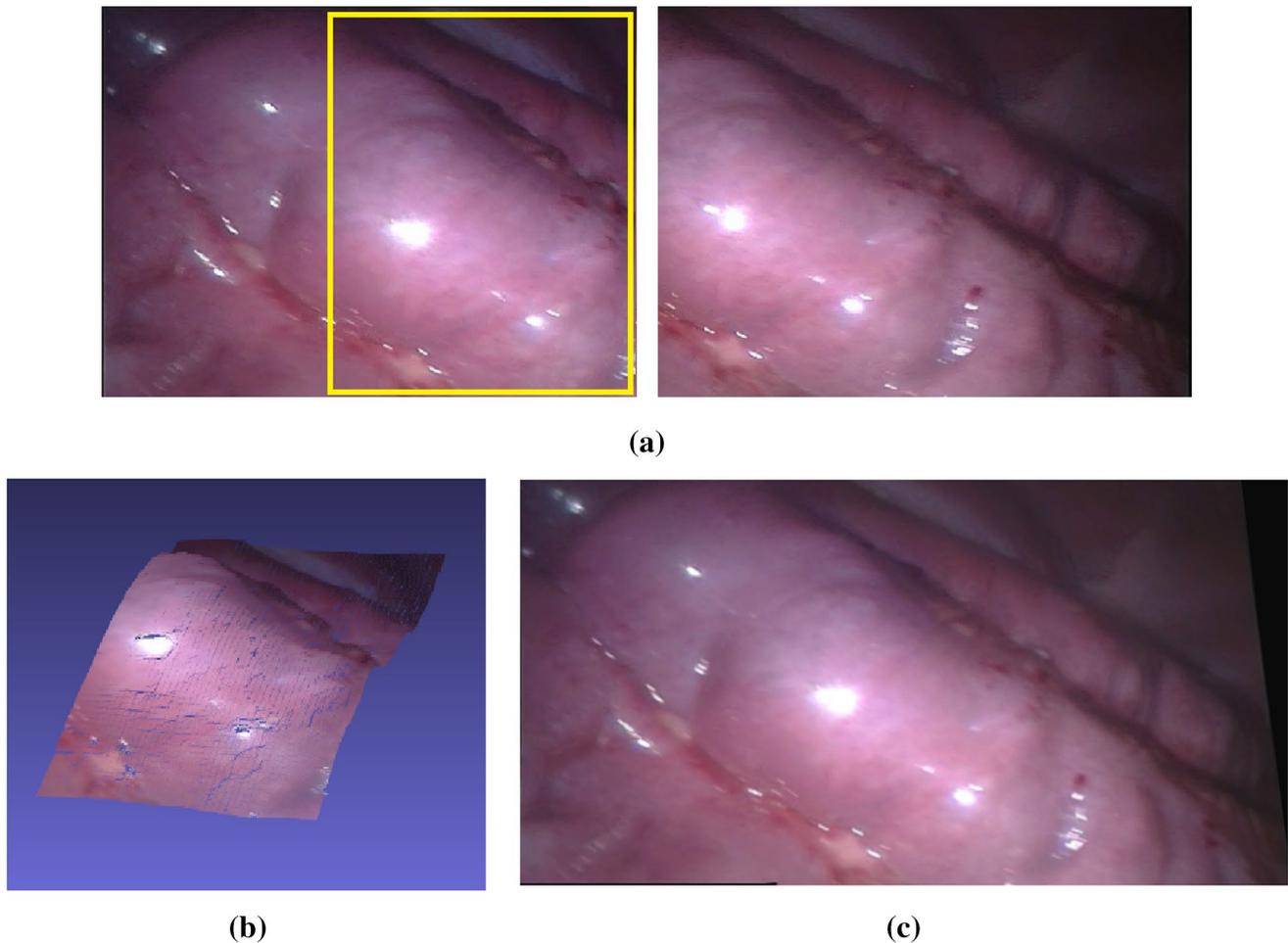


Fig. 6 Results in the in vivo animal experiment (sample 2): **a** two input images with an overlapped area (yellow), **b** 3D image and **c** Stitched image

Table 2 The parameters used in the evaluation

Parameters	Function	Value
BlockSize	StereoBM	9×9
Lambda	WLS	8000
SigmaColor	WLS	1.5
Sigma_spatial	FBS	8
Sigma_luma	FBS	8
Sigma_chroma	FBS	8

due to the risk of having the disparity of certain areas contaminated by the neighboring ones. Moreover, Figs. 7c and 8c show color discontinuities in the disparity image. Finally, the FBS improved this result through the confidence map and made the disparity map much smoother and more continuous, as shown in Figs. 7d and 8d.

Then, the 3D reconstruction (point cloud) of each disparity map was created using the triangulation technique. Thus,

the point cloud, as shown in Figs. 7, 8e–g, was significantly improved and seems to provide a realistic 3D picture of the real scene by the proposed method.

4.2.2 Evaluation of Distance Measurement

To confirm the reliability of the reconstructed 3D image, we conducted the evaluation of distance measurement on the phantom model.

According to Eq. (2), we determined the Euclidean distance from the central point A to 8 locations around it (B, C, D, E, F, G, H, I) and compared these values with the actual measuring distances. Figure 9a shows the estimated distance in mm (yellow) and the actual distance (green) at each side of AC, AD, AE, AF, AG, AH, AI, and AK. It can be observed that these results are quite close to the actual value with an error that is within 1 (mm), which indicates that our method reconstructed the 3D surface of

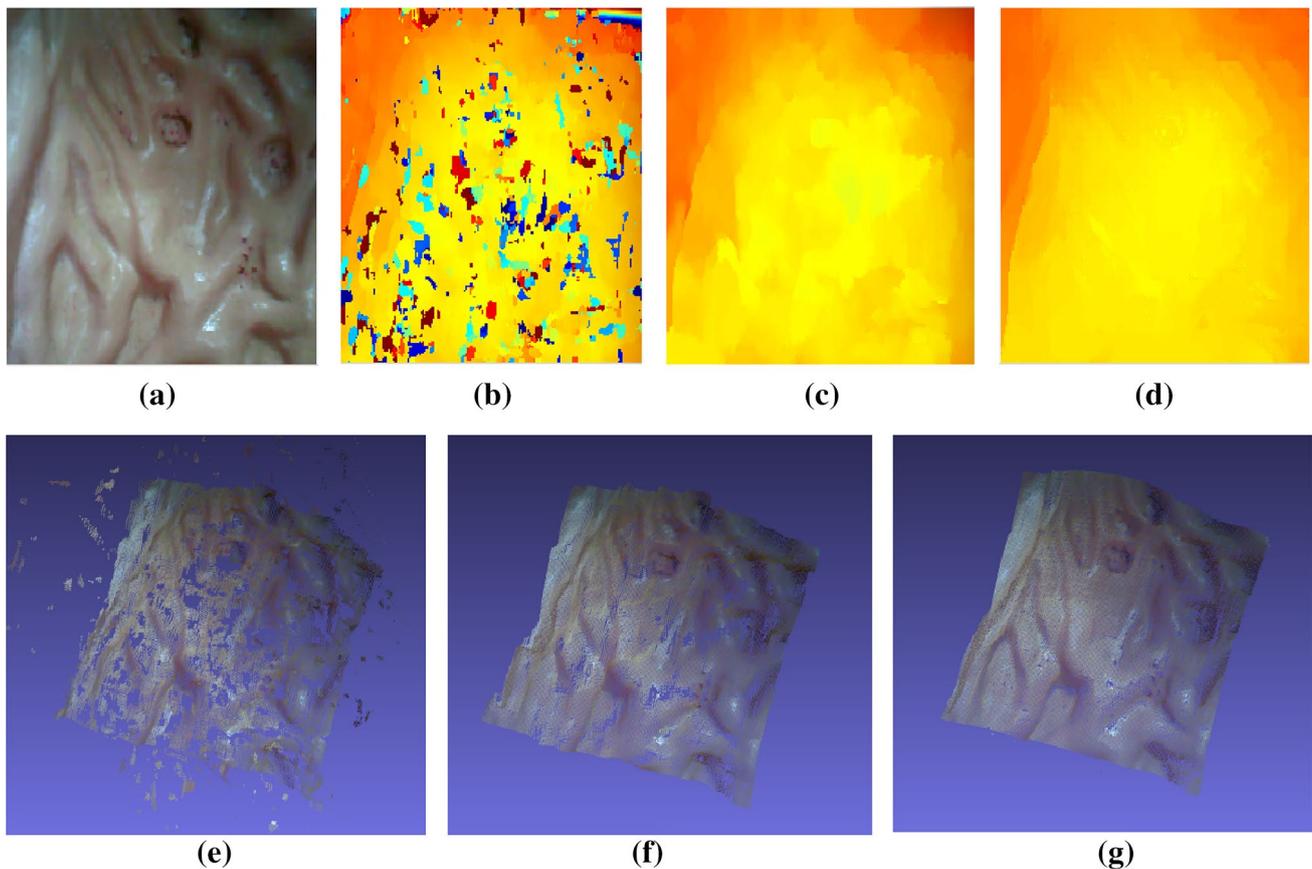


Fig. 7 Evaluation of the disparity map and 3D reconstruction on sample-1. The first row shows: **a** ROI image, **b** raw disparity map, **c** WLS disparity map and **d** WLS-FBS disparity map. The second row rep-

resents the corresponding 3D images: **e** point cloud of raw disparity map, **f** point cloud of WLS disparity map and **g** point-cloud of FBS-WLS disparity map

the phantom model quite accurately, and our system can also serve as a 3D measurement tool during MIS.

Afterward, we calculated the depth of point A according to Eq. (1). We moved the camera baseline at various distances to the model point. Figure 9b shows the estimated depth in comparison with the actual depth. These results demonstrate that both the curves of the depths were almost the same and in the range between 3 and 18 cm. These results are consistent with that of MIS, where the camera's position to the operating area was quite close.

4.3 Evaluation of the Stitching Result

To evaluate the effectiveness of the proposed algorithm, we performed video stitching for both of our datasets. Moreover, we compared our method with the SURF-based stitching method employed in our previous study [13].

4.3.1 Qualitative Comparison

In order to perform the qualitative comparison for both the methods, we avoided the seam cutting because it would have removed “ghosting” in the results. Figure 10 shows that the result achieved by SURF appear more “ghostly” in the area highlighted yellow, while our method significantly reduced the errors. In Fig. 11, the SURF method appeared “ghosting” in the area highlighted yellow, while our approach seems to have aligned correctly.

These results were present because the image registration of the two methods was different. The SURF method relies on “sparse” feature pairs to evaluate a homography matrix. The homography matrix is used to transform the position of the feature points on the right image to the location of the corresponding features on the left image with the smallest re-projection error. Additionally, the scene in MIS is not a plane and in the near focus. Therefore, when the matched feature pairs have failed or are unevenly distributed, it may lead to

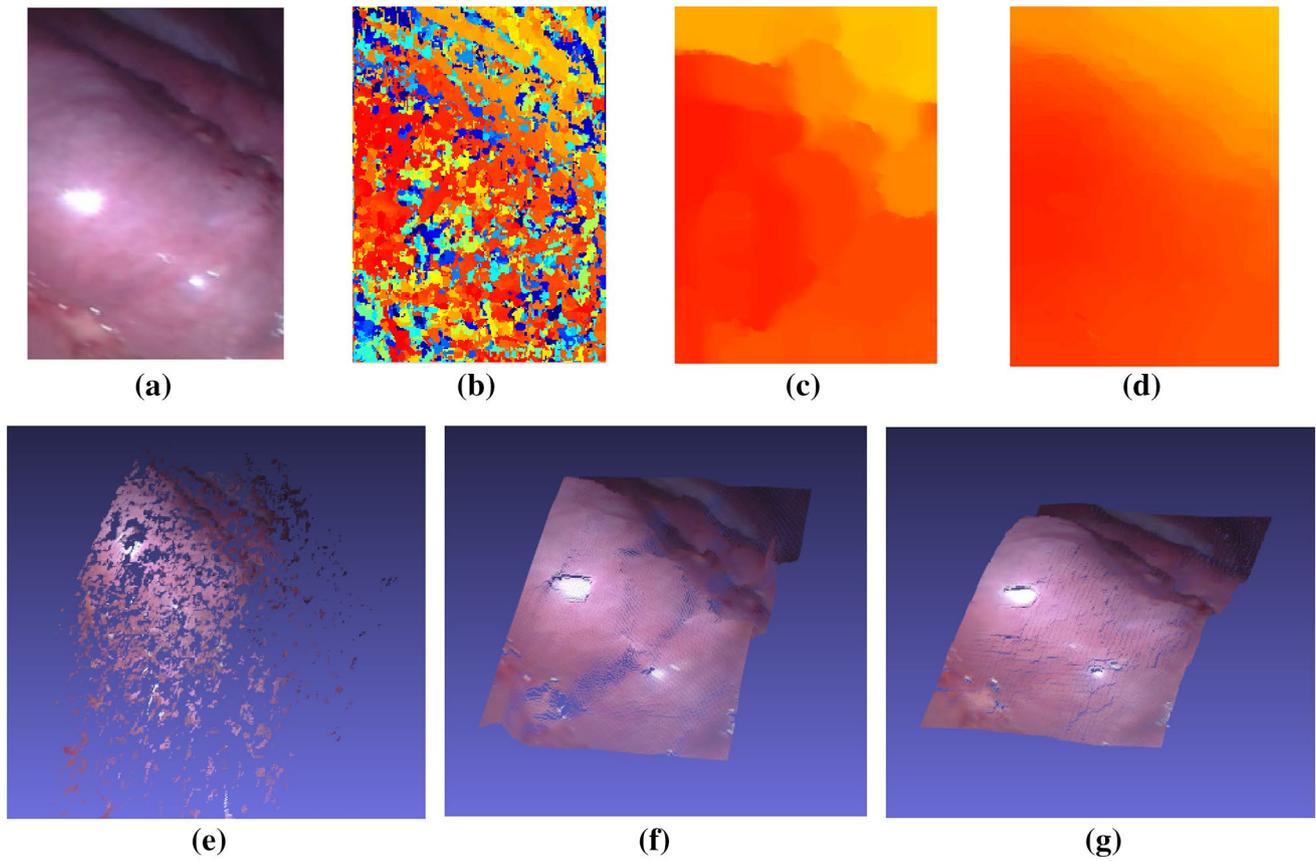
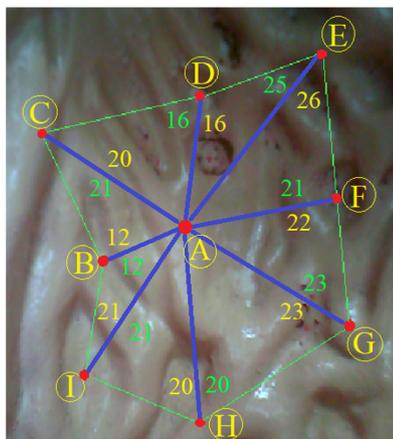
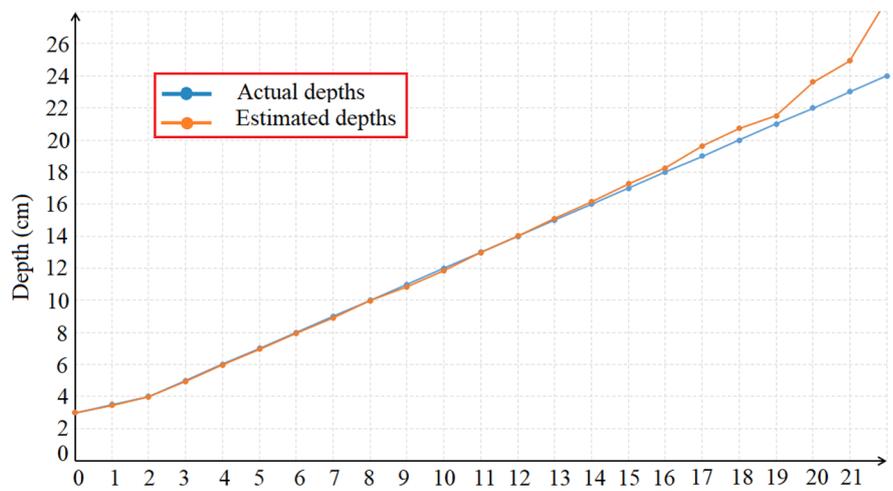


Fig. 8 Evaluation of the disparity map and 3D reconstruction on sample-2. The first row shows: **a** ROI image, **b** raw disparity map, **c** WLS disparity map and **d** WLS-FBS disparity map. The second row rep-

resents the corresponding 3D images: **e** point cloud of raw disparity map, **f** point cloud of WLS disparity map and **g** point-cloud of FBS-WLS disparity map



(a)



(b)

Fig. 9 Evaluation of distance measurement on sample-1. **a** Comparison of the estimated distance with the actual distance. Each side of AC, AD, AE, AF, AG, AH and AK represents the estimated distance

(yellow) and the actual distance (green); **b** comparison of the estimated depth with the actual depth in the phantom model experiment

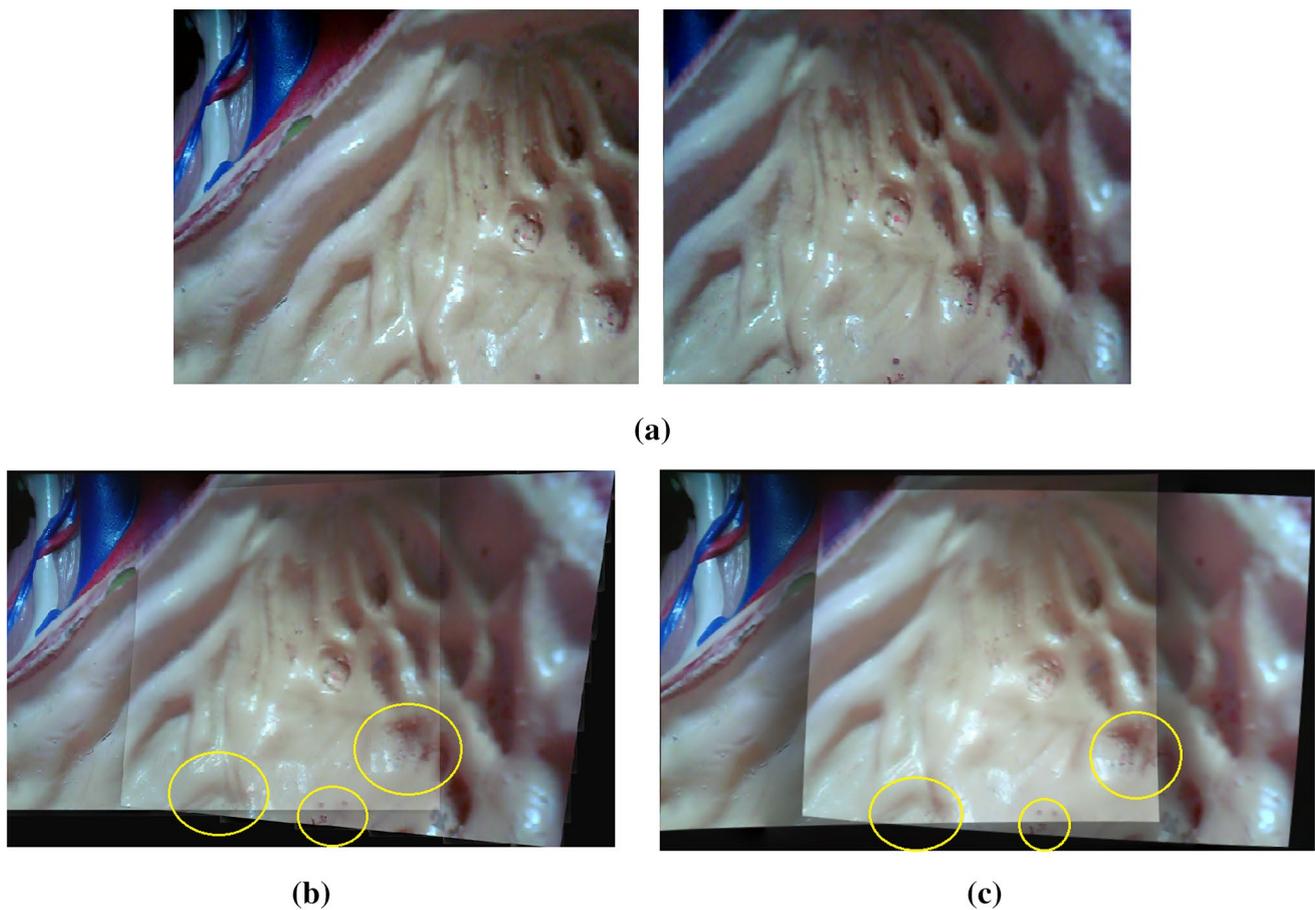


Fig. 10 Qualitative comparison on sample-1. **a** Input images, **b** stitched image by SURF method and **c** stitched image by our method. Yellow circle highlight errors

significant alignment errors in regions where no feature pair is detected. As a result, the stitched images by SURF can be distorted or deformed. Otherwise, our method used a large number of matching pairs that were evenly distributed in the overlap region so as to align. Therefore, our approach reduced the alignment errors, and the stitched images by our method were as “natural” as the ground truth.

4.3.2 Quantitative Comparison

To determine the alignment accuracy of both the methods, we computed the pixel difference between the two warped images in the overlap area. The alignment error of the stitched image was defined as the average intensity difference of the pixels in the same position within the overlap area of the two warped images. We assume that (f) and (g) were the two images after the warping transformation.

$$\text{alignment error} = \frac{\sum_{\text{overlap}} |f(x, y) - g(x, y)|}{\sum \text{pixels in overlap}} \quad (4)$$

Here, $f(x, y)$ and $g(x, y)$ are the grayscale pixel values for a point at coordinate (x, y) in the overlap area of the two images (f) and (g).

The alignment error of the stitched video is determined as the average value of the alignment error of the stitched frames. Table 3 shows the alignment error for both the video datasets. The alignment error for the proposed method was smaller than those for SURF.

4.3.3 Run Time Comparison

We evaluated the computational time for both the methods. The video stitching time is the average of the image stitching time of all the frame pairs from the two input videos.

In order to make comparisons, we implemented the video stitching for two rectified videos at a medium resolution of 640×480 . The program was executed in two hardware systems, one for the evaluation using the CPU only and the other using the same CPU in combination with an additional GPU (CPU + GPU).

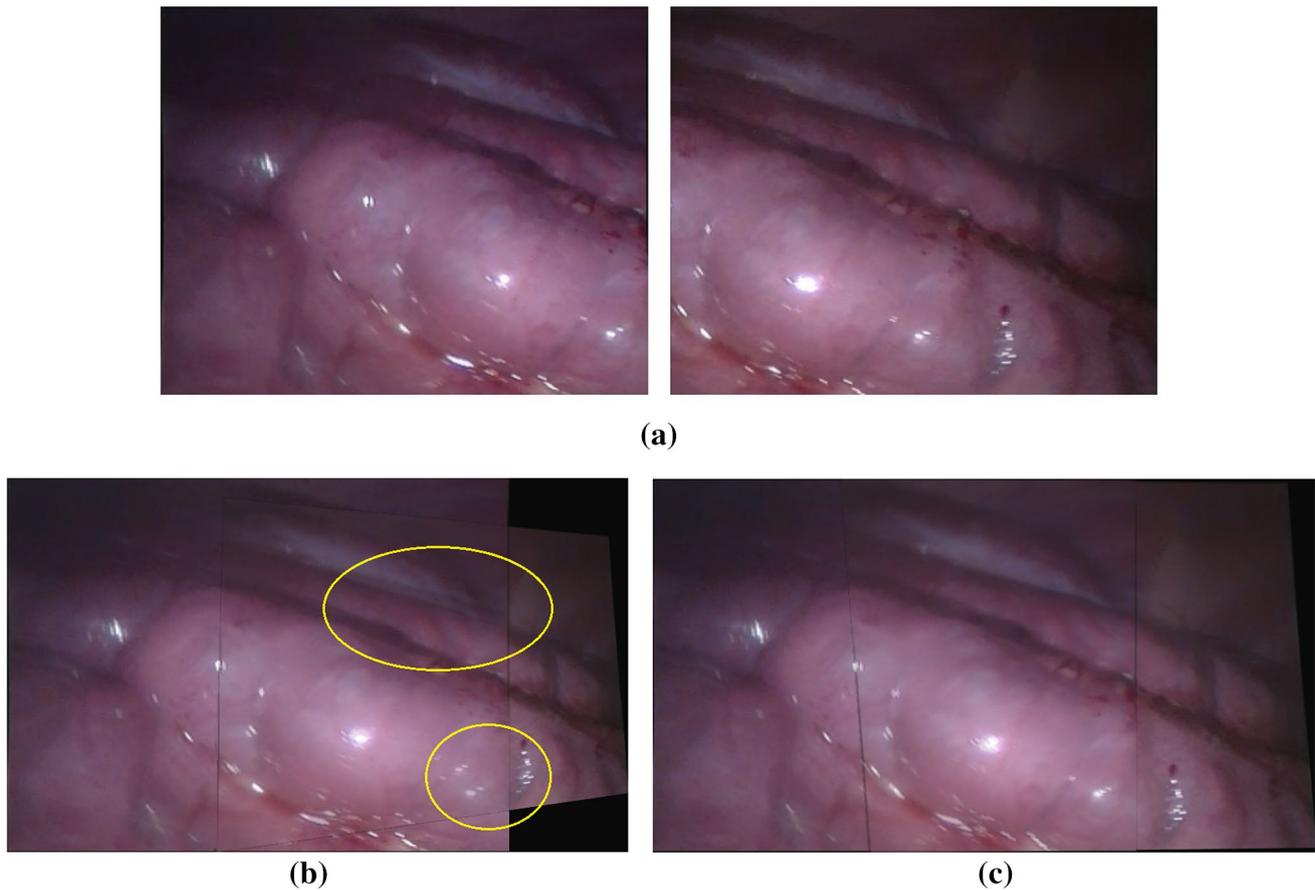


Fig. 11 Qualitative comparison on sample-2. **a** Input images, **b** stitched image by SURF method and **c** stitched image by our method. Yellow circle highlight errors

Table 3 Alignment error of SURF and the proposed method on the same dataset

Alignment error	Sample 1	Sample 2
SURF [13]	6.31	7.26
Ours	5.16	5.45

Table 4 The computational time of SURF, the proposed method, and the whole system

Time (ms)	Sample 1		Sample 2	
	CPU	CPU+GPU	CPU	CPU+GPU
SURF stitching method [13]	125	74	122	70
Proposed stitching method	87	53	83	52
Our system	90	57	86	56

Table 4 shows the stitching time for both the stitching method as well as the processing time for the entire system using the two datasets. On average, the SURF method took 123.5 ms on CPU and 72 ms on CPU + GPU, while the

proposed stitching method took 85 ms on CPU and 52.5 ms on CPU + GPU.

Therefore, our method was 1.45 times faster than the SURF method on the computer where only a CPU was used and 1.38 times faster on a computer where a CPU was used in combination with an additional GPU.

Thus, these results of the evaluation have proven that the proposed method would improve both the quality and speed of the video stitching process when compared to the SURF-based approach.

Furthermore, the rectification process was performed offline, and the execution time for reconstructing the 3D images after the disparity calculation was quite short. These processes would not degrade significantly on the entire performance of our system. Hence, the performance of the whole system execution was considerably high for the practical MIS situations. The results at the last row in Table 4 shows that, on average, our system took 88 ms on CPU and 56.5 ms on CPU + GPU, which indicates that the frame rates of the entire system was about 11.3 fps for a PC where only a CPU was used and 17.6 fps for a PC where the same CPU was used in combination with an additional GPU.

4.4 Discussion

This study presented the proposed endoscope system with the results described in the section above. These results confirm that our system promises to improve the existing limitations in contemporary laparoscopic surgery such as the limited FOV and the lack of depth perception.

However, this study has certain limitations of its own, which are as follows: first, both the sizes of the 3D image and the stitched image depend on the overlap's percentage of the two cameras. For example, as the two cameras move closer to the operating area, the overlap's rate becomes smaller. In this case, our system will increase the camera's FOV with a higher expansion rate, while the 3D image of the overlap area will show less information. Although this is a limitation, when the camera is close to the operating area, the expansion of the camera's FOV is more necessary. Furthermore, in cases where the distance from the cameras to the operating area is less than 2 cm, there may not be an overlap between the two cameras. Therefore, the proposed

algorithm cannot be implemented. Figure 12 shows an example of our endoscope, which was placed near the surgical area with a distance of 2 cm. Moreover, as Fig. 12a shows, the overlap percentage was about 12%. In such a case, our method can expand the FOV of the input image by 188%, while the SURF method cannot be completed because through it no matching feature pairs are correctly detected.

Second, the accuracy of the output results depends on the quality of the disparity map. Furthermore, the accuracy of the disparity calculation depends not only on the stereo matching algorithm used but also on the accuracy of the rectification process. In our experiments, the rectification process performed well when the distance from our endoscope to the operating area was within 3–15 cm. This distance range is suitable during MIS because cameras cannot be placed too close or too far from the surgical area. For the stereo matching algorithm, we only used the stereoBM algorithm in combination with recent edge-aware filters available in OpenCV for calculating the disparity owing to the system's real-time requirement. Besides, a new proposal to

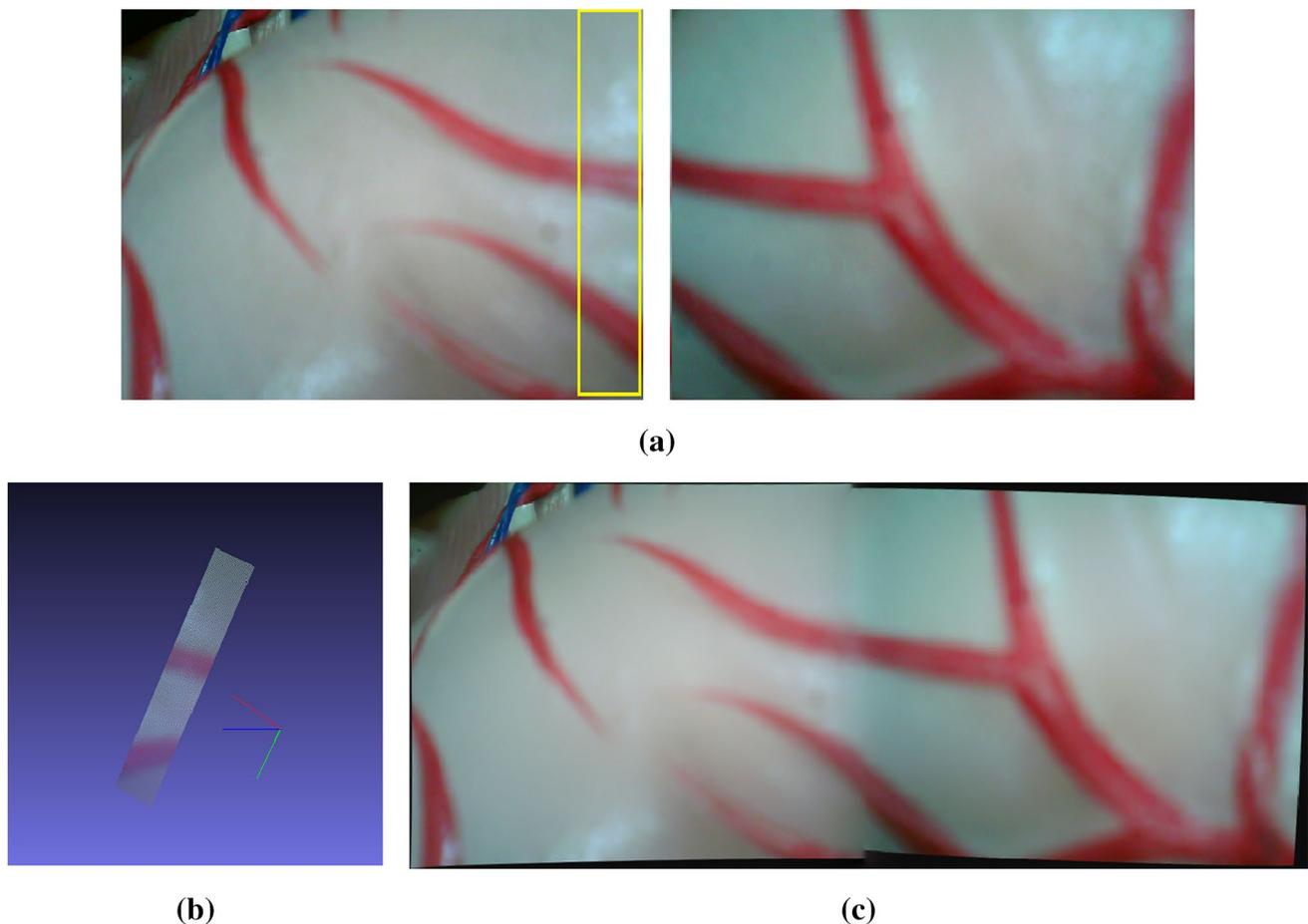


Fig. 12 Our endoscope is located about 2 (cm) from the surgical area. **a** Two input images with an overlapped area (yellow), **b** 3D image of the overlapped area shows less information, while **c** the stitched image can expand the FOV of input images by up to 188%

improve the quality of disparity maps and comparisons with state of the art need to be investigated in the future.

5 Conclusion

In this study, we proposed a New Endoscope for Panoramic-View with Focus-Area 3D-Vision (3DMISPE) to provide the surgeons with broad areas of view and the 3D surface image of the surgical field while ensuring real-time execution. The experimental results showed that 3DMISPE could combine two camera's FOV into one larger FOV. Moreover, the overlap area of the two cameras was also displayed in the 3D space with sufficiently good quality. Moreover, our system could also be a 3D measurement tool for endoscopic surgery. When the distance from the camera to the operating area was about 3 cm to 18 cm, the proposed system produced a distance measurement result with an error of about 1 mm. Furthermore, our system's frame rate for two endoscopic cameras at a resolution of 640×480 was determined to be 17.6 fps, which was good enough for practical uses for the MIS.

We have proposed a novel algorithm for the stitching video. The proposed stitching algorithm was based on the stereo vision theory and, thus, also supports 3D reconstruction. The experimental results obtained showed that our method could accelerate 1.4 times faster than the SURF-based stitching method. Furthermore, our approach also improved the stitched image quality by reducing alignment errors or "ghosting" when compared with the SURF method.

In the future, we plan to further improve the performance of our system in terms of both quality and speed. Further, we intend to develop an object-tracking module that is based on deep learning so as to perfect the current system.

Acknowledgements The authors would like to express our special thanks for the research Grant supported by the Ministry of Science and Technology (MOST) of Taiwan, Rep. of China (ROC) under the contact number of 106-2221-E-035 -095.

Compliance with Ethical Standards

Conflict of interest The authors declare that they have no conflict of interest.

Ethical Approval This article does not contain any studies with human participants or animals performed by any of the authors.

Informed Consent This article does not contain patient data.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing,

adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. Sinha, R. Y., Raje, S. R., & Rao, G. A. (2017). Three-dimensional laparoscopy: Principles and practice. *Journal of Minimal Access Surgery*, 13(3), 165–169.
2. Huang, X., Abdalbari, A., & Ren, J. (2013). 3D surface reconstruction of stereo endoscopic images for minimally invasive surgery. *Biomedical Engineering Letters*, 3(3), 149–157.
3. Lurie, K. L., Angst, R., Zlatev, D. V., Liao, J. C., & Ellerbee Bowden, A. K. (2017). 3D reconstruction of cystoscopy videos for comprehensive bladder records. *Biomedical Optics Express*, 8(4), 2106–2123.
4. Penza, V., Ortiz, J., Mattos, L. S., Forgiione, A., & De Momi, E. (2016). Dense soft tissue 3D reconstruction refined with superpixel segmentation for robotic abdominal surgery. *International Journal of Computer Assisted Radiology and Surgery*, 11(2), 197–206.
5. Schoob, A., Kundrat, D., Kahrs, L. A., & Ortmaier, T. (2016). Comparative study on surface reconstruction accuracy of stereo imaging devices for microsurgery. *International Journal of Computer Assisted Radiology and Surgery*, 11(1), 145–156.
6. Wang, C., Cheikh, F. A., Kaaniche, M., & Elle, O. J. (2018) Liver surface reconstruction for image guided surgery. In *Medical imaging 2018: image-guided procedures, robotic interventions, and modeling* (p. 105762H). International Society for Optics and Photonics.
7. De Zanet, S., Rudolph, T., Richa, R., Tappeiner, C., & Sznitman, R. (2016). Retinal slit lamp video mosaicking. *International Journal of Computer Assisted Radiology and Surgery*, 11(6), 1035–1041.
8. Liu, J., Wang, B., Hu, W., Sun, P., Li, J., Duan, H., et al. (2015). Global and local panoramic views for gastroscopy: An assisted method of gastroscopic lesion surveillance. *IEEE Transactions on Biomedical Engineering*, 62(9), 2296–2307.
9. Soper, T. D., Porter, M. P., & Seibel, E. J. (2012). Surface mosaics of the bladder reconstructed from endoscopic video for automated surveillance. *IEEE Transactions on Biomedical Engineering*, 59(6), 1670–1680.
10. Takada, C., Suzuki, T., Afifi, A., & Nakaguchi, T. (2017). *Hybrid tracking and matching algorithm for mosaicking multiple surgical views* (pp. 24–35). Cham: Springer.
11. Totz, J., Fujii, K., Mountney, P., & Yang, G.-Z. (2012). Enhanced visualisation for minimally invasive surgery. *International Journal of Computer Assisted Radiology and Surgery*, 7(3), 423–432.
12. Kim, D. T., & Cheng, C. H. (2016) A panoramic stitching vision performance improvement technique for Minimally Invasive Surgery. In *2016 5th International Symposium on Next-Generation Electronics (ISNE)* (pp. 1–2).
13. Kim, D. T., Nguyen, V. T., Cheng, C.-H., Liu, D.-G., Liu, K. C. J., & Huang, K. C. J. (2018). Speed improvement in image stitching for panoramic dynamic images during minimally invasive surgery. *Journal of Healthcare Engineering*, 2018, 1–14.

14. Peng, C.-H., & Cheng, C.-H. (2014). A panoramic wireless endoscope system design for the application of minimally invasive surgery. *Smart Science*, 2(2), 91–95.
15. Szeliski, R. (2006). Image alignment and stitching: A tutorial. *Foundations and Trends in Computer Graphics and Vision*, 2(1), 1–104.
16. Cattin, P. C., Bay, H., Van Gool, L., & Székely, G. (2006). Retina mosaicing using local features. *International conference on medical image computing and computer-assisted intervention* (pp. 185–192). Berlin: Springer.
17. Prokopetc, K., & Bartoli, A. (2017). SLIM (slit lamp image mosaicing): Handling reflection artifacts. *International Journal of Computer Assisted Radiology and Surgery*, 12(6), 911–920.
18. Bay, H., Ess, A., Tuytelaars, T., & Van Gool, L. (2008). Speeded-up robust features (SURF). *Computer Vision and Image Understanding*, 110(3), 346–359.
19. Behrens, A., Bommers, M., Stehle, T., Gross, S., Leonhardt, S., & Aach, T. (2011). Real-time image composition of bladder mosaics in fluorescence endoscopy. *Computer Science-Research and Development*, 26(1), 51–64.
20. Yang, L., Wang, J., Ando, T., Kubota, A., Yamashita, H., Sakuma, I., et al. (2016). Towards scene adaptive image correspondence for placental vasculature mosaic in computer assisted fetoscopic procedures. *The International Journal of Medical Robotics and Computer Assisted Surgery*, 12(3), 375–386.
21. Hu, W., Zhang, X., Wang, B., Liu, J., Duan, H., Dai, N., et al. (2016). Homographic patch feature transform: A robustness registration for gastroscopic surgery. *PLoS ONE*, 11(4), e0153202.
22. Ali, S., Daul, C., Galbrun, E., Amouroux, M., Guillemin, F., & Blondel, W. (2015). Robust bladder image registration by redefining data-term in total variational approach. In *Medical imaging 2015: image processing* (p. 94131H). International Society for Optics and Photonics.
23. Hamzah, R. A., & Ibrahim, H. (2016). Literature survey on stereo vision disparity map algorithms. *Journal of Sensors*, 2016, 1–23.
24. Scharstein, D., & Szeliski, R. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1), 7–42.
25. Maier-Hein, L., Mountney, P., Bartoli, A., Elhawary, H., Elson, D., Groch, A., et al. (2013). Optical techniques for 3D surface reconstruction in computer-assisted laparoscopic surgery. *Medical Image Analysis*, 17(8), 974–996.
26. Malti, A. (2014). Variational formulation of the template-based quasi-conformal shape-from-motion from laparoscopic images. *International Journal of Advanced Computer Science and Applications (IJACSA)*. <https://doi.org/10.14569/IJACSA.2014.050323>.
27. Chen, L., Tang, W., John, N. W., Wan, T. R., & Zhang, J. J. (2018). 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.
28. Totz, J., Mountney, P., Stoyanov, D., & Yang, G.-Z. (2011). Dense surface reconstruction for enhanced navigation in MIS. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 4, 89–96.
29. Collins, T., & Bartoli, A. (2012). Towards live monocular 3D laparoscopy using shading and specular information. In P. Abolmaesumi, L. Joskowicz, N. Navab, & P. Jannin (Eds.), *Information processing in computer-assisted interventions* (pp. 11–21). Berlin: Springer.
30. Stoyanov, D., Scarzanella, M. V., Pratt, P., & Yang, G.-Z. (2010). Real-time stereo reconstruction in robotically assisted minimally invasive surgery. In T. Jiang, N. Navab, J. P. W. Pluim, & M. A. Viergever (Eds.), *Medical image computing and computer-assisted intervention* (pp. 275–282). Berlin: Springer.
31. Bernhardt, S., Abi-Nahed, J., & Abugharbieh, R. (2012). Robust dense endoscopic stereo reconstruction for minimally invasive surgery. In B. H. Menze, G. Langs, L. Lu, A. Montillo, Z. Tu, & A. Criminisi (Eds.), *International MICCAI workshop on medical computer vision* (pp. 254–262). Berlin: Springer.
32. Rohl, S., Bodenstedt, S., Suwelack, S., Dillmann, R., Speidel, S., Kenngott, H., et al. (2012). Dense GPU-enhanced surface reconstruction from stereo endoscopic images for intraoperative registration. *Medical Physics*, 39(3), 1632–1645.
33. Münzer, B., Schoeffmann, K., & Böszörményi, L. (2018). Content-based processing and analysis of endoscopic images and videos: A survey. *Multimedia Tools and Applications*, 77(1), 1323–1362.
34. Kaehler, A., & Bradski, G. (2016). *Learning OpenCV 3: Computer vision in C++ with the OpenCV library*. Sebastopol: O'Reilly Media Inc.
35. Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11), 1330–1334.
36. Konolige, K. (1998). Small vision systems: Hardware and implementation. In Y. Shirai & S. Hirose (Eds.), *Robotics research* (pp. 203–212). London: Springer.
37. Min, D., Choi, S., Lu, J., Ham, B., Sohn, K., & Do, M. N. (2014). Fast global image smoothing based on weighted least squares. *IEEE Transactions on Image Processing*, 23(12), 5638–5653.
38. Barron, J. T., & Poole, B. (2016). The fast bilateral solver. In B. Leibe, J. Matas, N. Sebe, & M. Welling (Eds.), *European conference on computer vision* (pp. 617–632). Cham: Springer.
39. Fischler, M. A., & Bolles, R. C. (1981). Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6), 381–395.
40. Kwatra, V., Schödl, A., Essa, I., Turk, G., & Bobick, A. (2003). Graphcut textures: Image and video synthesis using graph cuts. *ACM Transactions on Graphics*, 22(3), 277–286.
41. Brown, M., & Lowe, D. G. (2007). Automatic panoramic image stitching using invariant features. *International Journal of Computer Vision*, 74(1), 59–73.