Hybrid Tracking and Matching Algorithm for Mosaicking Multiple Surgical Views

Chisato Takada[1(✉)], Toshiyuki Suzuki[2], Ahmed Afifi[3], and Toshiya Nakaguchi[4]

[1] School of Engineering, Chiba University, Chiba, Japan
takadachisato@chiba-u.jp
[2] Graduate School of Engineering, Chiba University, Chiba, Japan
[3] Faculty of Computers and Information, Menoufia University, Shebin El-kom, Menoufia, Egypt
[4] Research Center for Frontier Medical Engineering, Chiba University, Chiba, Japan

Abstract. In recent years, laparoscopic surgery has become major surgery due to several advantages for patients. However, it has disadvantages for operators because of the narrow surgical field of view. To solve this problem, our group proposed camera-retractable trocar which can obtain multiple surgical viewpoints while maintaining the minimally invasiveness. The purpose of this study is to obtain a wide visual panoramic view by utilizing image mosaicking of camera-retractable trocar viewpoints videos. We utilize feature points tracking in different videos to generate panoramic video independent of inter-cameras overlap and to increase mosaicking speed and robustness. We evaluate tracking accuracy according to several conditions and mosaicking accuracy according to overlap size. In contrast to the conventional mosaicking approach, the proposed approach can produce panoramic image even in the case of 0% inter-cameras overlap. Additionally, the proposed approach is fast enough for clinical use.

1 Introduction

Laparoscopic surgery, one of the minimally invasive surgeries (MIS), has several advantages for patients. For example, patients would feel less postoperative pain because of the small surgical wound, can early discharge and can early return to their social activities. However, this surgery has disadvantages for operators because of the narrow surgical field of view. Also, the safety improvement of this surgery is strongly required owing to concern for the medical accidents occurred in recent years. As one of the countermeasures for this problem, it has been demanded to realize a wide visual field such as abdominal surgery maintaining the minimally invasiveness that is an advantage of laparoscopic surgery. In the case of endoscopic surgery or robotic assisted surgery, image mosaicking and image mapping are proposed to achieve a wide visual field [1, 2].

In the present laparoscopic surgery, operators insert a laparoscope into a port and display a single viewpoint video on a monitor. It is a major operative procedure. In recent years, several mosaicking methods are proposed to expand the surgical view. These methods usually use monocular tracking [3–6], or stereo imaging devices [7]. However, these methods extend the field of view using static panorama images and do not provide a real extended view of the operation site. To achieve a real wider visual field, we must

observe intraperitoneal conditions from new ports other than the laparoscope port. Camera-retractable trocar is proposed by Okubo et al. [8] to invasively provide multiple surgical views. Trocar is a surgical instrument that is inserted through abdominal wall to secure forceps ports and to keep abdominal air pressure. Camera-retractable trocar, shown in Fig. 1(a), (b), has a miniature camera which can be retracted or expanded at the end of the trocar. It is possible to obtain several videos of different viewpoints from the camera-retractable trocar. Therefore, it is possible to obtain multiple surgical viewpoints videos while maintaining the minimally invasiveness which is an advantage of laparoscopic surgery. Although these advantage of camera-retractable trocar, observing multiple views at the same time may cause confusion specially in the case of overlapped views.

(a) retracted   (b) expanded
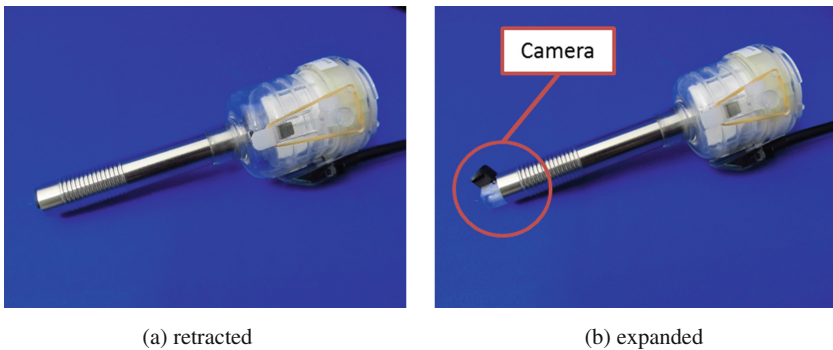
Fig. 1.   Camera-retractable trocar.

In this study, therefore, we tend to utilize camera-retractable trocar views to provide a more realistic expanded surgical view. Supposing the situation that two camera-retractable trocars are placed at two different ports, the purpose is to perform image mosaicking of these viewpoints videos to obtain a wide visual panoramic video of the operation site.

In traditional image mosaicking, an overlap between images is necessary for generating panoramic image. However, in the case of trocar-retractable cameras, an enough large overlap between cameras is not necessarily preserved during the operation because of trocar movement caused by operation of forceps. Therefore, in this work, feature points tracking in different videos is utilized to increase mosaicking speed and robustness. Moreover, by combining mosaicking and tracking, it is possible to generate panoramic video regardless the overlap between different cameras. The speed and efficiency of the proposed approach is evaluated in this study. From this evaluation, we can deduce that using feature tracking reduces the required number of free view point mosaicking. And then, the computational cost of the whole approach is reduced. Moreover, the mosaicking robustness can be improved.

2   Proposed Method

The general diagram of the proposed mosaicking approach using two cameras is shown in Fig. 2.
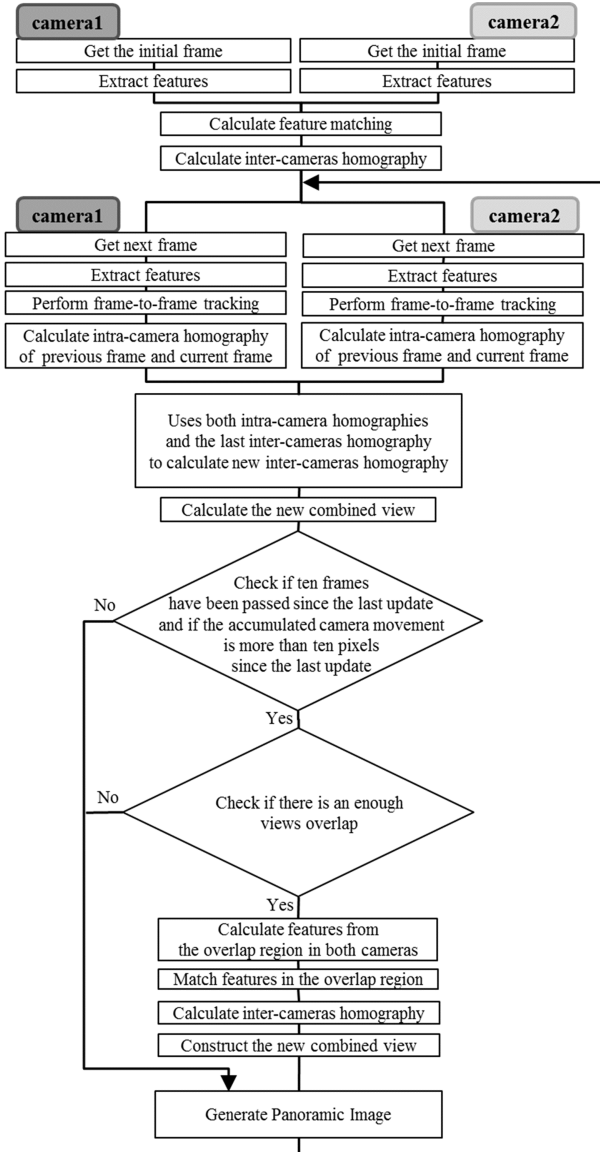

Fig. 2. The general diagram of the proposed mosaicking approach.

At the beginning, an initial panorama image is required. This initial panoramic view is constructed when an enough overlap is found at the exploration time. To construct this view, Speed-Up Robust Feature (SURF) algorithm [9] is utilized to extract feature points from the initial frames acquired from different trocar-retractable cameras. A robust feature matching is then performed by applying a ratio test and double matching from view1 to view2 and vice-verse. Consequently, the inter-cameras homography is calculated from a set of inliers matches found using random sample concise (RANSAC) algorithm [10].

After initialization, continues tracking is performed from frame to frame in each video. In this work, a set of feature points extracted using Good-Feature-to-Track technique [11] are tracked using Pyramidal Lucas-Kanade Optical Flow tracking [12]. These tracked points are utilized to estimate intra-camera homography, which models the relationship of consequent video frames. The current expanded view is then calculated using both intra-camera homographies and the last updated inter-cameras homography. By considering the last updated inter-cameras homography as H_{pano}, the intra-camera homography of the first and second view as H_{view1} and H_{view2}, the current expanded homography view is estimated as in Eq. (1). Figure 3 show an illustration of the estimation process.

$$H_{current} = H_{view1} \times H_{pano} \times H_{view2}^{-1} \tag{1}$$
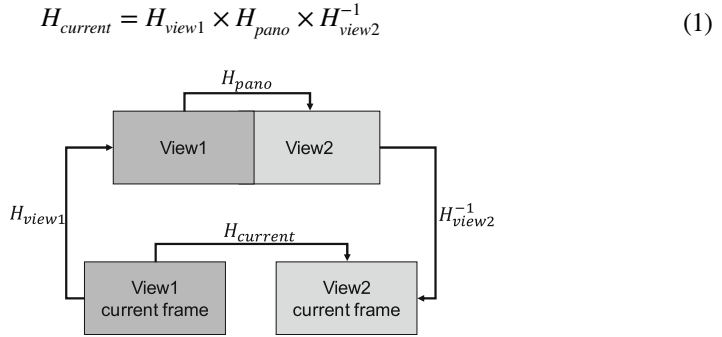
Fig. 3. Estimation of the expanded view.

By using the above mentioned homography estimation methodology, the relationship between different views can be maintained regardless the overlap size. However, the homography error accumulated from frame to frame which cause large estimation error in time. To alleviate this problem and to enhance the overall estimation, an update method is performed if one of the following conditions is satisfied:

(a) The accumulated camera movement is more than 10 pixels since the last update and there is an enough views overlap.
(b) Ten frames have been passed since the last update and there is an enough views overlap.

This update process utilizes the estimated homography to determine the overlap of view1 and view2, and warp the overlap area of view2 to view1 frame. Consequently, the update process is performed using SURF feature points detection and matching.

The matching process is performed locally around the detected points and then the matching time and error are reduced. A correction homography is then calculated from a set

of inliers using RANSAC algorithm, and the final corrected homography is calculated as in Eq. (2).

$$H_{final} = H_{correction} * H_{corrent} \tag{2}$$

where, H_{final} is the corrected current expanded view homography, $H_{correction}$ is the correction homography calculated from the overlap area in view1 and view2 and $H_{current}$ is the initial current view estimated homography. The inter-camera homography is now updated using H_{final} and all update conditions are reset.

In the proposed method, it is possible to generate panoramic image using frame-to-frame feature detection and temporal tracking independent of spatial overlap size, as shown in Fig. 4. Also, if there is an enough large overlap between two cameras, we can obtain more accurate panoramic image with direct mosaicking.
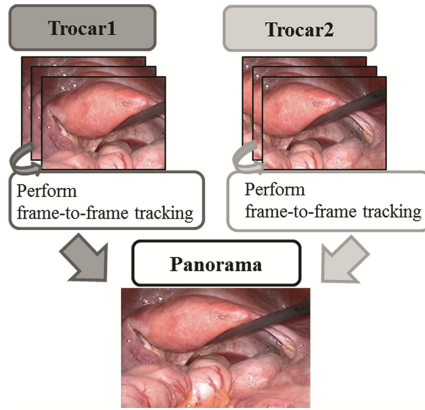

Fig. 4. Panoramic image by temporal tracking.

3   Evaluation Experiments

This section describes the evaluation of the proposed approach. We describe accuracy evaluation of tracking according to camera types and imaging conditions in Sect. 3.1, accuracy evaluation of mosaicking according to overlap size between two cameras in Sect. 3.2. Finally, the comparison of the proposed approach and the conventional mosaicking approach in provided in Sect. 3.3.

All experiments in this study were performed using OpenCV toolkit [13] on a PC with the following specifications; OS is Windows8.1 professional 64 bit, CPU is Intel® Core™ i7-2600 K, RAM is 8 GB, and GPU is NVIDEA GeForce GTX 560 Ti. Moreover, GPU-based features extraction, matching, tracking and image warping were utilized to accelerate the process.

3.1 Tracking Accuracy Evaluation According to Camera Types and Imaging Conditions

In this work, *in vivo* and *in vitro* videos are used to asses feature points tracking accuracy. The trocar-retractable camera is used to capture intra-operational videos of organs which have smooth and specular surfaces. Additionally, a blurring effect may happen during the operation.

3.1.1 Experimental Setup

In this experiment, we use three videos as shown in Fig. 5. These videos are captured at 30 fps for 10 s with a total number of 300 frames. The shelves video shown in Fig. 5(a) is captured by RGB camera (Lumenera Lu170C) which has a resolution of 640×480. The intra-operational video of a pig abdomen shown in Fig. 5(b) is captured by trocar camera which has a resolution of 640×480, and the intra-operational video shown in Fig. 5(c) is captured with the same trocar camera when blurring and turbulence occurs.

(a)The shelves video   (b)The pig abdomen   (c) The blurred video
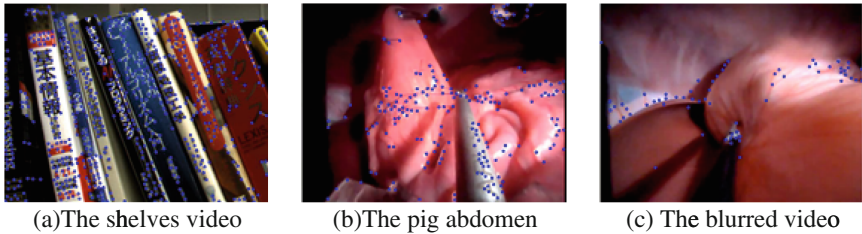
Fig. 5. Videos used for tracking accuracy evaluation, blue dots represent the tracked features. (Color figure online)

3.1.2 Results and Discussion

The tracking methodology of the proposed approach is applied to the videos described in the previous section and the results are evaluated. Figure 6 shows the result of feature tracking for all videos. For the video shown in Fig. 5(a), large number of feature points, more than 400, can be always tracked. In comparison with this video, in the intra-operational trocar videos, smaller number of features can be tracked specially in the case of blurred video shown in Fig. 5(c). In this video, the number of tracked features becomes almost 0 when high blurring effect occurs. This fluctuation is caused by noises of video under the influence of using surgical diathermy.

In the proposed approach, intra-camera homography can be calculated if the number of tracked feature points more than seven. Accordingly, we can perform tracking and calculate intra-camera homography in all three videos. However, more accurate intra-camera homography can be calculated when the number of tracked feature points is as large as possible. Therefore, we must examine the feature detection and tracking method for *in vivo* videos in more details.
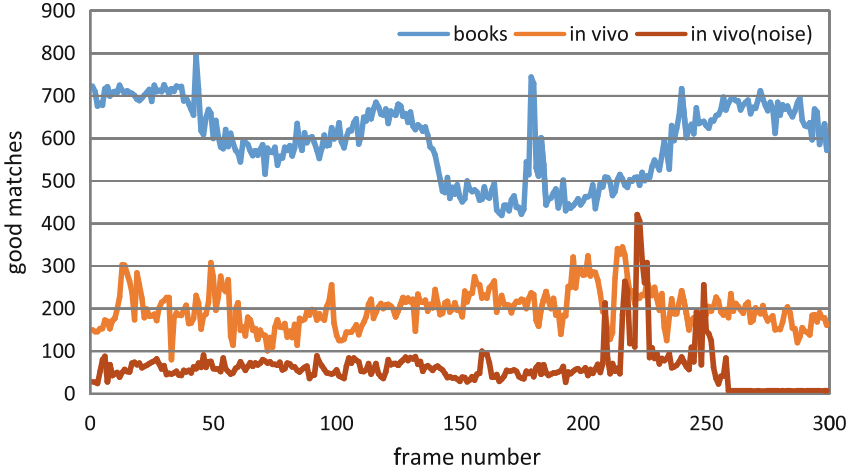
Fig. 6. Evaluation of the number of tracked features in different video frames.

## 3.2 Mosaicking Accuracy Evaluation According to Overlap Size Between Two Cameras

The proposed approach can maintain the expanded view regardless the overlap size. However, the mosaicking accuracy may be affected by the views overlap size because the update process is affected by overlap size. Therefore, we created test videos of different percentages of overlap range from 20% to 90% of frame size at an interval of 10%. Then, we implement the proposed approach and compare the results using these videos.

3.2.1 Experimental Setup

To create videos for this evaluation, we cut out two 640×480 rectangles from high resolution video captured by the "Stryker 1188 HD" monocular laparoscope, the resolution of which is 1280×720. These rectangles are considered as the viewpoint of camera-1(V_1) and the viewpoint camera-2(V_2), as shown in Fig. 7. The video captured by laparoscope mainly shows serosa of pig stomach. To change the overlap size as a percentage of the whole frame size, we translate V_1 and V_2 in a parallel direction and create eight types of videos, as shown in Table 1.

Table 1. Comparison of mosaicking accuracy for different views overlap size.

Overlap(%)	20	30	40	50	60	70	80	90
Overlap size (# of pixels)	128×480	198×480	256×480	320×480	384×480	448×480	512×480	576×480
Ideal values (X_r, Y_r)	(512, 0)	(448, 0)	(384, 0)	(320, 0)	(256, 0)	(192, 0)	(128, 0)	(64, 0)
Measured values(X_c, Y_c)	(569.9, 16.2)	(483.6, 21.9)	(383.6, −0.4)	(319.6, −0.3)	(255.6, −0.3)	(191.6, −0.4)	(127.6, −0.4)	(63.6, −0.5)
error(pixel)	69.2	49.9	4.5	3.9	3.2	2.5	1.8	1.7
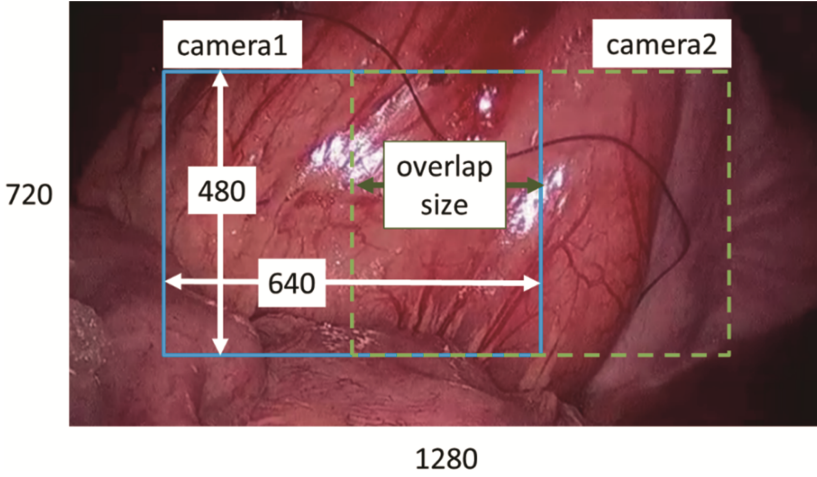
Fig. 7. Evaluation of the number of tracked features in different video frames.

To quantify the accuracy of the estimated panoramic image, we set the central coordinate of V_1 as a relative central position and calculate a relative position of the central coordinate of V_2, regarded as (x, y). We use the error between ideal values (x_r, y_r) and measured values (x_c, y_c) as Euclidean distances for accuracy evaluation as in Eq (3)

$$error = \sqrt{(x_c - x_r)^2 + (y_c - y_r)^2} \tag{3}$$

3.2.2 Results and Discussion

As noticed from Table 1, the expanded view can be obtained in all cases. However, in the cases of 20% and 30% percentage of overlap, the panoramic image is generated by tracking only and no update is performed. Accordingly, the error is accumulated from frame to frame and the mosaicking accuracy is degraded. In all other cases, when the overlap size is enough for update process, a very good mosaicking accuracy is achieved. Therefore, we deduce that we need to improve the tracking accuracy in order to further improve the mosaicking accuracy specially in the case of small overlap size.

Figure 8(a), (b) shows the viewpoints of camera-1(V_1) and camera-1(V_1) when the overlap size is 40%, and Fig. 8(c) shows the result of mosaicking. We can get very accurate panoramic image. On the other hand, Fig. 8(d) shows a case when errors occur.

Figure 9 shows the error measured for every frame in each video in the interval of 40% to 90% of overlap size. As can be noticed from this figure, the error accumulates between the update process and it is greatly reduced when the update is performed. It is also noticed that, the larger overlap percentages produces higher mosaicking accuracy. The results of this experiment, shows the importance of update process in reducing the accumulated error. However, as the tracking is an important component of the proposed approach, we must analyze of the tracking errors deeply and try other feature detectors in order to improve its accuracy. In this experiment, we only consider the parallel translation. Thus, we have to consider the rotation movement and its accuracy.

(a) Camera-1(V_1)   (b)Camera-2(V_2)

(c) Result of mosaicking
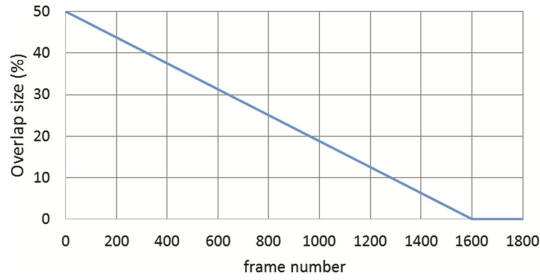

(d) Result of mosaicking (error)
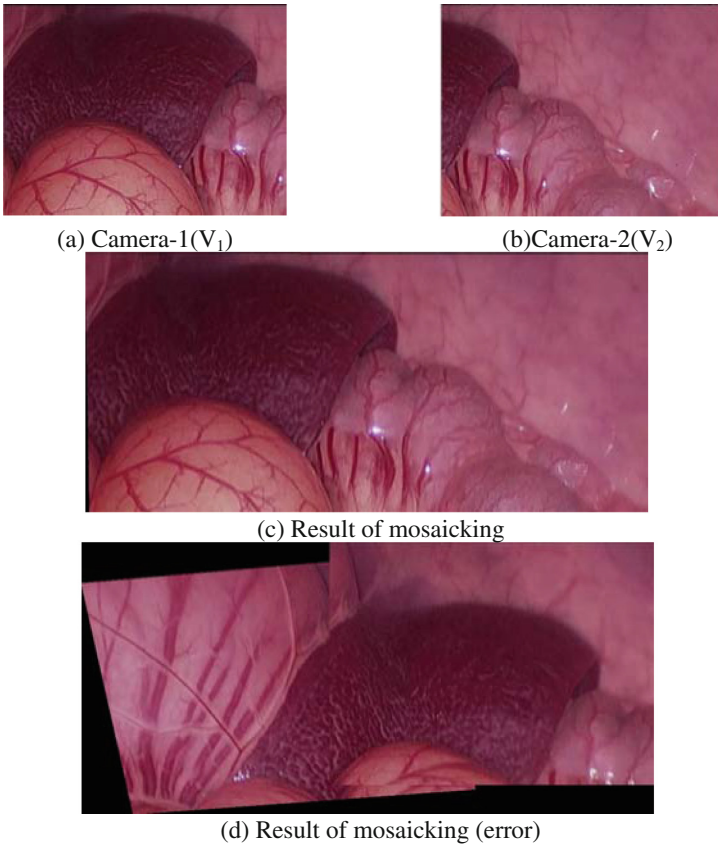
Fig. 8. Result of mosaicking (overlap size: 40%)
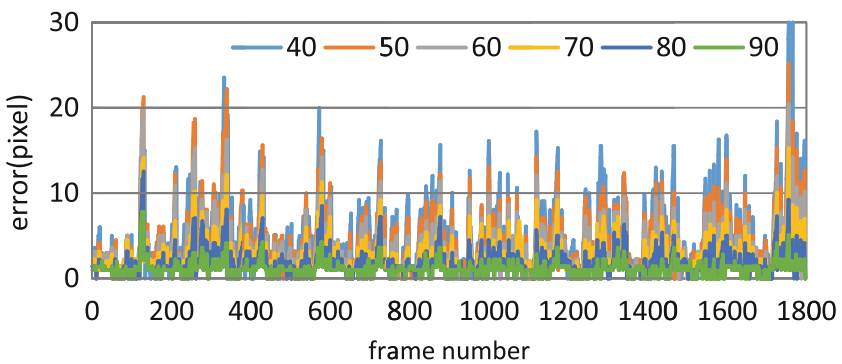
Fig. 9. Comparison of mosaicking accuracy at each video frame.

Fig. 10. Changes of the overlap size over video frames.

In addition to the tracking accuracy, physiological motion, forceps motion and tissue deformation would affect the result of mosaicking. We must distinguish these movements from the camera motion in the future.

3.3   Comparison of the Proposed Approach and the Conventional Approach

In conventional image mosaicking approach, an overlap between images is necessary for generating panoramic image. On the other hand, the proposed approach can utilize both tracking and direct mosaicking to generate panoramic image independent of presence of overlap between cameras. To check the efficacy of the proposed approach, we create a video in which the overlap size becomes smaller over time and we perform the comparative evaluation.

3.3.1   Experimental Setup

Similar to 3.3, to create videos for evaluation, we cut out two 640×480 rectangles from a high resolution video captured by the "Stryker 1188 HD" laparoscope. In this experiment, we do not fix the percentages of overlap size; however, we translate V_1 and V_2 while reducing the percentages of overlap, as shown is Fig. 10. The percentages of the overlap size of the first frame is set to 50%, we reduce the percentages at a regular speed until frame number 1600. At the frame number 1600, the percentage of overlap between V_1 and V_2 becomes 0% and we fix V_1 and V_2 until frame number 1800.

We implement the proposed approach and the conventional approach using these videos and the mosaicking accuracy and processing speed are evaluated. We use the error between ideal values (x_r, y_r) and measured values (x_c, y_c) as Euclidean distances for mosaicking accuracy evaluation as in Eq (3).

3.3.2   Results and Discussion

Figure 11 shows the evaluation results of the proposed approach and the conventional approach over time. As can be noticed from this figure, the mosaicking error of the conventional approach is low when an enough overlap is found. However, it becomes unstable from about frame number 1200, and completely stopped at frame number 1400 because of lack of an enough overlap size. On the other hand, the proposed approach can continue the process after frame number 1600 of which percentage of overlap
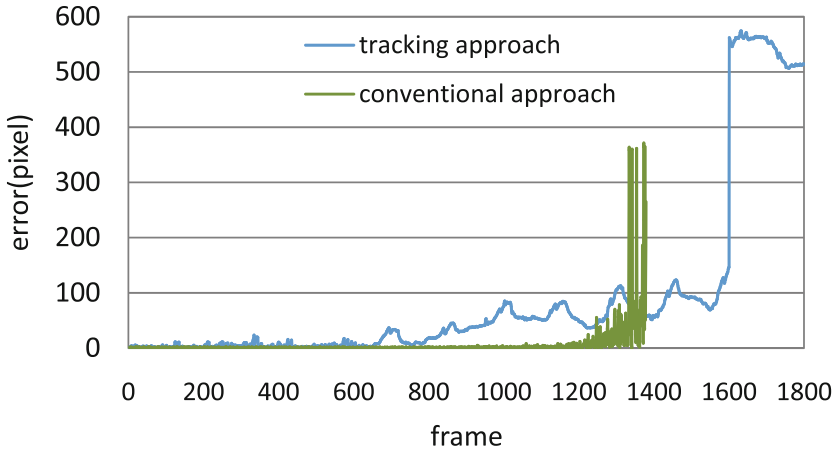
Fig. 11. Comparison proposed approach with conventional approach.

becomes 0%. The error increase from about frame number 800 because of the accumulated error of tracking. Additionally, the proposed approach achieves a frame rate of 18.7 fps while the conventional approach run at 10.1 fps.

From these experiments we deduce that, the proposed approach can provide the expanded view even in the case of 0% overlap, and we can also obtain the advantage in terms of the processing speed.

4   Conclusion

In this work, an approach for abdominal view expansion is proposed. This approach can utilize multiple trocar-retractable camera, image mosaicking and tracking. In contrast to the traditional mosaicking approach, the proposed approach can produce panoramic image even in the case of 0% inter-cameras overlap. Additionally, the proposed approach is about 9 frames per second faster than the conventional approach. The evaluation performed in this work shows that it is difficult to detect the adequate amount of features from trocar camera at the moment; however, the trocar camera is under active development and will be enhanced in the future. Moreover, we found that the overlap size affects the final mosaicking accuracy in the proposed approach. This limitation is mainly caused by the tracking accuracy and we tend to improve the tracking algorithm in the future. In this paper, we used videos created from the laparoscope video; however, we will examine the results using the actual trocar videos in the future.

References

1. Miranda-Luna, R., et al.: Mosaicing of bladder endoscopic image sequences: distortion calibration and registration algorithm. IEEE Trans. Biomed. Eng. **55**(2), 541–553 (2008)
2. Warren, A., et al.: Horizon stabilized—dynamic view expansion for robotic assisted surgery (HS-DVE). Int. J. Comput. Assist. Radiol. and Surg. **7**, 281–288 (2012)

3. Lerotic, M., Chung, A.J., Clark, J., Valibeik, S., Yang, G.-Z.: Dynamic view expansion for enhanced navigation in natural orifice transluminal endoscopic surgery. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 467–475. Springer, Heidelberg (2008)
4. Vemuri, A.S., Liu, K., Ho, Y., Wu, H.: Endoscopic video mosaicing: application to surgery and diagnostics. In: Living Imaging Workshop, 1–2 December, IRCAD, Strasbourg, France (2011)
5. Bergen, T., Ruthotto, S., Munzenmayer, C., Rupp, S., Paulus, D., Winter, C.: Feature-based real-time endoscopic mosaicking. In: Proceeding of 6th International Symposium on Image and Signal Processing and Analysis, pp. 695–700 (2009)
6. Bergen, T., Wittenberg, T.: Stitching and surface reconstruction from endoscopic image sequences: a review of applications and methods. IEEE J. Biomed. Health Inform. **20**, 304–321 (2014)
7. Tamadazte, B., Agustinos, A., Cinquin, P., Fiard, G., Voros, S.: Multi-view vision system for laparoscopy surgery. J. Comput. Assist. Radiol. Surg. **10**, 195–203 (2015)
8. Okubo, T., Nakaguchi, T., Hayashi, H., Tsumura, T.: Abdominal view expansion by retractable camera. J. Signal Process. **15**, 311–314 (2011)
9. Bay, H., Tuytelaars, T., Gool, L.: SURF: speeded up robust features. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3951, pp. 404–417. Springer, Heidelberg (2006). doi:10.1007/11744023_32
10. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM **24**, 381–395 (1981)
11. Shi, J., Tomasi, C.: Good features to track. In: 9th IEEE Conference on Computer Vision and Pattern Recognition, pp. 593–600 (1994)
12. Bouguet, J.Y.: Pyramidal implementation of the Lucas Kanade feature tracker: description of the algorithm, Intel Corporation Microprocessor Research Labs (2000)
13. OpenCV 2.4.13.0 documentation (2016). http://docs.opencv.org/2.4/index.html# Accessed 20 June 2016