

RESEARCH ARTICLE

Open Access



An image processing method for changing endoscope direction based on pupil movement

Yang Cao^{1*}, Satoshi Miura¹, Quanquan Liu², Yo Kobayashi¹, Kazuya Kawamura³, Shigeki Sugano¹ and Masakatsu G. Fujie¹

Abstract

Increased attention has been focused on laparoscopic surgery because of its minimal invasiveness and improved cosmetic properties. However, the procedure of laparoscopic surgery is considerably difficult for surgeons, thus paving the way for the introduction of robotic technology to reduce the surgeon's burden. Thus, we have developed a single-port surgery assistive robot with a master–slave structure that has two surgical manipulators and a sheath manipulator for the alteration of endoscope direction. During the development of the surgical robotic system, achieving intuitive operation is very important. In this paper, we propose a new laparoscope manipulator control system based on the movement of the pupils to enhance intuitive operability. We achieve this using a webcam and an image processing method. After the pupil movement data are obtained, the master computer transforms these data into an output signal, and then the slave computer receives and uses that signal to drive the robot. The details of the system and the pupil detection procedure are explained. The aim of the present experiment is to verify the effectiveness of the image processing method applied to the alteration of endoscope direction control system. For this purpose, we need to determine an appropriate pupil motion activation threshold to begin the sheath manipulator's movement. We used four kinds of activation threshold, measuring the time cost of a particular operation: to move the image of the endoscope to a specific target position. Moreover, we identified an appropriate activation threshold that can be used to determine whether the endoscope is moving.

Keywords: Non-rigid face tracking, Single-port endoscopic surgery, Master–slave structure, Image processing, Double-screw-drive mechanism, Activation threshold

Background

Laparoscopic surgery is a technique whereby a laparoscope and surgical instruments are inserted into the patient's body through an artificial or natural body cavity, followed by the surgeon operating the instruments based on the monitor image captured by the laparoscope [1]. Laparoscopic surgery has many advantages, such as shorter hospitalization times, lower physical burden on patients and cosmetic improvement [2] compared with open surgery. Although laparoscopic surgery has many

advantages for patients as described above, the difficulty of performing the technique is so high that surgeons need to carry out long-term training at a special medical training center. Even if they have experienced professional training, they still suffer from high mental stress during operations, which may reduce their dexterity or judgment [3]. One of the causes of such problems is that laparoscopic surgery requires another operator to hold the laparoscope for the surgeon. Therefore, the coordination between operators has a significant effect on the process and result of laparoscopic surgery [4].

To tackle the problems mentioned above, robotic technology is an effective solution. Naviot [5] provides surgeons with the possibility of solo surgery, and requires the surgeon to use one hand to hold a controller when

*Correspondence: a15caoyang@fuji.waseda.jp

¹ Graduate School of Creative Science and Engineering, Waseda University, 3-4-1 Ohkubo, Shinjuku-ku, Tokyo 169-8555, Japan
Full list of author information is available at the end of the article

they need to alter the direction of the laparoscope. The Da Vinci Surgical System [6] can change the control mode between the surgical manipulator and the endoscopic manipulator for the operator using a foot pedal. However, such methods do not allow the surgeon to alter the operative field while simultaneously manipulating tissue [7].

Other systems, like the automatic endoscope optimal positioning system (AESOP) from Computer Motion Inc. [8] and ViKY EP [9] use a voice control system to allow the operator to control the robotic endoscopic holder. However, rather than being helped by voice control, surgical time is actually often increased because of its slow response [10].

Alongside voice control, eye-tracking is a further intuitive hands-free method that can be used as an input signal for manipulating the laparoscope holder. As a practical example, Ubeda et al. [11] used the electrooculography signal from electrodes attached around the user's eye to control a manipulator. In the video-oculography field, Noonan, et al. [12] used a stand-alone eye tracker to alter the laparoscope direction based on gaze.

Single port surgery (SPS) is one of form of laparoscopic surgery that requires only one incision port. SPS has attracted increasing attention from patients because of its cosmetic advantages [13]. Although robots enhance the performance of standardized laparoscopic techniques [14], current systems, including those mentioned in the previous paragraphs, are not suitable for SPS because they require multiple incisions. Therefore, our laboratory has developed two prototypes of an SPS assistive robot system. Prototype 1 [15] uses respective controllers for tool manipulators and laparoscope direction; in Prototype 2 [16], the control mode can be changed between tool manipulators and endoscope direction by pushing a foot pedal like in the Da Vinci Surgical System [6].

The purpose of this paper is to introduce a control method for the alteration of endoscope direction using pupil-tracking into the Prototype 2 [16] system, which is achieved via image processing and the use of a webcam. The system will translate the obtained image data into an output signal. We propose a threshold for pupil movement distance: the sheath manipulator, which is used for the alteration of endoscope direction, activates when the user's pupil movement distance exceeds a threshold value; the manipulator remains in a static state when the pupil movement distance is below the threshold. Therefore, an appropriate threshold for the output signal needs to be determined to judge whether the movement state of the sheath manipulator is dynamic or static. To determine this, we tested four threshold values one by one in a horizontal movement experiment. The experimental outcome variable was the completion time of moving the

field of view to a specific target, so that the most appropriate threshold could be judged from the minimum completion time. The first part of this paper will describe how we obtain the pupil movement data. Then, the general system framework, including how the system integration is achieved, will be discussed. The second half of this paper describes an experiment to verify the effectiveness of the system and obtain a proper activation threshold value, which is important for the operability of pupil tracking.

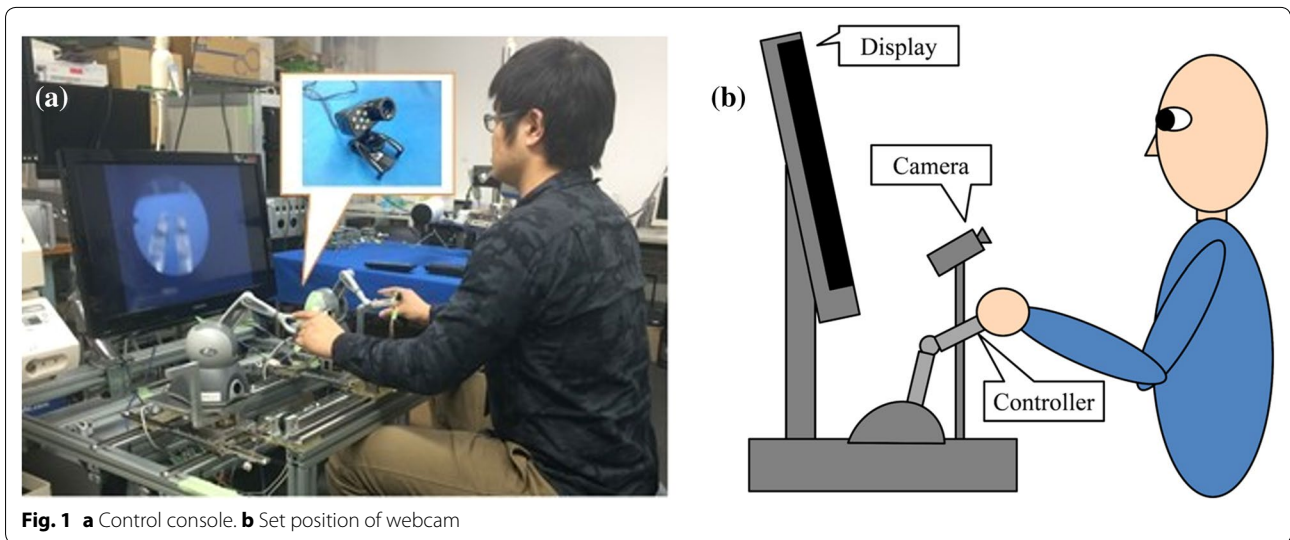
Methods

Acquisition method of pupil movement

In this study, a webcam was used to perform eye tracking. Webcams have an advantage compared with wearable gaze tracking systems, as they need not be mounted on the user's head. Such head-mounted displays or glasses-type head-wearable devices may induce additional physical and mental stress on the surgeon [17]. In contrast, the presence of a webcam has little or negligible burden on surgeons. During the development of this system, a near infrared (IR) LED webcam was eventually selected for one reason: the color of East Asian people's pupils in visible light is the same as that of the iris, but reflects white in near infrared light [18]. So the near infrared reflection features from the pupil were used in the image processing of the pupil movement data. Additionally, it is worth mentioning that near infrared light is harmless to the eye; this kind of light has been used in retina identification technology for a long time [19]. The infrared LED webcam (DC-NCR13U, Hanwha Q CELLS Japan Co., Ltd.) is used in this research, and its resolution is 1.3 million pixels. Figure 1a shows the control console of the prototype [20] and the set position of the webcam. Generally, the webcam is set on top of the display, but this may cause problems in capturing the whole eye because eyelashes or bangs may interfere. Thus, we set up the webcam in the position shown in Fig. 1b.

Applying non-rigid face tracking

An image processing method, non-rigid face tracking, was used to collect pupil movement data in this study. This method involves machine learning. The main merit of non-rigid face tracking, compared with traditional face tracking methods, is that it is non-rigid in that it can detect an individual face from the entire image and can also measure the relative distance between the feature positions, e.g., corners of the mouth, canthi and pupils. Thus, this method was used for pupil movement detection. The program used for image processing was OpenCV 2.4.4. The specific realization of non-rigid face tracking refers to [21]. An overview of the image processing is shown in Fig. 2. The complete procedure



consists of image capture, annotation, geometrical constraints, training detector, pupil tracking and setting the threshold.

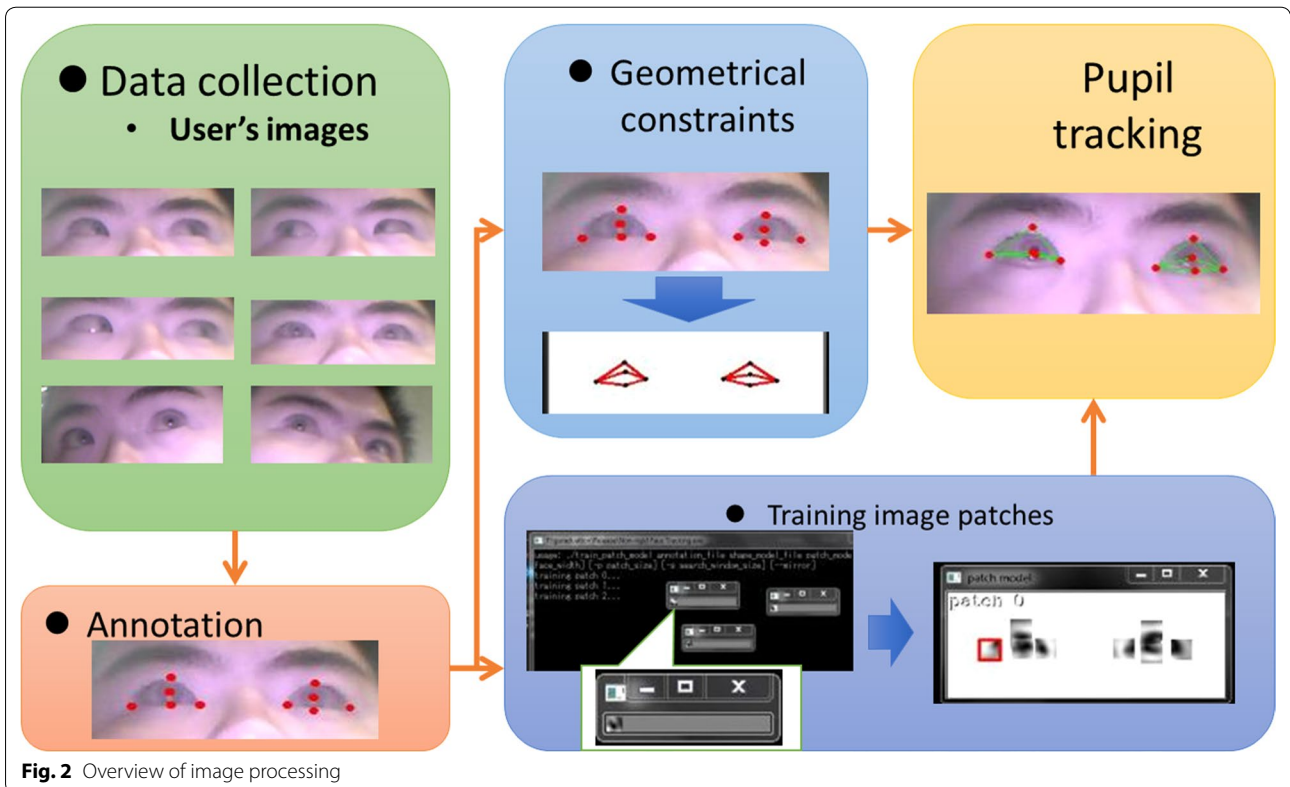
Image capture

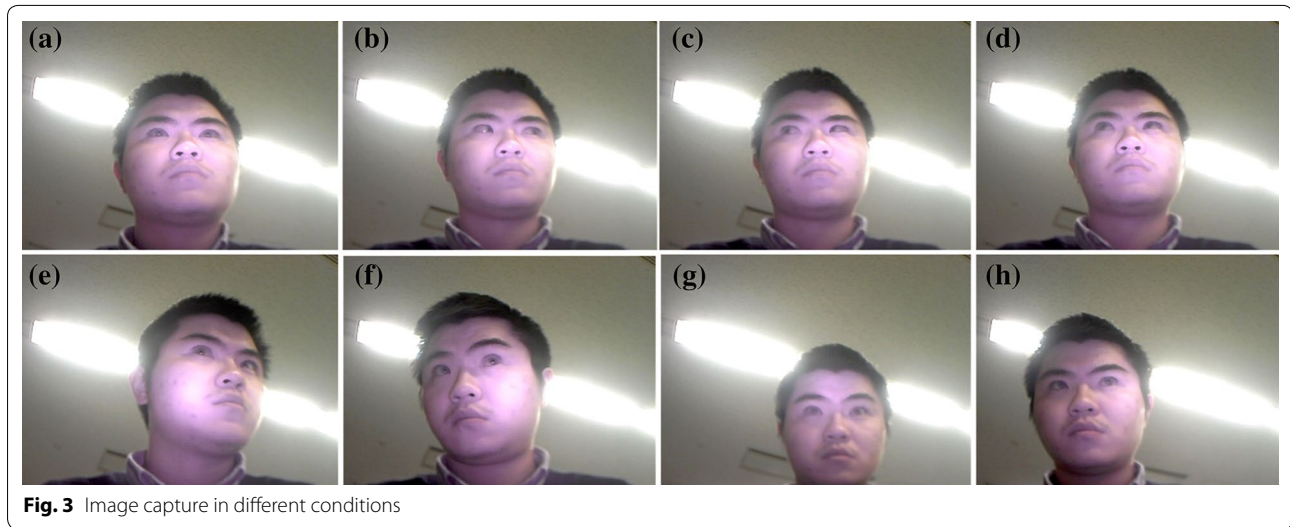
The first step was to use the IR LED webcam to capture as many of the operator’s facial images as possible (Fig. 3). The more images the webcam captures, the more

plentiful data the system has, which aids in training a robust face detector because the data include different lighting conditions, facial positions and eye directions.

Annotation

The second step was to annotate the tracking feature points on the facial image that could highlight the movement of the eyes. In computer vision, corner

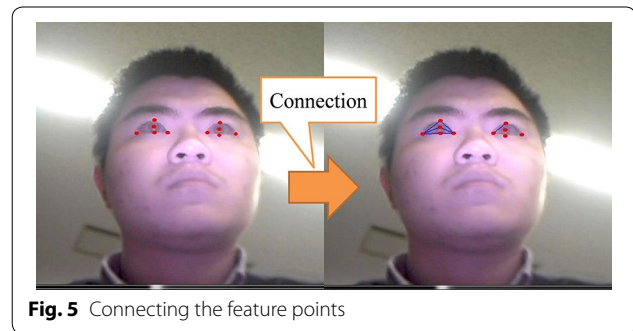
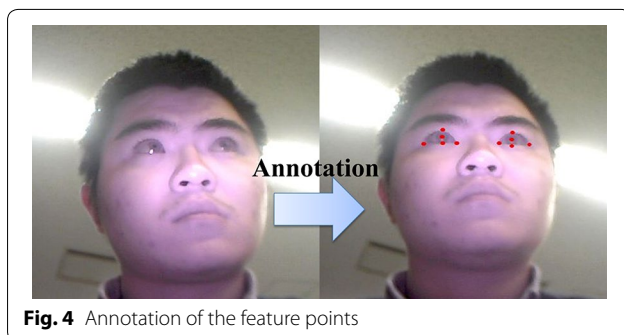




points are easily detected, so we annotated the eye corners. Moreover, our system needs to be able to track pupil points. Hence, we also annotated the pupils as feature points (Fig. 4). After annotation, we indexed the symmetry points. For example, the two inner eye corners and two pupils can be regarded as symmetry points. Symmetry indices can be used to mirror the captured images, which can increase the training dataset. Then, we connected the annotated points (Fig. 5), which are used for visualizing the pupil-tracking effect. Finally, the remaining images need to be processed using the previous steps. The size and resolution of these images are the same. The annotation points of the previous image will remain and appear on the next image. There is a deviation between the annotation points and the facial feature points; therefore, we need to use a mouse pointer to drag the annotation points to the facial feature points (Fig. 6).

Geometrical constraints

The set of annotations should correspond to physically consistent locations on the human face. In face tracking,



the face placement on an image will translate, rotate and scale. However, the corresponding relationship of the facial organs' positions, also known as facial geometry, will not change. Therefore, the corresponding relationship of the annotation points should also be constrained according to the facial geometry. These annotated points were combined as a geometrical model using Procrustes Analysis [22] to handle several variations on the pose of the user's face in the image such as translation, rotation and scale, and a shape model was established (Fig. 7).

Training image patches

To make annotation points to track the facial features, we need to train discriminative patch models [21]. The annotated images mentioned in the Annotation section comprise the training dataset. The image patch can be independently trained from every annotated point, and the trained patch is cross-correlated with an image region containing the annotated point. Then, the image patch will strongly respond to the part of the image containing the facial feature, and weakly respond everywhere else. Figure 8 shows the training process: the patches for

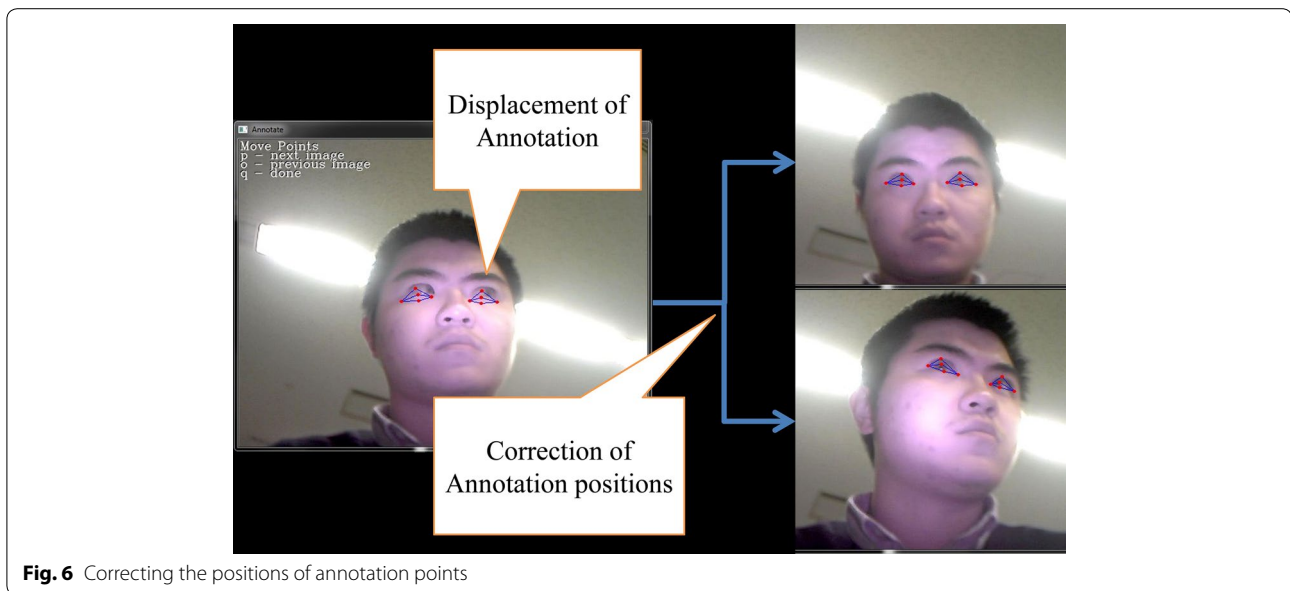


Fig. 6 Correcting the positions of annotation points

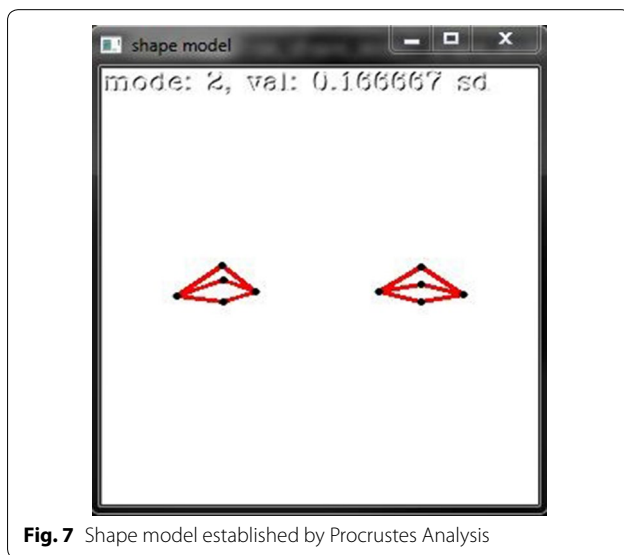


Fig. 7 Shape model established by Procrustes Analysis

detecting eye corners and pupils are being trained. Then, the result of the patch training is shown in Fig. 9.

Pupil tracking

Before processing pupil tracking, the last step is to combine the shape model and the image patches into one tracker model (Fig. 10). The reason for this step is that the image patch may respond strongly at an incorrect facial feature point. An accurate estimate of the facial feature position can be determined via the geometrical model, and then the facial features can be detected with the cross-correlation image patches, producing

good robustness and efficiency. Therefore, we need the shape model to restrict the matching region of the image patches. Moreover, initialization of pupil tracking should result in the ability to place the tracker model on the user's face in the image. In this step we use OpenCV's built-in face detector to confirm the box region of the user's face in the captured images. Then, the tracker model can start the tracking operation from the box region of the user's face.

After completing the above steps, the face tracker was able to achieve pupil tracking because the positions of both intraocular angles and pupils could be obtained from the tracking trajectory. As shown in Fig. 11a, the midpoint of both pupils is labeled as the active point (red circle), while the midpoint of both inner canthus angles is labeled as the static point (blue circle). As shown in Fig. 11b, the variation in distance between the active point and the static point was taken as the output signal.

Besides up–down movement and left–right movement, human eye movement also includes some unconscious movements such as blinking, saccade and tremble [23]. The output signal caused by these movements is regarded as noise by the control system. Therefore, the moving average method was used to filter the noise signal. The average value was calculated from 10 samples and the window size of the filter was 400 (ms). As a result, the frequency of pupil tracking was 25 Hz.

Figure 12 shows the internal processing of pupil tracking for the control system. Two parameters, D (pixel) and ΔX (pixel), can be quantifiably obtained via image processing. D represents the distance between both inner canthus, and ΔX (pixel) represents the horizontal

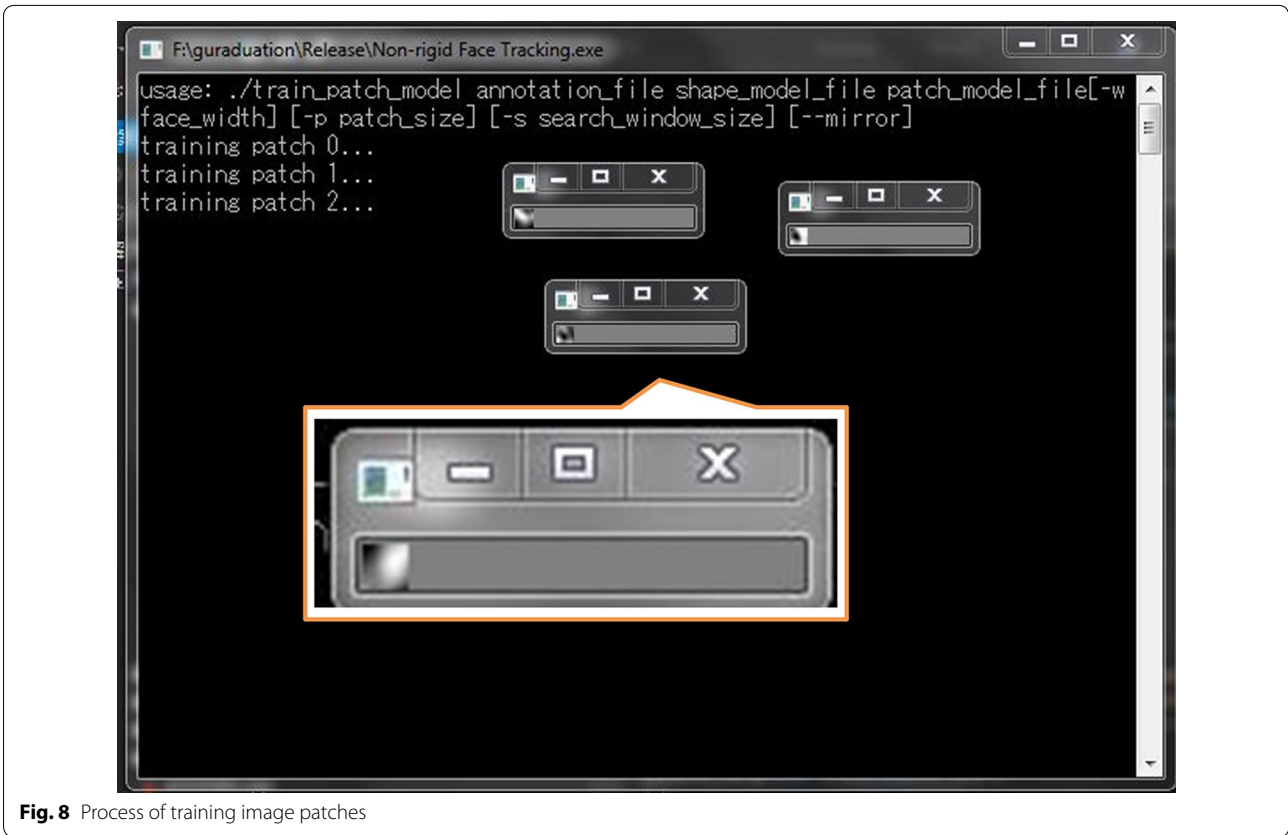


Fig. 8 Process of training image patches

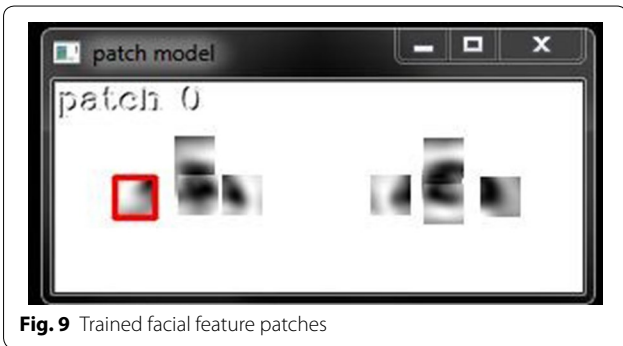


Fig. 9 Trained facial feature patches

distance between the static point and the active point. With the introduction of parameter t , D/t can be set as the changeable activation threshold. If the absolute value of ΔX exceeds the activation threshold value D/t , as shown in Eq. (1), the sheath manipulator will start to move.

$$|\Delta X| > \frac{D}{t} \tag{1}$$

Conversely, if the absolute value of ΔX does not exceed the activation threshold value D/t , as shown in Eq. (2), the sheath manipulator will remain static.

$$|\Delta X| \leq \frac{D}{t} \tag{2}$$

In this control system, Eq. (3) is used to control two servomotors, which are the actuators of the sheath manipulator.

$$V = \frac{TargetAngle - PresentAngle}{T_{cycle}} \tag{3}$$

In Eq. (3), V indicates the angular velocity of the servomotors, $TargetAngle$ indicates the target angle for the servomotors, $PresentAngle$ indicates the current angle of the servomotors, and T_{cycle} indicates the sampling period of the servomotors. ΔX can be thought of as the input signal to control the servomotors according to Eq. (4).

$$P \cdot (TargetAngle - PresentAngle) = \Delta X \tag{4}$$

P indicates the gain value and $P \cdot T_{cycle} = 800$.

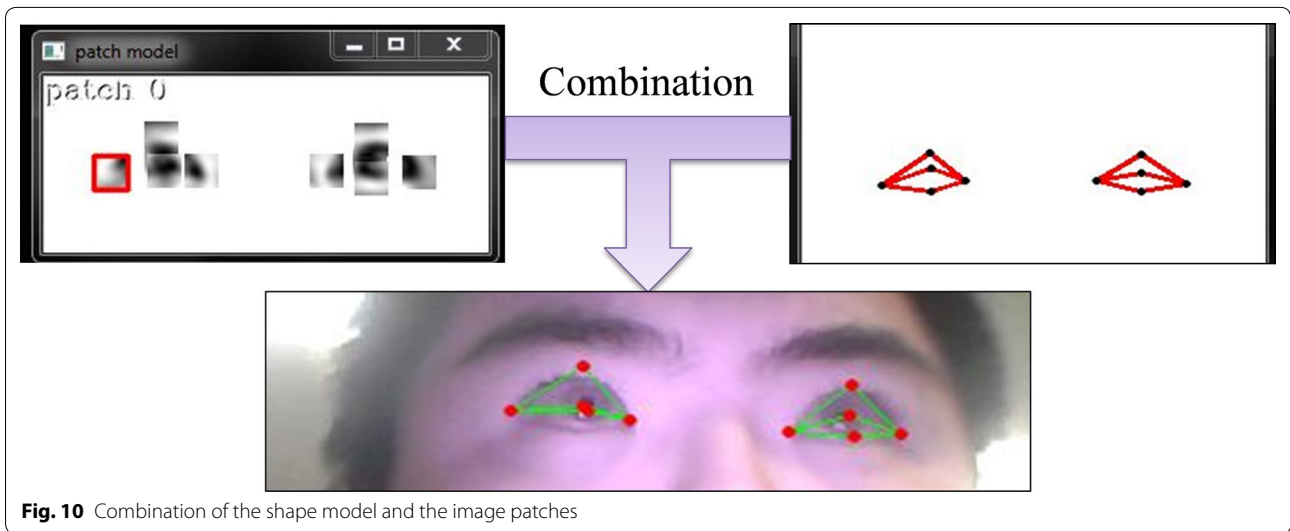


Fig. 10 Combination of the shape model and the image patches

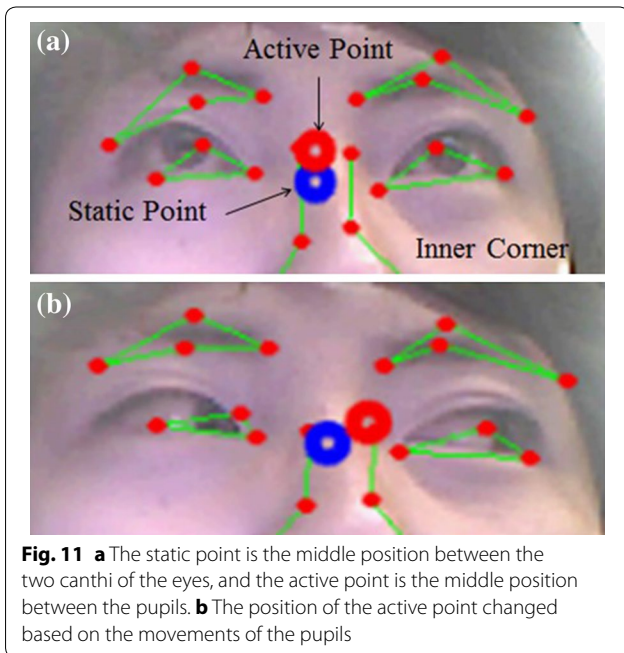


Fig. 11 **a** The static point is the middle position between the two canthi of the eyes, and the active point is the middle position between the pupils. **b** The position of the active point changed based on the movements of the pupils

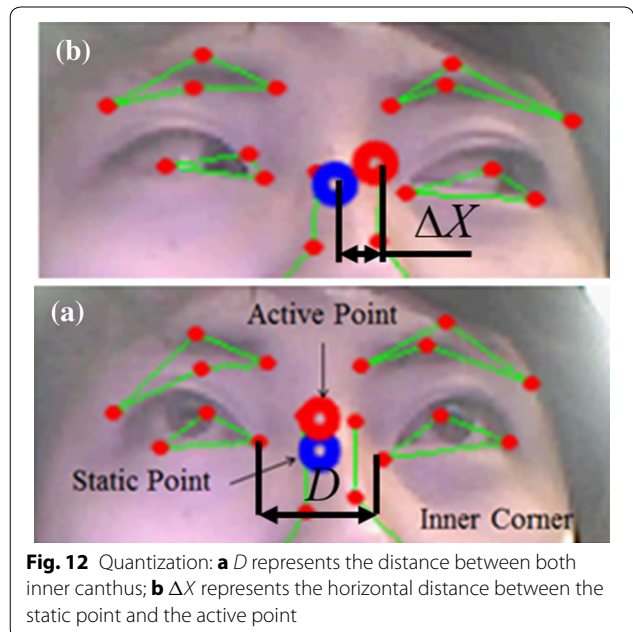


Fig. 12 Quantization: **a** D represents the distance between both inner canthus; **b** ΔX represents the horizontal distance between the static point and the active point

Through conversion from Eq. (1) to Eq. (4), Eq. (5) is obtained.

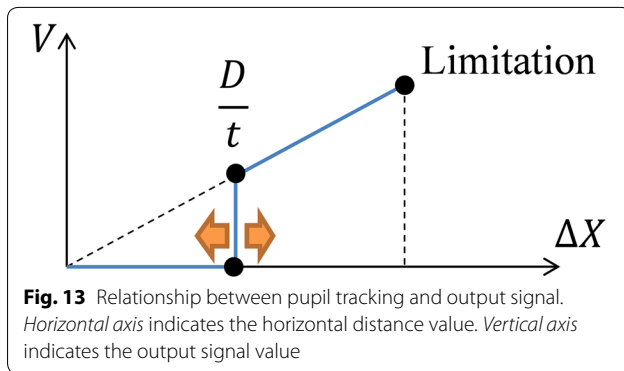
$$V = \begin{cases} \alpha \Delta X \left(|\Delta X| > \frac{D}{t} \right) \\ 0 \left(|\Delta X| \leq \frac{D}{t} \right) \cdot \left(\alpha = \frac{1}{P \cdot T_{\text{cycle}}} \right) \end{cases} \quad (5)$$

As shown in Eq. (5), ΔX is sent to the servomotors in a proportional manner, when the absolute value of ΔX exceeds the activation threshold value D/t . Conversely, no signal will be sent to the servomotors when

the absolute value of ΔX does not exceed the activation threshold value D/t . Additionally, Eq. (5) can be graphed as in Fig. 13: *Limitation* indicates the limits of the pupil movement range.

Performance analysis

As mentioned in the previous research, the frequency of image processing was 25 (Hz), and therefore the time delay was 40 (ms). A frequency equal to or greater than 25 (Hz) is regarded as real-time [24, 25]. Moreover, the response of the manipulator was less than 100 (ms) [15]. Therefore, the overall delay was acceptable because it did



not exceed 330 (ms), which was estimated as the maximum time delay compatible with the safe performance of surgical manipulations [26].

Before we design an experiment for determining the proper activation threshold value, we need to decide the range of D/t . If D/t is too big, the user has to rotate their eyes to move their gaze out of the screen; if D/t is too small, the sheath manipulator will activate even when the user keeps their eyes static. Therefore, three participants, engineering graduate students without glasses, tested the performance of the pupil-tracking method. During this test, we requested that the participants perform three kinds of eye motions: rotating the eyes to a maximum angle as much as possible, gazing at the edge of the monitor, and keeping the eyes static and gazing straight. Meanwhile, we recorded the variations in D and ΔX .

First of all, we made personal tracker models for each of the three participants as in the procedures mentioned in the previous sections. Then, we requested the participants to perform the three kinds of eye motions.

For rotating the eyes to a maximum angle as much as possible, we instructed each participant to move left and right twice. The purpose of this task was to obtain the maximum value of $\Delta X/D$. The result is shown as a graph (Fig. 14). On the vertical axis, with units of pixels, Nos. 1–4 represent the results for Participant 1, Nos. 5–8 represent the results for Participant 2, and Nos. 9–12 represent the results for Participant 3. The recorded values of D and ΔX are attached to the bottom of the graph. The average $\Delta X/D$ was calculated as $1/5$.

Then, we requested participants to gaze at the edge of the monitor shown in Fig. 1a. The purpose of this task was to observe the maximum value of $\Delta X/D$ when the range of visibility was within the screen, because the user's gaze should not leave the monitor screen when controlling the endoscopic manipulator via pupil tracking. The distance between the screen and the participants' eyes remained at 600 [mm]. The participants had to gaze at the edge of the monitor for 3 s every trial, and

were instructed to gaze at the left and right sides of the edges twice. The results are shown as a graph (Fig. 15). On the horizontal axis, Nos. 1–4 represent the results of Participant 1, Nos. 5–8 represent the results of Participant 2, and Nos. 9–12 represent the results of Participant 3. The units of the vertical axis are pixels. The recorded values of D and ΔX are attached to the bottom of the graph. The D and ΔX for each trial are averaged over the 3 s. Finally, the average $\Delta X/D$ was calculated as $1/7$.

Ideally, ΔX is equal to 0 when the eyes remain static and look straight. However, ΔX has some variation that needs to be observed. In this task, every participant was requested to keep their eyes static and look straight for 3 s. The results of the three participants are shown in Fig. 15. The average $\Delta X/D$ values shown in Fig. 16a, b are $1/40$, and the average $\Delta X/D$ value shown in Fig. 16c is $1/25$.

In the next section, we needed to identify an appropriate activation threshold that could be used to determine whether the sheath manipulator was moving or not. Using the above results, we confirmed that the activation threshold should be selected from a range between $D/7$ and $D/25$.

Experiment

Purpose of the experiment

The aim of the present experiment is to verify the effectiveness of the image processing method applied to the sheath manipulator control system. Moreover, to judge whether the sheath manipulator is dynamic or static, an appropriate output signal threshold needs to be obtained. In this experiment, the activation thresholds in four conditions were evaluated using operation time.

Sheath manipulator

In this experiment, we used the sheath manipulator of SPS robot prototype 2 [15] as an experimental platform. The sheath manipulator, designed for adjusting the direction of the endoscope, bends through double-screw-drive mechanisms (Fig. 17) to change the view orientation. The bending portion consists of three double-screw-drive mechanisms so that it can achieve several kinds of movement: up–down, left–right and diagonal turning. Furthermore, the screws between each double-screw-drive mechanism are connected by universal joints. Figure 18 shows that the upper two universal joints allow the endoscope to bend, while the one underneath is for support. If the upper two joints rotate at the same speed and in the same direction, the sheath manipulator will bend in the vertical direction. Conversely, if they rotate at the same speed but in opposite directions, the sheath manipulator will bend in the horizontal direction. In addition, if only one upper joint rotates, the sheath manipulator will bend in the diagonal direction.

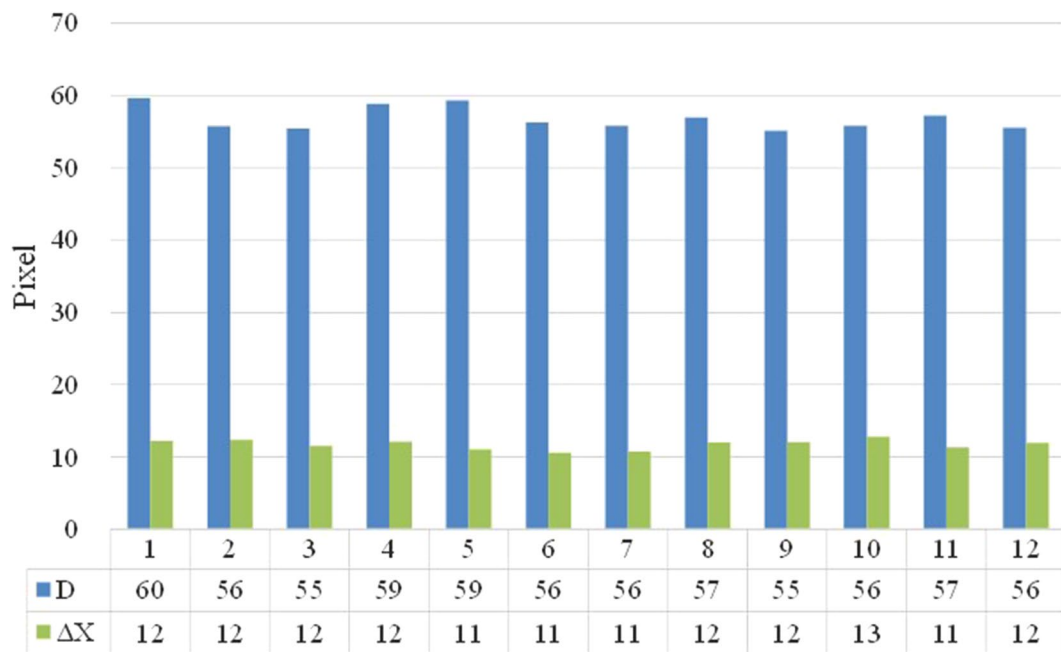


Fig. 14 The values of D and ΔX shown in the *graph* were obtained when participants were rotating their eyes to the maximum angle

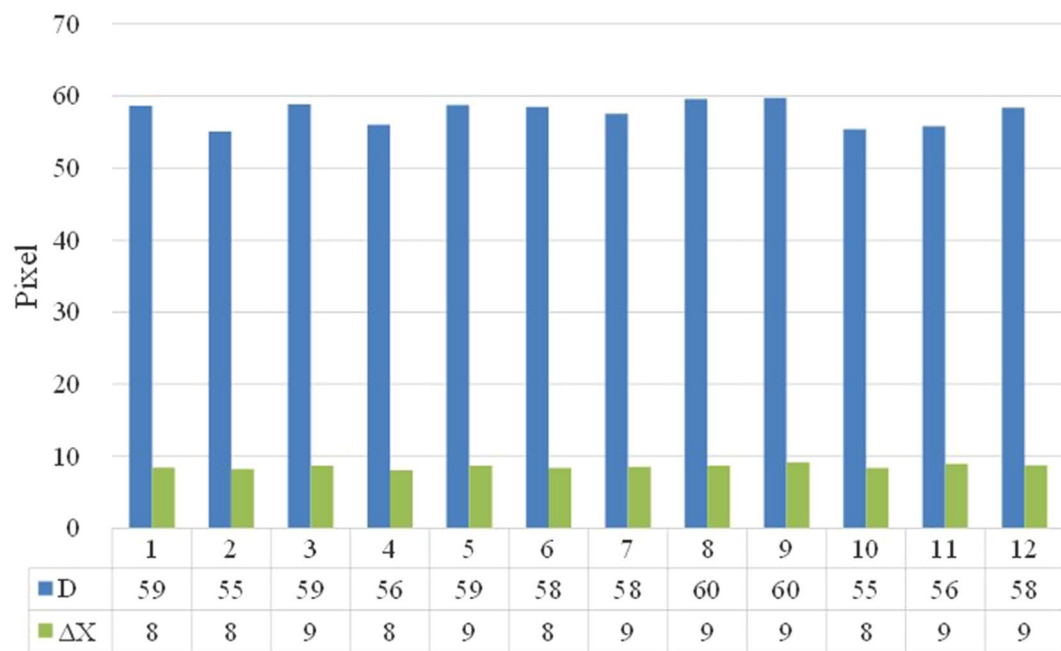
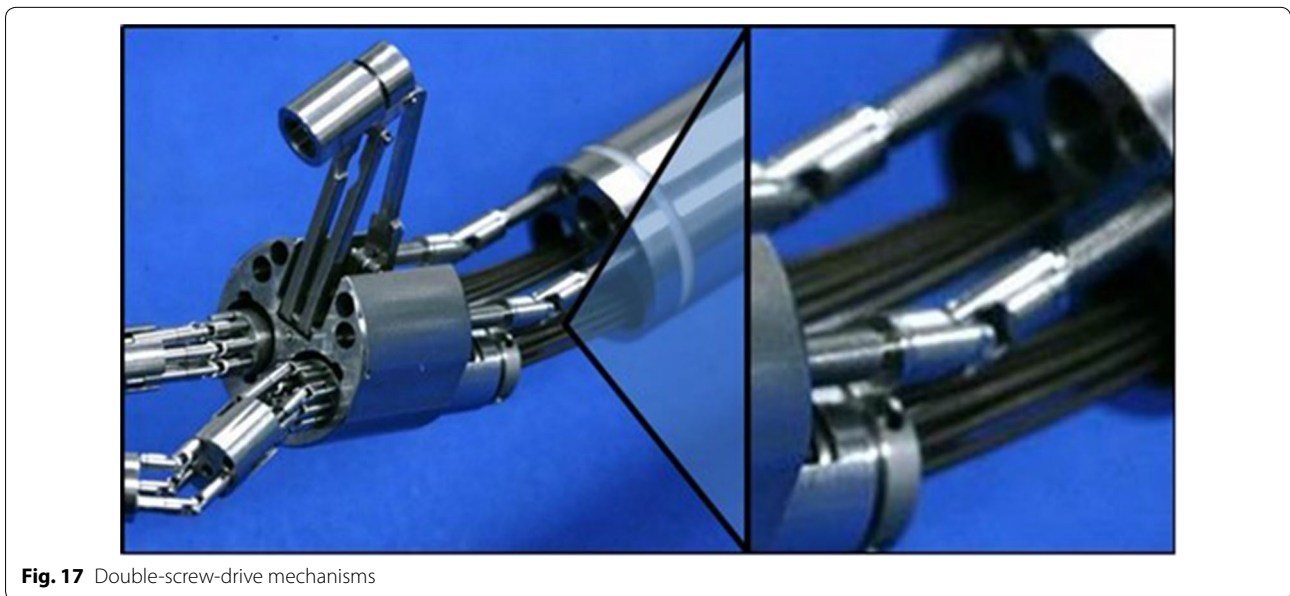
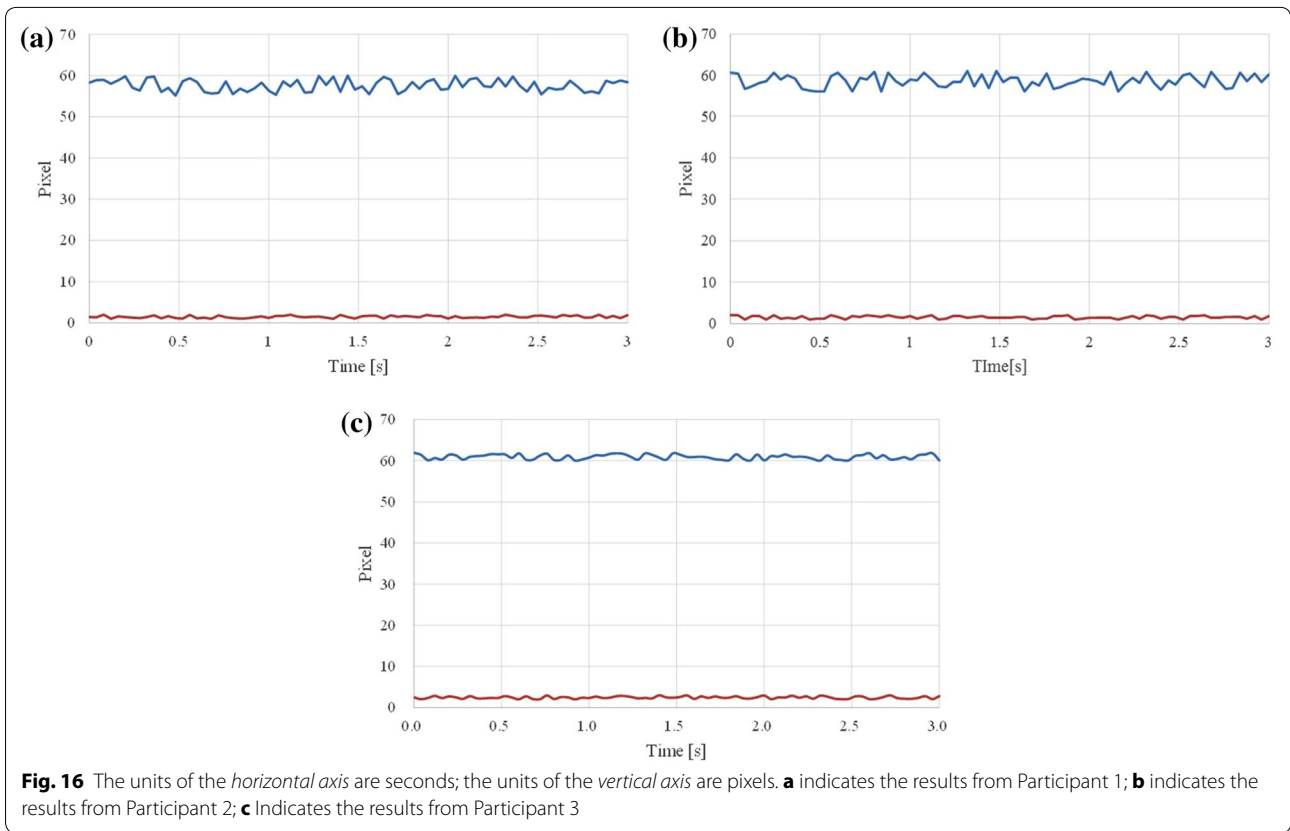


Fig. 15 The values of D and ΔX shown in the *graph* were obtained when participants were gazing at the edge of the monitor screen

Communication between master and slave

The SPS robot has a remote control function based on the master–slave structure. The master is a control console programmed using C/C++ on a Windows PC. The

slave is a dedicated computer that reads the encoders of the servomotors in real time and activates these servomotors according to the received signal from the master PC. The communication between master and slave is



realized using user datagram protocol (UDP) over Ethernet. Figure 19 shows the complete program architecture of this system.

The master computer was installed with Windows XP. The pupil-tracking program we proposed was merged into the operating system of prototype 2 and was

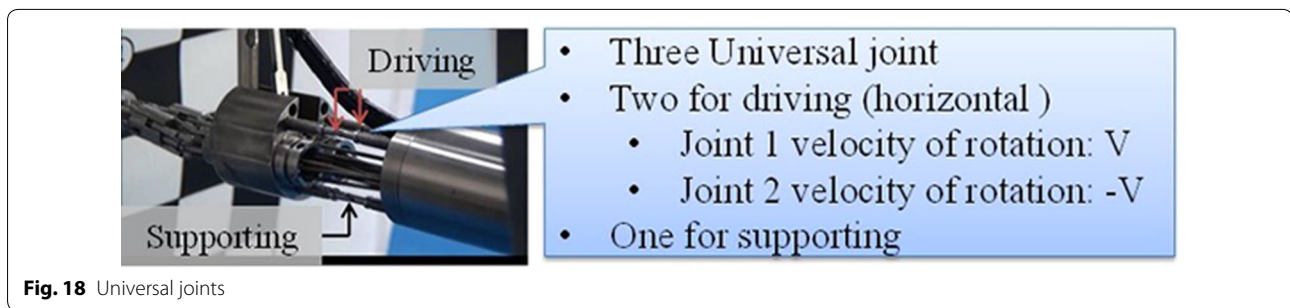


Fig. 18 Universal joints

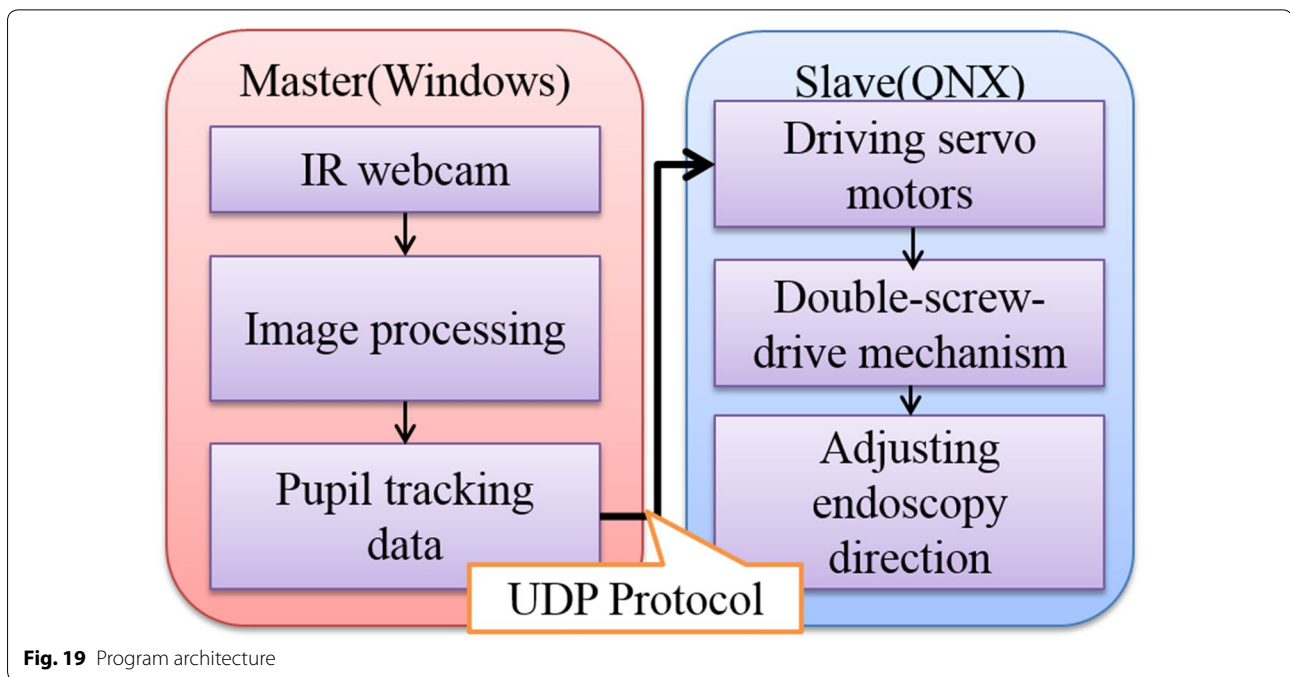


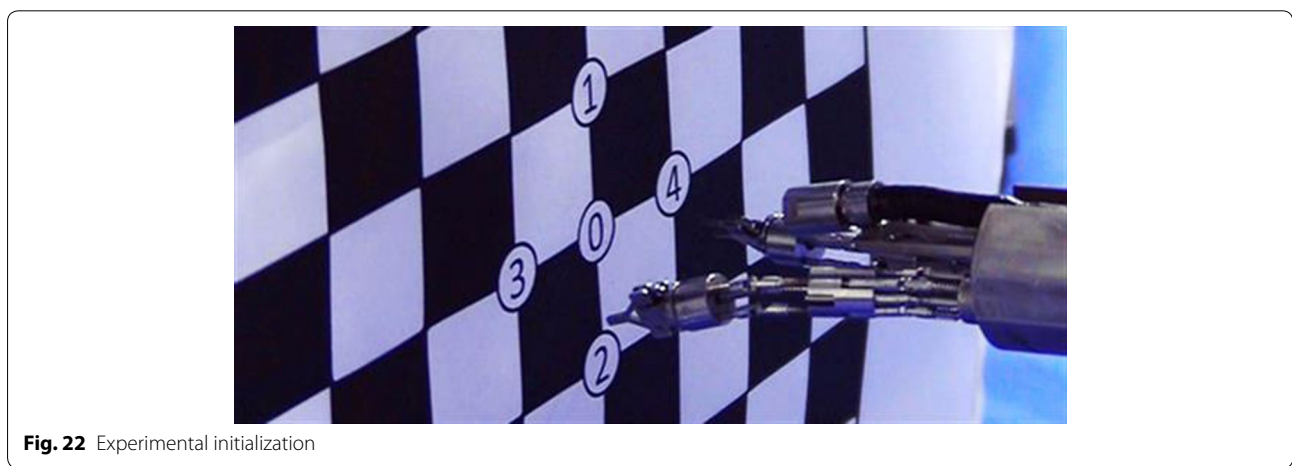
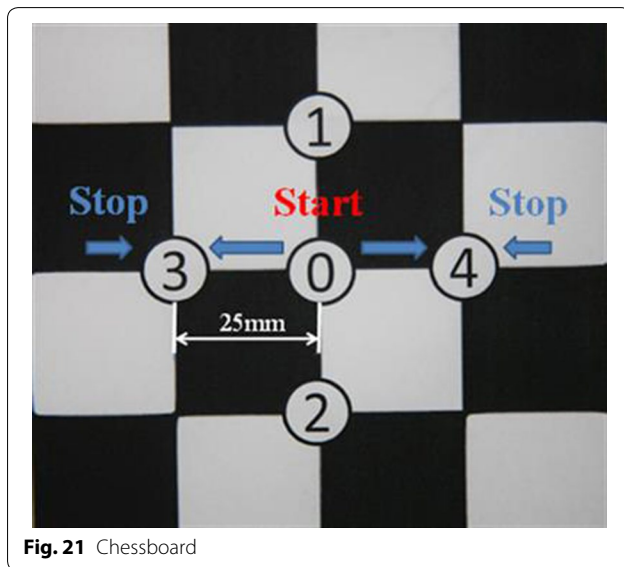
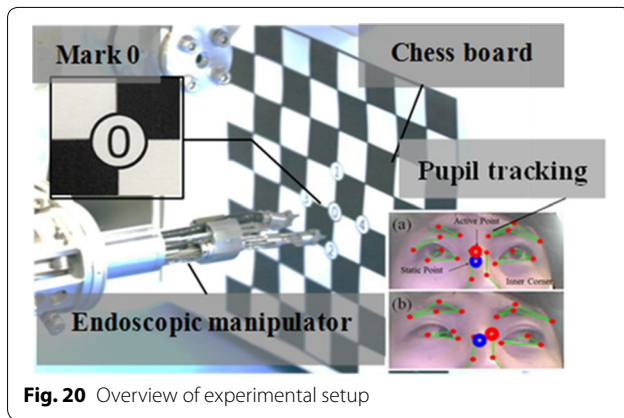
Fig. 19 Program architecture

compiled with Visual Studio 2008. As shown in Fig. 19, the Infrared LED webcam mounted on the master side initially captures the images. After this, the pupil movement data are extracted from the captured images by the pupil-tracking program and are then converted into an output signal for controlling the servomotor. Subsequently, the output signal is sent to the slave via UDP. After receiving the output signal, the slave computer activates the servomotors to drive the double-screw-drive mechanisms [15], thus adjusting the direction of the endoscope.

Experimental conditions

The participants in this experiment were five graduate students who had no experience of endoscopic surgery. Figure 20 shows the experimental setup. The distance between the display and the participant's face was set between 50 and 70 cm according to the most ergonomic view [27]. As shown in

Fig. 21, the operators use their pupils to operate the sheath manipulator, moving the image center point from "point 0" to "point 3" or "point 4", which were displayed on a chessboard with a $25 \times 25 \text{ mm}^2$ lattice. The reason for using the chessboard was to facilitate the observation of the image changes. The image center of the endoscope matching "point 0" was set as the initial state. The activation threshold value D/t was determined in four conditions: when t was equal to 14, 18, 22, or 24. As mentioned in the Session Performance analysis, the four conditions were selected from the range between $D/7$ and $D/25$. Therefore, we selected the first condition as $D/14$, which was half the value of $D/7$, and obtained the value $D/25$ when participants kept their eyes static and looking straight. The manipulator could not be static if we had chosen a condition smaller than $D/25$. Therefore, we chose $D/24$ as the smallest condition. Beyond that, we selected two additional conditions.



Experimental procedure

Before starting the experimental procedure, the personal eye tracker model was established for each of the five participants as mentioned in the (“Methods”) section. Then, the flow of the experiment proceeded as follows:

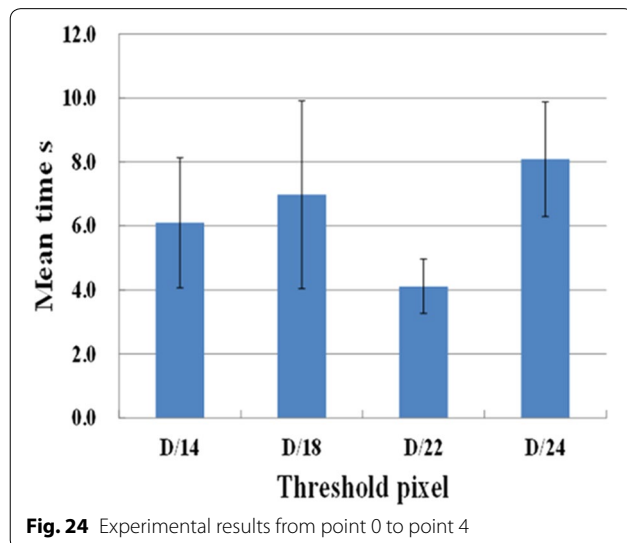
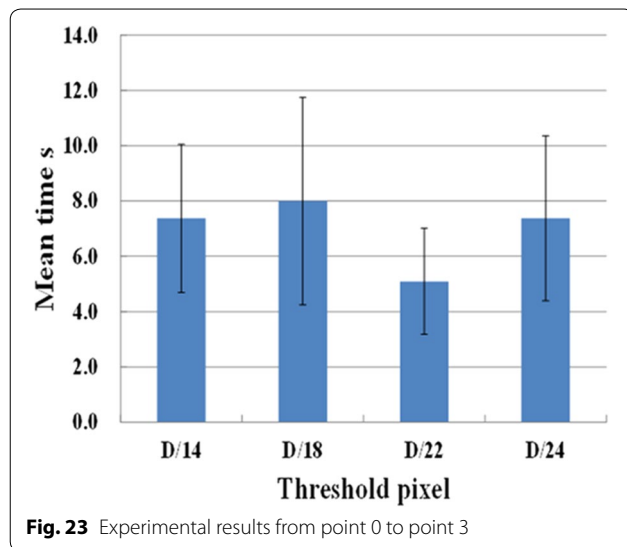
- The activation threshold was set for the control system.
- The system was tested as to whether it altered direction with the participant’s pupil movement (Additional file 1).
- Initialization: the image center of the endoscope was changed to match “point 0” (Fig. 22).
- Each participant moved the image center of the endoscope from “point 0” to “point 3” four times, and from “point 0” to “point 4” four times. The time cost of every trial was recorded.
- The threshold value was changed and the above flow from step *a* to step *c* was repeated.

The change order was D/14, D/18, D/22 and D/24.

For the present experiment, it was important that the system reflected the operator’s viewing intention so that the participants could confirm their location during the experiment. To judge whether the sheath manipulator was in the static state, each participant verbally confirmed when the operation was completed, i.e., when the image center of the endoscope had reached the target and stopped. Thus, the time measurement ended when the participant replied that the movement was complete.

Results and discussion

All participants could alter the direction of the sheath manipulator with their pupil movement. Figures 23 and



24 show the results of the experiment. In particular, Fig. 20 shows the results when moving the image center of the endoscope from “point 0” to “point 3” (Figs. 25), and 21 shows the results when moving the image center of the endoscope from “point 0” to “point 4” (Fig. 26). The horizontal axis represents the operation time and the vertical axis represents the activation threshold D/t (pixel). As shown in Fig. 23 and 24, the average operation time was shortest and the standard deviation of operation time was smallest when the activation threshold was set as $D/22$. In the $D/22$ condition, the average operation time was 5.2 s and the standard deviation was 1.9 s when the endoscope turned to the left (Fig. 25). In the other conditions, the average operation time was 4.2 s and the standard deviation was 0.8 s when the endoscope turned to the left (Fig. 26).

For the alteration of the endoscope direction operation system based on pupil position tracking, the shortest operation time was when the activation threshold value was equal to $D/22$. Similarly, the standard deviation values were smallest in both experiments when that threshold value was selected. Therefore, it is appropriate to use $D/22$ as an activation threshold for the operation system of the proposed SPS robot. In addition, we found that operating the sheath manipulator via pupil tracking can provide good stability and response when an appropriate threshold value is used. Also, the period between blinks was 6–8 s [28], which prevents the eye from fatiguing and from drying out. We observed an obvious difference for operation time and standard deviation in the $D/22$ condition in the two experiments. [29, 30] suggested that the center of monitor image should be aligned to the center of operative field. If the target points: “point 0”, “point 3” or “point 4” is projected on the center of monitor image, ΔX will not exceed the activation threshold value. Therefore, the operator can stop the desired target position. The limitation of this experiment is that the numbers of participants and conditions were both rather small. Therefore, there is limited evidence to support the optimal threshold value. Moreover, the accurate relationship between the motion of the manipulator and the movement of the pupils was not confirmed. After causal analysis, we found that the difference arose from the intertwining flexible shafts that are set between the manipulators and servomotors; such a situation affects the stability and speed of rotation. As a solution to this problem, these flexible shafts need to be sheathed in pipes or fixed on a fixation device so as to avoid intertwining. Furthermore, we hope to achieve better pupil tracking in the vertical direction by using a higher resolution webcam.

To invite surgeons or medical trainees to participate the manipulation experiment in the future, we gave a presentation of the system and the experimental results to two endoscopic surgeons. After the presentation, the surgeons agreed with the usefulness of the proposal and stated that:

1. The proposed system is more intuitive than voice-control and pedals. The surgeons certainly need a hand-free strategy to manipulate the endoscope. Using webcam and image processing is a good approach because it does not require the surgeons to be attached to additional devices, which may increase their burden during an operation.
2. The proposed system is useful because the manipulator will stop when the center of the visual field aligns with the target. One of the fundamentals of manipulating the endoscope is aligning the center of the visual field with the center of operation.

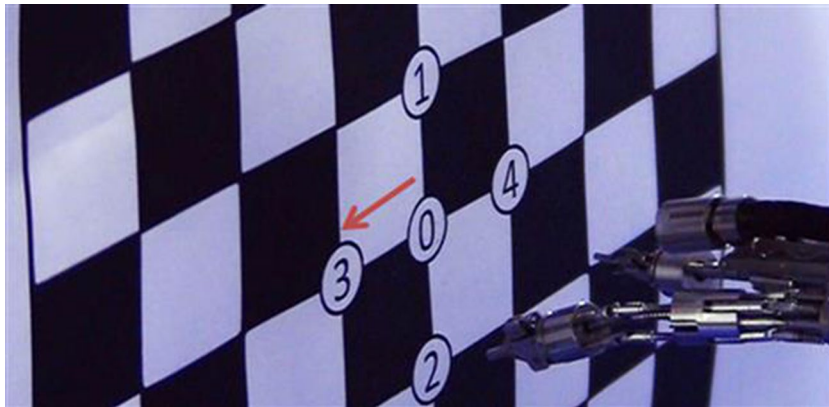


Fig. 25 Sheath manipulator turning to the *left*

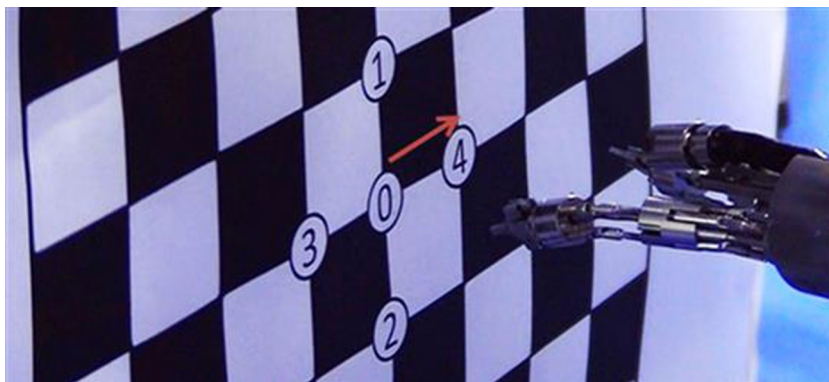


Fig. 26 Sheath manipulator turning to the *right*

3. The horizontal alternation of the visual field is greater than the vertical, but the vertical alternation is still indispensable.
4. A zoom function is indispensable for an endoscopic control system when the surgeon is performing a delicate operation. For example, the surgeon would ideally like to zoom in the visual field when they are peeling the tissue from around a vessel. Moreover, in a future system, it would be better to be able to adjust the rate of visual field alternation according to the magnification of the lens.
5. An emergency stop button is an indispensable part of the system, which is used to avoid a collision between the endoscope and tissue.

Conclusions

In this paper, a hands-free technique for controlling the alteration of endoscope direction using a pupil-tracking method via an image processing method was introduced. The novelty of the proposed method is its ability to achieve pupil tracking because the variation in distance

of both intraocular angles and pupils could be obtained from the tracking trajectory. In this method, an appropriate output signal threshold needs to be obtained for judging whether the sheath manipulator, which alters the endoscope direction, is dynamic or static. An experiment was performed to verify the effectiveness of the image processing method applied to the sheath manipulator control system, and the activation threshold of the control system had to be determined and used for the horizontal direction movement of the sheath manipulator. We found an activation threshold value that fulfils stability and response simultaneously. This time, we only verified the horizontal direction because of the limitations of our method. At present, it is quite difficult for the sheath manipulator to make vertical movements, because the vertical movement range of the eye is much less than its horizontal movement range. To realize vertical direction movement, we need a higher resolution webcam to detect the relatively small vertical movement of the eyes. As surgeons pointed out, a zoom function is indispensable for endoscope manipulation. Thus, using only the pupil

position parameter as shown in this experiment is not sufficient to achieve the zoom function. In future work, we aim to develop an algorithm that includes a large number of operating conditions to judge the movement state and achieve more types of movements. Also, we will improve the current control system based on the surgeons' suggestions. Furthermore, we also plan to invite surgeons and medical trainees to be participants in the manipulation experiments.

Additional file

Additional file 1. Animation related to the verification of the effectiveness of the image processing method applied to the laparoscope manipulator control system.

Abbreviations

AESOP: automatic endoscope optimal positioning system; GUI: graphical user interface; UDP: user datagram protocol.

Authors' contributions

YC derived the basic concept of the overall system, technically constructed the system and drafted the manuscript. All authors read and approved the final manuscript.

Author details

¹ Graduate School of Creative Science and Engineering, Waseda University, 3-4-1 Ohkubo, Shinjuku-ku, Tokyo 169-8555, Japan. ² Graduate School of Advanced Science and Engineering, Waseda University, 2-2 Wakamatsu-cho, Shinjuku-ku, Tokyo 162-8480, Japan. ³ Graduate School and Faculty of Engineering, Chiba University, 1-33 Yayoi-cho, Inage-ku, Chiba 263-8522, Japan.

Acknowledgements

The authors sincerely thank the volunteers for participating in our experiments. The work was supported in part by a research grant from JSPS Global COE Program: Global Robot Academia, JSPS Grant-in-Aid for Scientific Research (A) No. 20339716, JSPS Grant-in-Aid for Scientific Research (S) No.25220005, JSPS Grant-in-Aid for Exploratory Research No. 15K12606 and the Program for Leading Graduate Schools, "Graduate Program for Embodiment Informatics" of the Ministry of Education, Culture, Sports, Science and Technology.

Competing interests

The authors declare that they have no competing interests.

Received: 29 March 2015 Accepted: 11 January 2016

Published online: 22 January 2016

References

1. ASCRS, Laparoscopic surgery—what is it?, Available from: http://www.fascrs.org/patients/treatments_and_screenings/laparoscopic_surgery/
2. Horgan S, Vanuno D (2001) Robots in laparoscopic surgery. *J Laparoendosc Adv Surg Tech A* 11:415–419
3. Matern U, Konecny S (2007) Safety, hazards and ergonomics in the operating room. *Surg Endosc* 21:1965–1969
4. Jihad K, George-Pascal H, Raj G, Mihir D, Monish A, Raymond R, Courtenay M, Inderbir G (2008) Single-Port laparoscopic surgery in urology: initial experience. *Urology* 71:3–6
5. Yasunaga T, Hashizume M, Kobayashi E, Tanoue K, Akahoshi T, Konishi K, Yamaguchi S, Kinjo N, Tomikawa M, Muragaki Y, Shimada M, Maehara Y, Dohi Y, Sakuma I, Miyamoto S (2003) Remote-controlled laparoscope manipulator system NaviotTM, for endoscopic surgery. *Int Congress Ser* 1256(2003):678–683
6. Intuitive Surgical Inc. [homepage on the Internet]. Available from: <http://www.intuitivesurgical.com>
7. Meyer A, Oleynikov D (2013) Surgical robotics, *Mastery of Endoscopic and Laparoscopic Surgery*, Vol 2, 4th edn. p 62–71
8. Kraft B, Jager C, Kraft K, Leibl B, Bittner R (2004) The AESOP robot system in laparoscopic surgery: increased risk or advantage for surgeon and patient? *Surg Endosc* 18:1216–1223
9. ViKY EP, available from: http://www.endocontrol-medical.com/viky_ep.php
10. Kolvenbach R, Schwiez E, Wasillijew S, Miloud A, Puerschel A, Pinter L (2004) Total laparoscopically and robotically assisted aortic aneurysm surgery: a critical evaluation. *J Vasc Surg* 39:771–776
11. Ubeda A, Ianez E, Azorin J (2011) Wireless and portable EOG-based interface for assisting disabled people. *IEEE/ASME Transact Mechatron* 16(5):870–873
12. Noonan D, Mylonas G, Shang J, Payne C, Darzi A, Yang G (2010) Gaze contingent control for an articulated mechatronic laparoscope. *Proceedings of the 2010 3rd IEEE RAS and EMBS International Conference on Biomedical Robotics and Biomechatronics*, The University of Tokyo, Tokyo, Japan, 26–29 September 2010
13. Autorino R, Cadeddu JA, Desai MM, Gettman M, Gill IS, Kavoussi LR, Lima E, Montorsi F, Richstone L, Stolzenburg JU, Kaouk JH (2010) Laparoendoscopic single-site and natural orifice transluminal endoscopic surgery in urology: a critical analysis of the literature. *Eur Assoc Urol* 59(1):26–45
14. Hubens G, Coveliers H, Balliu L, Ruppert M, Vaneerdeweque W (2003) A performance study comparing manual and robotically assisted laparoscopic surgery using the da Vinci system. *Surg Endosc Other Intervent Tech* 17(10):1595–1599
15. Kobayashi Y, Tomono Y, Sekiguchi Y, Watanabe H, Toyoda K, Konishi K, Tomikawa M, leiri S, Tanoue K, Hashizume M, Fujie GM (2010) The international journal of medical robotics and computer assisted surgery. *Int J Med Robotics Comput Assist Surg* 6:454–464
16. Kobayashi Y, Sekiguchi Y, Noguchi T, Liu Q, Oguri S, Toyoda K, Konishi K, Uemru M, leiri S, Tomikwa M, Ohdaira T, Hashizume M, Fujie GM (2015) The international journal of medical robotics and computer assisted surgery. *Int J Med Robotics Comput Assist Surg* 11:235–246
17. Rassweiler J, Gözen A, Frede T, Teber D (2011) Laparoscopy vs. robotics: ergonomics—does it matter? *Robotics Genitourin Surg* 2011:63–78
18. Ohno T, Mukawa N (2004) A free-head, simple calibration, gaze tracking system that enables gaze-based interaction. *ETRA 2004, Eye Tracking Research and Applications Symposium*, p 115–122
19. Hill B (1996) Retina identification, *Biometrics*, p 123–141
20. Liu Q, Kobayashi Y, Zhang B, Noguchi T, Takahashi Y, Nishio Y, Cao Y, leiri S, Toyoda K, Uemura M, Tomikawa M, Hashizume M, Fujie M G (2014) Development of a smart surgical robot with bended forceps for infant congenital esophageal atresia surgery. *2015 IEEE International Conference on Robotics and Automation*, Hong Kong, p 2430–2435
21. Emami S, Ievgen K, Saragih J (2012) Non-rigid face tracking, mastering openCV with practical computer vision projects. *Chapter 6:189–233*
22. Gower J (1975) Generalized procrustes analysis. *Psychometrika* 40(1):33–51
23. Robinson D (1964) The mechanics of human saccadic eye movement. *J Psychol* 1964(174):245–264
24. Marchand E, Chaumette F (2002) Virtual visual servoing: a framework for real-time augmented reality. *Comput Graph Forum* 21(3):289–297
25. Daniilidis K, Krauss C, Hanse M, Sommer G (1998) Real-time tracking of moving objects with an active camera. *Real-Time Imaging* 4(1):3–20
26. Marescaux J, Leroy J, Gagner M, Rubino F, Mutter D, Vix M, Butner SE, Smith MK (2001) Transatlantic robot-assisted telesurgery. *Nature* 413:379–380
27. Rempel D, Willms K, Anshel J, Jaschinski W, Sheedy J (2007) The effects of visual display distance on eye accommodation, head posture, and vision and neck symptoms. *Hum Factors J Hum Factors Ergon Soc* 49(5):830–838
28. Barbato G, Ficca G, Muscettola G, Fichelle M, Beatrice M, Rinaldi F (2000) Diurnal variation in spontaneous eye-blink rate. *Psychiatry Res* 93(2):145–151
29. Hashizume M (2005) Fundamental training for safe endoscopic surgery, *Innovative Medical Technology*, Graduate School of Medical Science Kyushu University, p 49 (in Japaeses)
30. Donnez J (2007) Human assistance in laparoscopic surgery, *Atlas of operative laparoscopy and hysteroscopy*, 3rd edn, p 409