

# Eye-Hand Coordination Analysis According to Surgical Process in Laparoscopic Surgery Training

Takafumi Marutani, Hiromi T. Tanaka, Nobutaka Shimada,  
Masaru Komori, Yoshimasa Kurumi and Shigehiro Morikawa

**Abstract** Laparoscopic surgery is a recent common surgical technique as one of minimally invasive surgery. This surgery is performed under the situation, such as the limited field of view egocentric video from a laparoscope, and poor force sensation from surgical instruments. Especially, the techniques called “Eye-hand coordination”, which represents the collaboration of gazing action and hand action (instrument manipulation), is very important technique in the surgery. In this paper, we analyze gazing actions and instruments manipulation for evaluating the “Eye-hand coordination” technique in order to support the training of Laparoscopic surgery.

## 1 Introduction

Recently, laparoscopic surgery has become a common surgical technique as one of minimally invasive surgery. There are a number of advantages (e.g. smaller incision, less pain, less hemorrhage, and shorter recovery time) to the patient with Laparoscopic surgery comparing to an open procedure. However, advanced surgical techniques are required in the surgery, because doctors must perform the surgery according to the limited field of view from an endoscope with poor force sensation from surgical instruments. Therefore, it is important to support the training of the laparoscopic surgical techniques.

Generally, the trainings of the surgical techniques are mainly supported by experts' one-by-one coaching. Through the training, trainee learn the techniques about collaboration of gazing action and hand action, which is called Eye-hand coordination. To support the learning of the technique, various VR Laparoscopic surgery simulators have been developed. Almost of all VR-based Laparoscopic surgery simulator has the function to assess the surgical techniques of trainees from various

---

T. Marutani (✉) · H.T. Tanaka · N. Shimada  
Ritsumeikan University, Kyoto, Japan  
e-mail: marutani@cv.ci.ritsumei.ac.jp

M. Komori · Y. Kurumi · S. Morikawa  
Shiga University of Medical Science, Otsu, Japan

measurements (e.g. operating time, path length, and so on). These measurements can evaluate instrument manipulation technique. However the gazing action cannot be evaluated in spite of its importance. Therefore, we aim to analyze the Eye-hand coordination technique of experts in order to evaluate the trainee's Eye-hand coordination technique.

## 2 Related Works

There are various researches to support the training of laparoscopic surgery. Ahmad et al. investigate various assessment methods about surgical techniques [1]. In that research, they support trainees by assessing trainees' surgical techniques manually. Tagawa et al. developed a VR-based Laparoscopic surgery simulator, which can display visual, and force information in order to achieve laparoscopic surgical training of basic tasks [2]. They support trainees by developing training environment. However, it is not sufficient to substitute experts' coaching task by using the environments.

Neumuth et al. proposed Surgical Process Model, which represent operational steps and actions in surgical operation for various reasons: the evaluation of surgical assistant system, the control of surgical robots, and so on [3]. Surgical process represents a part of the operational step of surgery. However, the surgical process cannot suggest how trainees do in each operational step. Then it is not sufficient to describe the detail of Eye-hand coordination techniques.

Thijssen et al. reported various measurements in commercial VR simulators [4]. From their reports, almost of all VR-based Laparoscopic surgery simulator has the function to assess the surgical techniques of trainees from various measurements (e.g. operating time, path length, and so on). These measurements can evaluate the result of trainee's operation. However the measurements cannot evaluate the Eye-hand coordination techniques, because these measurements do not focus on gaze actions.

About detailed technical assessment research using electronic devices, Datta et al. analyzed technical differences between novice and expert according to operating time, path length of instruments, and number of movements in Vein patch insertion [5]. Egi et al. analyzed technical differences between novice and expert according to the trajectories of forceps [6]. These researches focused on how "instruments" are operated. Then, their systems also cannot evaluate the Eye-hand coordination techniques, because these measurements do not focus on gaze actions.

As for the research about gaze action analysis in Laparoscopic surgery, Ibbotson et al. [7] investigate when the doctor, who is performing the laparoscopic operation, look at object during the surgery. Tien et al. [8] analyze the difference of gaze action among in performing the surgery and watching the surgery video. They report some specific tendency of gaze actions about laparoscopic surgery, but focus on only the gaze movements. Then the relationship between gaze action and hand movements is

not consider in their research. Therefore, to support the training of Eye-hand coordination, the framework for analyzing both of gaze and hand movements are necessary.

In this paper, at the first step of realizing the framework, we try to analyze the Eye-hand coordination technique of experts in order to reveal the important points about Eye-hand coordination technique.

### 3 Eye-Hand Coordination Analysis

#### 3.1 Overview of Our Training Support Framework

In our research, we try to analyze trainee’s techniques according to surgical processes. Figure 1 shows an overview of our laparoscopic surgical training support framework.

At first, we archive the expert’s operation according to surgical process. In concrete, we record expert’s operations of instruments and movements, gazing actions, and transformations of organs through surgical simulator. Then, we can describe a surgical process as time-sequential data of expert’s actions about gaze and instruments, and transformations of organs. At the next time, by analyzing the trainee’s action according to expert’s surgical process, and by displaying visual and force information of expert’s surgical techniques, trainees can train their techniques without expert’s assistances. In this paper, we analyze the Eye-hand coordination techniques according to surgical processes, as the first step of our framework.

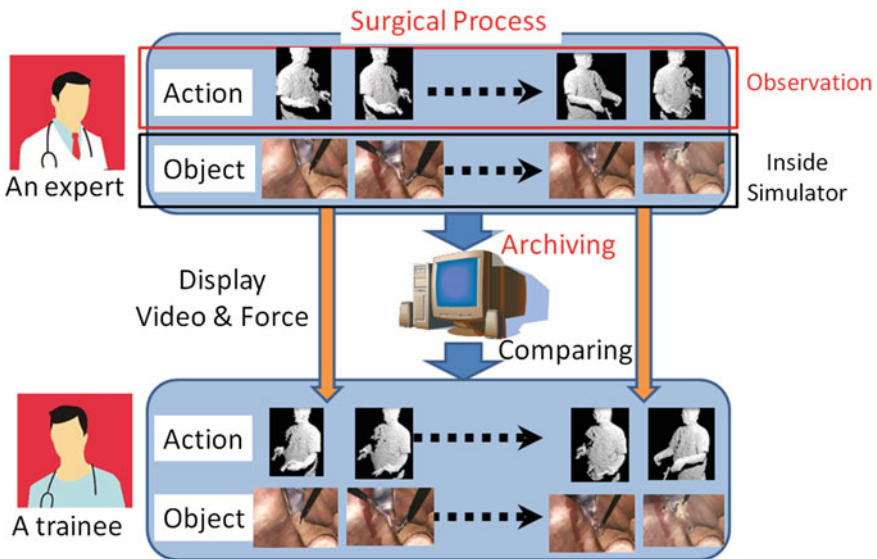


Fig. 1 Framework of our surgical techniques training support

### 3.2 Surgical Process Model

In our research, we represent surgical process by using Neumuth's Surgical Process Model [9]. Neumuth's Surgical Process Model represents surgical steps as computational representation in order to record and analyze surgical procedures. Neumuth's surgical process consists from following elements.

- *functional*:  
What is done in the work step (e.g. dissect, position).
- *organizational*:  
Who is performing the work (e.g. surgeon, right hand).
- *operational*:  
Which instruments, devices, or resources are used (e.g. forceps, scalpel).
- *spatial*:  
Where the step is performed (e.g. dura, cranial nerve).
- *behavioral*:  
When the step is performed (e.g. start, end).

In this paper, we focus on “*spatial*” element, because the element represent the target object in the process, and the position of the target object becomes the basis of position analysis.

### 3.3 Analysis Procedures

In order to analyze Eye-hand coordination, we perform following procedures. Figure 2 shows the overview of our analysis. At first step, we calculate the positions of feature points (target, instrument's tip, and target) in each frame of ego-centric video. By this procedure, we can acquire many sample data of gaze and hand movement through a surgical process. A sample data consists of followings features acquired from a frame; target position  $(x_t(t), y_t(t))$ , tip's position  $(x_f(t), y_f(t))$ , gaze point  $(x_g(t), y_g(t))$ , and timestamp  $t$ . At the second step, we normalize the sample data by translating the tip's position and gaze point based on the position of target, and transforming the value range of tips and gaze in  $[-0.5;0.5]$  and  $t$  in  $[0;1]$ . At the last step, we perform clustering to the normalized sample points. Through the clustering, the data, in which the gaze and tip are closely related, may be revealed. The details of these procedures are explained below.

#### Position detection

In order to start analysis, we must acquire the position data of target, tip, and gaze point. In this research, we acquire the gaze point data by eye tracking device, and target position by manual operation, because the target changes drastically about its position and shape, and is hard to extract by computational method. In opposite, the

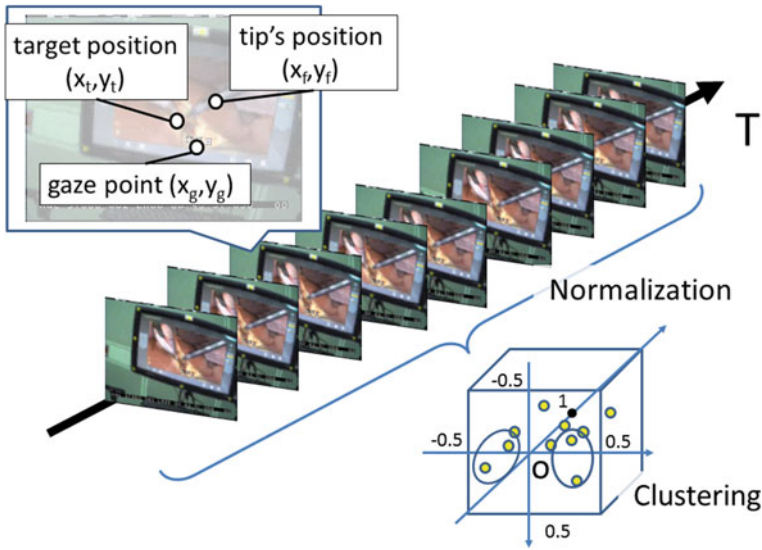


Fig. 2 Overview of Eye-hand coordination analysis

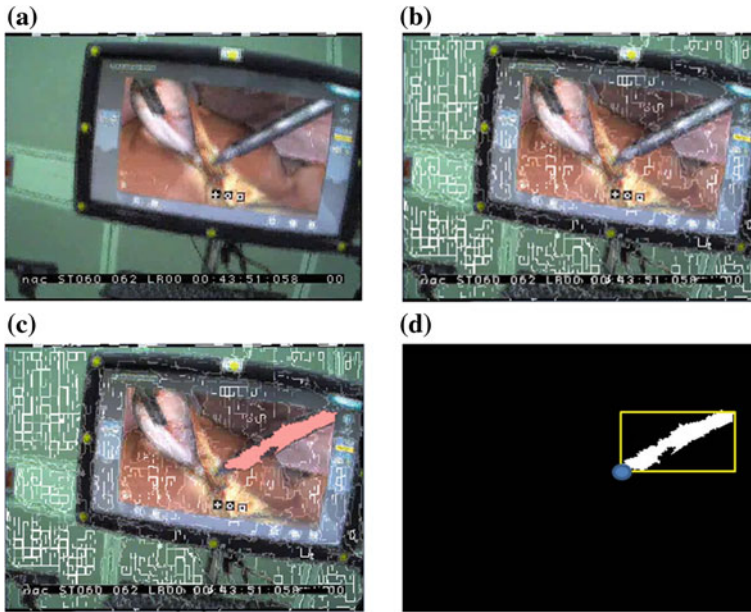
tip's position is relatively easy to detect, because the instruments are rigid objects. To detect the tip's position, we apply Fathi's method [10]. Figure 3 shows the position detection process. At first, we construct superpixel based on similarity among neighboring pixels. After that, we segment the instrument area based on harmonic solution by selecting some superpixels as instruments, and some other superpixels as background. At last, we detect a position as tip's position based on the segmented region and instrument constraint (its' long and narrow shape and it's direction in the working area).

### Normalization

In surgical process, the target position and the operating time differ by various reasons (manipulation, camera angle, skill level, and so on). Therefore, we analyze gaze position and tip's position by relative position from target position and reduce the scale differences among position data and time data. Then, we suppose that there are sample points  $\{\mathbf{P}(t) = (x_g(t), y_g(t), x_f(t), y_f(t), x_t(t), y_t(t), t)\}$  ( $0 \leq x_g(t), x_f(t), x_t(t) \leq X_{max}, 0 \leq y_g(t), y_f(t), y_t(t) \leq Y_{max}, t_s \leq t \leq t_e$ ), we calculate normalized sample point  $\hat{\mathbf{P}}(t) = (\hat{x}_g(t), \hat{y}_g(t), \hat{x}_f(t), \hat{y}_f(t), \hat{t})$  by followings;

$$\hat{x}_{\{g,f\}}(t) = \frac{x_{\{g,f\}}(t) - x_t(t)}{X_{max}}$$

$$\hat{y}_{\{g,f\}}(t) = \frac{y_{\{g,f\}}(t) - y_t(t)}{Y_{max}}$$



**Fig. 3** Forceps' tip detection process (a) Original data (b) Superpixel (c) Segmentation based on manual labeling (d) Tip detection based on instrument's constraint

$$\hat{t} = \frac{t - t_s}{t_e - t_s}$$

## Clustering

By clustering the data points  $\{\hat{\mathbf{P}}(t)\}$ , we aim to concentrate some specific combination of gaze movements and tip's movements. In this paper, we choose  $k$ -means algorithm as the clustering algorithm. After the clustering, we analyze each cluster by comparing the datapoints in the cluster and the egocentric video of corresponding duration.

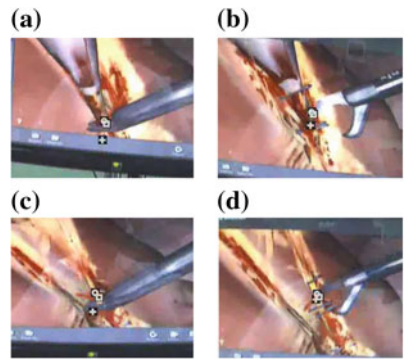
## 4 Experiment

To analyze Eye-hand coordination by our method, We measured the 3 experts' eye movement under VR laparoscopic cholecystectomy surgery training. We use commercial VR surgical simulator (Lap-Mentor [11]) and eye-tracking device (EMR-9 [12]) for the measurement. Figure 4 shows the experimental settings. The subjects demonstrate a series of surgical processes about laparoscopic cholecystectomy surgery by VR simulator. The details of the training surgical

**Fig. 4** Experimental settings (the case of expert clinician)



**Fig. 5** Example of screenshot about surgical processes. **a** Clipping Cystic Duct. **b** Cutting Cystic Duct. **c** Clipping Cystic Artery. **d** Cutting Cystic Artery



processes are as follows; ablating fats, clipping cystic duct, cutting cystic duct, clipping cystic artery and cutting cystic artery. In these processes, we select 4 process for analysis, because there are clear target position in these processes. (The screenshots of the selected processes are shown in Fig. 5.) All subjects demonstrate through operation by 7–15 min. The sampling rate for eye movements is 60[Hz]. The effective number of pixels of head mounted camera view is 640[H] by 480[w], and the horizontal angle of view is 62°. We divided the eye movement data into each process manually. And the target position in each surgical process are attached manually.

Tables 1, 2 and 3 are the details of acquired data. In each table, “ClipCD”, “CutCD”, “ClipCA” and “CutCA” represent the surgical processes “clipping cystic duct”, “cutting cystic duct”, “clipping cystic artery” and “cutting cystic artery”. Table 1 shows the elapsed time of subjects for completing each surgical process ([sec]). In addition, we measure the distance of eye movements and instrument’s movement, because these measurement are used for technical assessment [4]. Table 2 shows the total eye movement distance in egocentric video of subjects for completing each surgical process([pixel]). Table 1 shows the total instrument movement distance in egocentric video of subjects for completing each surgical process([pixel]).

From the Table 1, novices spent less time to complete surgical processes than experts in many cases. The reasons for this may be as follows; All subjects in this experiment use the simulator for the first time. Therefore, novices can be familiar with the simulator soon. On the other hand, experts have trouble with controlling

**Table 1** Elapsed time for completing surgical processes ([sec])

subject	Process			
	ClipCD	CutCD	ClipCA	CutCA
expertA	0:21	0:15	0:45	0:09
expertB	0:30	0:25	0:26	0:15
expertC	0:31	0:10	0:27	0:07
noviceA	0:28	0:15	0:13	0:13
noviceB	0:19	0:07	0:46	0:06
noviceC	0:17	0:03	0:17	0:02

**Table 2** Total eye movement distances in surgical processes ([pixels])

subject	Process			
	ClipCD	CutCD	ClipCA	CutCA
expertA	7,706	5,748	8,987	1,539
expertB	4,013	2,137	1,887	1,402
expertC	3,488	800	2,053	460
noviceA	5,397	2,567	2,821	2,165
noviceB	4,265	3,362	10,858	816
noviceC	2,245	314	2,390	175

**Table 3** Total instrument movement distances in Surgical Processes ([pixels])

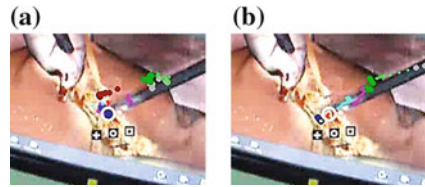
subject	Process			
	ClipCD	CutCD	ClipCA	CutCA
expertA	3,735	2,465	4,919	433
expertB	2,316	1,879	1,420	1,096
expertC	2,518	682	1,761	404
noviceA	2,590	1,012	1,895	2,145
noviceB	1,784	1,121	3,106	498
noviceC	1,500	243	1,863	102

the simulator by feeling gap of real surgery. Especially, expertA failed in holding gallbladder many times. It is also be revealed that all subjects spent more time in clipping process than that in cutting process. This is because the clipped positions in clipping process affects the difficulty of cutting process. Therefore subjects tend to spend much time for deciding clipping positions.

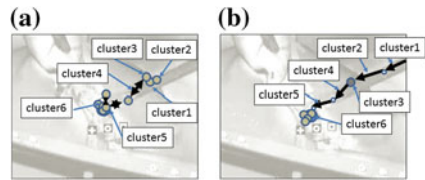
From the Tables 2 and 3, expertA moved his eye and hand more than other subjects. However, as mentioned in Table 1, he had trouble to control the simulator. Therefore, we exclude his data for analysis. Comparing movement distances among experts and novices, the distances of experts are relatively shorter than that of novices



**Fig. 6** Point distributions of gaze and forcep’s tip about one subject in a surgical process (cutting cystic artery). **a** Gaze points distribution. **b** Tip’s positions distribution



**Fig. 7** Movements of gaze and forcep’s tip about one subject in a surgical process (cutting cystic artery) **a** Gaze points movement. **b** Tip’s positions movement

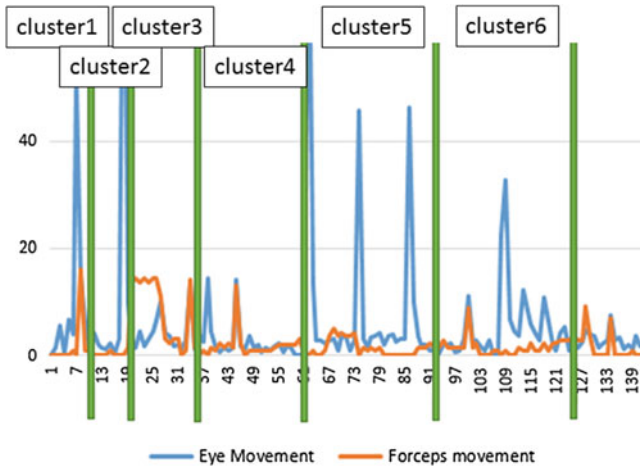


when we take into account of elapsed times. Especially, the eye movement distances of experts are obviously less than that of novices. This is because the experts can fix their gaze in particular points. Experts understood where to watch based on their surgical experiences.

Figure 6 shows the clustering result ( $k = 10$ ) in cutting cystic artery process of a subject. In this figure, the left figure shows the position distribution of gaze point (relative position to target position.) And the right figure shows the position distribution of forceps’ tip (relative position to target position.) The points of same color in both figures are belonging to same cluster. From this result, there are some cases that the tips’ position approach to the gaze fixation points. This means that experts fix their attention to particular point, and then approach the instruments towards the point in a straight line.

Figure 7 shows the detailed movements of gaze and instruments in cluster 1–6. (In this experiments, we choose  $k = 10$  for clustering. However the movements of gaze and instrument are very small in cluster 7–10. Then we omit their data in Fig. 7). In this figure, arrows show the movement and large circle represent the fixation of gaze and instrument. At first, expert look at the halfway position to the target and bring the instruments to the position in cluster 1 and 2. After he checked the condition of instrument tip in cluster 3, he checked the next position (around target) and move the instruments to the position in cluster 4. Then, he repeat checking the condition of the instrument tip and target position, and move the instrument to the target position in cluster 5. At last, he checked the condition around the target. This means that he did not move the instrument to the target position in one breath. This is because he avoided puncturing other organs. In Laparoscopic surgery, the workspace is very narrow and the distance among target and other organs is very short. Therefore, large movement of instrument may be dangerous.

For more detailed analysis, we make a graph, that shows the velocity transitions of gaze and instruments of same data, as shown in Fig. 8. In this graph, the vertical axis represents the velocity of gaze point and tip of instrument ([pix/frame]), and the horizontal axis represents the time ([frame]) from start time. The vertical lines



**Fig. 8** Velocity transitions of gaze and forceps (closeup) [ $\text{pix}/\text{frame}$ ]

in the graph show the borderline among clusters. From this graph, the clusters are mainly divided according to time sequences, and the width of cluster depends on the value of variances. In the graphs of cluster 1, the velocity of gaze make a peak before the border. This timing is that the expert looks at the target position. After that, the expert operate the instrument with stable speed (at the first half of cluster 2.) This means that the expert can control the instrument smoothly, and move it to appropriate position courteously. In addition, in the graph of cluster 4 and 5, there are some peaks about velocity of gaze. These timing are that the expert checks the state of instrument's tip or checks the state around the target position. During the timing, the velocity of tip's position are almost 0. This means that the expert can fix the instrument while checking surgical situation by their eyes. These tendency are also found in other processes and other experts.

## 5 Conclusions

In this paper, we analyze Eye-hand coordination based on egocentric video during laparoscopic surgery training according to surgical process. In order to analyze eye-hand coordination, we calculate the positions of feature points (target, instrument's tip, and target), normalize them based on the target position and value range, and perform clustering by using conventional clustering algorithm. After that, we compare the distance of eye movement and instrument among 3 experts and 3 novices, and analyze the cooperative movements of gaze and instrument in experts' training.

From the result of analysis, we could find some relationship between gaze movements and instrument movements. However, this analysis is based on only limited surgical processes. Therefore, we will apply our method to other surgical processes and check the differences among the experts and the novices for further analysis.

## References

1. Ahmmad, S.N.Z., Ming, E.S.L., Fai, Y.C., Harun, F.K.C.: Assessment Methods for Surgical Skill, World Academy of Science, Engineering and Technology, vol. 58, pp. 752–758 (2011)
2. Tagawa, K., Tanaka, H.T., Kurumi, Y., Komori, M., Morikawa, S.: Laparoscopic Surgery Simulator Using First Person View and Guidance Force. *Medicine Meets Virtual Reality*, vol. 20, pp. 431–435 (2013)
3. Neumuth, D., Loebe, F., Herre, H., Neumuth, T.: Modeling surgical processes: a four-level translational approach. *Artif. Intell. Med.* **51**, 147–161 (2011)
4. Thijssen, A.S., Schijven, M.P.: Contemporary virtual reality laparoscopy simulators: quicksand or solid grounds for assessing surgical trainees? *Am. J. Surg.* **199**, 529–541 (2010)
5. Datta, V., Mackay, S., Mandalia, M., Darzi, A.: The use of electromagnetic motion tracking analysis to objectively measure open surgical skill in the laboratory-based model. *Am. Coll. Surg.* **193**, 479–485 (2001)
6. Egi, H., Okajima, M., Yoshimitsu, M., Ikeda, S., Miyata, Y., Masugami, H., Kawahara, T., Kurita, Y., Kaneko, M., Asahara, T.: Objective assessment of endoscopic surgical skills by analyzing direction-dependent dexterity using the hiroshima university endoscopic surgical assessment device (HUESAD). *Surg. Today* **38**, 705–710 (2008)
7. Ibbotson, J.A., MacKenzie, C.L., Cao, C.G.L., Lomax, A.J.: Gaze patterns in laparoscopic surgery. In: *Medicine Meets, Virtual Reality*, pp. 154–160 (1999)
8. Tien, G., Atkins, M.S., Jiang, X., Khan, R.S.A., Zheng, B.: Identifying eye gaze mismatch during laparoscopic surgery. In: *Medical Meets, Virtual Reality*, pp. 453–457 (2013)
9. Neumuth, T., Loebe, F., Jannin, P.: Similarity metrics for surgical process models. *Artif. Intell. Med.* **54**, 15–27 (2012)
10. Fathi, A., Balcan, M.F., Ren, X., Rehg, J.M.: Combining self training and active learning for video segmentation. *Proc. BMVC* 78.1–78.11 (2011)
11. Lap-Mentor. <http://symbionix.com/simulators/lap-mentor/>
12. EMR-9. [http://www.eyemark.jp/product/emr\\_9/index.html](http://www.eyemark.jp/product/emr_9/index.html)