Surgery Recording without Occlusions by Multi-view Surgical Videos

Tomohiro Shimizu[1], Kei Oishi[1], Ryo Hachiuma[1][a], Hiroki Kajita[2], Yoshihumi Takatsume[2] and Hideo Saito[1][b]

[1]*Faculty of Science and Technology, Keio University, Yokohama, Kanagawa, Japan*
[2]*School of Medicine, Keio University, Shinanomachi, Tokyo, Japan*

Keywords: Camera Scheduling, Dijkstra's Algorithm, Multi-viewpoint Camera.

Abstract: Recording surgery is important for sharing various operating techniques. In most surgical rooms, fixed surgical cameras are already installed, but it is almost impossible to capture the surgical field because of occlusion by the surgeon's head and body. In order to capture the surgical field, we propose the installation of multiple cameras in a surgical lamp system, so that at least one camera can capture the surgical field even when the surgeon's head and body occlude other cameras. In this paper, we present a method for automatic viewpoint switching from multi-view surgical videos, so that the surgical field can always be recorded. We employ a method for learning-based object detection from videos for automatic evaluation of the surgical field from multi-view surgical videos. In general, frequent camera switching degrades the video quality of view (*QoV*). Therefore, we apply Dijkstra's algorithm, widely used in the shortest path problem, as an optimization method for this problem. Our camera scheduling method works so that camera switching is not performed for the minimum frame we specified, and therefore the surgical field observed in the entire video is maximized.
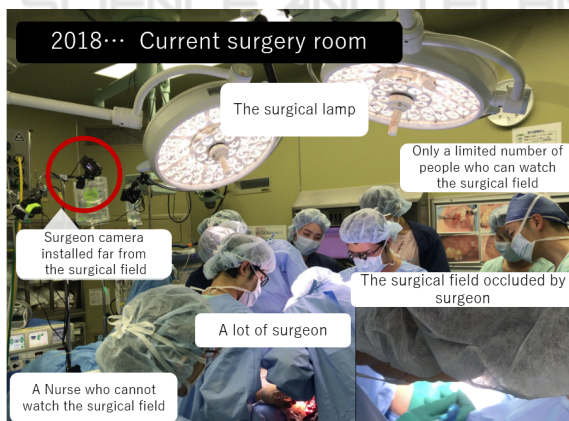
1 INTRODUCTION

Figure 1: The current surgery room.

Recording surgery is an indispensable task for a variety of reasons, such as education, sharing surgery technologies/techniques performing case studies of diseases, and evaluation of the medical treatment, etc. Video recording is one of the most important ways of recording surgery, so a number of cameras have been used for recording surgery.

The most important target of video recording is the surgical field to which medical doctors operate with medical tools and their hands, such as an abdominal operation and an orthopedic surgery. However, as shown in Figure 1, recording the surgical field with the camera is often difficult because there are usually several medical doctors around the surgical field.

We may be able to use cameras mounted on the medical doctors head, but the videos captured with such cameras are always affected by motion blur because of fast and wide head movements, so the video is not always useful for recording purposes. It is assumed that camera installed right above the surgical field is suitable to record it. Therefore we might be able to put a camera at the best position to capture the target surgery field, but such cameras will prevent the operation of medical doctor and cannot actually be installed since a surgical lamp is installed right above the most surgical field. After all, there is almost no position where such recording cameras can be placed for the recording surgery target area in most of surgery.

Even in such difficult situations regarding the placement of recording cameras, we turn our attention to surgical lighting systems, which have multiple light

[a] https://orcid.org/0000-0001-8274-3710
[b] https://orcid.org/0000-0002-2421-9862

bulbs for illuminating the surgical field from multiple directions, so that shadows are reduced. This implies that at least one of the multiple light will always illuminate the surgical field. Therefore, in this study, we first create a surgical lamp with a camera in which multiple cameras are attached to the surgical lamp. At that time, by attaching one camera to the light unit of the surgical lamp, it is guaranteed that any camera always captures the surgical field as long as the surgical field is illuminated. The created surgical lamp had five light units, so five cameras were attached to the surgical lamp.

We propose a method to automatically switch the surgical video of multiple viewpoints taken using the created surgical lamp. At that time, it is known that frequent switching of the camera video will reduce the quality of view (QoV) of the video. Switching video creation is divided into two processes: scoring and scheduling.

First, in scoring, segmentation of the surgical field was performed using the method of Li *et al.*(Li and Kitani, 2013) and the ratio of the area other than the surgical field of each frame was used as the frame score. After that, in scheduling, we used a graph for selecting best view, and Dijkstra's algorithm, usually used in the shortest path problem, was applied. In the graph generation for scheduling, each frame is a node, and we changed the edge connection between when the camera sequence was switched and when it was not switched. The obtained score was used as the edge weight. By optimizing this graph using Dijkstra's algorithm, camera switching is not performed for the minimum frame we specified, and camera scheduling is performed to maximize the area of the surgical field observed in the entire video.

In the experiment, the effectiveness of the proposed method is verified by automatic switching of multiple cameras. This system was installed at Keio University School of Medicine and surgery was recorded. A camera switching video was created from the captured video and it is confirmed that the surgical field can be observed throughout the surgery. After that, we conducted a questionnaire on viewing quality to 14 active doctors using camera switching videos and video shot with one camera, and verified the minimum number of frames with the highest viewing quality.

Our contributions are as follows:

1. We present a novel surgery recording system in which multiple cameras are attached to the surgical lamp.

2. We propose a method which switches multi-view videos considering QoV using Dijkstra' algorithm.

3. Qualitative evaluation shows that the video created by our method does not select the occluded frame while keeping the QoV. Moreover, the user test of the doctors on the quality of the switching video quantitatively verify the effectiveness of our proposed method.

The rest of the paper is organized as follows: we first present related work in Section 2. Next, we present details of surgical recording systems in Section 3 and our proposed method in Section 4. We then conduct experiments on creating switching video to validate our method. At the School of Medicine in our university, we recorded multi-view surgical videos with multiple cameras mounted on the surgical lamp. Then, we presented our experiment, results and discussions in Section 5. Last, we conclude this paper in Section 6.

2 RELATED WORK

In this paper, we present a new surgery recording system using multiple cameras attached to the surgical lamp. In Section 2.1, we introduce the conventional surgical recording systems. In addition, we proposed a novel camera switching method to automatically select the best view. In Section 2.2, we introduce the conventional camera switching method to clarify the novelty of the proposed method.

2.1 Surgical Recording Systems

As doctors have a duty to teach their surgical skills to future generations, it is important to record the surgery and generate video for trainees. Moreover, the usefulness of surgery recording has been recognized in terms of reviewing. The surgery, such as laparoscopic surgery, which is performed through the endoscope camera can be easily recorded. However, the surgery that the doctor directly sees, such as the surgery that involves dissection, it is difficult to record due to the presence of the surgeon and spatial restrictions.

Kumar *et al.*(Kumar and Pal, 2004) designed a camera arm system with a camera mounted on the arm to record the surgery. The camera arm is set to the position that does not get occluded by the doctor and is often set to the position far from the surgical field. In addition, it is troublesome to position the camera according to the surgical situation and the environment. Therefore, Byrd *et al.*(Byrd et al., 2003) presented a system of mounting a camera on the surgical lamp. However, the view is occluded by the doctor's head

or body and it is difficult to observe the surgical field with a single camera constantly.

Other attempts have also been made to record surgery with a surgical field camera placed between the eyes of a doctor. The camera of such recording systems were not high resolution and did not produce good video quality (Matsumoto et al., 2013; Murala et al., 2010) because of the limited hardware system. In addition, doctors had to perform surgery with interference by the surgical camera itself and its code. Nair *et al.*(Nair et al., 2015) recorded the surgery by putting a high-resolution camera (GoPro Hero 4) on the doctor's head. The doctor's head moves greatly during the surgery, and camera cannot always shoot the surgical fields, and video is always shaking. Therefore, video recorded by it can be offending for viewers.

In our proposed system, multiple cameras are attached to multiple lights mounted the surgical lamp, and the surgery is recorded by them. While one of the lights illuminates the surgical field, we assume that the surgical field can be observed one of the attached cameras. Therefore, the surgical field is recorded without disturbing the surgery.

2.2 Multiple Camera Switching

In recent years, multiple cameras are introduced in any place, such as office environments, sports stadiums, and downtown areas. Instead of its convenience, it is difficult to extract only the necessary information from the huge amount of video sequences from a lot of cameras. Therefore, camera self-control technology, such as the automatic viewpoint switching video generation, and highlight video generation technology, are regarded as important issues (Chen and Carr, 2014).

Liu *et al.*(Liu et al., 2001) interviewed pro editors to gain knowledge of video editing, and implemented the camera switching rules. Based on the rules, they switched the viewpoint of three cameras shooting the speaker, the audience, and the entire in conference video. Doubek *et al.*(Doubek et al., 2004) observed moving objects using multiple fixed cameras in an office environment. Selection of a camera was performed based on the score of each camera, and the resistance coefficient was introduced so that the switching of a camera may be performed only when the score changed significantly. However, such camera switching strategies may occur frequently if cameras with competing scores existed, which may reduce the overall *QoV*.

In order to suppress camera switching frequency, Jiang *et al.*(Jiang et al., 2008) proposed a cost func-

tion calculated based on the size, posture, orientation, etc. of the target, and controlled the frequency of the camera switching while considering *QoV*. Daniyal *et al.*(Daniyal and Cavallaro, 2011) calculated the visibility score of an object using a multivariate Gaussian distribution model, and used the partial observation Markov decision process for the camera switching to maximize the visibility score while suppressing the camera switching frequency. Although these methods selected the optimal view using past sequential information, the switching should be conducted using not only past but also future information. Compared to the conventional switching method, our method uses both past and future frames to switch cameras so that a higher *QoV* can be generated.

Also, *QoV* and user-specified weights for camera switching may change depending on the target scene, because they detected the event for calculating *QoV*. For that reason, the hyper-parameter of the method which determines camera switching depends on the surgical scene, and it is difficult for non experts to determine. On the other hand, in the proposed method, since the minimum frame during that camera switching does not perform is specified and optimization is performed for the entire frame, the camera switching frequency does not change depending on the target surgery scene.

3 MULTI-CAMERA RECORDING

As shown in Figure 2, we attached multiple cameras to multiple lights mounted the surgical lamp. Thereby, as long as the surgical field is illuminated by one of the lights, our proposed camera recording system shoots the surgical field. Compared to the previous camera recording system (Matsumoto et al., 2013) which attached cameras to the doctor's head, our system does not bother the doctors during surgery while maintaining visibility of the surgical field.

4 PROPOSED METHOD

Figure 3 shows the overview of the proposed method which consists of two components: camera scoring and camera switching. The multiple surgical videos are captured from the our capturing system (Section 3). To switch between camera sequences to generate the best video quality, the frame in each sequence has to be scored. The score represents how the surgical field can be seen in the image. In our methods, first, the score is estimated against each frame in each sequence. Next, the frame is selected sequentially using the score.

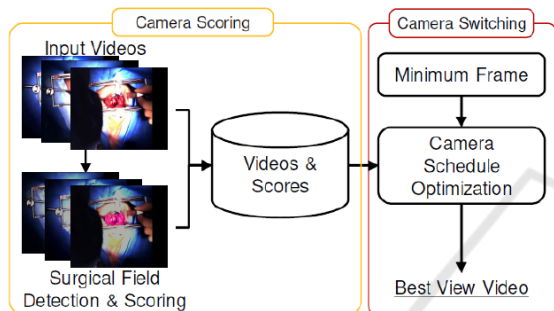Figure 2: Multiple cameras mounted the surgical lamp.

Figure 3: The overview of the proposed method.

4.1 Camera Scoring

To generate a single video from videos recorded by multiple cameras, the frame captured from each camera sequence is scored which represents how the surgical field can be seen. The score is used as the switching criterion. In the proposed method, after segmenting the surgical field, the camera is scored based on the number of pixels of the segmentation mask. The segmentator F that performs segmentation of the surgical field is defined as follows:

$$F(I(i,j)) = \begin{cases} 1 & surgical\,field \\ 0 & otherwise \end{cases} \qquad (1)$$

Here, I denotes the input RGB image, and (i,j) is the pixel coordinate in I.

In this paper, we chose the method proposed by Li *et al.*(Li and Kitani, 2013) as the segmentator F. They used hand color and texture information for learning, and performed hand segmentation from the learning model. The surgical field changes shape over time, and it is difficult to keep detecting it. However, the area other than the surgical field is covered with cloth and the color contrast between the surgical field and other parts is large. Moreover, although their method is often difficult to detect the target due to environmental changes and high contrast shadows, the sur-
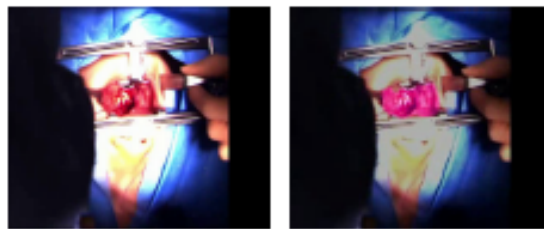

Figure 4: The segmentation result of the surgical field; (left): input; (right): output.

gical field is always illuminated, and the surgery is always performed in the surgical room. Therefore, their method, which was performed only from color and texture information, is well suited to detect surgical fields. Figure 4 shows the segmentation result of the surgical field.

In the proposed method, the score is a ratio of non segmented pixels in the image I. A score s_t^c at timestep t of a camera c is defined as follows:

$$s_t^c = 1 - \frac{\sum_i^w \sum_j^h F(I_t^c(i,j))}{wh}, \qquad (2)$$

where w and h are the width and height of the image I_t^c. The point is, the larger the area of the surgical field is, the smaller the score is.

4.2 Camera Switching

The simple approach to achieve camera switching is taking the minimum score at every timestep. However, this approach does not consider the scores at the previous/next frames so that the selected camera may change over and over during the sequence; the QoV of video will be decreased.

Hence, the camera switching should be considered using previous/next frames. In the proposed method, the scores of all camera sequence were calculated in advance. Then, the switching video is generated so that the sum of the scores of the selected sequences was minimized. However, as above, frequent camera switching occurs in this approach. For this reason, video sequence is selected by using a graph so that camera switching is suppressed. We changed the edge connection between when the camera sequence was switched and when it was not switched. Then, by optimizing so that the score is minimized throughout the graph, we created the video sequence in which camera switching is suppressed. Therefore, we propose a combined optimization method that applies to Dijkstra's algorithm (Ahuja et al., 1990).

4.2.1 Graph Generation

In the proposed method, the edge connection was changed between when the camera sequences was
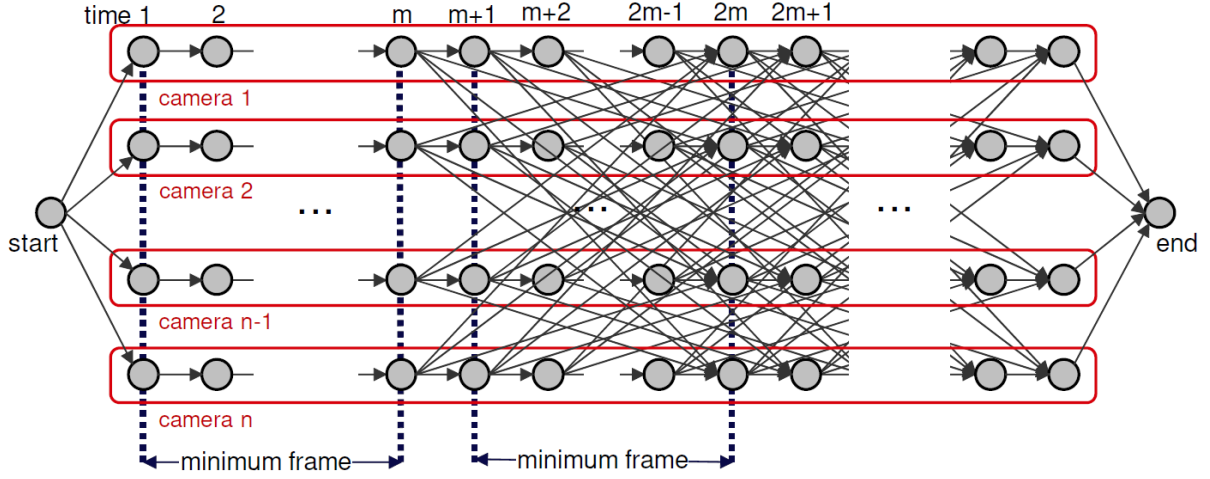
Figure 5: The expected result of generating graph.

switched and when it was not switched. Figure 5 shows the expected graph generation results. Each node $V_{c,t}$ has information on camera number c and the timestep t. It determines whether or not to connect edges between nodes, and its weight. Between each node, if the edge E is connected to the node of the same camera sequence, it is connected to the next node. On the other hand, if it is connected to the node of the different camera sequence, it is connected to the node ahead of the minimum number of frames. Each edge E is defined as follows:

$$E(V_{c1,t1}, V_{c2,t2}) =$$
$$\begin{cases} 1 & (V_{c1,t1} = start) \\ & \cup((c_1 = c_2) \cap (t_1 + 1 = t_2)) \\ & \cup((c_1 \neq c_2) \cap (t_1 + m = t_2) \\ & \quad \cap (t_2 \leq 2m)) \\ & \cup(V_{c2,t2} = end) \\ 0 & otherwise \end{cases} \quad (3)$$

m is the minimum number of frames specified by the user, and $E(V_{c1,t1}, V_{c2,t2})$ denotes the edge that connects between node $V_{c1,t1}$ and node $V_{c1,t1}$. Edges are connected when $E = 1$.

At this time, the weight W of edge E is defined as follows:

$$W(E(V_{c1,t1}, V_{c2,t2})) =$$
$$\begin{cases} 1 & (V_{c1,t1} = start) \\ & \cup((c_1 = c_2) \\ & \quad \cap((t_1 + 1 = t_2)) \\ \sum_{i=t_1+1}^{i=t_2} s_i^{c_2} & ((c_1 \neq c_2) \cap (t_1 + m = t_2) \\ & \quad \cup(t_2 \leq 2m)) \\ 0 & V_{c2,t2} = end \end{cases} \quad (4)$$

4.2.2 Optimization

We apply Dijkstra's algorithm to the graph generated in Section 4.2.1. Dijkstra's algorithm is the search algorithm for solving the single-source shortest path problem when the weight of the edge in graph is non-negative.

In the proposed method, the camera number array is obtained by acquiring the information of the nodes, excluding the start and end nodes, after optimization by Dijkstra's algorithm. However, since there are edges connected to the node ahead of the minimum number of frames, the array is smaller than the number of frames of the actual video. Therefore, at the position where the camera number in the array has changed, the skipped camera number is added to the array.

5 EXPERIMENTS

In this section, we describe the details of the experiments we conducted to verify the effectiveness of the proposed method. At the School of Medicine in our university, we verified the effectiveness of the proposed method using multi-view surgical videos shot with multiple cameras mounted on the surgical lamp. We introduced our surgical lamp system in the actual surgery of jaws, and we recorded surgery by using it.

5.1 Example Result of Automatic Camera Switching

We set five cameras on the surgical lamp and performed experiments using a surgical image of the jaw. The images are shown in Figure 6. Each row shows
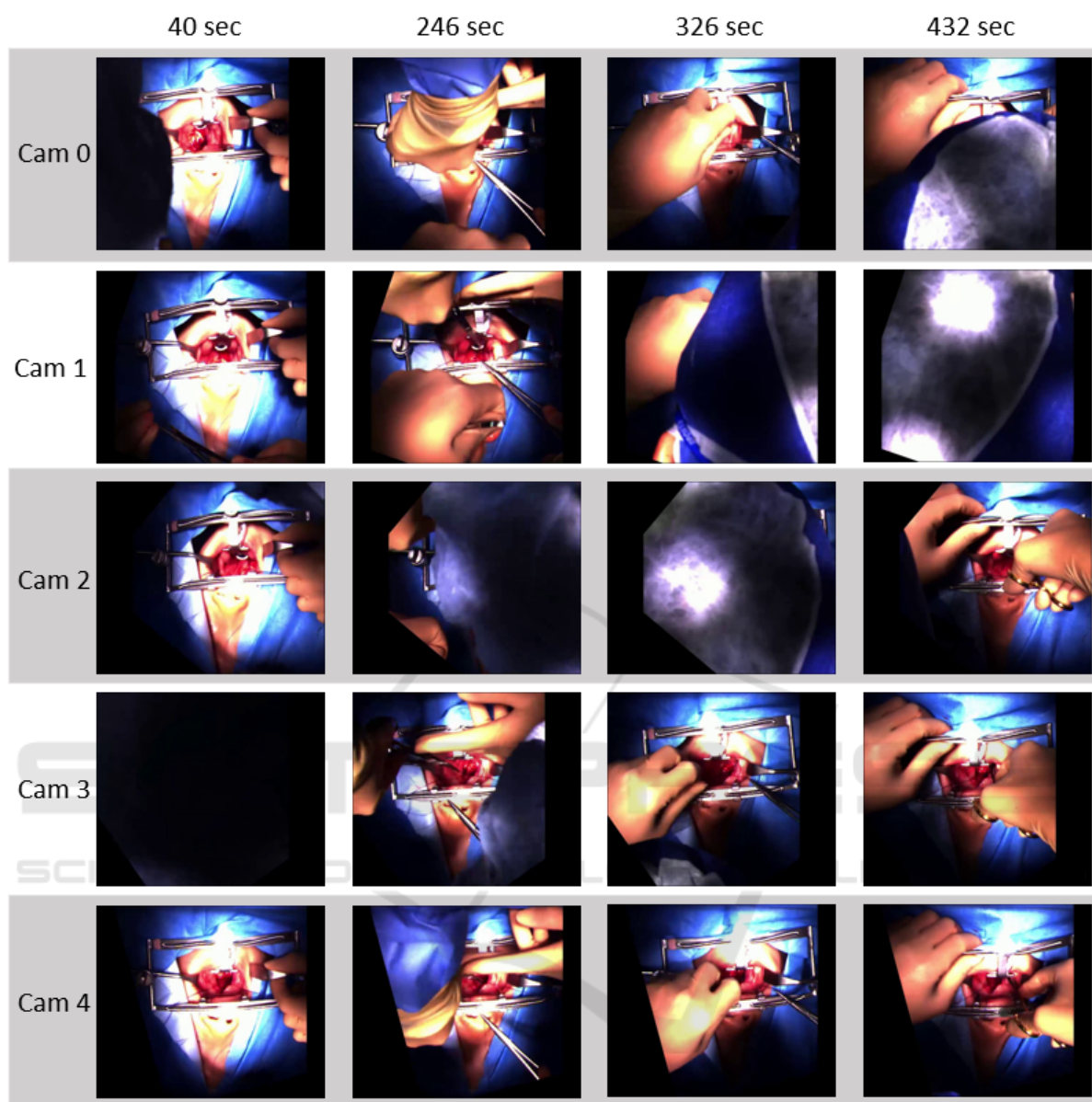
Figure 6: Images recorded by the multiple cameras mounted the surgical lamp.

the images captured by each camera, each column shows the timestamp.

In Figure 6, it can be seen that at least one camera always captures the surgical field. The segmentator was trained using about 100 images randomly extracted from the videos. We manually annotate the surgical field. In the experiment, we set the minimum frame to one second.

The generated videos are shown in Figure 7. In Figure 7 (a), the surgical field cannot be observed due to the occlusion of the head, etc., in the image obtained from one camera. On the other hand, in Figure 7 (b), the video created by our proposed method was

the video with diminishing occlusion.

As seen from the upper graph in Figure 8, when the camera switching is not scheduled, it is frequently performed. However, when the proposed method is applied, it can be seen that the camera switching is suppressed.

As shown in the lower graph, even when camera switching is suppressed, the score of the selected camera is almost the same as before the suppression. Therefore, we can say that QoV was improved while the observation of the surgery was maintained.

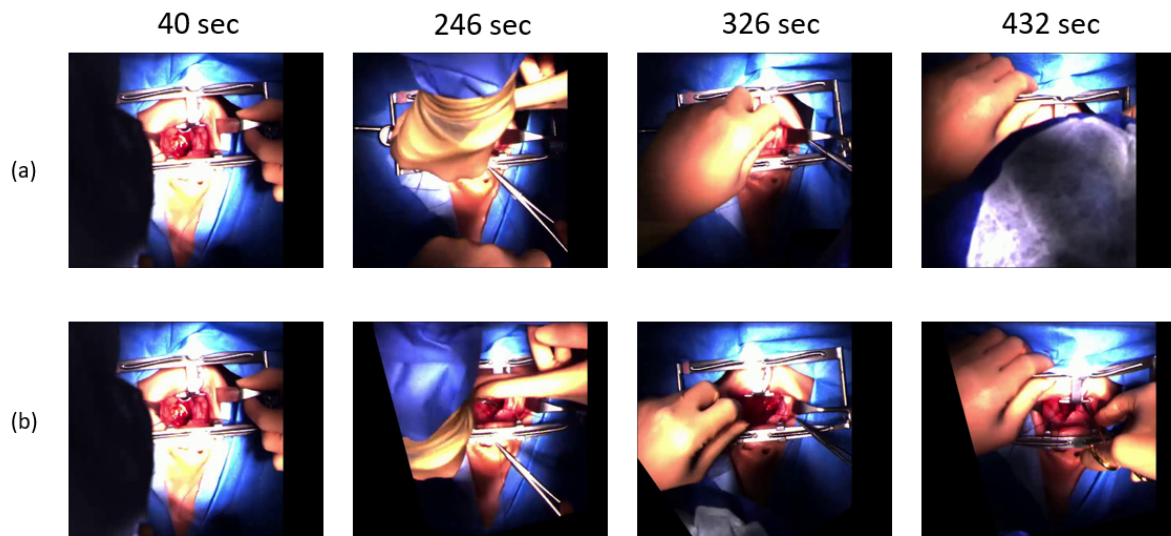40 sec   246 sec   326 sec   432 sec

(a)

(b)

Figure 7: Comparison between a video taken with one camera and a video created by the proposed method; a: images of each time of the video taken with one camera; b: images of each time of the video created by the proposed method.
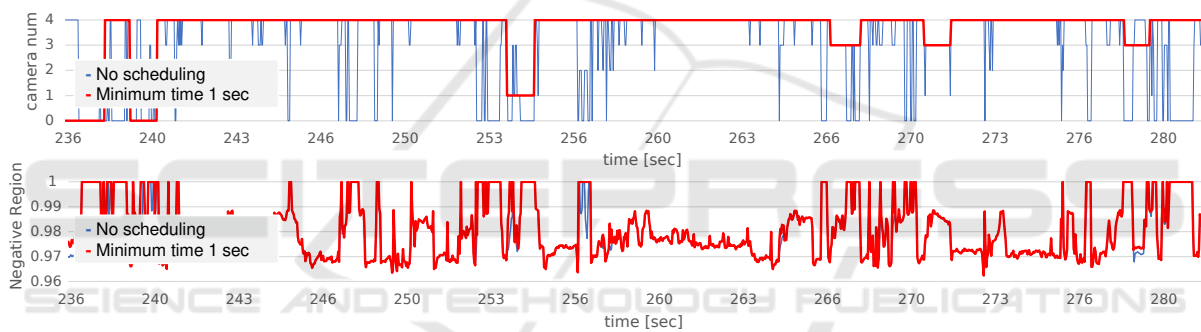
Figure 8: Camera switching and scoring results; (upper graph): the result of the camera switching; (lower graph): the result of camera scoring.

5.2 Assessment of Switching Video by Medical Doctors

We asked 13 doctors about the usefulness of the created video. Three videos of one camera, no schedule, and the proposed method were shown, and we asked two questions (Q1: whether switching cameras is not bothersome, Q2: whether it is possible to recognize the surgical operation.). Their responses were collected on a five-stage scale. (1: Strongly disagree, 2: Disagree, 3: Neither agree nor disagree, 4: Agree, 5: Strongly agree). Responses of Q1 and Q2 are shown in Table 1.

As a result, it can be seen that the recognition rate of the surgical operation is improved compared to using one camera from the result of Q2, and the QoV is improved compared to the video without scheduling from the result of Q1.

In addition to the assessment by doctors, we evaluated the performance of the automatic switching

Table 1: The result of questionnaire.

video	question	average	standard deviation
one camera	Q1	4.23	0.80
	Q2	1.69	0.91
without scheduling	Q1	1.38	0.49
	Q2	3.08	1.00
proposed method	Q1	3.77	0.80
	Q2	4.15	0.77

video by checking if each selected frame captures the target surgery region or not by actually watching the video. According to the visual examination, we have confirmed that the ratio of missing the target region is less than 5 %.

6 CONCLUSION

In the proposed method, multiple cameras were installed corresponding to the multiple light sources provided for the surgical lamp, and it became possible to switch the camera and record the operation automatically while diminishing the doctor's head and body. As a result, doctors can record surgery without being aware of the presence of a camera during surgery. In addition, we experimented with the proposed method and evaluated its usefulness. In the future, we would like to switch the multi-view videos without determining the minimum frame manually, and to generate camera switching video more in accordance with the preference of the doctor.

ACKNOWLEDGEMENT

REFERENCES

Ahuja, R. K., Mehlhorn, K., Orlin, J., and Tarjan, R. E. (1990). Faster algorithms for the shortest path problem. *Journal of the ACM (JACM)*, 37(2):213–223.

Byrd, R. J., Ujjin, V. M., Kongchan, S. S., and Reed, H. D. (2003). Surgical lighting system with integrated digital video camera. US Patent 6,633,328.

Chen, J. and Carr, P. (2014). Autonomous camera systems: A survey. In *Workshops at the Twenty-Eighth AAAI Conference on Artificial Intelligence*.

Daniyal, F. and Cavallaro, A. (2011). Multi-camera scheduling for video production. In *2011 Conference for Visual Media Production*, pages 11–20. IEEE.

Doubek, P., Geys, I., Svoboda, T., and Van Gool, L. (2004). Cinematographic rules applied to a camera network. In *Omnivis2004: The fifth Workshop on Omnidirectional Vision, Camera Networks and Non-Classical Cameras*, pages 17–29. Prague, Czech Republic: Czech Technical University.

Jiang, H., Fels, S., and Little, J. J. (2008). Optimizing multiple object tracking and best view video synthesis. *IEEE Transactions on Multimedia*, 10(6):997–1012.

Kumar, A. S. and Pal, H. (2004). Digital video recording of cardiac surgical procedures. *The Annals of thoracic surgery*, 77(3):1063–1065.

Li, C. and Kitani, K. M. (2013). Pixel-level hand detection in ego-centric videos. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3570–3577.

Liu, Q., Rui, Y., Gupta, A., and Cadiz, J. J. (2001). Automating camera management for lecture room environments. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 442–449. ACM.

Matsumoto, S., Sekine, K., Yamazaki, M., Funabiki, T., Orita, T., Shimizu, M., and Kitano, M. (2013). Digital video recording in trauma surgery using commercially available equipment. *Scandinavian journal of trauma, resuscitation and emergency medicine*, 21(1):27.

Murala, J. S., Singappuli, K., Swain, S. K., and Nunn, G. R. (2010). Digital video recording of congenital heart operations with surgical eye. *The Annals of thoracic surgery*, 90(4):1377–1378.

Nair, A. G., Kamal, S., Dave, T. V., Mishra, K., Reddy, H. S., Della Rocca, D., Della Rocca, R. C., Andron, A., and Jain, V. (2015). Surgeon point-of-view recording: using a high-definition head-mounted video camera in the operating room. *Indian journal of ophthalmology*, 63(10):771.