Automatic viewpoint switching for multi-view surgical videos

Tomohiro Shimizu* Keio University
Kei Oishi † Keio University
Hideo Saito ‡ Keio University
Hiroki Kajita § Keio University
Yoshifumi Takatsume ¶ Keio University

ABSTRACT

Recording medical surgery operations is important for sharing the various operating techniques. In most operating rooms, fixed surgery 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 in real surgery operations, we proposed the installation of multiple cameras in a surgical lighting system, so that at least one camera can capture the target surgical field even when the surgeon's head and body occlude other cameras. In this paper, we present a method for automatic viewpoint selection from multi-view surgical videos, so that the surgical field can always be recorded in the output video. We employ a method for learning-based object detection from videos for automatic evaluation of the surgical field area from multiple input videos. By selecting the viewpoint with the largest area of the surgical field, we can virtually reduce the area of the surgeon's head and hands. In general, frequent camera switching degrades the video quality of view (QoV). Therefore, we apply the Dijkstra method widely used in the shortest path problem as a combinatorial optimization method for this problem. Our camera scheduling method is that the camera switching is not performed for a certain period of time, and the surgical field observed in the entire video is maximized.

Index Terms: Human-centered computing—Human computer Interaction(HCI)—Interaction paradigms—Mixed/augmented reality; Artificial Intelligence—Computer vision—Computer vision tasks—Video summarization

1 INTRODUCTION

While the usefulness of videotaping surgery has been recognized for a long time, it is difficult to always record the surgical field because the surgeon's head gets between the camera and the surgical field [2, 3]. In order to avoid such occlusion by the surgeon's head, we have designed a novel surgical lighting system with multiple cameras, assuming that at least one camera can capture the surgical area, as at least one of the multiple lights on the surgical lighting system illuminates the surgical field. In this paper, we propose a system to automatically switch the camera (camera switching) so that the size of the imaging range of the surgical field becomes as large as possible for multi-view videos captured by our camera system. Our system realizes video which diminishes an occlusion such as the surgeon's head.

2 METHOD

The proposed system consists of two frameworks: camera scoring and camera switching. The input is a multi-view surgical video shot with multiple cameras mounted on the surgical lamp, and the score calculated by each camera in each frame is stored in the data storage. Lastly, the final output video is obtained by scheduling the camera.

*e-mail: tomy1201@hvrl.ics.keio.ac.jp
†e-mail: oishi@hvrl.ics.keio.ac.jp
‡e-mail: saito@hvrl.ics.keio.ac.jp
§e-mail: jmbrx767@keio.jp
¶e-mail: tsume@keio.jp

Figure 1: Multiple cameras mounted the surgical lamp

Multiple cameras mounted the surgical lamp As shown in Figure 1, we installed multiple cameras attached to the position corresponding to the multiple lights mounted the surgical lamp. Thereby, as long as the surgical field is illuminated, we propose a camera system that creates a situation in which at least one camera shoots the surgical field. With this camera system, the doctor knows that surgery is being recorded as long as the surgical field is illuminated. This frees the doctor from bothersome camera operations and allows them to focus on surgery without being aware of the presence of a camera.

Camera scoring Detection of the surgical field is performed using the method of Li et al. [4]. Considering that the area other than the surgical field is covered with cloth and the color contrast between the surgical field and other parts is large, detection of the surgical fields in the surgical video is compatible with the method of Li et al., which is performed only from color and texture information.

In the proposed system, the score is a rate of detected area other than the surgical field in the image.

Camera switching When the score calculated at a certain time continues to output the best camera image at a certain time, camera switching is frequently performed, so the QoV of the video decreases remarkably and the user feels uncomfortable. Therefore, it is necessary to consider the scheduling of the camera to control the camera switching. In the proposed system, camera switching is not performed for the minimum number of frames specified by the user, and the detected surgical field is maximized. Therefore, we propose a combined optimization method that applies Dijkstra's method [1]. widely used in the shortest path problem.

In generating the graph, each node has information about the camera number and time. If the camera does not switch, the edge is connected to the node one node ahead, and the weight of the edge is the score of each image in each frame. On the other hand, if the camera switches, the edge is connected to the node with the minimum number of frames ahead, and the weight of the edge is the total score of each camera in the minimum number of frames.

In the proposed system, the camera number array is obtained by acquiring the information of the nodes, excluding the start and end nodes, after optimization by the Dijkstra method. However, since there is an edge connected to the node of the minimum frame ahead, the array is smaller than the number of frames of the actual video.

978-1-7281-4765-9/19/$31.00 ©2019 IEEE
DOI 10.1109/ISMAR-Adjunct.2019.00037

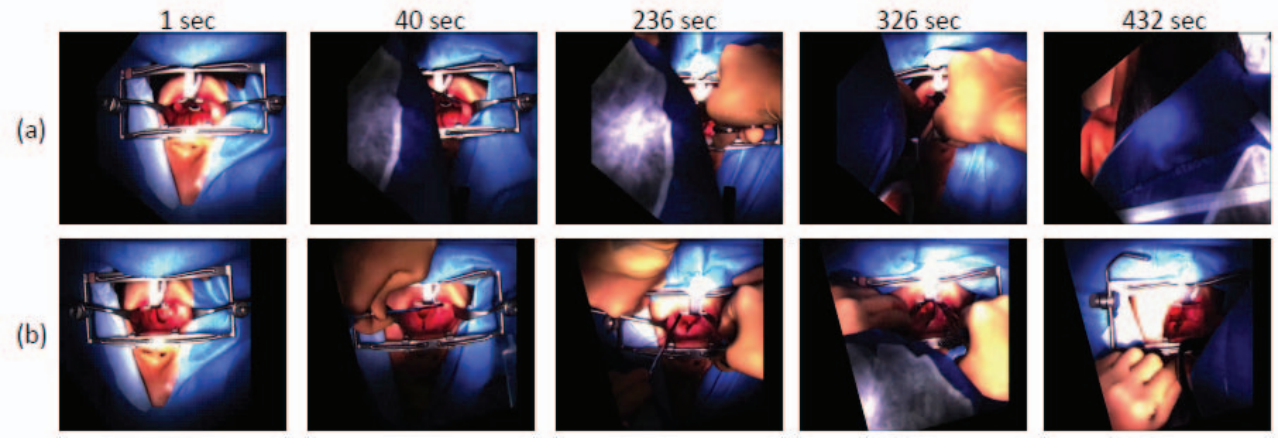

Figure 2: Comparison between a video taken with one camera and a video created by the proposed method; a: images of each time of the video taken with one camera; b: images of each time of the video created by the proposed method

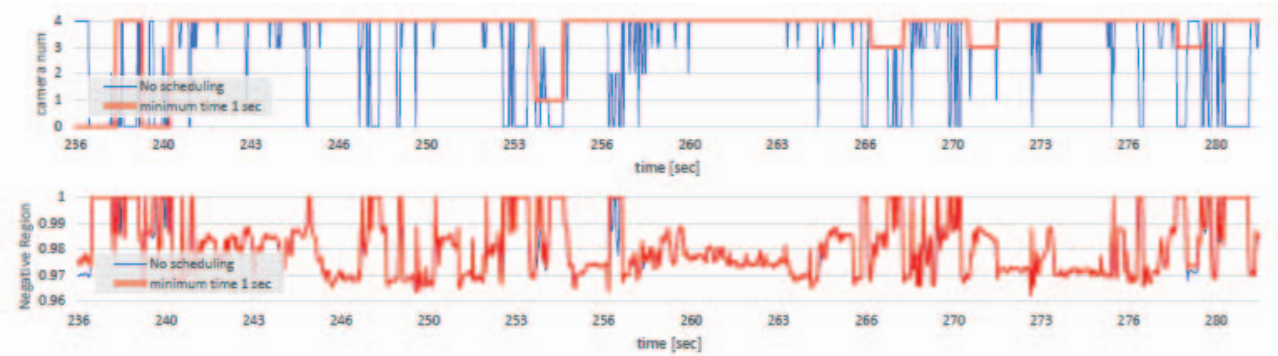

Figure 3: Camera switching and scoring result; (upper graph): the result of camera switching; (lower graph): the result of camera scoring

Therefore, at the position where the camera number in the array has changed, the skipped camera number is added to the array.

3 Results

In this paper, we describe the details of the experiments we conducted to verify the effectiveness of the proposed method. At the Keio University School of Medicine, we verified the effectiveness of the proposed method using multi-view surgical videos shot with multiple cameras mounted on the surgical lamp.

We set up five surgical field cameras on the surgical lamp and performed experiments using a surgical image of the jaw. The learning of the detector was performed using about 100 images randomly extracted from the image. In the experiment, the minimum frame was 1 second. The results are shown in figure2 and 3. In Figure 2, there is a frame where the field cannot be observed due to the occlusion of the head, etc., in the image obtained from one camera, but the image created by the proposed method creates the image with diminishing occlusion. As seen from the upper graph in Figure 3, when the camera switching is not scheduled, it is frequently performed. However, when the proposed method is applied, it can be seen that camera switching is suppressed. And, as shown in the lower graph, even when the camera switching is suppressed, the score of the selected camera is almost the same as before the suppression. Therefore, it can be said that QoV was improved while the observation of the operation was maintained.

Finally, 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). The average of Q1 responses for one camera, no schedule, and the proposed method was 4.23, 1.38 and 3.77 respectively. And, the average of Q2 responses was 1.69, 3.08 and 4.15 similarly. As a result, it can be seen that the recognition rate of the surgical operation is improved compared to using one camera, and the QoV is improved compared to the video without scheduling.

4 Conclusion

In the proposed system, 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 system and evaluated its usefulness. In the future, we would like to study the multi-dimensional score calculation method considering the angle of the camera and the operation process.

Acknowledgments

This research was funded by AMED research expenses (task number JP18he1902002h0001) and JSTCREST (JPMJCR14E1, JPMJCR14E3).

References

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

[2] R. J. Byrd, V. M. Ujjin, S. S. Kongchan, and H. D. Reed. Surgical lighting system with integrated digital video camera, Oct. 14 2003. US Patent 6,633,328.

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

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