2N-07

Multi-view Imaging System to Support Surgical Operations

Yinghao Wang[a] Hidehiko Shishido[b] Shinji Hashimoto[c] Tatsuya Oda[c] Itaru Kitahara[b]

[a] Humanics Program, School of Integrative and Global Majors
[b] Center for Computational Sciences, University of Tsukuba
[c] Department of Medicine, University of Tsukuba

1. Introduction

With the improvement of medical technology, surgical treatment makes progress. However, there are some issues with the current surgical imaging systems. For example, the cameras above the surgeon are often obscured by his head. Hence, as shown in Fig.1, we propose to deploy a camera system stationed between surgeons and patients. The position of surgeon's hands and surgical instruments change frequently during the process of a surgery. That could lead to occlusion effects, thus making a recording unusable for further analysis. The traditional video recording system that uses only one camera, has a fixed point of view, therefore, the occlusion effect is unavoidable. To solve this problem, we designed a system that uses multiple cameras to record the surgical procedure from various perspectives. This way, if a certain camera has a poor field of view, the system will automatically switch to another one that captures the target region properly.

By installing a deep learning algorithm (Yolov4), our system could identify the surgical instrument and the surgical area. Then, essential parameters (the recognition rate of the surgical area or scalpel) are extracted to serve as a basis for switching the viewpoints. The proposed system can capture high-quality surgical videos using the multi-view cameras. This allows for the proper post analysis.

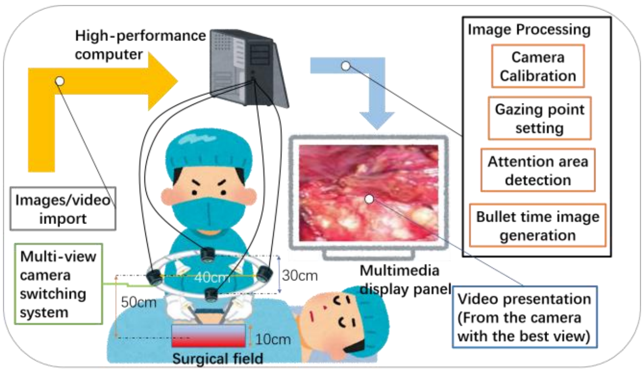

Figure 1. The overview of our proposed system.

2. Related research

2.1 Detection algorithm

You Only Look Once Version4 (YOLOv4) [1] algorithm is mainly used for object detection. The algorithm combines recognition and localization, using simple structure and fast processing speed.

2.2 Multi-view video technology

Multi-viewpoint video technology uses multiple cameras to capture images or videos [2]. It allows for the surround view of a target object and can also perform tasks such as complete matching, depth estimation, 3D reconstruction.

2.3 Surgical recording system

The main function of the imaging system is to record the surgical process, which is important to the evaluation of the medical treatment, education for improving surgical skills. Although, the recording system installed on the shadowless lamp [3], there is a problem that the surgical area is blocked by surgeons' head. As the result, the quality of the video captured by the system is reduced.

3. Multi-view surgical imaging system

To provide a solution to the problems mentioned above, we proposed a novel surgical recording system. As shown in Fig.2, the system has 20 USB camera (recording module), LED strip lights (60 LEDs) and 20 spotlights (lighting module), and it is divided into two parts (i.e., recording system and image processing system).

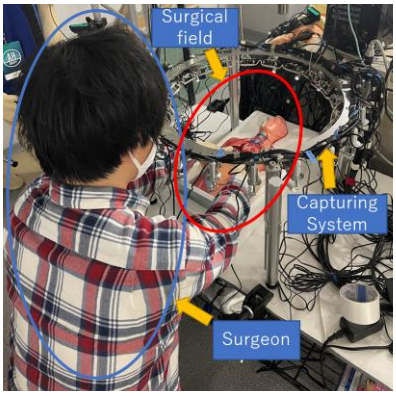

Figure 2. Our prototype system.

3.1 Calibration of multi-view capturing system

The system uses the SfM method to perform camera calibration. First, by importing the images taken by the multi-view camera as shown in Figure 3, the camera parameters are obtained, and the image feature points are extracted. The position and order of the camera are estimated based on the corresponding point information, and the 3D position of the corresponding point is estimated through stereo vision. After that, the information attained can be used to refine the estimated 3D positions of the corresponding points and camera parameters.

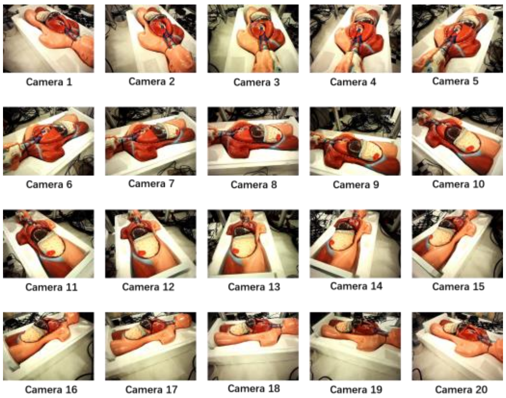

Figure 3. Captured multi-view images.

3.2 Capturing and camera switching system

For setting the gaze point on the surface of the object, our system employ target detection algorithms to track specific objects (such as scalpels) in surgery. While detecting the target object, we will set a gaze point on its surface for tracking. By comparing the parameters in image processing, (e.g. the confidence level of the detection target object, the position of the gaze point) the program switches the camera views smoothly. And we also develop a multi-view camera switching interface using OpenCV and QT5.

3.3 Object detection

We used the same version of python for the camera switching system as the opensource code used for YOLOv4. This is to assure the smoothness of the transfer procedure of the dataflow. the required parameters for the camera switching system is the camera ID with the recognition status and confidence of each target. In order to be able to identify the target (e.g., surgeon's hand, surgical instrument) more accurately, of the YOLOv4 algorithm was trained on a special dataset that has the following specifications.

- Resolution: 1920 pixels x 1080 pixels
- Number of frames: 1218
- The size of the data set: 250,037 KB
- Marking tool: yolo_mark[4]

4. Experiments and Results

Yolov4 is a new version algorithm, and there has been almost no research about surgical detection based on it before. So we designed this experiment to test its effect in recognizing several objects in the surgical scene.

In the experiment, we produced an specialized data set. The described dataset is marked with yolo_mark manually and be used to target detection. About the images, we capture a video stream (21-seconds). Using the FFmpeg python library, from each second of the footage, 8 frames were extracted based on IPB method to get the key frames. Totally 168 images were attained. The experimental process and the general appearance of the target to be detected are shown in Figure 5.

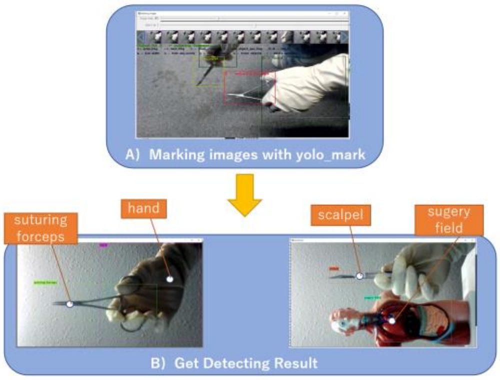

Figure 4. Dataset generation and the detection results.

Our experimental results reach an average accuracy of 67.5%. Table 1 shows the detailed results of our tool detection task. Among all the objects, the surgeon's hand is the highest performing object, mainly due to its apparent features that are easy to detect. However, the low performance of suturing forceps is mainly due to posture's diversity in actual use and the manual marking errors.

Table 1. The detection frequency and recognition confidence for each object type.

Categories	Frames appearing(num)	Confidence(%)
Hand	133	85
Scalpel	42	57
Suturing forceps	52	49
Sugery field	102	79
Mean	82.25	67.5

5. Conclusion

We proposed a surgical video capturing system, that achieves a higher recording quality. Thanks to a multi-view camera-based approach, it is possible to realize real-time point of view adjustment with avoiding occlusion. At the same time, because the system is deployed between the surgeons and the patient, the problem where surgeon's head is blocking the surgery area is be avoided. This work was supported by JSPS KAKENHI grant 17H01772, Japan.

References

[1] Bochkovskiy Alexey, et al. "YOLOv4: Optimal Speed and Accuracy of Object Detection." arXiv preprint arXiv:2004.10934 (2020).

[2] Nagai, Takasuke, et al. "An On-site Visual Feedback Method Using Bullet-Time Video." Proceedings of the 1st International Workshop on Multimedia Content Analysis in Sports. 2018.

[3] Shimizu, Tomohiro, et al. "Surgery Recording without Occlusions by Multi-view Surgical Videos." VISIGRAPP (5: VISAPP). 2020.

[4] YoloMark, [online] Available: https://github.com/AlexeyAB/Yolo_mark.