Multiple Angle Viewer for Remote Medical Training

Adriana Olmos
Department of Electrical and Computer Engineering
McGill University
3480 University Street
Montreal, QC H3A 2A7
aolmos@cim.mcgill.ca

Kevin Lachapelle
Arnold and Blema Steinberg Medical Simulation Centre
McGill University
3575 Parc Ave, Suite 5640
Montreal, QC QC H2X
kevin.lachapelle@mcgill.ca

Jeremy R. Cooperstock
Department of Electrical and Computer Engineering
McGill University
3480 University Street
Montreal, QC H3A 2A7
jer@cim.mcgill.ca

ABSTRACT

We present the design of an interface for a camera array that will enable mentoring and monitoring of dissections and surgical procedures for medical instructors and students. While considerable research has investigated the recording and broadcasting of surgical procedures and dissection sessions for medical instruction, little work has been reported on the integration of an interface able to display multiple viewpoints within a medical context. The interface presented here allows a designated individual, the instructor, to provide the best viewing point to observe and execute a procedure, and simultaneously, offers the remote viewer the freedom to change viewpoints.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation**]: User Interfaces—*Graphical user interfaces*

General Terms

Design, Human Factors

Keywords

remote education, medical collaboration, telegrading, camera array, interface design, user-centered design

1. INTRODUCTION

The operating room of the nineteenth century was a surprisingly collaborative environment [1]. Surgeons, nurses, consultants and other members of the healthcare team, as well as medical students, nurses-in-training, and other learners were free to come to the operating room and observe, learn and interact during a medical procedure. Back then, the operating room was a theatre where people gathered around the process of a surgery to learn and sometimes contribute. Medical knowledge and practice have since evolved significantly. However, the advent of aseptic technique, while improving the percentage of successful treatments and in turn, the lifespan of patients, has also fundamentally altered the interactions between people involved in a medical procedure. The operating room (OR) of the twentieth century might be seen, in most circumstances, as "anti-collaborative" with respect to the relationship between those inside and outside the space. To enter an OR today, individuals must change clothes and wear a layer of sterile clothing. This greatly impedes the ability to collaborate freely and has motivated increased interest in finding ways to improve collaboration between medical practitioners inside the OR with those outside.

The introduction of local intercoms and telephones began to reconnect the OR with the outside world. Perhaps representing the state of the art in this respect, the Barrow Neurological Institute installed a MedPresence telepresence system `http://www.medpresence.com/mcr400.cfm` consisting of a modular classroom-auditorium for medical training, with a 5 m wide video wall that allows users to see the entire horizontal landscape of the operating room. In parallel, images are acquired via surgical scopes and displayed on high-resolution LCD monitors embedded in desktops at each seating station within the classroom. Telecollaboration in general has become a particularly valuable practice in isolated areas, where access to major centres and expertise is often difficult to achieve.

2. LITERATURE REVIEW

Knowledge sharing is one of the primary benefits of ICT. An on-line search for videos with the keywords "surgical procedures" turns up over 5,600,000 results as of this writing. Recognizing the potential for such knowledge dissemination, many medical schools have equipped their labs with cameras in order to record procedures, often making use of the same technology to record students' procedures for pedagogical value, in particular, assisting in subsequent debriefing sessions [4]. The structure of the video recording varies with the intended audience and the procedure being presented. For example, recordings of sophisticated surgical procedures intended for review by qualified surgeons might offer a close up view of the body part being operated upon [6]. For junior medical students, the ability to see a global view that includes other medical staff intervening as a team in the operation might also be beneficial.

Significant research has been conducted on the recording and broadcasting from multiple viewpoints using an array of cameras [3, 10, 8, 11, 9]. Related examples include youtubedoubler, in which two synchronized video streams

MTDL'10, October 23, 2010, Florence, Italy.

are displayed to convey a message and iFoxCam, a live video surveillance solution for home and business use, recently targeting the childcare sector.

Although the use of camera array architectures and video synthesis from multiple views have been studied extensively in the context of commercial broadcasting applications, little has been mentioned with regard to the design of applications within a defined context of use. Specifically, a medical procedure can be seen as a highly attentive task in which the participants may need to move virtually around the operating table effortlessly when something is occluding their view. The need to support this simple interaction during remote viewing for medical training is thus the core design problem considered in this paper. In Section 3, we begin with the presentation of a field study in the context of medical education. This is followed by a summary of the design implications that must be taken into account in the design of a system for remote viewing of dissections and surgical procedures. Our design proposals are presented in Section 4 with results from our experiment in Section 5, representing an effort to address the implications above. Section 6 concludes the paper with a discussion of directions for further research to enhance the work presented here.

3. FIELD STUDY OVERVIEW

The objective of our study was to gain insight into the practices of surgeons, medical instructors, staff, and students before, during and after a teaching session involving a medical procedure or cadaveric dissection [5].

3.1 Fly on the wall at the dissection room

Fly on the wall observations are one of the many techniques used as an ethnographic field method in the design process. The main objective is to observe and record behavior and actions in their context, without interfering with peoples' activities. The motivation is to see what people actually do within real contexts and time frames, rather than simply accept their post hoc verbal recounting of the event [7]. We obtained permission from the head of the McGill Medical Simulation Centre (MSC) to attend some of their sessions, mostly addressing resident students. The different types of surgical and dissection procedures hosted by the MSC vary from dissection sessions involving a full cadaver to specialized surgical procedures, in some cases involving an endoscope. Each station is equipped with an overhead ceiling-mounted monitor that displays video of the instructor's activities at another table, as captured by an overhead camera. The camera is typically operated by an assistant, using a wireless remote control, to provide a close-up of the instructor's actions and the specimen.

Endoscopic sessions: One of the observed sessions involved the setup of an endoscope and a human foot. The instructor started by explaining the procedure and the objectives to the students. During the endoscopic session, the students initially directed their attention to the hands of the instructor, but later, focused mostly on the overhead monitor projecting the live endoscopic feed. Once the explanation of the procedure was completed, the students practiced at their own stations.

Generic dissections: Another session observed involved a spine dissection. At the beginning of the session, the instructor explained the procedure and the objectives to the students. While most of the laboratories are equipped with ceiling-mounted fluorescent lights, an extra lamp is occasionally required as an additional source of illumination. On many occasions, this became a problem as another person had to manipulate the lamp while the instructor was giving instructions.

To complement these observations, we also conducted several interviews with second year medical students from the medical program. These students are encouraged to review study material prior to attending various laboratory sessions. Some of the students keep the study material handy during the lab session. At the beginning of each session, the instructor explains the dissection procedure and anatomy of a certain body part. The instructor is also responsible for controlling the overhead camera that provides a video feed of his actions to various monitors around the lab. This video is recorded and subsequently made available to the students through the university's internal network. After the talk, the students proceed to locate the anatomical parts in their assigned cadaver, while the instructor circulates among the various stations, guiding the students. The cadavers are also available for the students to review later.

One of the shortcomings mentioned with regard to studying only from the video recordings is that only a single viewing angle is available, as obtained from the overhead camera controlled by the instructor. This is not always ideal, in particular since anatomy examination questions are always posed from a different viewpoint from the one shown in the videos. This serves as the motivation to provide views of the body part from multiple angles.

3.2 Design implications

We introduce a series of design factors inspired by our field study, which considered the activities conducted before, during, and after a surgical procedure or a dissection session.

Seamless, continuous navigation between guided and free viewpoints: This comes from the need to provide the instructor's recommended viewpoint, but at the same, allow the students the freedom to explore other viewpoints at their own pace. This ability is considered key to supporting the students' learning and understanding, at least in their first year. Moreover, supporting a continuously updated display between viewpoints is anticipated to increase the degree of student engagement as this would avoid introducing discontinuities in their experience of the session.

Flexibility in setting up a session: The staff members require the freedom to set up the tables with the cadavers and instruments at any orientation, according to the requirements of the procedure. For the design of a camera array, a symmetrical configuration does not require the staff members to explicitly choose "best" orientation of the setup with respect to the perspective of the cameras, which may vary from case to case.

Maintaining global context: Easy access to an overview of the operating table is required for the cases when an endoscopic session is presented. In addition, this global viewpoint can be useful for orienting oneself at the beginning of a session or throughout the session.

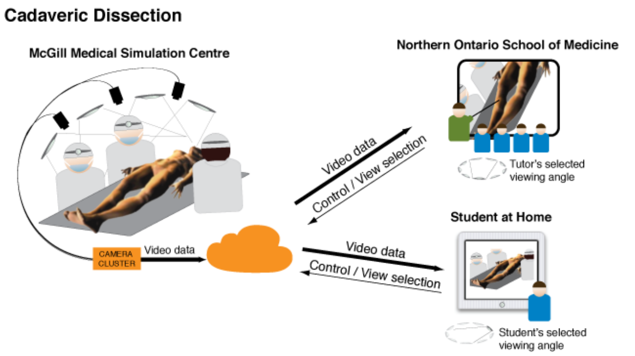

Figure 1: Overview of a dissection session scenario as experienced by instructors and students, both locally and remotely.

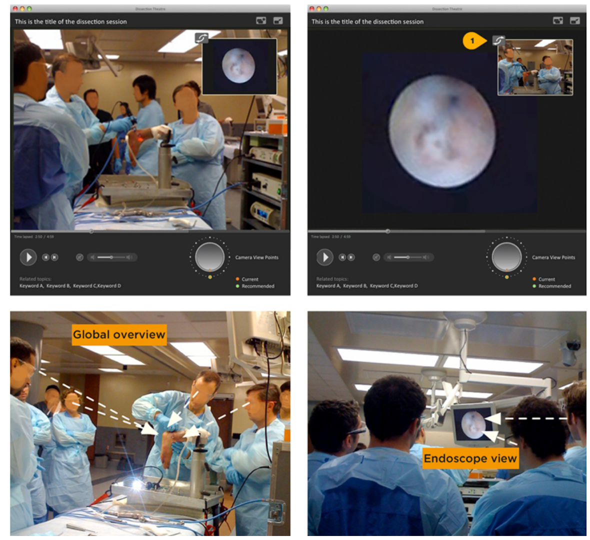

Figure 2: Top: an overview of the proposed interface as experienced by a remote student. Callout "1" indicates the button that swaps the picture-in-picture views. Bottom: the corresponding perspective of students in the lab.

Supporting material: The ability to contextualize the presented content allows one to relate the session to other relevant material, whether in video or another form. This also permits review of supporting material during observation of the video of a session, for example, facilitating comparison with an anatomical drawing or photograph from a (digitalized) textbook.

Integrated lighting: Integrating a lighting system with the camera array would alleviate many of the sources of distraction we observed during the sessions. Conceivably, the instructor may also recognize, and hence, avoid, blocking some of these light sources which would in turn result in fewer head occlusions blocking the view. Planned testing following deployment of our design will verify whether the validity of this hypothesis.

Communication: Predominantly for students in remote locations, who lack the ability to attend a local session with a live cadaver, an audio channel is imperative to support their interaction with the local team and ability to ask and answer questions.

4. MULTIPLE ANGLE VIEWER (MAV)

Based on the field research described in the previous section, MAV was designed to support the following remote medical training scenarios (Fig. 1):

Scenario A: A cardiac surgeon is presenting a sophisticated procedure on the coronary arteries to a group of residents. While carrying out the procedure, he can move his head freely without worrying about the camera position. If he does occlude a camera, the residents are able to switch to an alternative, obstruction-free view at any time. At the same time, the assistant is able to select and broadcast the best viewing point from the perspective of the surgeon. Using this mechanism, the surgeon can ask the students to observe from his viewpoint, and indicate the cues to look for during surgery.

Scenario B: The surgeon is working with less advanced students. He relies in part on the overview of the operating session so that the students can observe the interactions with other members of the surgical team. This also allows them to see how the instruments are laid out, picked up, and passed to the surgeon.

Scenario C: A knee surgeon is demonstrating an endoscopic procedure. The video feeds show how he manipulates the knee of the patient while inserting the endoscope and also the output from the endoscope itself once it is inserted. Availability of both viewpoints at a glance is an important asset, allowing for observation of the small manipulations performed with the tool.

Scenario D: The staff from a medical simulation centre are setting up for a surgery class involving a thorax. A large set of instruments is required for this session and needs to be laid out in a specific order. Fortunately, the staff need not concern themselves with the visibility of these instruments to the camera given that we used a symmetric configuration of cameras around the ring.

To satisfy the requirements of these scenarios, our design involves a set of approximately 17 cameras, rigidly mounted with respect to one another (Fig. 3), and whose position can be adjusted easily to cover whatever section of the operating table is required. The system indicates the recommended viewpoint from the perspective of the instructor, but as all cameras are accessible as sources of streaming video, remote viewers can select their viewpoint dynamically to obtain their own desired views of the procedure. In addition, an overview of the people interacting around the operating room is displayed. If desired, the system can also synthesize intermediate views from a perspective anywhere in between the physical cameras, offering not only a greater choice of viewing position but also a sense of continuous movement between cameras.

The MAV user interface (Fig. 2) provides remote viewers with the ability to select from three simultaneously received video streams:

1. the selected viewpoint, dynamically changeable

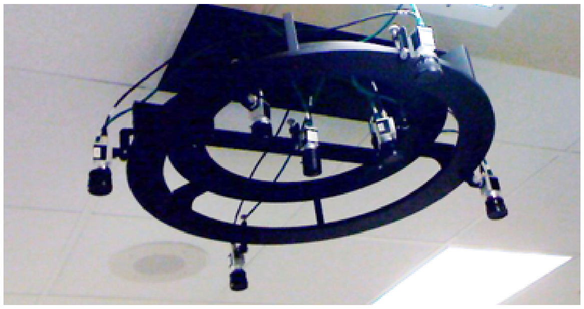

Figure 3: Prototype of the camera array structure.

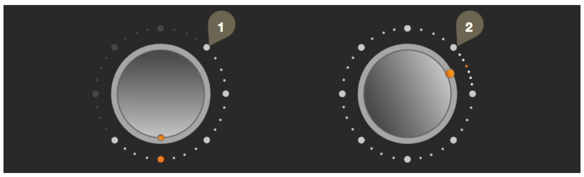

Figure 4: Different camera configurations, as made accessible through the knob control. 1) Only a subset of camera viewpoints are available. 2) Several viewpoints can be added or rendered in a specific section of the array depending on the anatomical part selected for the presentation or operation.

2. a thumbnail view of each camera to enable rapid selection of other viewpoints
3. a wide-angle overview of the entire procedure, or a view from an endoscope or other medical device, depending on the procedure being viewed

In addition to the video feeds, students also receive information regarding the viewpoint recommended by the instructor, as well as keywords that link to "topics" or content related to the session being observed. A virtual knob is used to select the desired viewpoint (Fig. 4); while rotating the knob, a thumbnail displays the real-time view of the camera at the selected position. When the user releases the mouse controlling the knob, the main view is updated to the newly selected video stream. The design choice of displaying the prospective view in a thumbnail rather than in the main window was deliberate, in order to allow users to maintain the current view in the periphery while deciding where to move. The described approach accommodates several objectives: First, it resembles moving around the specimen under observation, mapping the symmetrical arrangement of the cameras looking down at the table. Second, it allows for easy and dynamic addition of extra viewpoints within the array at any given time. Third, it maintains some sense of a fluid transition, even in the cases when few cameras are available, or when the distance between the cameras is large. Finally, it overcomes some of the constraints of lower bandwidth networks. The thumbnail of the potential "next view" can be updated faster, at a lower resolution or frame rate, while maintaining the delivery priority of the two main video streams being observed.

Our initial design for the presentation of video streams gave priority to the most important one and relegated the

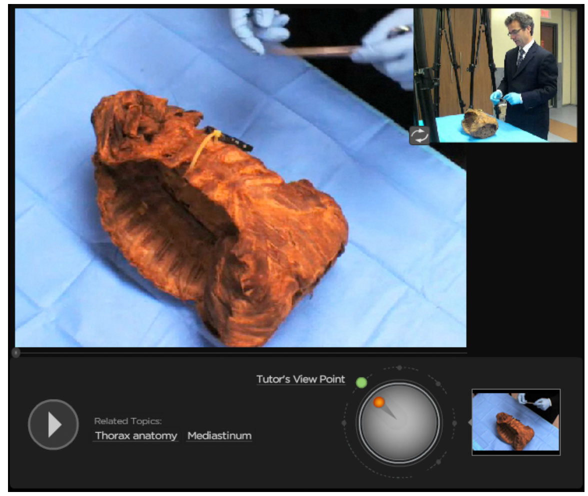

Figure 5: Picture-in-picture layout used for the user test sessions (interface treatment *P*).

secondary stream to the periphery, using a picture-in-picture layout as shown in Figure 5. However, we did not find any conclusive evidence in the literature to suggest that this configuration is superior, from a pedagogical perspective, to a side-by-side arrangement with the two video windows of equal size, as shown in Figure 6. The latter design, in fact, follows the emerging trend among telepresence systems of providing several screens, often wall-mounted, that display multiple viewpoints simultaneously. However, in the case of remote observation of a surgical procedure using a personal computer, screen size can be a constraint.

During our observations and discussions with surgeons, it was suggested that a single screen displaying the primary viewpoint may be sufficient, and perhaps, even preferable, as multiple video displays could potentially be regarded as distracting. On the other hand, several individuals suggested that for some, access to alternative viewpoints in parallel could be beneficial. The conflicting perspectives on this question led us to investigate the choice of layout of the video feeds in the context of viewing a surgical procedure. At the same time, we considered the usability of the proposed viewpoint selection control.

5. USABILITY EXPERIMENT

We designed an experiment that elicited interactions with the interface while viewing a dissection session, of approximate duration of 220 seconds, related to the anatomy of a human thorax. Six different viewpoints were recorded, focusing on the manipulation of the thorax, along with a global view including the instructor. The recording was done at the Medical Simulation Centre where the anatomy session was given by an expert surgeon of the same university. The interface was programmed using ActionScript 3.0.

Eight second-year medical students were invited to participate in this experiment. The students were asked to answer ten anatomy-related questions while watching the video footage. Each question required the student to name the anatomical part at which the instructor was pointing and to select the best viewpoint from which to observe that

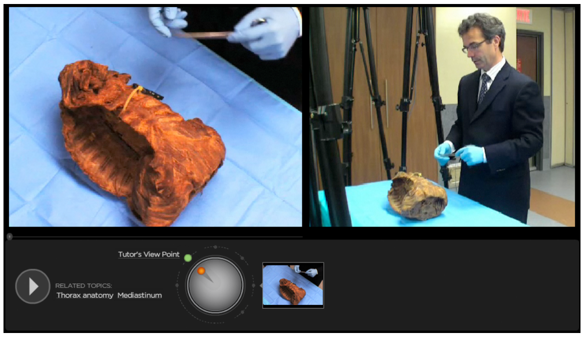

Figure 6: Two video windows side-by-side (interface treatment *W*).

part. Each question was prompted automatically by a number appearing on the screen. The students entered their answers on a separate sheet of paper, and had full freedom to pause, advance and rewind the video in order to select any viewpoint. Each student completed two trials with the same dissection session, one with each of the two interface layouts, or treatments, *P* for picture-in-picture, and *W* for side-by-side. A different set of anatomical questions was used for each trial. The presentation order of the interface treatments and set of anatomical questions were balanced across subjects.

The user interaction, mouse clicks, comments and facial expressions were recorded using the *SilverBack* software tool (`http://silverbackapp.com`). At the end of the testing session the students were asked to complete the following questionnaire and invited to express their views and comments regarding the interface: A) Describe the purpose of the controls in the interface. If there is a control that you did not notice while testing, or you are not sure what it is for, please indicate so. B) Which of the controls listed above was the most difficult to use? Why? C) Please rate the ease of use of this control. (This question was introduced with a picture of the control to select one of the many viewpoints) D) If you had to use any of the two players again to watch another dissection session, which layout would you prefer to use? Why? E) From your previous experience in the lab, could you list any advantages and/or disadvantages of this interface compared to a real-life situation?

Paired t-tests were used to compare the two trials. Students completed the second trial significantly faster than the first trial (t(7)= 3.0834, p = 0.0081), as shown in Figure 7. This can be attributed to the increased familiarity with the task, video footage, and interface controls in the second round. Interestingly, the time it took the students to start interacting with the viewpoint selector demonstrated the opposite effect, with students taking more time before doing so on the second trial (t(7)=-2.6031, p = 0.0353), as shown in Figure 8. This might be explained by a shift from first-time exploration mode in the first trial, in which students experimented with the interface controls, to a goal-directed mode in the second round, in which students used the viewpoint selector only when it was deemed useful to find a better view for the task. These observations relate to another study [2], concluding that learners who can control the presentation of an anatomical model consistently perform better than learners who have only fixed views, even though they actually spend most of the time viewing from (or near) the standard (top, side, front) viewpoints.

Finally, we did not find any statistically significant difference (t(7)=0.1865, p = 0.8573) in time between the two interface treatments (*P* or *W*). Nevertheless, all subjects indicated their preference for the picture-in-picture interface (*P*) if they had to watch another session, typified by the following response: "The video of the instructor is less important than the one of the prosection. Interface *P* allows you to focus on [dissection of the cadaver] and switching to the video of the instructor whenever is needed. In interface *W*, you see both videos at once and it is harder to concentrate on the one you want to watch; it is distracting."

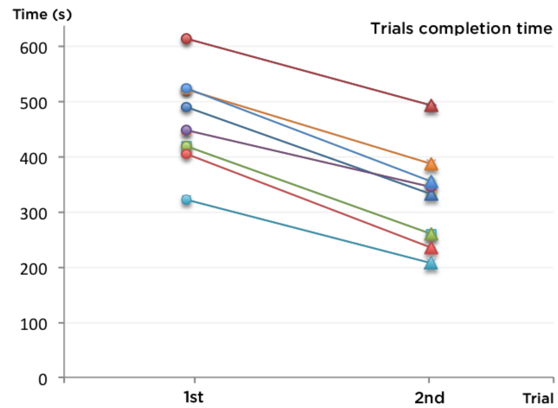

Figure 7: Completion times for first two trials

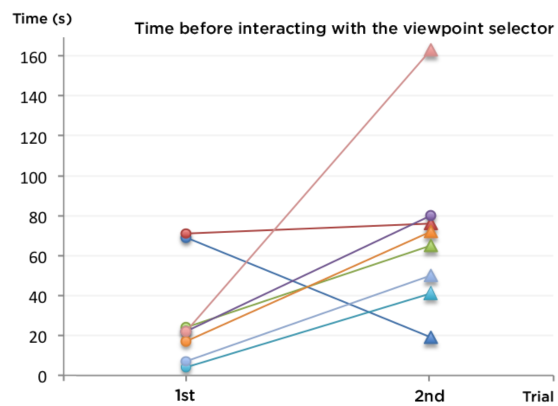

Figure 8: Time before students started interacting with the viewpoint selector

Of the eight students who participated in the experiment, six noticed and understood the button to swap the positions of the two video windows. Although they considered this a potentially useful control for surgical procedures, the students mentioned that for the experimental anatomical session this functionality was not really important. There was unanimous agreement on the need to support zoom into the prosection being presented, a feature not yet provided.

The purpose of the knob, or viewpoint selector, was well understood by the students at the start of interaction. On a seven-point rating scale, 1 being easy to use and 7 difficult to use, the mean rating given to this control was 4.1. Criticisms included the requirement for precise manipulation to access the next viewpoint, and a perceived lack of sensitivity when the knob was not reacting to commands.

The ability to pause and control the viewpoint selector at any given point was regarded as one of the major assets of the interface. This went along with the option of having "related topics" at hand. However, only one of the users naturally explored these links. When the users were asked if they noticed the corresponding links area, most thought that these were simply part of a title, keywords, or more videos. The users were then invited to explore the links, resulting in a positive surprise, especially when they discovered that one of the links was the image of a textbook drawing of the prosection being studied. Clearly, the lack of awareness of these links is also indicative of the need for improvement of the related UI content. Similarly, during the experiment, we observed users returning to the instructor's viewpoint while interacting with the interface. Nevertheless, users did not understand that the green dot was not only an indicator but also offered a shortcut to return to that position.

Overall, the purpose and usage of the controls were clear for the users, but various refinements of the interface are required: improvement of the knob control or viewpoint selector, presentation of the "instructor viewpoint", and the visualization of the "related topics", better labeled as "related material" or "support material". Despite its shortcomings, the first generation interface and its ability to present multiple viewpoints were well received. Among the advantages listed were: accessibility from home; video theoretically available immediately; possibility of viewing different angles, especially for exam preparation. The two consistent disadvantages noted by the students were not being able to zoom in, and that this should not replace the actual dissection and contact with the cadaver. More recent iterations of the interface can be seen at `http://www.cim.mcgill.ca/sre/projects/hsvo/`.

6. CONCLUSIONS AND FUTURE WORK

We have presented the design of an interface for a camera array that enables the mentoring and monitoring of dissections and surgical procedures for medical instructors and students at a distance. The interface allows a designated individual, the instructor, to provide the best viewpoint to observe and execute a procedure, and simultaneously, offers the remote viewer the freedom to change viewpoints. Through our user experiment, we learned the preference for a picture-in-picture layout with the context of anatomical sessions, despite the lack of statistical benefit to performance.

Within the context of anatomical dissections, half of the students commented that they would benefit from having the various body parts highlighted, and even labeled, as the instructor is pointing to them. For instance distinguishing veins and arteries can be challenging for a second year student. This could imply additional video editing or post-processing workload. There is also an opportunity to consider new input devices within this medical teaching context, for example, a specialized pointer or pen, whose button might trigger the highlighting of the part being pointed at in the video.

During the demonstration of a surgical procedure or dissection session, having an assistant is a luxury. Among other duties, this individual covers any material required by the instructor, manipulates the camera to target the body parts that the instructor wants to illustrate. While there always seems to be a helping hand available during these teaching procedures, it would be interesting to automate the switching of the instructor's point of view as he performs the operation. Future work could investigate the advantage of automatic determination of the surgeon's viewpoint during the operation and providing this information to the interface.

7. ACKNOWLEDGEMENTS

This body of work was developed as part of the Health Services Virtual Organization project, supported by funding from Canada's Advanced Research and Innovation Network (CANARIE). The design portion of the camera array presented here was carried out by Jeff Blum, Josh Redel, Haijian Sun and Stéphane Pelletier.

8. REFERENCES

[1] K. Cleary, A. Kinsella, and S. K. Mun. Or 2020 workshop report: Operating room of the future. *Computer Assisted Radiology and Surgery*, 1281:832 – 838, 2005.

[2] A. X. Garg, G. R. Norman, K. W. Eva, L. Spero, and S. Sharan. Is there any real virtue of virtual reality? the minor role of multiple orientations in learning anatomy from computers. *Academic Medicine: Supplement*, 77(10):S97–S99, 2002.

[3] W. W. Gaver, A. Sellen, C. Heath, and P. Luff. One is not enough: multiple views in a media space. In *Proc. SIGCHI Conf.Human Factors in Computing Systems*, pages 335–341, New York, NY, USA, 1993. ACM.

[4] G. G. Hamad, M. T. Brown, and J. A. Clavijo-Alvarez. Postoperative video debriefing reduces technical errors in laparoscopic surgery. *American Journal of Surgery*, 194(1):110 – 114, 2007.

[5] Kouprie and S. Visser. framework for empathy in design: stepping into and out of the user's life. *Journal of Engineering Design*, 20(5):437–448, 2009.

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

[7] D. Schuler and A. Namioka. *Participatory Design: Principles and Practices*. Lawrence Erlbaum Associates, 1993.

[8] C. Z. Tsuhan, C. Zhang, and T. Chen. A self-reconfigurable camera array. In *In Eurographics Symposium on Rendering*, pages 243–254, 2004.

[9] B. Wilburn, N. Joshi, V. Vaish, M. Levoy, , and M. Horowitz. High-speed videography using a dense camera array. In *Proc. Computer Vision and Pattern Recogntion*, volume 2, pages 294–301, 2004.

[10] K. Yamaashi, Y. Kawamata, M. Tani, and H. Matsumoto. User-centered video: Transmitting video images based on the user's interest. In *Proc. SIGCHI Conf. Human Factors in Computing Systems*, pages 325–330, New York, NY, USA, 1995. ACM.

[11] J. C. Yang, M. Everett, C. Buehler, and L. McMillan. A real-time distributed light field camera. In *Eurographics Workshop on Rendering*, pages 1–10, 2002.