An Interactive Augmented-Reality Video Training Platform for the da Vinci Surgical System

Jaimie Carlson[1] and Katherine J. Kuchenbecker[2]

***Abstract*— Teleoperated surgical robots such as the Intuitive da Vinci Surgical System facilitate minimally invasive surgeries, which decrease risk to patients. However, these systems can be difficult to learn, and existing training curricula on surgical simulators do not offer students the realistic experience of a full operation. This paper presents an augmented-reality video training platform for the da Vinci that will allow trainees to rehearse any surgery recorded by an expert. While the trainee operates the da Vinci, they see their own instruments overlaid on the expert video. Tools are identified in the source videos via color segmentation and kernelized correlation filter tracking, and their depth is calculated from the da Vinci's stereoscopic video feed. The user tries to follow the expert's movements, and if any of their tools venture too far away, the system provides instantaneous visual feedback and pauses to allow the user to correct their motion. This combined and augmented video provides the user with an immersive and interactive training experience.**

I. INTRODUCTION

Robotic surgery systems such as the da Vinci Surgical System (Intuitive Surgical, Inc.) combine the small incisions of laparoscopy with the precision and mobility of a physician's hands in open surgery [1]. However, the da Vinci can have a steep learning curve; surgeons may require 15 to 50 cases to master a specific operation [2] [3].

Existing methods of learning how to use the da Vinci all have drawbacks. Many training models focus on simplified tasks like peg transfer or knot tying; once students have mastered these, they turn to shadowing physicians or passively watching videos of experts to learn about real operations. Some robotic surgery simulators exist, but they either provide a notably different user interface than the da Vinci itself, or they use the da Vinci controllers to interact with simulated materials, providing a less realistic experience limited to a few preprogrammed operations [4]. Furthermore, these systems cannot realistically simulate possible anatomical variations among patients or demonstrate how a surgeon handles an unforeseen complication on the spot.

In the proposed system, trainees use the da Vinci, while viewing their own tool movements superimposed over a previously recorded stereoscopic video from an actual operation performed by a skilled surgeon (teacher video). The system encourages the student to follow along with the movements of the teacher in three dimensions and alerts them when they stray from the teacher's path. In this way, students can realistically and easily learn the beneficial motion patterns performed by experts (no dithering, efficient movement, use of both tools) for a variety of real procedures.

[1] Jaimie Carlson is with the Departments of Bioengineering and Computer Information Science, University of Pennsylvania, Philadelphia, PA 19104. jaimiec@seas.upenn.edu

[2] K. J. Kuchenbecker is with the Departments of Mechanical Engineering and Applied Mechanics and Computer and Information Science, University of Pennsylvania, Philadelphia, PA 19104, USA and with the Haptic Intelligence Department, Max Planck Institute for Intelligent Systems, Stuttgart, BW 70569, Germany kjk@is.mpg.de

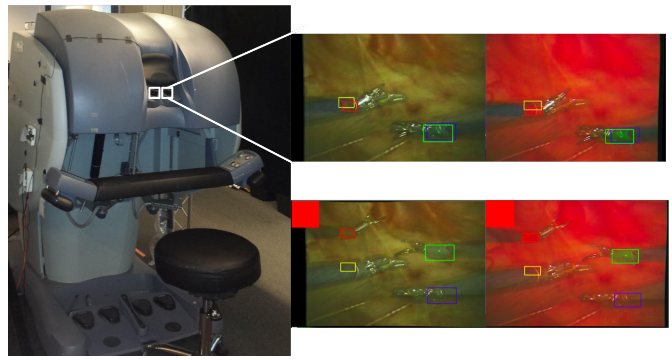

Fig. 1. System setup. Main image shows the da Vinci master console; top inset shows what a user would see in the left and right eyepieces when properly mirroring the teacher; bottom inset shows the view of a user not correctly following. Red and green rectangles outline student tools; yellow and blue rectangles outline teacher tools.

II. SYSTEM OVERVIEW

The system was designed to provide simple and instantaneous feedback for users wishing to practice any operation. It was developed on a da Vinci Standard, but the approach should work on other robot models as well. After sitting at the console shown in Fig. 1 and logging into the system, a user can select stereo-paired teacher videos on which to practice. In the da Vinci eyepieces, an augmented video is displayed showing the student's tools superimposed over the teacher's video. Highlighting of each tool (as shown in the inset to Fig. 1) enables the user to identify their own movements as well as the teacher's against a realistic surgical background. If the user fails to follow the teacher, the system instantly warns them and pauses the video to allow them to correct their path. The trainee can later track their progress by viewing statistics on metrics such as number of warnings given, total time taken, and path efficiency.

III. DESIGN

A. Hardware

Live videos from the left and right stereoscopic cameras are extracted from the da Vinci's camera control unit via two S-Video to USB converters, attached to a Linux machine on which software processing was performed. HDMI outputs from the machine are fed through HDMI-SDI converters into the left and right visual inputs to the da Vinci console, so that the user sees an augmented camera view.

B. Software

The student's tools are marked with green (left) and blue (right) tape. Color segmentation locates the tools in the left and right videos and constructs a mask eliminating their non-surgical background.

Because teacher video is prerecorded from a real procedure, markers cannot be added to the tools. Instead, they are initially located by simple color segmentation of grey, black, or white tools against a pink internal environment; images are converted to HSV format and thresholded by low saturation. In the event of a tool leaving the frame or two tools coming together, which often happens when the two tools are suturing or gripping the same area, this color segmentation is not robust. Thus, to identify specific tools, left and right tool tips are tracked with a Kernelized Correlation Filter after first being found. To reduce the real-time computational load, teacher videos are preprocessed.

The 3D locations in pixels of both student and teacher tools are analyzed by comparing their pixel positions in the left and right videos. We tested our da Vinci Surgical System's stereoscopic camera with OpenCV's stereoCalibrate() function, which returned the camera matrix and distortion coefficients for each camera, plus the rotation and translation matrices between the two cameras. OpenCV's stereoRectify() function was used to ensure that in each video frame, pixels corresponding to the same physical point would be placed at the same y-position in the left and right images. With OpenCV's stereo block matching tool, disparities were found between specific locations in both images, creating a disparity map of distances between corresponding points in pixels. These dimensions are converted to mm using the reprojection matrix calculated during rectification, making a depth map of the entire image.

Each tool is modeled as a spherical RRP robot with all joints at the origin, where the tool is placed so that its remote center aligns with a specific hole in a training shell used to model the abdominal cavity. This origin relative to the camera frame for each shift of the camera is found by locating the tools in each frame via color segmentation, calculating a line which approximately runs through the shaft of each tool, and finding the point closest to the intersection of these lines using singular value decomposition [5]. All Cartesian coordinates relative to the picture's origin in the depth map are recalculated relative to this new origin.

Spherical coordinates are calculated from these Cartesian coordinates, since their motion would not be independent. To remove noise created by the color segmentation method, a three-dimensional Kalman filter is run on spherical coordinates. The value of the Kalman filter after the update stage is then converted back to Cartesian coordinates. This is done for both student and teacher video, and the locations of both are recorded, both to update path length taken by student tools and to measure the distance in mm between corresponding student and teacher tools.

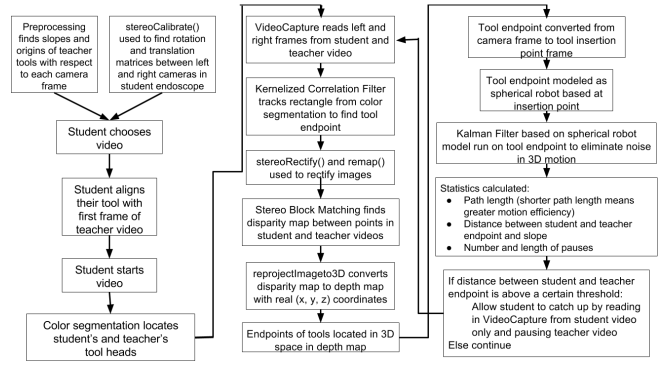

Fig. 2. Image processing algorithm for augmenting da Vinci video.

After identifying locations of the student and teacher tools, the system surrounds each tool end with a separate box for easy identification by the user, adds the two images, and displays the result. If either student tool strays too far from its teacher (an adjustable distance presently set to 50 mm), a warning is displayed, and the teacher video is paused to give the student a chance to realign their tools. Once the tools are aligned again, processing of the augmented video resumes, following the algorithm stated above and shown in Fig 2.

IV. DISCUSSION

This platform will be tested by users in order to get feedback about the realism of the expert video overlaying, the helpfulness of the distance feedback, and the optimal distance threshold for correcting the user.

Future improvements to this system will work on increasing the accuracy of motion tracking. Color segmentation and kernelized correlation filters work fairly well for identifying tool location, but their robustness over full-length surgical videos must be confirmed.

Currently, the system tracks only two tools at once, with static color labels required for each student tool. However, complex operations typically require one more robotic tool, along with additional handheld laparoscopic tools. Distinguishing tool identity is important as well: a da Vinci tool performing electrocautery is functionally distinct from a basic gripper, even if both are in the same location. Relatedly, the platform will move beyond location tracking to identify the pose of each tool tip, so the student can also learn to perform specific operations like grasping or releasing.

V. CONCLUSION

This platform augments live feed from a student's da Vinci Surgical System with video recorded from an expert surgery and provides instantaneous feedback on the student's performance. This platform could teach any variety of operation, from explanatory training videos to lengthy surgeries with unforeseen complications. The application of augmented reality to existing robotic surgery videos may help train a new generation of minimally invasive surgeons.

ACKNOWLEDGMENTS

The authors thank Jeremy Brown for his advice on this project.

REFERENCES

[1] S. Maeso, M. Reza, J. A. Mayol, J. A. Blasco, M. Guerra, E. Andradas, and M. N. Plana, Efficacy of the da Vinci surgical system in abdominal surgery compared with that of laparoscopy: a systematic review and meta-analysis, Ann. Surg., vol. 252, no. 2, pp. 254-62, 2010

[2] M. B. Bokhari, C. B. Patel, D. I. Ramos-Valadez, M. Ragupathi, and E. M. Haas, Learning curve for robotic-assisted laparoscopic colorectal surgery, Surg. Endosc., vol. 25, no. 3, pp. 855-60, 2011.

[3] J. P. Lenihan, C. Kovanda, and U. Seshadri-Kreaden, What is the learning curve for robotic-assisted gynecologic surgery?, J. Minim. Invasive Gynecol., vol. 15, no. 5, pp. 589-94, 2008.

[4] A. Moglia, V. Ferrari, L. Morelli, M. Ferrari, F. Mosca, and A. Cuschieri, A Systematic Review of Virtual Reality Simulators for Robot-assisted Surgery, Eur. Urol., vol. 69, no. 6, pp. 1065-1080, 2015.

[5] L. Han and J. C. Bancroft, Nearest approaches to multiple lines in n-dimensional space, CREWES Research Report, vol. 22, pp. 1-17, 2010.