Gaze-contingent Robotic Assistance in the Operating Theater

Alexandros A. Kogkas, Ahmed Ezzat, Rudrik Thakkar, Ara Darzi, George P. Mylonas, *Member, IEEE*

I. Introduction

Shortages in operating theatre personnel may increase patient mortality risk [1], therefore integration of new technologies could enhance patient safety and improve work flow. Robotic scrub nurses have been developed to support the surgeon in surgical instruments selection and delivery.

We introduce a novel gaze controlled robotic scrub nurse which is supported by a perceptually-enabled Smart Operating Room platform [3]. This is based on a novel real-time framework for theatre wide and patient wise 3D gaze localization, to allow the surgeon to operate in a mobile fashion, providing a "third hand" and facilitating practical and user-friendly human-technology interaction intra-operatively

II. System Overview

The core functionality of the real-time framework presented here is to provide the user's 3D point of regard (PoR) in a world coordinate system (WCS), defined by multiple co-registered RGB-D sensors fixed in the theater. It relies on estimating the pose of the scene/RGB camera of the wearable eye-tracking glasses (ETG) in the WCS and tracing the gaze ray provided by the ETG on the head frame of reference, on the 3D reconstructed space. The ETG scene camera pose is estimated with the employment of a motion capture system (MCS) and spherical markers mounted on the ETG. The user's head pose (equivalent to the ETG's scene camera pose) can be used to map the 2D gaze to 3D fixation in the WCS. The 3D gaze ray can be used to detect fixations on a screen fixed in space (micro fixation) and after a certain dwell time trigger the robot routine. The robot moves towards a surgical instrument selected by the user, grasps it with a magnetic gripper and transfers it to the user. When the force/torque sensor mounted on the robot end-effector senses the instrument is picked up, it returns to its homing pose (Fig. 1).

III. Materials and Methods

Equipment: The wireless SMI (SensoMotoric Instruments GmbH) glasses are used for *eye-tracking*. Microsoft Kinect v2 cameras are employed for *RGB-D sensing* and the OptiTrack MCS, with 4 Prime 13 cameras for *head pose tracking*. The *robot* is a UR5 (Universal Robots), with the Robotiq FT-300 *F/T sensor* attached. A 42″ LG *screen* with 1920x1080 px resolution is used for the instrument selection GUI.

Offline Calibration: In an offline calibration routine, the rigid transformations between the ETG rigid body–ETG scene camera and MCS CS – WCS are calibrated. The screen corners 3D coordinates in the WCS are manually selected on the Kinect RGB image and the 3D points are generated. The surgical instruments are positioned in fixed positions on a tray. The robot is manually moved over each instrument and the target pose is calibrated. 9-point eye-tracking calibration is performed before every task.

Interface Design: The GUI displayed on the screen consists of two parts: instrument selection (left 2/3) and the image navigation (right 1/3). *Left*: Six blocks equally split demonstrate common surgical instruments. Micro fixation on any of the blocks initiates a traffic light sequence (red-amber-green) followed by relevant audio feedback. Starting with red block borders, dwell time of 0.6 *s* into the same block turns the borders into orange and another 1 s turns the borders into green. The design is based on pilot experiments aimed to allow the user to have sufficient feedback (audio/visual) for the estimated micro fixation (red), be warned before finalizing the instrument selection (amber) and confirm the action (green). *Right*: Three slides are presented to provide information necessary for the task workflow. The user can navigate through them by fixating on the top and bottom 1/6 parts of the screen for previous and next slide respectively. Dwell time here is 1 *s*.

Robot Control: The selection of an instrument on the screen (server) triggers the robot (client) to handle the corresponding instrument to the user. TCP/IP is used to transmit the instrument ID to the robot client. The robot client has predefined poses for homing, instruments grasp and instruments delivery. The robot moves towards instruments grasping pose, grasps the instrument with the magnetic gripper and delivers it to the user. Then the robot stays idle until the F/T sensor senses the instrument collection by the user and 2 *s* extra time to ensure proper instrument collection. Finally, it returns to its homing position.

IV. Experimental Setup

TABLE I. Abbreviations for the experimental setup

Ab.	*Explanation*	*Ab.*	*Explanation*
HT	Human nurse only task	RN	Robotic scrub nurse
RT	Robot and human nurse task	SA	Surgeon assistant
ST	Surgical trainee	NA	Scrub nurse assistant
HN	Human scrub nurse		

*Research supported by the NIHR Imperial Biomedical Research Centre (BRC).

Alexandros A. Kogkas, Ahmed Ezzat and George P. Mylonas are with the HARMS Lab, Department of Surgery and Cancer, Imperial College London, UK (e-mail: a.kogkas15@imperial.ac.uk).

Rudrik Thakkar is with the St George's, University of London, UK.

Ara Darzi is with the Department of Surgery and Cancer, Imperial College London, UK.

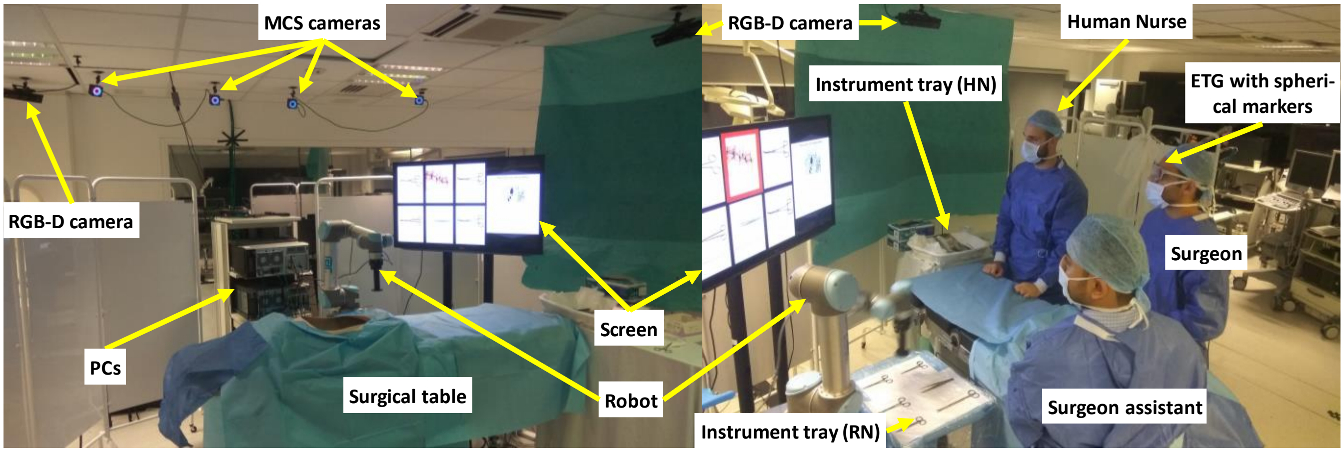

Fig. 1: The experimental setup.

The abbreviations stated in Table I are used to describe the experimental setup. STs were recruited to perform ex vivo resection of a pig colon and hand sewn end-to-end anastomosis. Each surgeon performed two experiments in randomized order: (1) a HT with the assistance of a HN and (2) a RT with the assistance of both RN and HN.

The instrument tray inventory consists of the 6 most frequently utilized instruments during this particular task. The RT starts with a 9-points eye-tracking calibration. Familiarization with the system setup is offered for 1 minute. During the task, the ST selects the desired instrument by gazing at the screen and once it is delivered and collected, the SA responds to verbal command or prior experience to return the instrument to its predefined position on the instrument tray. The ST addresses to the HN for further instruments. In case of wrong instrument delivery, the ST expresses the error verbally. If eye-tracking recalibration is necessary, the task continues after recalibration. During the HT the setup is identical, but the ST relies entirely on the HN to deliver instruments based on verbal commands. During both experiments, distractions are introduced to the HN. The NA asks the HN to stop and perform an instrument count twice and solve a puzzle at specific task stages.

10 STs participated (7 male and 3 female). Two had corrected vision. STs were between 30-40 years with at least 6 years surgical experience. 5 trained theater HN were recruited. One ST, with 2 years surgical experience, acted as SA and one medical student acted as NA for all experiments.

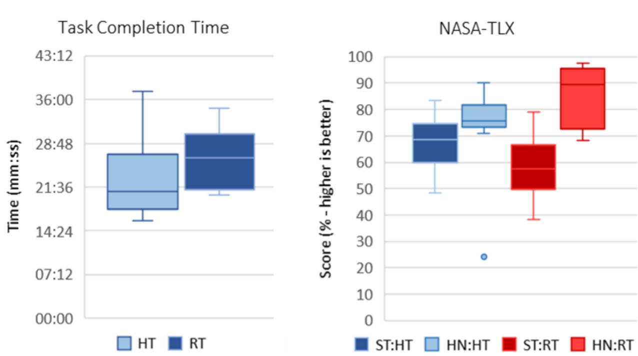

Fig. 2: Task completion time for HT and RT (left). NASA-TLX results for HT and RT responded by ST and HN (right).

V. Results

After each task, the ST and HN were asked to complete a NASA-TLX questionnaire and the results are depicted in Fig. 2. Overall time to complete the task was 22:35±6:30 *min* vs 26:04±4:50 *min* (HT vs RT, respectively) and no statistically significant difference was found. ANOVA was used to compare preference of ST and HN on HT over RT. Perception of HT and RT by ST and HN is also investigated. The only significant difference is shown on the perception of the RT by the ST (58±12%) and the HN (86±12%).

VI. Discussion

A novel robotic scrub nurse, responsive to surgeon gaze, has been proposed. This platform allows the surgeon to visually select an instrument, using an ETG device, pick it up and deliver to complete a task. We tested the RN with 10 different surgical teams in a simulation of a common operative scenario, with similar theater staff representation and operative field set up. Objectively, RT and HT showed no significant difference in overall experiment duration and all surgical experiments were completed. Subjectively, RN received positive feedback. NASA-TLX data demonstrated no significant difference between HN vs RN across perceptions.

We aim to further develop our gaze-guided RN by enabling real-time recognition and tracking of the surgical instruments and screen position in space. Practical aspects such as the RN delivering and returning the instruments will be implemented. Our ultimate goal is to develop a system to imitate the human nurse's most significant attribute, anticipation of next instrument selection. This involves work flow segmentation and task phase recognition.

References

[1] C. A. Velasquez, R. Mazhar, A. Chaikhouni, T. Zhou, and J. P. Wachs, "Taxonomy of communications in the operating room," in *Advances in Intelligent Systems and Computing*, 2018.

[2] M. Jacob, Y. T. Li, G. Akingba, and J. P. Wachs, "Gestonurse: A robotic surgical nurse for handling surgical instruments in the operating room," *J. Robot. Surg.*, 2012.

[3] A. A. Kogkas, A. Darzi, and G. P. Mylonas, "Gaze-contingent perceptually enabled interactions in the operating theatre," *Int. J. Comput. Assist. Radiol. Surg.*, pp. 1–10, 2017.