Gaze-Driven Human-Robot Interaction in the Operating Theatre

A.A. Kogkas, A. Darzi, G.P. Mylonas

Department of Surgery and Cancer, Imperial College London, UK
a.kogkas15@imperial.ac.uk

INTRODUCTION

A safe operating room has to constantly adapt to the increasing complexity of introduced new technologies and surgical procedures. Although new technologies may add complexity to the surgical workflow, at the same time they offer unique opportunities to improve patient safety, operational workflow and clinical outcome. The operating theatre is an environment where unintentional patient harm is most likely to happen, with most influential factors relating *to suboptimal communication* among the staff, *poor flow of information, staff workload and fatigue* and environment *sterility* [1].

For communication in particular, 30.6% of all team exchanges in the operating room are classified as failures, with one third resulting to immediate effects that can imperil patients [2]. Main cause for such failures is the lack of familiarity between the surgeon and the nurses, causing team instability and incoordination [3].

Therefore, keeping the surgeon in the loop of the decision making and task execution process is likely to reduce communication errors. Moreover, it is expected to improve the performance and efficiency of the surgeon. For example, a hand-gestures and voice-driven robotic nurse introduced by Jacob et al. has been shown to reduce the number of movements without significantly affecting task execution time compared to collaboration with human nurses [4]. Hands-free interactions could prove more beneficial.

Eye-tracking methodology has the potential to provide a "third hand" and a seamless way to allow "perceptually enabled" interactions with the surgical environment. Previous work demonstrated screen-based gaze control of surgical instruments [5] and improved collaboration among staff during surgery [6].

More recently, we have introduced a novel framework for theatre-wide and patient-wise 3D gaze localisation in a simultaneous and unrestricted/mobile fashion [7]. An extension of this framework is presented here, that allows hands-free gaze-driven interactions with the environment and a robotic manipulator. The framework is expected to facilitate seamless and meaningful integration of human and technology in the theatre for improved safety, collaboration and clinical outcome.

MATERIALS AND METHODS

The original framework presented in [7] uses wearable eye-trackers and their integrated scene cameras to provide 2D gaze information from one or more users. Concurrently, RGB-D cameras are used for real-time 3D reconstruction of the theatre environment. The Parallel Tracking and Mapping (PTAM) methodology [8] is employed to estimate the user's head pose within the reconstructed theatre. The pose is then used to map the 2D gaze information reported by the eye-tracker to a unique 3D fixation in the world frame-of-reference. For the work presented here, the framework is complemented by an articulated collaborative robotic arm, which is also co-registered with the reconstructed theatre environment. In a nutshell, depending on the required behaviour and a decision making procedure, the robot can perform tasks modulated by one or more users' 3D fixations and gaze behaviour (Figure 1).

Based on the eye-tracker's monocular scene camera, PTAM first generates a 3D keyframe and scaled map of the unknown environment. Then it updates the keyframe map and tracks the relative camera pose in parallel.

For eye-tracking, the SMI (SensoMotoric Instruments GmbH) glasses are used, with a stated accuracy of 0.5° of visual angle and a scene camera with a resolution of

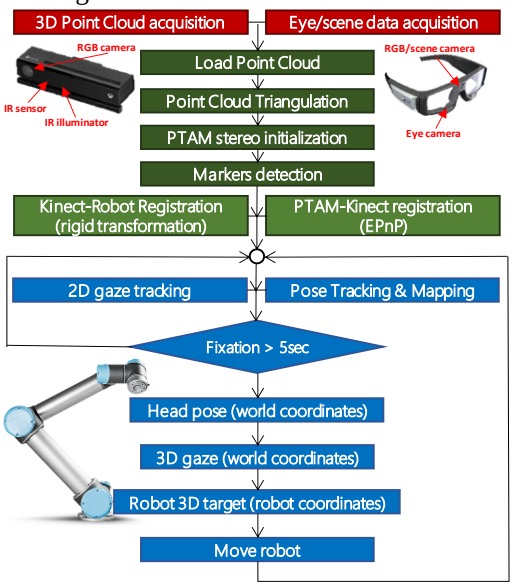

Figure. 1. The implementation consists of three main phases: data acquisition (in red), initialization (green), run-time (blue). PTAM is initialized by two keyframes. The first camera pose in PTAM coordinate system is estimated on the second keyframe, where we use 4 fiducial markers and EPnP [9] to estimate the real head pose (world coordinates) and then the transformation matrix between the two coordinate systems. Similar transformation matrix is estimated by the rigid transformation of the fiducial markers coordinates to register the RGB-D camera to the robot coordinate system. When the subject fixates for more than 5 sec, the head pose is transformed to world coordinates, the intersection of the 2D gaze vector with the triangulated point cloud provides the 3D fixation and the robot moves to these coordinates.

1280x960 pixels. For RGB-D sensing, the Microsoft Kinect 2 is used, with an RGB resolution of 1920x1080 pixels at 30Hz, 512x424 depth resolution with 70° (horizontal) and 60° (vertical) field of view, infrared sensor, time-of-flight technology and 30ms latency. The robot is a UR5 by Universal Robots, with 6 DOF, ±360° joint ranges, a reach radius of up to 850mm and payload of up to 5 kg.

EXPERIMENTAL SETUP

For the evaluation of our framework an experimental setup is used, including four fiducial markers (for the initialization phase) and three objects of different sizes (Figure 2). The task involves gazing in a random order at marked points on the objects. Four subjects, 2 males and 2 females, 25-60 years old, normal uncorrected vision, took part in the study. Off-line processing is performed and recorded 2D fixations are mapped to 3D. Fixations are transmitted to the robot if the dwell time is over 5sec. The robot then approaches the resolved 3D coordinates from above and according to the order of fixation, in a simulated gaze-guided manipulation task.

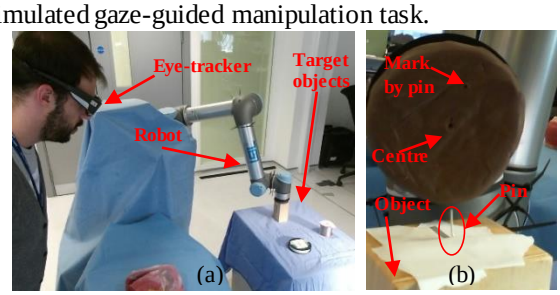

Figure. 2. (a) The experimental setup as seen from Kinect camera. (b) Plasticine used on the robotic tool for validation.

RESULTS

The accuracy of 3D fixation and gaze-guided manipulation task is evaluated. The error is calculated as the Euclidean distance between an object marker's actual coordinates and the respective recovered fixation coordinates. For task related error evaluation, a thick layer of plasticine is pasted on the robot end-effector. A thin pin is positioned on each of the three target objects. The distance from the centre and the depth of the pin imprint on the plasticine provides a measure of task accuracy.

The results summarized in Table 1 show the error averaged over all subjects. The 3D gaze error is 4.22cm and the gaze-guided task error is 4.20cm. While these values are expected identical and refer to theoretical and actual results respectively, a slight difference is caused by inaccurate robot to world registration.

Table 1: Mean error, std deviation and max. error for all subjects.

	Mean	SD	Upper
3D Gaze error (cm) (compounded by eye-tracking error)	4.22	0.37	4.56
Gaze-guided task error (cm)	4.20	0.39	4.51

DISCUSSION

A novel framework has been presented that allows gaze-driven interaction with an operating theatre environment and a collocated robotic manipulator. This is achieved by the combination of unrestricted wearable gaze-tracking, theatre 3D reconstruction and advanced computer vision concepts. The investigation supports our hypothesis that human vision can be used to achieve seamless and accurate collaboration between a surgeon and a table-side robot, but further studies are required for practical evaluation. The 3D gaze error is not uniformly distributed over the entire theatre as it depends on many parameters. Application specific considerations can be used for error minimisation. For instance, task accuracy is expected to improve with the integration of a robot force/torque sensor. Immediate work will focus on *safety, ergonomics, object and body recognition*, as well as *real-time* aspects. Further plans involve the implementation of *workflow segmentation* and *task-phase recognition* functionalities, based on the plurality of real-time data provided by the framework. *Context-awareness* will reveal a vast array of perceptual information and unlock new applications.

This work is supported by NIHR Imperial Biomedical Research Centre (BRC) award P61946. We would like to thank Mikael Sodergren for providing the eye-tracking hardware.

REFERENCES

[1] C. K. Christian, M. L. Gustafson, E. M. Roth, T. B. Sheridan, T. K. Gandhi, K. Dwyer, M. J. Zinner, and M. M. Dierks, "A prospective study of patient safety in the operating room," *Surgery*, vol. 139, no. 2, pp. 159–173, 2006.

[2] L. Lingard, S. Espin, S. Whyte, G. Regehr, G. R. Baker, R. Reznick, J. Bohnen, B. Orser, D. Doran, and E. Grober, "Communication failures in the operating room: an observational classification of recurrent types and effects.," *Qual. Saf. Health Care*, vol. 13, no. 5, pp. 330–4, 2004.

[3] J. Carthey, M. R. De Leval, D. J. Wright, V. T. Farewell, and J. T. Reason, "Behavioural markers of surgical excellence," *Saf. Sci.*, vol. 41, no. 5, pp. 409–425, 2003.

[4] J. P. Wachs, M. Jacob, Y.-T. Li, and G. Akingba, "Does a robotic scrub nurse improve economy of movements?," *Proc. SPIE - Int. Soc. Opt. Eng.*, vol. v 8316, p. 83160E, 2012.

[5] D. P. Noonan, G. P. Mylonas, A. Darzi, and G.-Z. Yang, "Gaze contingent articulated robot control for robot assisted minimally invasive surgery," *2008 IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, pp. 1186–1191, 2008.

[6] A. S. A. Chetwood, K.-W. Kwok, L.-W. Sun, G. P. Mylonas, J. Clark, A. Darzi, and G.-Z. Yang, "Collaborative eye tracking - a potential training tool in laparoscopic surgery," *Surg. Endosc.*, vol. 26, no. 7, pp. 2003–2009, 2012.

[7] A. A. Kogkas, M. H. Sodergren, A. Darzi, and G. P. Mylonas, "Macro- and Micro-Scale 3D Gaze Tracking in the Operating Theatre," in *The Hamlyn Symposium on Medical Robotics 2016 (accepted for publication)*.

[8] G. Klein and D. Murray, "Parallel tracking and mapping for small AR workspaces," *2007 6th IEEE ACM Int. Symp. Mix. Augment. Reality, ISMAR*, 2007.

[9] V. Lepetit, F. Moreno-Noguer, and P. Fua, "EPnP: An Accurate O(n) Solution to the PnP Problem" *Int. J. Comput. Vis.*, vol. 81, no. 2, pp. 155–166, 2009.