Macro- and Micro-Scale 3D Gaze Tracking in the Operating Theatre

A. A. Kogkas, M. H. Sodergren, A. Darzi, G. P. Mylonas

Department of Surgery and Cancer, Imperial College London, UK

a.kogkas15@imperial.ac.uk

INTRODUCTION

By approaching the eyes as the only visible part of the brain, we can consider them not just in the conventional sense as receptors of visual stimuli, but also as perception- and cognition-rich actors within the operating theatre. Using eye-tracking, several concepts have been developed for harnessing the power of the human visual system [1], for assessment of expertise [2], for objective measurement of surgical skills and for training [3].

In robotic [4] and laparoscopic [5] surgical settings, screen-based collaborative eye-tracking of multiple collaborators was shown to significantly improve verbal and non-verbal communication, task cooperation and efficiency, by allowing to share gaze information between the collaborators. However, screen-based eye-trackers restrict the surgeon’s head to a fixed position and can only provide 2D gaze information on a screen (micro-scale). On the other hand, wearable eye-trackers can provide unrestricted theatre-wide (macro-scale) 2D gaze information on the head frame-of-reference (Fig. 1). Simultaneous use of both modalities is mutually exclusive due to IR interference of the eye cameras. Moreover, 3D world coordinates would be desirable and beneficial.

We propose a novel framework that can provide unrestricted and simultaneous 3D eye-tracking on both the macro- and micro-scales, by using just a wearable eye-tracker. Preliminary investigations are carried out to assess the feasibility of the proposed framework.

Fig. 1 (a) Screen based eye-tracker. (b) Wearable eye-trackers provide 2D gaze coordinates on the scene camera frame-of-reference, which is equivalent to the head frame-of-reference.

MATERIALS AND METHODS

The proposed framework combines the use of a wearable eye-tracker and its respective RGB/scene camera, RGB-D cameras for real-time 3D reconstruction of the environment and the Parallel Tracking and Mapping (PTAM) methodology [6]. The objective is to estimate the head pose within the 3D reconstructed operating theatre and use this information to map the 2D fixations reported by the eye-tracker to macro-scale 3D fixations in the world frame-of-reference.

Based on a monocular camera, PTAM first generates an initial 3D keyframe scaled map of an unknown environment using the five-point stereo algorithm [7]. Then it builds/updates the keyframe map and tracks the camera pose relative to it, in parallel. For the macro-scale requirements, we use PTAM to estimate the eye-tracker’s RGB/scene camera pose (equivalent to the head pose) in the world frame-of-reference. This information and the 2D fixation point reported by the eye-tracker provide the gaze vector. In a final step, the sparse 3D point cloud generated by PTAM is triangulated and the 3D fixation is estimated by the intersection of the gaze vector with a triangular facet (Fig. 2). For the micro-scale requirements, the 3D model of the laparoscope screen is detected and tracked. Gazing on the screen is automatically detected and provided as 3D fixations on the screen model. As a final step, the screen plane position can be refined using perspective transformation in order to obtain more accurate micro-scale fixations.

Fig. 2 (Top) Views from the eye-tracker’s scene camera, showing the 2D fixations (red dot), the reprojected 3D fixations (green circle), as well as the PTAM generated 3D map and point cloud. (Bottom) The 3D fixation estimation.

Capturing more accurate real world coordinates and dynamic environment updates would require co-registration of a dense 3D point cloud from an RGB-D camera with the corresponding sparse 3D features provided by PTAM. Due to the dissimilarity of the corresponding geometric features, this is not implemented in the current off-line implementation. Instead, a static 3D map was generated once by PTAM, explicitly scaled to real world coordinates, pre-allocated during an initialization step and used thereafter. Also, at the moment, the 3D model of the laparoscope screen is detected and tracked with the assistance of fiducial markers at its four corners. For eye-tracking, the SMI (SensoMotoric Instruments GmbH) glasses are used, with a stated accuracy of 0.5° of visual angle and an RGB/scene camera with a resolution of 1280x960 pixels.
EXPERIMENTAL SETUP

For the experimental setup, an image with ten explicitly positioned and numbered markers is displayed on the monitor. Similarly, ten markers are placed on objects on a surgical table (Fig. 3). The task involves gazing in a predefined order at the monitor and table markers by using variable body and head poses. Eight subjects, 5 males and 3 females between the ages of 24 and 27 with normal uncorrected vision, were invited to take part in the study. After informed consent, the subjects were taken through the experimental setup and given time to familiarize themselves with it. Subjects underwent a 3-point eye-tracking calibration by fixating on a printed pattern at 1 m distance. Each subject was then instructed to fixate at the twenty markers in ascending order and for at least 3 seconds each, from a predefined standing position, allowing free head-pose. After completion of one round, subjects were asked to repeat the process from a different position. 2D fixations in the head frame-of-reference were recorded by the eye-tracker.

Off-line analysis using the proposed framework was undertaken and the recorded 2D fixations were mapped to 3D fixations in the world frame-of-reference (macro-scale) and in the monitor frame-of-reference (micro-scale) as appropriate.

RESULTS

The actual 2D screen and 3D world marker coordinates are compared with those recovered with the proposed framework. The error is calculated as the average Euclidean distance between a marker’s actual coordinates and the respective recovered fixation coordinates. One subject was excluded because of corrupted raw eye-tracking data. The results summarized in Table 1 show the error averaged over all subjects. In micro-scale mode (screen), the error is reported as a percentage of the screen dimensions (51.3x29 cm). In the macro-scale mode (world), the error is reported in cm. The errors are 1.53% and 8.57 cm respectively.

These values are compounded by the intrinsic error of the eye-tracker due to ocular micro-movements during fixation (drift, microsaccades, physiological nystagmus), the parallax effect [8] and subject-specific cornea irregularities. The intrinsic error is calculated as the Euclidean distance as a percentage of the eye-tracker’s scene camera frame, between a marker’s 2D coordinates and the respective 2D fixation coordinates. A comparison of the intrinsic error with the overall framework error is performed by estimating the framework reprojection error. On the eye-tracker camera frame, the 2D reprojection of a 3D fixation is estimated along with the respected marker’s 2D coordinates. Their Euclidean distance as a percentage of the camera frame size provides the framework error. The results indicate the framework error at 2.37% and the eye-tracker error at 1.42%. This signifies that 59.92% of the overall framework error is attributed to the eye-tracker.

Table 1. Mean error, standard deviation and maximum error for all subjects.

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Micro-scale (%)</td>
<td>1.53</td>
<td>0.98</td>
<td>3.74</td>
</tr>
<tr>
<td>Macro-scale (cm)</td>
<td>8.57</td>
<td>6.85</td>
<td>20.87</td>
</tr>
<tr>
<td>Framework reprojection error (%)</td>
<td>2.37</td>
<td>1.89</td>
<td>5.92</td>
</tr>
<tr>
<td>Eye-tracker error (%)</td>
<td>1.42</td>
<td>1.03</td>
<td>3.37</td>
</tr>
</tbody>
</table>

DISCUSSION

An early framework that can provide unrestricted theatre-wide and patient-wise 3D eye-tracking was presented. Preliminary evaluation of its accuracy is deemed satisfactory. The study supports our hypothesis that combined macro- and micro-scale eye-tracking can provide accurate information on the surgeon’s fixations both globally and on a laparoscope monitor, which opens the way for a number of practical applications. The proposed framework also promises to solve the parallax problem, which will significantly increase the overall 3D eye-tracking accuracy.

Our immediate plans involve the generation of a real-time and dynamically updated 3D model of the theatre by the use of RGB-D cameras (e.g., Kinect), and implicitly generated and co-registered PTAM maps. This is also expected to improve the head pose estimation.

REFERENCES