

A Study of Saccade Transition for Attention Segregation and Task Strategy in Laparoscopic Surgery

Marios Nicolaou, Adam James, Ara Darzi, and Guang-Zhong Yang

Royal Society/Wolfson Medical Image Computing Laboratory & Department of Surgical Oncology and Technology, Imperial College London, London, United Kingdom.
{m.nicolaou, a.james, a.darzi, g-z.yang}@imperial.ac.uk

Abstract. The advent and accelerated adoption of laparoscopic surgery requires an objective assessment of both operative performance and perceptual events that lead to clinical decisions. In this paper we present a framework to extract the underlying strategy through the analysis of saccadic eye-movements that lead to visual attention, and identification of intrinsic features central to the execution of basic laparoscopic tasks. Markov modeling is applied to the quantification of the saccadic eye movements for elucidating the intrinsic behaviour of the participants and the spatial-temporal evolution of visual search and hand/eye coordination characteristics. It has been found that participants adopted a unified strategy but the underlying disparity in saccadic behaviour reflect temporal and behavioural differences that could be indicative of the mental process by which the task was executed.

1 Introduction

Minimal Invasive Surgery (MIS) has been an important technical development in surgery in recent years. It achieves its clinical goals with reduced patient trauma, shortened hospitalisation and improved diagnostic accuracy and therapeutic outcome. Laparoscopic surgery is a subset of the general field of MIS and relates to most procedures performed in the abdomen in which the surgeon is required to operate by remote manipulation, using specially designed, elongated instruments inserted into port sites that are located through small incisions at specific points in the abdominal wall. The operative field is viewed by means of a laparoscope in which a small camera relays a video signal to a 2D monitor. During laparoscopic surgery, however, the surgeon's direct view is often restricted, thus requiring a higher degree of manual dexterity. The complexity of the instrument controls, restricted vision and mobility, difficult hand-eye co-ordination, and the lack of tactile perception are major obstacles in performing laparoscopic procedures. To date, a number of techniques have been developed for objective assessment of operative skills during laparoscopic surgery [1]. Most existing techniques are concentrated on the assessment of manual dexterity and hand-eye coordination with the combined use of virtual and mixed reality simulators. These environments offer the opportunity for safe, repeated practice and

for objective measurement of performance. Limited research, however, has been carried out in assessing how attention and different visual cues are integrated. There are a number of adverse events that go unnoticed in the operating room and many of them are not only associated with dexterity but also attention and visual-spatial perception. Both performance and safety are affected due to constraints placed upon the surgeon in laparoscopic surgery [2]. These events are a growing problem for healthcare organizations around the world [3].

During laparoscopic procedures, the primary constraint is vision as the surgeon is limited by the narrow monoscopic field of view and the representation of the operative scene provided by the laparoscope. To navigate the procedure and execute the appropriate motor commands, the surgeon is reliant upon limited cues through perspective changes in anatomy, occlusion, lighting variations and surface textures [4]. The procedure also involves three entirely distinct, spatial environments or "domains" that include internal (the point of operation), external (the external surgical field and the operating theatre) and the 2-D laparoscopic video [5]. To understand the underlying perceptual and cognitive processes involved, it is necessary to examine in detail the visual search behavior during different surgical steps.

Visual search is the act of searching for a target within a scene that involves both eye and head movement. It is a reactive rather than deliberative process for most normal tasks. The best visual acuity of the human eye falls within a visual angle of one to two degrees. This is called foveal vision, and for areas that we do not direct our eyes towards when observing a scene, we have to rely on a cruder representation of the objects offered by non-foveal vision, of which the visual acuity drops off dramatically from the centre of focus. When we try to understand a scene we fixate our eyes on particular areas and move between them. The intrinsic dynamics of eye movements are complex and saccadic eye movements are the most important to consider when studying visual search. It has a fast acceleration at approximately 40,000 degrees/second² and a peak velocity of 600 degrees/second². The objective of a saccade is to foveate a particular area of interest in a search scene [6].

The manner in which a surgeon assimilates information from different visual cues has been studied at a basic level mainly through video examination and interview. Experienced surgeons seem to have unparalleled capabilities in mastering visual-spatial capabilities by integrating a number of visual and motor cues. They make extensive use of ancillary and corroborative landmarks to plan for each maneuver and compensate for the loss of depth perception due to video projection. The study of the exact mechanism that underpins these behaviours, however, has proved to be difficult as many actions and perceptual filtering that occurs, are rapid and happen at an unconscious level. The process is further hampered by the fact that visual features are difficult to describe and assimilation of near-subliminal information is cryptic. These drawbacks call for the use of eye tracking for assessing detailed visual search behaviour during surgery. It is well understood that the main function of the oculomotor system is to keep the centre of gaze very close to the point of greatest interest being temporally and spatially coupled to the task at hand [7] and analysis of saccades can reveal much about the underlying cognitive mechanisms that guide them [8]. The use of eye tracking has already enjoyed a certain level of success in decision support in radiology [9], and the purpose of this paper is to present a new paradigm of

analyzing surgical competence by extracting underlying task strategies through the analysis of saccadic eye movements that lead to visual attention whilst performing laparoscopic tasks.

2 Materials and Methods

2.1 Surgical Task and Experiment Setup

A homogenous sample group of seven second year medical students (3 male, 4 female, mean age=20) were recruited to participate in this study. Ethics approval was granted by St Mary's local research ethics committee (LREC) with informed consent being obtained from all participants. All participants had no prior laparoscopic experience. The elementary surgical task required participants to grasp a colored section of a simulated blood vessel fixed on a plastic skin pad and transect it at a subsequent colored section using real laparoscopic instruments including grasper and scissors, as shown in Figure 1. The skin pad was placed within a custom-made laparoscopic box trainer with laparoscopic ports for the insertion of the instruments. The operative field was recorded using an analog camera fixed within the box and live video footage was streamed via a computer onto a 2D video screen. The participants were required to perform the task by manipulating the instruments and obtaining visual feedback of their actions on the screen. The displayed video was recorded in digital format on the hard disk of a computer for later analysis. The two targets on the simulated blood vessel and both tool tips were differently colored as to allow the automatic extraction and tracking of these 4 features from the recorded video footage. Gaze positions were also recorded during the procedure using an eye tracker. A diagram and a photo of the overall hardware set-up can be seen in Figure 1(a) and (b).

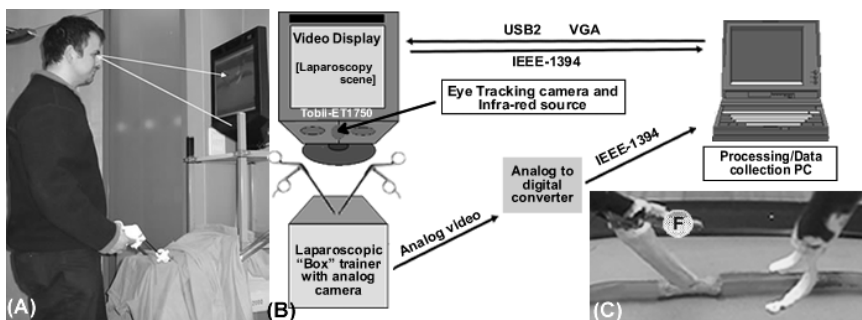


Fig. 1. (a) Experimental setup of the laparoscopic trainer used for this study, and (b) a schematic diagram showing the main data/command flow during real-time eye tracking with a Tobii remote eye tracker. (c) Shows an example of the laparoscopic view of the setup with foveation points super-imposed.

Gaze tracking was performed using a Tobii ET 1750 eye tracker. This is an infra-red video-based binocular eye-tracking system recording the position of gaze in the work

plane (screen) at up to 38 samples per second. A fixed infra-red light source is beamed towards the eye whilst a camera records the position of the reflection (known as the Purkinje reflection) on the cornea surface relative to the pupil centre. The infra-red images are real-time digitized and processed. Following a calibration procedure, the point of regard can then be determined with an accuracy of 1 degree across the work plane. This allows accurate tracking of the position of gaze of subjects standing approximately 60cm away from the equipment looking at the video screens as indicated in figure 1a. The system allows a certain amount of head movement, thus providing a realistic setting for the laparoscopic procedure. In this study, we used the screen to feed live video footage of the procedure. Good motion compensation combined with binocular eye tracking allows a seamless integration of eye tracking into the training procedure, and the average calibration time required is twenty seconds approximately [10].

2.2 Analysis of Fixation and Visual Search Strategy

The areas of interest (AOI) for a basic simulated laparoscopic task have been previously assessed [11] and we used this to determine the AOI for the study. The instrument tip (A), the target anatomy (B) and the feature space in-between (C) represented the AOI. For a detailed analysis of visual search and hand/eye strategy, the surgical task was decomposed into two steps: vessel grasping (task 1) and vessel transection (task 2). Each AOI was defined by different colours and automatically tracked by colour segmentation by hue-saturation space filtering previously described [12]. The co-ordinates of each AOI were combined with the gaze data and using developed software we determined which AOI foveal vision (taken as 2° visual angle) was centered. In order to extrapolate intrinsic information of each individual's dynamic eye movements, Markov modeling was used to investigate the sequence of temporal fixations [13] [9]. In this paper we focus on a discrete time, discrete state space model (DTMC) based on a first-order Markov process. The evaluation of a DTMC [14] is achieved by an observation at time i as a random variable X_i , variables $X_0, X_1, X_2, \dots, X_m$. For all states x_n the following applies:

$$\begin{aligned} P \{ X_{n+1} = x_{n+1} \mid X_0 = x_0, X_1 = x_1, \dots, X_n = x_n \} \\ = P \{ X_{n+1} = x_{n+1} \mid X_n = x_n \} \end{aligned} \quad (1)$$

In this study, areas A,B,C were designated as the states for the Markov model for both the grasping and the transection task. In the first instance, we examined the total number of fixations and transitions between all of the states for i and j for each participant. To further analyse the transition probabilities p_{ij} , between the states i and j were calculated by determining the number of fixations per AOI. The total numbers of transitions were normalized for both the task and for each sub task. To observe the transition probabilities between states intrastate transitions were excluded. Only three independent states were considered. The following approach was taken (2):

$$\begin{pmatrix} * & t_{12} & t_{13} \\ t_{21} & * & t_{23} \\ t_{31} & t_{32} & * \end{pmatrix} \rightarrow \begin{pmatrix} * & p_{12} = \frac{t_{12}}{t_{12} + t_{13}} & p_{13} = \frac{t_{13}}{t_{12} + t_{13}} \\ p_{21} = \frac{t_{21}}{t_{21} + t_{23}} & * & p_{23} = \frac{t_{23}}{t_{21} + t_{13}} \\ p_{31} = \frac{t_{31}}{t_{31} + t_{32}} & p_{32} = \frac{t_{32}}{t_{31} + t_{32}} & * \end{pmatrix} \quad (2)$$

The transition probabilities for the group were further analysed by averaging the sum of the matrices for each task to obtain the arithmetic mean. To demonstrate the spatial-temporal behavior of the individuals in terms of how strategy evolved over the course of the task, we compared the performances of two participants for a single task. The slowest and fastest participants were selected for temporal analysis.

3 Results

For each of the participants we calculated probability transitions matrices for the task as a whole, and the sub-tasks, task 1: grasping and task 2: transection of the vessel. We focused on transitions between states but not self-transitions. All of the other features in the space (C) were classified as a single state.

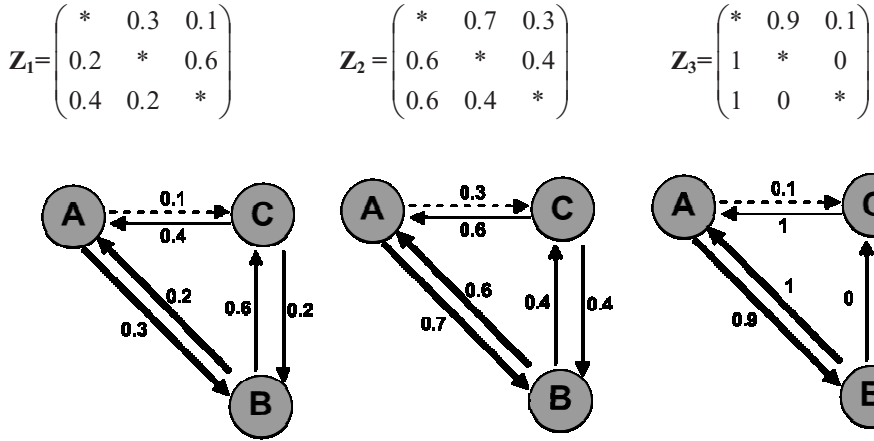


Fig. 2. Example Markov matrices derived for a single participant where Z_1 = Task (1&2) Z_2 = Task 1 (Grasping), Z_3 = Task 2 (transection). Only interstate transitions were considered with the thickness of the line representing the strength of the relationship between states.

The transition data for the participants represents the feature states that formed the underlying strategy. A dominant strategy was observed for the proportion of individuals, with key transitions between $A > B$ and $B > A$ corresponding to both the

instrument and the vessel. The eye saccades between the two states, with each feature (A,B) receiving the proportion of fixations. Consequently, each feature receives maximum attention, as the fovea, the area of the eye with highest acuity, is centered on the feature. The mean transitions of the group followed this underlying strategy. Figure 4 illustrates the mean strategy for the task with two states being more dominant more than the rest.

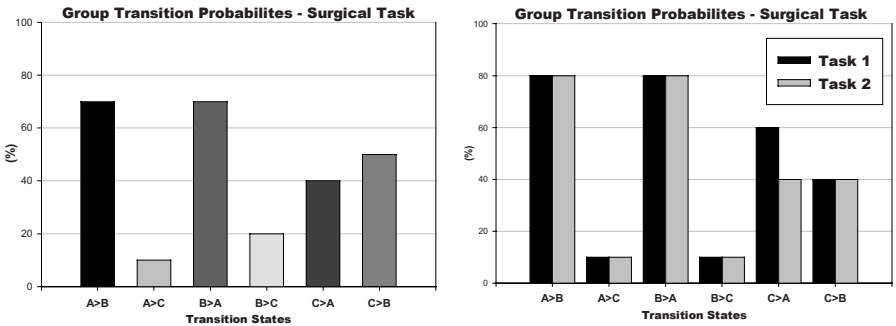


Fig. 3. (a) Mean transition probabilities for the participants performing the surgical task. Transitions between the vessel and instrument were the clear method adopted. There are further saccades between the vessel, instrument and surrounding feature space. (b) Decomposed mean transition probabilities for the group for task 1 and 2. Data clearly illustrates the similarities in strategy for both task one and two.

By decomposition of the task into two subtasks: grasping (1) and transection task (2), it is shown that there is little change in strategy with predominant saccadic activity continuing between the states A>B and B>A, as illustrated in Figure 3.

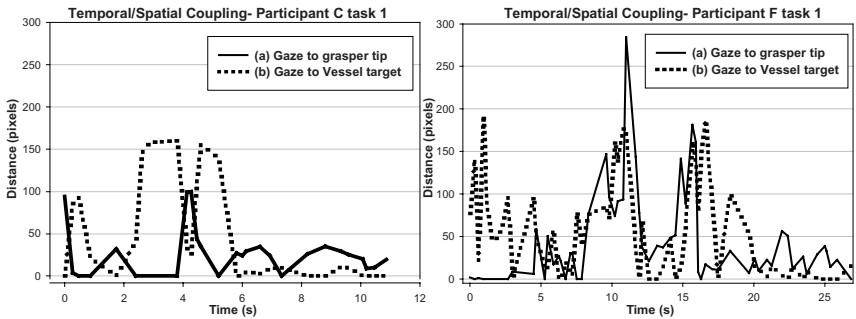


Fig. 4. Data representing the temporal and spatial coupling of eye behaviour for two participants (C) and (F).

The temporal and spatial coupling of saccadic behaviour of the individuals draws out characteristics differences in their underlying strategy and execution of the task. For participant C gaze alternates between instrument (A) and vessel (B) as shown by the

waveform of the graph (with a predominant strategy of gazing on the instrument). Six seconds in, fixation concentrates on the vessel whilst the grasper tip is fine tuned into position. A distinct change in behaviour occurs (6s) coupled to the task of executing the grasping of vessel once having attuned the grasper to the correct distance. The behaviour of participant F is distinctly different with most time at the beginning of the task being spent fixating on the instrument. Between 6-16 seconds gaze shifts between the two with some alternate behaviour at 16-20s, guiding the instrument to the vessel. From 20 seconds, like participant C, fixation is concentrated on the vessel, whilst the instrument homes in, void of gaze supervision.

4 Discussions and Conclusions

In conclusion, we have presented a framework for extracting visual search behaviour by the analysis of eye movements and transitions between areas of interest (AOI) within a simple laparoscopic surgical task. The results suggest that there is a definite pattern of transitions within a homogeneous group of laparoscopic novices: a sequence of transitions from the instrument to the vessel and from the vessel to the instrument. This may seem an obvious strategy but it was clear to see that some individuals clearly formed this strategy early on. More importantly temporal analysis revealed intrinsic differences in the process by which the eye of individuals adopts and dictates the strategy. This implies that the underlying mental processes controlling the eye guide the execution of the task. These processes subsequently determine the timely nature by which the task is conducted. Temporal analysis distinguishes the strategic approach and importantly highlights the disparities between individuals that are important to understanding the exact nature of how the task was conducted.

As the surgical profession is in need of a reliable and valid method of assessing operative skill that is distinct from most subjective assessments [7], an understanding of the implicit strategies and features that guide an individual may provide information on the disparities in performance, ultimately quantifying the precise differences between individuals. Understanding the exact nature of these differences may provide a basis by which to attain procedural knowledge critical to optimize task performance. A system that can provide unbiased and objective measurement of surgical precision could help training, complement knowledge based examinations, and provide a benchmark for certification [15] Such a system would have wider implications in guiding the development of safety protocols and the current study is an important step towards the development of such an integrated framework.

Acknowledgements. We would like to thank our colleagues, Rachel King, Michael Rans and Benny Lo from the Royal Society/Wolfson Medical Image Computing Lab. for their support and contribution to the study. We would also like to thank the second year medical students of Imperial Medical School for their support and enthusiastic participation in the study.

References

1. Moorthy K, Munz Y, Sarker SK, Darzi A. Objective assessment of technical skills in surgery. *BMJ*, 1, 1032-37, 2003.
2. Cao CGL, Mackenzie CL. Task and motion analyses in endoscopic surgery. 1996 ASME IMECE Con Proceed: 5th Ann Sym on Haptic Environment and Teleoperator systems, 583-90.
3. Vincent C, Neale G, Woloshynowych M. Adverse events in British hospitals: preliminary retrospective record review. *BMJ*, 3, 517-18, 2003.
4. Tendick F, Bhoyrul S, Way L. Comparison of laparoscopic imaging systems and conditions using knot tying task. *Computer Aided Surg.*, 2(1), 24- 33, 1993.
5. Tchalenko J, James A, Darzi A, Yang G-Z. *Proceedings ECEM* 12, 2003.
6. Yang G-Z, Dempere-Marco L, Hu X-P, Rowe A. Visual Search: psychophysical models and practical applications. *Image and Vision Computing*, 20, 291-305, 2002.
7. Land M, Mennie N, Rusted J. The roles of vision and eye movements in the control of activities of daily living. *Perception*, 28, 1311-1328, 1999.
8. Ballard DH, Hayhoe MM, Li F, Whitehead SD. Hand-eye coordination during sequential tasks. *Philosophical Trans. of the Royal Society of London* , Series B 337, 331-39, 1992.
9. Dempere-Marco L, Hu X-P, MacDonald SLS, Ellis SM, Hansell DM, Yang G-Z. The use of visual search for knowledge gathering in image decision support. *IEEE Trans. on Medical Imaging*, 21(7), 741-54, 2002.
10. Tobii technology. User Manual, 2003, <http://www.tobii.se>.
11. Law B, Atkins SM, Lomax AJ, Wilson J. Eye trackers in virtual laparoscopic training environment. *Medicine Meets Virtual Reality: J.D Westwood et al (Eds)*, IOS Press, 184-186, 2003.
12. Lo BP, Darzi A, Yang G-Z. Episode classification for the analysis of tissue/instrument interaction with multiple visual cues. *MICCAI*, 2878, 230-237, 2003.
13. Hacisalihzade SS, Stark LW, Allen JS. Visual perception and sequences of eye movement fixations: A stochastic modeling approach. *IEEE Trans. Syst., Man, Cyber.*, , 22, 474-81, May/June 1992.
14. Meyn SP, Tweedie RL. *Markov chains and stochastic stability*. London, UK: Springer-Verlag, 1996.
15. Taffinder N, Smith S, Darzi A. Assessing operative skill. *BMJ*, 318(3), 887-888, 1999.