

Eye-Gaze Driven Surgical Workflow Segmentation

A. James, D. Vieira, B. Lo, A. Darzi, and G.-Z. Yang

Royal Society/Wolfson Medical Image Computing Laboratory & Department of Biosurgery and Surgical Technology, Imperial College London, London, United Kingdom
{a.james, d.vieira, benny.lo, a.darzi, g.z.yang}@imperial.ac.uk

Abstract. In today's climate of clinical governance there is growing pressure on surgeons to demonstrate their competence, improve standards and reduce surgical errors. This paper presents a study on developing a novel eye-gaze driven technique for surgical assessment and workflow recovery. The proposed technique investigates the use of a Parallel Layer Perceptor (PLP) to automate the recognition of a key surgical step in a porcine laparoscopic cholecystectomy model. The classifier is eye-gaze contingent but combined with image based visual feature detection for improved system performance. Experimental results show that by fusing image instrument likelihood measures, an overall classification accuracy of 75% is achieved.

Keywords: Eye-tracking, surgical workflow, neural networks, minimal invasive surgery, surgical simulation.

1 Introduction

Over the past decade, the operating room has undergone significant transformation to evolve into a highly complex and technologically rich environment. Increasingly, surgeons find themselves immersed amidst the sophistication of surgical workflow and procedures. Minimally Invasive Surgery (MIS) is an important advance in surgical technology with proven benefits to the patient and care providers. They include reduced hospitalization, shorter rehabilitation, less pain and decreased hospital costs. However, as the procedures and devices used have become more specialized, the need for a greater range of supporting equipment has arisen. The increase in equipment not only obstructs the area around the surgical field but also increases the burden on the surgeon who has to adapt to and integrate with these systems during the operation.

For a surgeon to assimilate all of the components during a surgical procedure, it adds further pressure above and beyond the clinical demands encountered during MIS. In addition, there is continual pressure for surgeons to demonstrate their surgical competence so as to improve standards and patient safety, in part by recognizing and reducing errors [1]. Thus far, the existing training paradigm has not adapted to the dynamic, technologically rich environment [2]. Existing work in the field has focused on improving training through the use of virtual reality trainers and simulated operating rooms [3,4]. Subsequent work has focused on recovering instrument metrics

in an attempt to classify the underlying surgical repertoire and evaluate surgical skill levels [5,6]. Time series analysis has also been used to classify the overall surgical workflow [7,8]. Current work has made progress in classifying and automating the recognition of surgical performance and surgical workflow. However, no one framework has the means of objectively assessing surgical skills and recovering surgical workflow with sufficient levels of detail to tease out subtle behaviors that lead to poor outcome.

Existing research has shown that the main focus in MIS should be to establish quantitative methods of assessing not only manual dexterity [9,10] but also the cognitive process that leads to potential surgical errors. In practice, the understanding of the cognitive and perceptual process that governs the surgical decision is a challenging problem. Existing research [11] has shown that by the use of eye-gaze tracking, it is possible to elucidate certain aspects of this process. Eye movements can provide a number of clues to the underlying cognitive processes and their interactions with the outside world [12]. During MIS, surgeons process a wide range of sensory information and certain aspects are derived through overt changes in gaze direction. Even for simple procedures, the myriad of eye movements in response to the underlying visual attention is so complex that most of them are sub-consciously applied. In this case, post recall of the exact process involved is difficult, and this represents the most challenging part of eliciting surgical workflow and the surgical sequence in MIS.

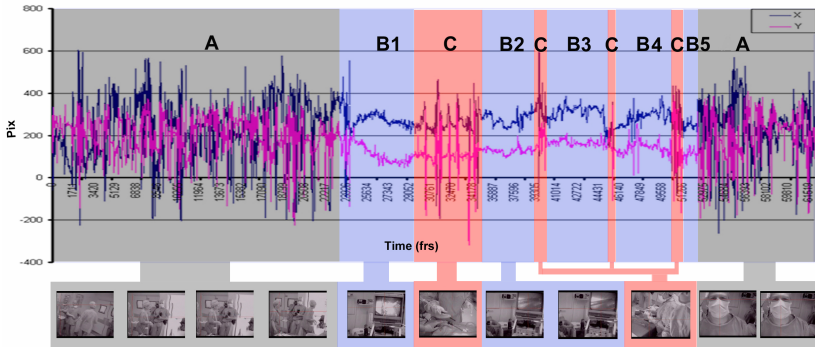


Fig. 1. Eye-gaze data captured from an in-house mobile eye-tracker system during a laparoscopic cholecystectomy. The eye movement amplitude has a syntactic structure that is indicative of underlying surgical behavior. Section A = operating theatre interaction, B = primary surgical maneuvers (B1 = Grasp Hartmann’s pouch, B2 = dissect cystic artery/duct, B3 = Clip cystic artery/duct, B4 = Cut cystic artery/duct and B5 = dissect gall bladder from liver) and C = tool transition periods.

To demonstrate this effect, Fig. 1 shows an example eye tracking trace for a laparoscopic cholecystectomy procedure performed in the operating theatre, demonstrating how visual attention changes throughout different stages of the procedure, including the interaction in the operating theatre, primary surgical steps and instrument transitions. It is apparent that basic characteristics of the eye

movement are different. Sequences of focused attention with prolonged fixation and smooth pursuit movements are purposeful but may not be consciously registered by the surgeon. Current research has shown that spatio-temporal recognition of eye-movements can provide a means of quantifying attention selection. The identification of characteristic maneuvers, however, is a prerequisite for quantifying surgical skills and identifying the underlying idiosyncrasy of different operators. The purpose of this paper is to propose an eye-gaze driven MIS workflow segmentation scheme. We will demonstrate the accurate segmentation of a primary surgical step from within a porcine laparoscopic cholecystectomy model. By combining this with traditional video sequence analysis techniques, the proposed method has shown to give significantly improved performance in workflow segmentation and the potential to accurately segment surgical sequences.

2 Materials and Methods

2.1 Experiment Setup

Three senior specialist registrars were recruited to participate in the study and perform a full laparoscopic cholecystectomy (gall bladder removal) procedure on a porcine model. Each of the three registrar's performed five procedures forming a total data set of fifteen procedures. It is well documented that approximately 95% of all gall bladder removals are performed laparoscopically. The frequency of the procedure in practice lends itself to the implementation as a model for skill and workflow analysis [8]. In a typical acute patient with normal anatomy, there are six primary steps to the surgical workflow once the surgeon has entered the peritoneal cavity. Due to the nature of the porcine anatomy, only five out of the six steps are achievable, due to the absence of the cystic artery. The five surgical steps of the porcine cholecystectomy are illustrated in Fig.2.

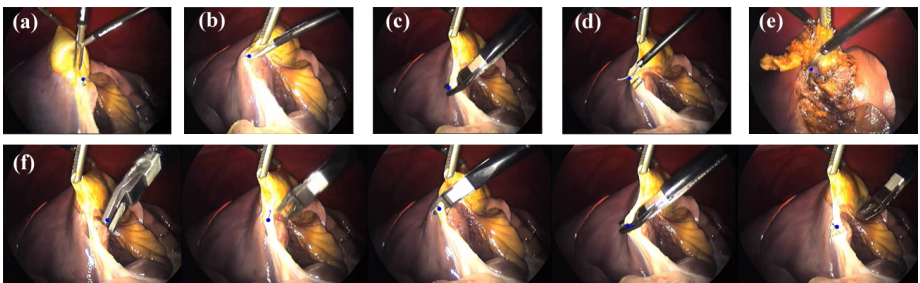


Fig. 2. The five main steps: (a) 1 = Grasp the Hartmann's pouch. (b) 2 = Dissect the cystic duct. (c) 3 = Clip the cystic duct (d) 4 = Transect the cystic duct (e) 5 = Dissect the gallbladder from the liver bed (f) Represents the primary surgical step classified in the experiment where each participant has to navigate with an endoclip instrument and clip the cystic duct in three places, one clip placed proximally and two distally.

The focus of the study was to develop a framework to classify and automate the recovery of the critical surgical step, step 3 (clipping of the cystic duct). The selection of this step is based on the fact that most iatrogenic injuries occur during this step, as the common bile duct (CBD) is often mistaken for the cystic duct, resulting in transection of the CBD and occasionally excision of the CBD and most of the hepatic duct. During the procedure eye-gaze was recorded using an eye tracker (Tobii ET 1750). 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 Hz. With this device, a fixed infra-red light source is beamed towards the eye whilst a camera records the position of the 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 one degree across the work plane [13].

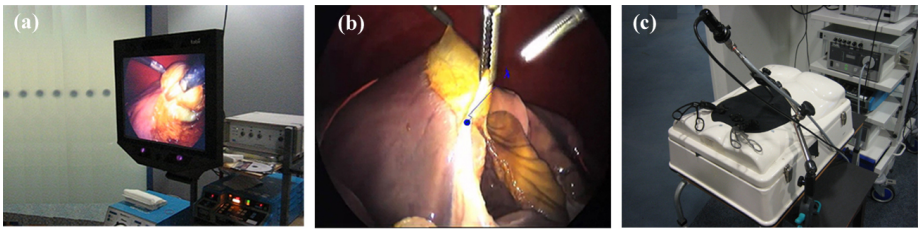


Fig. 3. (a) Tobii ET 1750 eye-tracker integrated into laparoscopic stack. (b) Video frame captured from the ET 1750 demonstrating the gaze transition (blue point). (c) Box trainer used to simulate laparoscopic conditions and house the porcine model.

Workflow segmentation was performed using eye-gaze vectors and image features. It is well known that video processing is computationally expensive so our approach employs eye-gaze to drive image feature selection to guide the search in the image space towards regions containing salient information. The outline of the classifier is illustrated in Fig. 4.

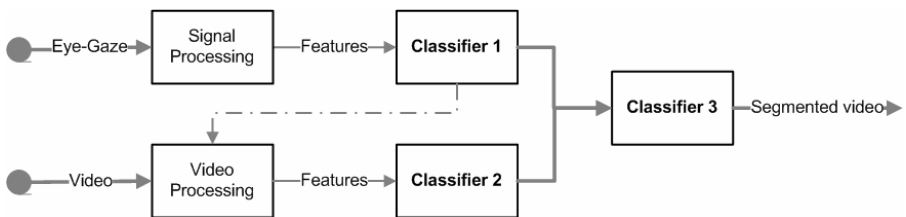


Fig. 4. Eye-gaze driven segmentation framework that calls on selected image features to classify surgical step 3 (clip cystic duct) in the procedure

Instrument transitions were selected to define the time window to segment the surgical step from the workflow. It is typical for the surgeon to exchange instruments between the primary steps of the procedure and the underlying gaze behavior is

indicative of the transitions. Gaze deviates away from the work plane (monitor), introducing a perturbation in the eye signal that can be detected and classified. Gaze off-screen periods can also be brought about by other factors. However, our work classifies the off-screen periods to determine their likelihood of instrument transitions, thus automatically defining the recognition window for surgical step 3.

The eye gaze data was processed and classification likelihood was given based on the off-screen duration and the length of the time window between two off-screens, to define Classifier 1 in Fig.4. For Classifiers 1 and 2, it is common to deal with a small data set, so a more robust learning machine, based on small sample theory is implemented.

2.2 Machine Learning with a Small Training Data Set

One of the challenges of MIS workflow analysis is the size of the training data set available for machine learning. Given a training set $S = \{(x_1, y_1), \dots, (x_T, y_T)\}$, where x_t is the input and y_t is the corresponding response, the main problem of machine learning is to find the optimal machine parameter w that maps x_t , *i.e.*, $f(x_t, w)$ to the supervisor's response y_t . One possible way to find is to use the empirical risk minimization (training error minimization):

$$R_{emp} = \sum_{t=1}^T L(y_t, f(x_t, w)). \quad (1)$$

It was observed in [14] that the empirical risk minimization is not consistent given problems when the data sample is small. For these problems the minimization of the structural risk is more adequate. The minimization of structural risk is a bi-objective optimization problem in which two conflicting objectives should be minimized. They are related to the empirical training error (2), and the machine complexity [15].

$$\min \begin{cases} f_1 = R_{emp} \\ f_2 = \Omega \end{cases}. \quad (2)$$

Mathematically, f_1 represents the minimization of some empirical risk function (2) and f_2 the minimization of the complexity, *i.e.*, the VC dimension, function gradient and high-frequencies energy. Using this formulation it can be theoretically proved that the upper bound of the learning machine error is minimized.

In this study, a neural network called Parallel Layer Perceptron (PLP) was used to classify the off-screen periods [16]. The PLP was trained using a bi-objective formulation that takes into account the minimization of the training error and machine complexity [17]. Decreasing the machine complexity is fundamental to achieve reliable results when the training set is small and/or is composed of many outliers. The problem this paper proposes to solve is affected by both, small sample sizes and outliers.

2.3 Video Analysis for Instrument Detection

In a laparoscopic cholecystectomy procedure, a select number of instruments are required for different surgical steps. As such, the segmentation of surgical sequences

from the workflow can be enhanced by identifying and detecting the specific instrument in use. In this study, we incorporated a simple visual based instrument detection method to recognize the change of instrument. To achieve motor control fidelity, surgeons generally position the laparoscopic instruments on the left and right parts of the laparoscopic view. To detect the change of surgical instruments, a region was defined empirically on the laparoscopic video, and a Sobel edge segmentation technique was applied to segment the instrument from the predefined region. Instrument likelihood measures were determined by using the threshold values related to the presence or absence of the instrument in a predefined target window. These likelihood values are augmented into the classification process through Classifier 2.

3 Results

For all experimental data collected for this paper, the off-screen periods were divided in two classes, where the 1st class (C_1) has the off-screen periods representing instrument transitions and the 2nd class (C_2) represents off-screen periods with no instrument transition. Kurtosis and the skewness were measured for the two classes. The Kurtosis of the class C , $k(C)$, measures the Gaussianity of the distribution, where for Gaussian distributions $k(C) = 0$ and $k(C) > 0$ indicates that the distribution of C is outlier-prone. The Skewness measures the asymmetry of the data around the sample mean, where $S(C) = 0$ represents symmetrical data $S(C) > 0$ means that the data is distributed to the left and $S(C) < 0$ that it is distributed to the right. For the classes C_1 and C_2 the following was found: $S(C_1) = 0.9$, $k(C_2) = 16.16$, $S(C_2) = 3.4$. These values discourage the using of Naïve Bayes classifiers since the distributions cannot be well represented by Gaussians.

One PLP, PLP 1, with one Gaussian non-linear neuron in parallel with a 1st order polynomial was trained using the Minimum Gradient Method (MGM) described in Reducing the machine complexity is fundamental to achieving a reliable result when the training set is small and/or it is composed of many outliers [17]. In the present case, both situations held since C_1 is composed of few samples and the skewness of the classes indicates that they are susceptible to outliers. The positive observations are the same as the sample set, fifteen. A second PLP, PLP 2, was used to train the time length between the two off-screen periods that define step 3. This class C_1 is composed of fifteen positive examples and 223 negative observations. A voting machine was used to consider the results of C_1 and C_2 for the final classification. The assembly of PLP 1 and 2 defines Classifier 1 in Fig.4. This distribution is also asymmetric, thus, outlier-prone. As the data set is small the error was evaluated using the leave-one-out cross validation method, which is an unbiased statistical measure that is recommended with small sample data. A recognition accuracy of 66% was obtained with eye-gaze and an accuracy of 75% with eye-gaze and instrument detection.

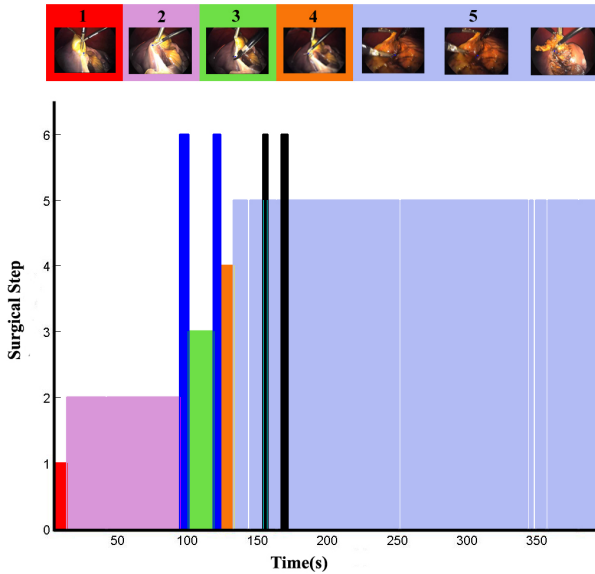


Fig. 5. The bar chart outlines the segmentation of surgical step 3. Step 3 was automatically recovered bounded by two off screen instrument transition periods (blue). All other steps 1, 2, 4 and 5 were manually segmented. The black bars highlight the false positive identification of off screen periods.

4 Discussion and Conclusions

Accurate segmentation of a surgical workflow is important for improving the general surgical standards and reducing errors in MIS. Thus far, little research has been conducted for investigating the cognitive processes that lead to surgical errors. In this study, we have demonstrated the promise of eye-gaze driven surgical workflow segmentation for a primary surgical step in a porcine laparoscopic cholecystectomy model. Experimental results have shown that the critical step can be recognized with an accuracy of 66% with eye-gaze alone and 75% with eye-gaze combined with image based workflow analysis. The use of PLP's and other associated neural networks techniques to elicit subtle behavior in inter- and intra-operative performance from eye-gaze has remained largely unexplored. The proposed method is novel in that it seamlessly integrates the perceptual behavior that is intrinsically linked to a fully developed workflow with the conventional video processing approach. This is essential for resolving surgical steps that are ambiguous in terms of visual feature representation, but cognitively carries a very different mental process.

Acknowledgements. We would like to thank our colleagues: George Mylonas and Marios Nicolaou and Smith & Nephew and Tyco Healthcare for their valuable support enabling us to advance our work.

References

1. Rattner, W.D., Park, A.: Advanced devices for the operating room of the future. *Seminars in Laparoscopic Surgery* 10(2), 85–89 (2003)
2. Taffinder, N., Smith, S., Darzi, A.: Assessing operative skill. *BMJ* 318, 887–888 (1999)
3. Gallagher, A.G., Satava, R.M.: Virtual reality as a metric for the assessment of laparoscopic psychomotor skills. *Surgical Endoscopy* 16, 1746–1752 (2002)
4. Aggarwal, R., Undre, S., Moorthy, K., Vincent, C., Darzi, A.: The simulated operating theatre: comprehensive training for surgical teams. *Qual Saf Health Care* 13, 27–32 (2004)
5. Rosen, J., Macfarlane, M., Richards, C., Hannaford, B., Sinanan, M.: Surgeon-tool force/torque signatures evaluation of surgical skills in minimally invasive surgery. *Medicine meets virtual reality- the Convergence of Physical & Informational technologies: options for a New Era in Healthcare* 62, 290–296 (1999)
6. Rosen, J., Solazzo, M., Hannaford, B., Sinanan, M.: Task decomposition of laparoscopic surgery for objective evaluation of surgical residents' learning curve using hidden markov model. *Comput. Aided Surg.* 7, 802–810 (2002)
7. Sielhorst, T., Blum, T., Navab, N.: Synchronizing 3d movements for quantitative comparison and simultaneous visualization of actions. In: *ISMAR. Proc. IEEE and ACM International on Mixed and Augmented Reality* (2005)
8. Ahmadi, S.-A., Sielhorst, T., Stauder, R., Horn, M., Feussner, H., Navab, N.: Recovery of surgical workflow without explicit models. In: *Larsen, R., Nielsen, M., Sporing, J. (eds.) MICCAI 2006. LNCS, vol. 4190, pp. 420–428. Springer, Heidelberg* (2006)
9. Martin, J., Regehr, G., Reznick, R., MacRae, H., Murnaghan, J., Hutchinson, C., Brown, M.: Objective structured assessment of technical skill (OSATS) for surgical residents. *Br.J.Surg.* 193(5), 479–485 (2001)
10. Datta, V., Mackay, S., Mandalia, M., Darzi, A.: The use of electromagnetic motion tracking analysis to objectively measure open surgical skill in the laboratory-based model. *J. Am.Coll.Surg.* 193(5), 479–485 (2001)
11. James, A., Tchalenko, J., Darzi, A., Yang, G.-Z.: *Proceedings ECEM, vol. 12* (2003)
12. Schmid, R., Zambrieri, D.: Strategies of eye-head coordination. In: *Schmid, R., Zambrieri, D. (eds.) Oculomotor control and cognitive processes, pp. 229–248. North Holland, Amsterdam* (1991)
13. Tobii technology. User Manual (2003), <http://www.tobii.se>
14. Vapnik, V.N.: *Statistical learning theory. Wiley, New York* (1998)
15. Caminhas, W.M., Vieira, D.A.G., Vasconcelos, J.A.: Parallel Layer Perceptron. *Neurocomputing* (55), 771–778 (2003)
16. Vieira, D.A.G., Vasconcelos, J.A., Caminhas, W.M.: Controlling the parallel layer perceptron complexity using multiobjective learning algorithm. *Neural Computing & Applications* 16(4/5), 317–325 (2007)
17. Vieira, D.A.G., Takahashi, R.H.C., Palade, V., Vasconcelos, J.A., Caminhas, W.M.: The Q-norm complexity measure and the Minimum Gradient Method: a novel approach to the machine learning structural risk minimization problem. *IEEE Transactions on Neural Networks* (2005)