33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Gaze-contingent autofocus system for robotic-assisted minimally invasive surgery

Neil T. Clancy, George P. Mylonas, Guang-Zhong Yang and Daniel S. Elson

Abstract—**A gaze-contingent autofocus system using an eye-tracker and liquid lens has been constructed for use with a surgical robot, making it possible to rapidly (within tens of milliseconds) change focus using only eye-control. This paper reports the results of a user test comparing the eye-tracker to a surgical robot's in-built mechanical focusing system. In the clinical environment, this intuitive interface removes the need for an external mechanical control and improves the speed at which surgeons can make decisions, based on the visible features. Possible applications include microsurgery and gastrointestinal procedures where the object distance changes due to breathing and/or peristalsis.**

I. INTRODUCTION

ADVANCES in minimally invasive surgery (MIS), particularly the robotic-assisted field, have made greater precision and diagnostic ability available to the surgeon through mechatronic and imaging improvements. The challenge is to integrate the surgeon with these new devices to reduce the burden of an extra layer of complex controls.

Eye-tracking technology is central to this as it provides an intuitive and natural interface with the user by tracking the location of the user's gaze. In the case of video-oculography, one of the less intrusive eye tracking techniques, this is achieved by illuminating the eye with infrared light and using a camera to image the reflection from the cornea ('glint'). Together with the position of the centre of the pupil, this defines a vector representing the gaze of the user [1,2]. In a binocular set-up, gaze vectors acquired from both eyes simultaneously can be used to calculate accurate coordinates of the point in space being observed (fixation point).

Previous work has shown that this technology can be integrated and used successfully in surgical robotics platforms by using the calculated 3D gaze position to visually stabilize motion in a beating heart operation [3, 4]. Furthermore, data acquired from the tracker also allowed the optimization of the stabilization algorithm by analyzing the effect of motion at the periphery on overall visual acuity during a surgical task [5, 6]. More recently, the concept of Gaze-Contingent Perceptual Docking presented in [7], makes use of eye tracking as a means of more effective human-machine interaction, improved hand-eye coordination and *in situ* description of safety boundaries without prior knowledge of tissue morphology.

While many MIS procedures will not require dynamic focusing due to the relatively static field of view, there are a number of applications where refocusing of the optics could be useful. In laparoscopy for example, the surgeon may want to quickly survey a large area in the abdomen in order to identify certain anatomical structures for navigational purposes which may appear at a large range of depths. In endoscopy of the upper gastrointestinal (GI) tract, there will be a demarcation in focus between the tissue closest to the tip of the endoscope, and that further along the lumen. Coupled with movement due to breathing and peristalsis, a dynamic scene is presented that requires constant refocusing. In the field of microsurgery the small field of view and relatively shallow depth of field of operating loupes place constraints on the surgeon's ability to see objects in the periphery. In each of the aforementioned examples, adjustment of the focal plane must be carried out using external controls such as a wheel or lever, or additional personnel (theatre camera operator) leading to potential for error and delays in decision-making. Finally, in cases where computer vision algorithms rely on high acuity visual features of the tissue, gaze-contingent autofocus can be vital when local processing around the fixation point is required.

We have implemented a dynamic refocusing system using a liquid lens system and the da Vinci surgical robot. This has been designed to test the usability of an eye-tracked autofocus system and a series of user trials were designed and performed. Besides the application on a stereo da Vinci robot, the system could also be adapted for other types of minimally invasive procedure and future directions are discussed in the conclusion.

II. MATERIALS AND METHODS

The gaze-contingent autofocus system was constructed

Manuscript received April 15, 2011. This work was supported by the UK EPSRC and Technology Strategy Board grants EP/E06342X/1 and DT/E011101/1.

N. T. Clancy is jointly with the Hamlyn Centre for Robotic Surgery, Institute of Global Health Innovation, and the Department of Surgery and Cancer, Imperial College London, SW7 2AZ, UK (phone: +44207-59-40865; e-mail n.clancy@imperial.ac.uk).

G. P. Mylonas is jointly with the Hamlyn Centre for Robotic Surgery, Institute of Global Health Innovation, and the Department of Computing, Imperial College London, SW7 2AZ, UK (e-mail: george.mylonas@imperial.ac.uk).

G. Z. Yang is jointly with the Hamlyn Centre for Robotic Surgery, Institute of Global Health Innovation, and the Department of Computing, Imperial College London, SW7 2AZ, UK (e-mail: g.z.yang@imperial.ac.uk).

D. S. Elson is jointly with the Hamlyn Centre for Robotic Surgery, Institute of Global Health Innovation, and the Department of Surgery and Cancer, Imperial College London, SW7 2AZ, UK (e-mail daniel.elson@imperial.ac.uk).

around a liquid lens (Arctic 416, Varioptic SA, France) and a commercially-available eye-tracking system (Tobii x50, Tobii Technology AB, Sweden). The liquid lens works on the principle of electrowetting, whereby the curvature of the interface between a drop of water and a drop of oil in a sealed capsule can be altered by changing the amplitude of an applied 1 kHz AC square wave voltage [8]. Varying this voltage between 0 and 60 V (RMS) adjusts the focal length of the lens from infinity to 6 cm. The voltage signal was provided using a manufacturer-supplied driver chip (DrivBoard LL3 I^2C), which was in turn computer-controlled using an interface that sent commands *via* a USB-connected microcontroller.

The lens was attached to the proximal end of a da Vinci Surgical System 30° stereo endoscope (Intuitive Surgical, Inc., USA) between the endoscope and the camera head using a rapid prototyped lens mount and adapter plate. The eye-tracker was mounted underneath a display that showed the output of one of the da Vinci's cameras, while this video feed was also directed to the computer through an S-Video to USB digitizer. The eye-tracker, which acquired gaze data at approximately 50 Hz, together with the computer (Dual core 2.40 GHz; 1.98 GB RAM; Microsoft Windows XP Professional) ensured real-time operation. A schematic of the system is shown in Fig. 1.

The da Vinci console's own focusing control is operated using a foot pedal that drives a motor, which in turn moves the lenses in the camera head along the optical axis. This pedal was positioned underneath the eye-tracker so that the test subject's gaze could also be monitored while using the pedal control.

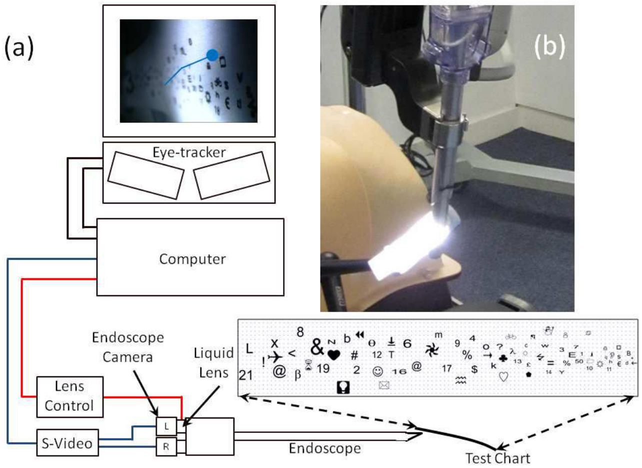

Fig. 1. (a) Schematic showing interface of the liquid lens, eye-tracker and endoscope camera. The monitor shows the test chart inclined to the endoscope tip with only the nearest symbols in focus and legible. (b) Photograph showing the system incorporated in the da Vinci surgical robot in a virtual operating theatre.

With an object at an incline with respect to the tip of the endoscope, the lens-control software was written so that two reference voltages could be programmed referring to the focal planes at the nearest and farthest points in the field of view. With the 'near' and 'far' points set, a linear transformation between the x-y plane of the display screen and the focal plane was implemented in the software.

To test the performance of the eye-tracking system against the da Vinci's in-built focusing mechanism, a search task was designed that required the user to find objects in different focal planes. This evaluates the total time, accuracy and search strategy employed as a way of inferring the usability and usefulness of the system. A total of 17 people with a non-clinical background were recruited to take part in the study, with no exclusion criteria applied. Approximately half of these had taken part in eye-tracking experiments previously, but had not had any experience with the focusing system described here. The search time was recorded by software that logged the timestamps of keystrokes marking the start and end of each search. Each test subject was asked to independently perform a search of six symbols from two different charts using the foot pedal control and the gaze-contingent system. To avoid bias, the order in which each focusing technique used was randomized.

The coordinates of the subject's fixations on the screen during the test were also recorded along with corresponding millisecond timestamps. This allowed detailed analysis of each subject's strategy in finding a particular symbol. In particular, it was possible to obtain qualitative data on whether or not subjects adopted different approaches when using each focusing technique and whether or not their visual behavior changed over the course of the experiment.

III. RESULTS

The total time taken to perform the search of the given symbols is shown in Fig. 2. In 12 out of 17 cases, the eye-tracking system enabled the test subject to perform a faster search.

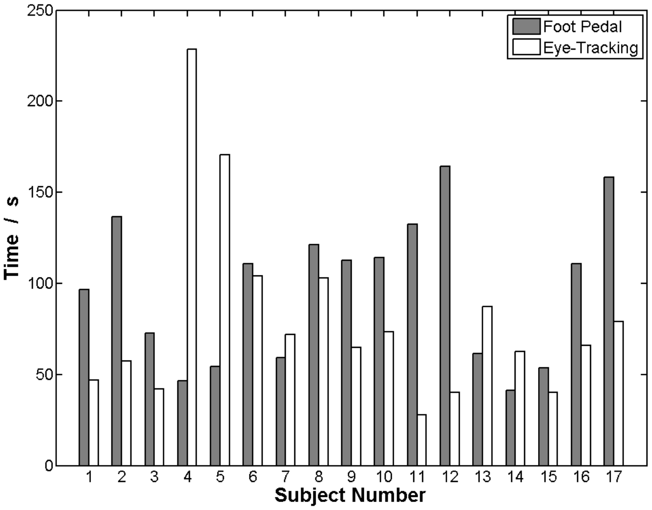

Fig. 2. Total time taken by each subject to perform search task.

Of the five results where the prototype system was slower, there were two outliers: Subjects 4 and 5 showed a significantly longer time taken when using the eye-tracker. For Subject 4, the 'near' and 'far' voltage settings for the lens were slightly inaccurate, meaning that the focal plane of the lens was not matched exactly to the user's fixation. This made the search considerably more challenging and time-consuming. Subject 5 wore glasses that interfered with the eye-tracker's ability to detect the eyes. For this reason, a reliable calibration could not be completed, fixation points showed large errors, and correct focusing was impossible. It should be noted that ten of the test subjects wore glasses but only one reported any difficulty. The results of Subjects 4 and 5 are omitted from the remaining analysis.

The results of Subjects 7 and 13 show a limitation in this initial study: the times are skewed by the search for two particular symbols that took much longer than the corresponding ones using the foot pedal control as the print quality made it difficult to determine if they were in fact the ones that were requested. It is worth noting that for the rest of that subject's results, where they were confident in identifying the symbol, the eye-tracking system was faster than the foot pedal.

When the two outliers are removed (Subjects 4 and 5) the gaze-contingent system is observed to be significantly faster than the foot pedal control ($p = 0.002$). When each chart is evaluated independently, the search times for the individual symbols may also be assessed. This data is presented as a box plot in Fig. 3.

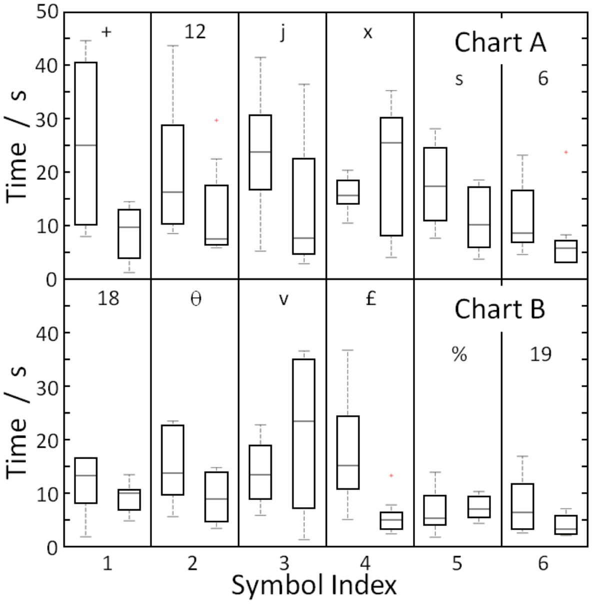

Fig. 3. Boxplots showing search times for each individual symbol on Chart A and B. The time taken using the foot pedal is shown on the left in each sub-panel, with the eye-tracking result on the right. The symbols being searched for are shown at the top of each panel. The top and bottom of each box mark the 25th and 75th percentiles, the central grey line the median and the error bars the extreme values. Outliers are defined as being 1.5 times the interquartile range and are marked by a red '+'.

The first noticeable feature of Fig. 3 is that the average search times for Chart B are lower overall than those of Chart A. Because Chart A was always the first one presented to the test subjects, it is speculated that an improvement in search strategy after completing the first test might be the reason. However, the times for the eye-tracking autofocus system are still broadly quicker than those of the foot pedal, with one exception in Chart A and two in Chart B.

Analysis of the motion path of the gaze gives clues as to the search strategy used by subjects and whether or not it changes based on the focusing method used. This is illustrated in Fig. 4, where Subject 1's search for a symbol in the lower right hand corner of the screen (near focus) is shown by plotting the fixation points recorded while using the foot pedal (Fig. 4 (a); Chart A) and the eye-tracking control (Fig. 4 (b); Chart B).

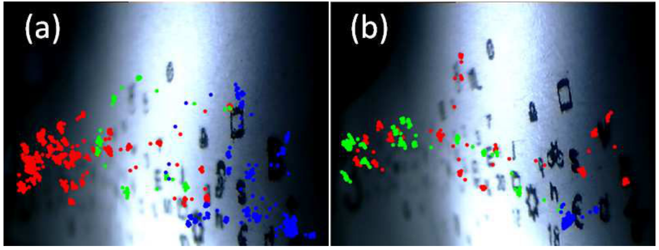

Fig. 4. Analysis of Subject 1's gaze behavior during a single search. The coordinates of each gaze point are color-coded by time in the order red, green and blue. The dense collection of blue dots indicates the symbol that the test subject found. (a) Pedal control. Red: 8404-12915 ms; Green: 12915-17425 ms; Blue: 17425-21936 ms. (b) Eye-tracking control. Red: 11107-19656 ms; Green: 19656-28204 ms; Blue: 28204-36753 ms.

In the case of the foot pedal control, the user employed a systematic search strategy. This involved setting the focus of the endoscope until a region of the chart became visible, searching that region, and then shifting the focus on to a neighboring region and repeating the process until the desired symbol was found. This is visible in Fig. 4 (a) as the test subject starts by focusing on and searching the dense arrangement of symbols at the left-hand side (red dots), before adjusting the foot pedal and searching the central part of the chart. Then, with one final adjustment of the pedal, the section of the chart nearest the endoscope tip is brought into focus and the symbol found.

Fig. 4 (b) shows the somewhat contrasting behavior of the eye-tracking autofocus system as the test subject scans their eyes rapidly across the field of view. Discreet sections of the chart are not examined systematically as before but instead, the eyes dart from one side of the screen to the other, sometimes revisiting a previously examined area. This is illustrated by the mixing of the colored dots in the gaze plot.

IV. DISCUSSION AND CONCLUSIONS

The gaze-contingent autofocus system described in this paper has been constructed with the intention of applying it in the clinic during robotic-assisted surgery. Due to the fast response time of the liquid lenses used (~ 30 ms) and the responsiveness of the eye-tracker (~ 50 gaze coordinates captured per second), the experimental system proved to be significantly faster than the da Vinci's in-built focus mechanism when tested by users in a search task. The speed of the system allowed for real time operation and users commented that the focusing ability was comfortable and allowed them to search the pattern in a natural way. This in turn allowed greater freedom in choice of search strategy allowing, for example, ‘random-access’ searching rather than just a sequential scan.

It should be noted at this point that it would be possible to replicate the results from this study using a fast mechanically-actuated lens [9] and an autofocus system using an algorithm commonly used in commercial digital cameras. However, the experimental gaze-contingent system described here retains a number of key advantages. In comparison to their mechanical counterparts, which need a separate piezo or micromotor attachment, the simplicity of the liquid lenses and the fact that they have no moving parts make them much easier to integrate into existing optical systems. Their small footprint (total outer diameter of 7.75 mm) makes them suitable to mount on the distal tip of a GI endoscope, for example. Existing digital camera autofocus algorithms use iterative optimization steps to attain some predefined measure of sharpness or contrast in the image. In the dynamic surgical environment however, homogenous or specularly-reflecting surfaces could cause errors. Furthermore at points where the algorithm’s region overlaps tissue at different depths, the focus may jump between one plane and another. The eye-tracking system offers the potential to estimate the absolute depth of point of interest to the surgeon. This would eliminate the computational overhead of an iterative focusing algorithm and reliably focus on only the region of direct interest to the surgeon.

Although the set-up of the experiments described in this paper concentrate on endoscopic surgical applications, there is also the potential to apply this technology to open surgery. In microsurgical tasks such as vessel grafting, the use of loupes is necessary to magnify the field of view. A head-mounted gaze contingent system could enable fast adjustments in focus to give the surgeon the ability to view objects in the periphery (tools, monitoring equipment) as well as the magnified tissue site, which lie in different focal planes.

The method used to adjust the focus of the endoscope had an influence on the test subject’s approach to the task. Analysis of the gaze data shows that when using the foot pedal, subjects tended to apply a systematic approach that is compartmentalized by the focused area set by the foot pedal. Using the gaze-contingent system however, the approach is somewhat different. Without the distraction of an additional external control the test subject performs rapid scans in the horizontal direction, moving quickly between the nearest and farthest points. This is more similar to the ‘feature ring’ approach used when performing search tasks in scenes by eye [2]. We believe that this type of behavior is an indication of the intuitiveness of the system. Such is the ease of use and the speed at which objects at varying depths can be interrogated, the system actually encouraged test subjects to be more adventurous in their searches rather than following a prescribed systematic approach (such as that imposed by the position of the foot pedal). The gaze-tracking data shows that on a number of occasions when a test subject was asked to find a particular symbol, they would locate it after a number of seconds, fixate on it, then search again to make sure that there were no others of a similar appearance before returning to and reporting the original discovery.

In the clinical environment the intuitive and natural scene viewing that will provide the best advantage for this system, besides the important speed advantage. During surgery, and minimally-invasive surgery in particular, it is often necessary to perform a visual search of an area of tissue to guide decision-making. For example, in a robotic-assisted prostatectomy, when the endoscope is inserted into the abdomen, the surgeon needs to quickly survey the entire cavity to identify anatomical landmarks for navigation before advancing the endoscope closer to the target tissue. Also, in microsurgery while focusing on a small region of interest, it may be necessary to monitor tissue at the periphery of the field of view. These are adjustments that must be done at a glance, without adding an extra layer of complexity in the form of an external focus control. The gaze-contingent system addresses both of these concerns through use of liquid lenses for speed, and eye-tracking technology for an intuitive interface directly linked to the user’s fixation point i.e., their focus of attention.

Future work will concentrate on adapting the system to be fully automatic. Instead of using pre-calibrated ‘near’ and ‘far’ points, the eye vergence estimation of depth will be used to calculate the object distance and hence, the lens focal length required.

REFERENCES

[1] T.E. Hutchinson. United States Patent no. 4,950,469, 1990.

[2] G.-Z. Yang, L. Dempere-Marco, X.-P. Hu, and A. Rowe, Image Vision Comput., vol. 20, (no. 4), pp. 291-305, 2002.

[3] G.P. Mylonas, A. Darzi, and G.-Z. Yang, Comput. Aided Surg., vol. 11, (no. 5), pp. 256-266, 2006.

[4] G.P. Mylonas, D. Stoyanov, F. Deligianni, A. Darzi, and G.-Z. Yang, Lect. Notes Comput. Sc., vol. 3749, pp. 843-850, 2005.

[5] G.P. Mylonas, D. Stoyanov, A. Darzi, and G.-Z. Yang, Lect. Notes Comput. Sc., vol. 4792, pp. 660-667, 2007.

[6] R. Aggarwal, K. Moorthy, and A. Darzi, Br. J. Surg., vol. 91, (no. 12), pp. 1549-1558, 2004.

[7] G.P. Mylonas, K.-W. Kwok, D.R.C. James, D. Leff, F. Orihuela-Espina, A. Darzi, and G.-Z. Yang, Med. Image Anal., In Press, Corrected Proof, 2010.

[8] B. Berge and J. Peseux, Eur. Phys. J. E, vol. 3, (no. 2), pp. 159-163, 2000.

[9] C. Park, S. Cha, Y. Lee, O. Kwon, D. Park, K. Kwon, and J. Lee, *IEEE Int. Symp. Circ. S.*, pp. 1284-1287, 2010.