Assessment of Perceptual Quality for Gaze-Contingent Motion Stabilization in Robotic Assisted Minimally Invasive Surgery

George P. Mylonas, Danail Stoyanov, Ara Darzi, and Guang-Zhong Yang

Institute of Biomedical Engineering and Department of Computing
Imperial College London, London, United Kingdom
{george.mylonas, danail.stoyanov, a.darzi,
g.z.yang}@imperial.ac.uk

Abstract. With the increasing sophistication of surgical robots, the use of motion stabilisation for enhancing the performance of micro-surgical tasks is an actively pursued research topic. The use of mechanical stabilisation devices has certain advantages, in terms of both simplicity and consistency. The technique, however, can complicate the existing surgical workflow and interfere with an already crowded MIS operated cavity. With the advent of reliable vision-based real-time and *in situ in vivo* techniques on 3D-deformation recovery, current effort is being directed towards the use of optical based techniques for achieving adaptive motion stabilisation. The purpose of this paper is to assess the effect of virtual stabilization on foveal/parafoveal vision during robotic assisted MIS. Detailed psychovisual experiments have been performed. Results show that stabilisation of the whole visual field is not necessary and it is sufficient to perform accurate motion tracking and deformation compensation within a relatively small area that is directly under foveal vision. The results have also confirmed that under the current motion stabilisation regime, the deformation of the periphery does not affect the visual acuity and there is no indication of the deformation velocity of the periphery affecting foveal sensitivity. These findings are expected to have a direct implication on the future design of visual stabilisation methods for robotic assisted MIS.

1 Introduction

One of the main challenges of robotic assisted Minimally Invasive Surgery (MIS) is to perform Totally Endoscopic Coronary Artery Bypass (TECAB) grafting on a beating heart. The complexity of such a delicate task is complicated by the destabilization of the heart due to cardiac and respiratory motion, compounded by a high degree of visual magnification through the use of immersive optics. This significantly affects precise hand-eye coordination and tissue-instrument interaction. In current surgical practices, it is common to use epicardial mechanical stabilizers to dampen the cardiac motion. However, the residual deformation may still be large enough to hinder tasks such as small vessel anastomosis [1][2]. To overcome this problem, a number of techniques have been proposed for introducing virtual

stabilisation to the surgical scene based on soft-tissue deformation tracking and image warping [3][4][5][6]. The application of these techniques to real-time *in situ, in vivo* settings, however, still remains a challenge due to the morphological complexity of the tissue. Furthermore, no detailed psychovisual experiments have been performed to assess the effect of virtual stabilization on foveal/parafoveal vision and general visual acuity during MIS.

The purpose of this paper is to evaluate the hypothesis of whether foveal motion stabilisation is sufficient for robotic assisted MIS and to examine the effect of peripheral motion on visual acuity. Detailed psychovisual experiments have shown that stabilisation of the whole visual field is not necessary for MIS, and it is sufficient to perform accurate motion tracking and deformation compensation within a relatively small area that is directly under foveal vision. Simple rigid body motion of the camera can therefore be used to provide a perceptually stable operating field-of-view. This also avoids the use of large area 3D tissue deformation recovery, which tends to be error prone and limited by the paucity of reliable anatomical landmarks. The finding is expected to underpin the synergistic use of computer vision based feature tracking and deformation recovery combined with real-time gaze tracking for robotic assisted MIS. Given the complexity of robotic control in surgical environments, this is expected to facilitate the effective hand-eye coordination for improved surgical performance. This research further extends the current work on real-time eye tracking and saccadic eye movement analysis for investigating gaze contingent approaches for robotic control in surgery [7][8].

2   Methods

2.1   Visual Acuity with Foveal and Parafoveal Stabilization

In order to simulate gaze-contingent motion stabilization, it is necessary to obtain a robust motion map around the fixation point. To this end, feature-based tracking using stereo laparoscopic images recorded from a daVinci robot were used in this study. Gradient-based landmarks using the Shi and Tomasi operator [11] were implemented for performing feature tracking using a variant of the Lucas-Kanade tracker. The method minimises the squared residual error ε between feature location $\mathbf{x}$ with feature template T_n for the left and right images I_n [4]

$$\varepsilon = \sum_{n=1}^{2} \sum_{\mathbf{x}} [I_n(\mathbf{W}_n(\mathbf{x};\mathbf{p})) - T_n(\mathbf{x})]^2 \tag{1}$$

where $\mathbf{p} = \begin{bmatrix} X & Y & Z \end{bmatrix}^{\mathsf{T}}$ and $\mathbf{W}_n$ is the warp function. The camera was calibrated with one of the standard calibration methods available [13]. The 3D coordinates of the points recovered from the calibrated camera matrices $\mathbf{P}_n = \mathbf{K}_n[\mathbf{R}_n \mid \mathbf{t}_n]$ describe the internal and external optics of the cameras. They were used directly to parameterize the translational warping function. To obtain a dense motion map, it is possible to combine feature detectors as shown in [4] and this approach can also be used to identify landmarks that are less prone to view dependent reflections. For performance considerations, however, only image derivative-based operators were

used and a threshold filter was applied to the correlation measurement during the tracking process, combined with the epipolar constraint. The features tracked with the above technique were used as the control points in the subsequent Radial Basis Function (RBF) texture warping. To stabilize the deforming tissue, the video texture was first mapped to a grid of regularly spaced vertices and the grid vertices were interpolated and warped according to RBF so that the target tissue area appeared static.

In order to assess the visual acuity with and without foveal and parafoveal stabilization, a pre-recorded video clip of a TECAB surgery with the daVinci surgical robot was used. A total of 13 fixation points (each lasted for 4s) were created to direct the observers' foveal vision. Motion compensation was then applied to each of the fixation points in sequence for cancelling out tissue deformation. In order to maintain strict visual fidelity of the foveal area, only rigid body motion cancellation was applied. This avoided the use of any image warping that could inadvertently introduce visual artefacts. Since the foveal area only corresponds to an area of 2° visual angle [9], this way of linear motion cancellation is generally sufficient for most *in vivo* applications. However, the parafoveal area that surrounds the fovea is usually large and non-linear motion compensation must be applied in order to stabilise the tissue motion in this region. To this end, sparse features were first tracked by using the aforementioned stereo tracking algorithm. They were subsequently used as the control points for RBF-based warping. The outer region surrounding the parafoveal area corresponds to the visual periphery. Controlled motion compensation, in terms of tracked sparse features, was not performed within this area. Only a global free warping was applied as the result of rigid transformation of the parafoveal boundary control points with respect to the stationary image boundary. It should be noted that the boundary points defining the fovea and parafovea were also considered as warping control points. In this way, smooth transitions were achieved from the area of translational motion cancellation (fovea), to the area of dampened motion with controlled warping (parafovea), to free transformation (far periphery).

For experimental validation, 13 subjects were asked to foveate to the suggested fixation points on the screen with and without the above motion stabilisation scheme being introduced. The videos of the stabilized and the non-stabilized tissue were played back-to-back in random succession. At random intervals and for a period of 8 video frames (320ms), a total of 16 epicardial stimuli were randomly introduced. For the stabilized tissue, the stimulus remained within the foveal region for the entire 320ms period. For the non-stabilized tissue, the stimulus may or may not remain within the fovea region depending on the amplitude and the velocity of the tissue deformation when the stimulus was first introduced. The observers were instructed to signal (press the spacebar) if they spotted a stimulus appearing in or close to the foveal window. Fig. 1 schematically defines the foveal, parafoveal, and peripheral regions used for motion cancellation and image warping. It also illustrates the visual stimuli introduced in the foveal window delineated in dotted circles that correspond to the foveal window.

To analyse the underlying visual behaviour of the subjects and to ensure that their actual fixation points corresponded to those suggested, gaze tracking was performed using a Tobii ET-1750 eye-tracker. This is an infrared video-based binocular eye-tracking system recording the position of gaze in the work plane (screen) at up to

50Hz, with an accuracy of 1° visual angle [10]. Fixations from the raw eye gaze data were identified and keyboard events were obtained and time-stamped using the Clearview software (Tobii Technology, Sweden). The minimum fixation duration filter was set to 100ms. Receiving Operating Characteristic (ROC) analysis was used to analyse the effect of the tissue stabilization in the fovea and the performance of the subjects towards identifying the introduced stimuli.
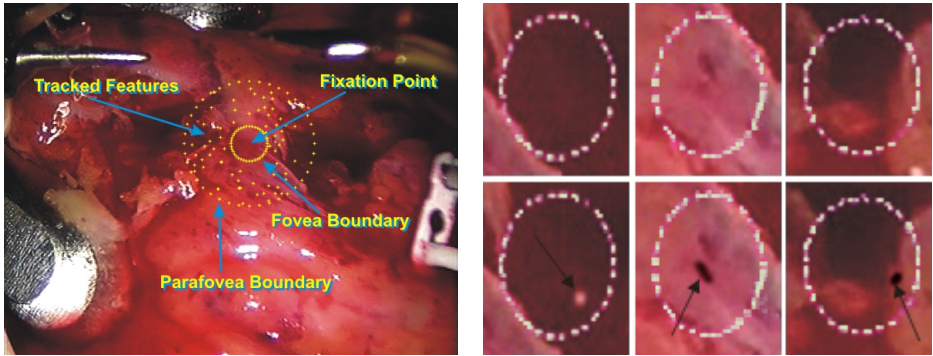

Fig. 1. (left) Illustration of the foveal, parafoveal, and peripheral regions used for translational, controlled warping and free transformation applied to the proposed motion cancellation scheme. (right) Example fovea regions with (bottom) and without (top) stimuli being introduced.

2.2 The Effect of Peripheral Motion on Visual Sensitivity

In the previous experiment, the peripheral vision was not stabilised. In order to assess the effect of peripheral motion on visual sensitivity, a further experiment was conducted with controlled velocity of peripheral motion. Images from a robotic assisted TECAB procedure were warped to simulate different amount of respiratory and cardiac induced tissue deformation. For simplicity, the respiratory component was assumed to only cause a rigid translation. The cardiac component was represented by a Gaussian mixture model. The two deformations were combined linearly with different weighing factors. The equations for the deformation applied to a vertex at time T are:

$$\mathbf{V}_{t+1} = \mathbf{V}_0 + \sin(\phi) \begin{bmatrix} W_x^s e^{-W_x^d(V_x - C_x)} \\ W_y^s e^{-W_y^d(V_y - C_y)} \\ W_z^s e^{-W_z^d(V_z - C_z)} \end{bmatrix} + \sin(\varphi) \begin{bmatrix} W_x^r T_x \\ W_y^r T_y \\ W_z^r T_z \end{bmatrix} \qquad (2)$$

in which the first sinusoidal product represents the cardiac deformation and the second one a time-varying translational motion. In the above equation, V_o represents the original vertex of the surface and C is the centre of deformation. The sinusoidal terms specify the frequency of oscillation in such a way that a full sinusoidal cycle is completed in N steps. A looped-over video from the sequence of N images was

created and used for assessing the effect of peripheral motion on visual acuity. The parameters used include:

$$C_x = 400, C_y = 300, W_x^s = W_y^s = 0$$
$$W_z^s = 2.0\,, W_z^d = -0.07$$
$$W_x^r = W_y^r = W_z^r = 0\,, N = 25$$

The above parameters are designed to eliminate the influence of breathing which is reasonable due to its relatively slow evolution. The resulting warped video effectively stabilises the central area with increasing deformation amplitude towards the periphery. Similar to the previous experiment, 30 suggested fixation points were prescribed and each fixation lasted for 4s. The fixations were presented in random order in such a way that 10 were displayed in the apparently stationary central area of the video and the remaining 20 in the deforming periphery.
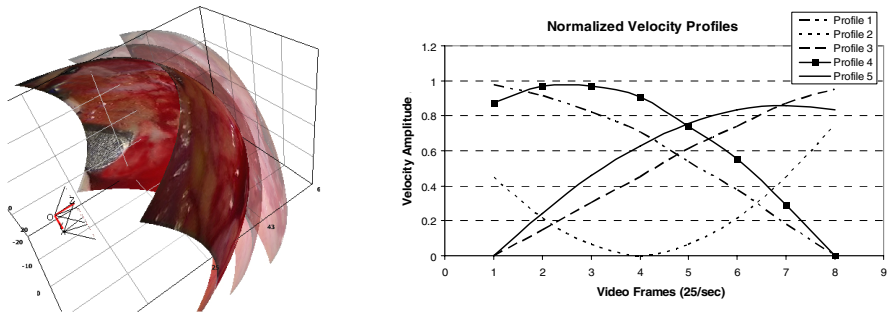

Fig. 2. (left) The prescribed peripheral motion for assessing the effect of different peripheral motion velocity on visual acuity, where three instants over a half sinusoidal cycle are depicted (~12 video frames at 25fps). (right) Five peripheral velocity profiles over which the central fovea stimuli were introduced.

Ten subjects were involved in this study and asked to foveate to each of the suggested fixation points. All subjects were eye-tracked as in the previous experiment. At random intervals a total of 20 visual stimuli were introduced inside some of the 30 parafoveal areas, with one stimulus per simulated fixation. The study was designed in such a way that the central 10 parafoveal areas are presented 5 times with a stimulus and 5 times without. Also, care was taken so that the stimuli appear under different peripheral deformation velocities. A total of five velocity profiles were used as shown in Fig. 2. Considering the apparent size of the video frame and the viewing distance maintained, the amplitude of deformation of a feature point in the far periphery over a full sinusoidal cycle was approximately 3.8° of visual angle.

Similarly to the previous study, the subjects were asked to press the spacebar on a computer keyboard when they saw a stimulus. In this study, only the data concerned with the stabilized central area are analyzed. The simulated fixations introduced in the deforming periphery are merely used as distracters. The effect of the deforming periphery in the stabilized fovea and parafovea was analysed for assessing the performance of the subjects towards identifying the introduced stimuli.

3   Results

For the experiment on foveal and parafoveal stabilisation, Fig. 3 depicts the specificity against the sensitivity data for the subjects observing the stabilized and the non-stabilized tissues. It is evident that motion stabilisation improves the foveal sensitivity significantly, which confirms that stabilization facilitates the identification of visual stimuli. On the other hand, it is also evident from the figure that stabilization didn't improve the specificity greatly, which was already relatively high for most subjects. This indicates that the subjects were generally competent in confirming the absence of a stimulus. Data from Subject 12 was discarded as he failed to adhere to the experiment protocol. Fig. 3 also shows a comparison of each subject's number of fixations for the stabilized and non-stabilized experiments. The total number of fixations for all the subjects studied is summarised in Table I, in which the average duration and standard deviation are also provided. It is clear that performing the same task on the non-stabilized tissue requires a considerable amount of short fixations. Data from Subject 10 was discarded because eyetracking was not perfectly stable.
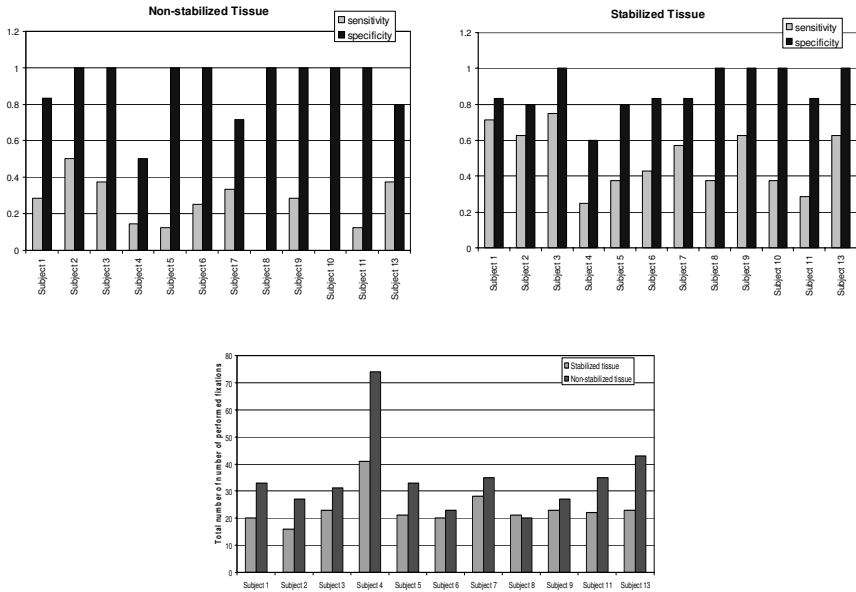

Fig. 3. (top) Sensitivity vs specificity of the subjects in identifying the presence/absence of the visual stimuli under non-stabilized (left) and stabilized (right) views. (bottom) The number of fixations for the stabilized and non-stabilized tissue experiments.

Similar analysis was performed for the second experiment concerning the sensitivity in the fovea and the parafovea as a function of the peripheral deformation velocity. The results are shown in Table II, which clearly demonstrates that all the subjects were consistent in identifying the introduced stimuli irrespective of the deformation velocity of the periphery.

Table I. Fixations statistics for all subjects over the stabilized and non-stabilized tissues

	Stabilized tissue	Non-stabilized tissue
Number of fixations	258	381
Average duration (ms)	2240.99	1593.66
Standard Deviation	476.37	511.49

Table II. The score of all subjects for the visual sensitivity study on a stabilized fovea and parafovea during peripheral warping

Subject	True Positive	False Positive	True Negative	False Negative	Sensitivity	Specificity
1	5	0	5	0	1	1
2	5	0	5	0	1	1
3	5	1	4	0	1	0.8
4	4	0	5	1	0.8	1
5	5	0	5	0	1	1
6	4	0	5	1	0.8	1
7	5	0	5	0	1	1
8	5	0	5	0	1	1
9	4	0	5	1	0.8	1
10	4	0	5	1	0.8	1
Totals	**46**	**1**	**49**	**4**	**0.92**	**0.98**

4 Discussion and Conclusions

In this study, we have assessed the effect of motion stabilisation in robotic assisted MIS. The results have shown that gaze-contingent soft-tissue stabilization can significantly increase the visual acuity and the method is relatively immune to peripheral motion. It is very interesting to note the reduced number of required fixations and their duration increase by almost 30% when operating on a virtually stabilized tissue. Research in psychophysiology has shown that an increase in the fixation duration is more efficient for performance improvement rather then darting around the visual field [12]. This is particularly relevant for identifying subtle, transient events. The findings have obvious consequences on performing micro-surgical tasks such as small vessel anastomosis. This study has also confirmed that under the current motion stabilisation regime, the deformation of the periphery does not affect the visual sensitivity of the stabilized foveal and parafoveal regions. Furthermore, there is no indication of the velocity of the deformation affecting the foveal sensitivity. These results should have a direct implication on the future design of effective visual stabilisation methods in robotic assisted MIS.

As a final point, we should mention that the above study does not consider the presence of surgical tools in the visual field. Under the virtual stabilization framework this would require accurate 3D instrument segmentation and AR rendering which is outside the scope of a study on perceptual quality like the one presented here.

References

1. Wimmer-Greinecker, G., Deschka, H., Aybek, T., Mierdl, S., Moritz, A., Dogan, S.: Current status of robotically assisted coronary revascularization. Am. J. Surg. 188(Suppl.4A), 76S–82S (2004)
2. Purkayastha, S., Athanasiou, T., Casula, R., Darzi, A.: Robotic surgery: a review. Hospital Medical 65(3), 153–159 (2004)
3. Cuvillon, L., Gangloff, J., de Mathelin, M., Forgione, A.: Toward Robotized Beating Heart TECABG: Assessment of the Heart Dynamics Using High-Speed Vision. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3749, pp. 551–558. Springer, Heidelberg (2005)
4. Stoyanov, D., Mylonas, G.P., Deligianni, F., Darzi, A., Yang, G.Z.: Soft-Tissue Motion Tracking in Robotic MIS Procedures. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3749, pp. 26–29. Springer, Heidelberg (2005)
5. Stoyanov, D., Darzi, A., Yang, G.-Z.: A practical approach towards accurate dense 3D depth recovery for robotic laparoscopic surgery. Comp. Aided Sur. 10(4), 199–208 (2005)
6. Stoyanov, D., Darzi, A., Yang, G.-Z.: Dense Depth Recovery for Robotic Assisted Laparoscopic Surgery. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) MICCAI 2004. LNCS, vol. 3216, pp. 41–48. Springer, Heidelberg (2004)
7. Mylonas, G.P., Stoyanov, D., Deligianni, F., Darzi, A., Yang, G.-Z.: Gaze-Contingent Soft Tissue Deformation Tracking for Minimally Invasive Robotic SUrgery. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3749, pp. 843–850. Springer, Heidelberg (2005)
8. Mylonas, G.P., Darzi, A., Yang, G.-Z.: Gaze Contingent Depth Recovery and Motion Stabilisation for Minimally Invasive Robotic Surgery. In: Yang, G.-Z., Jiang, T. (eds.) MIAR 2004. LNCS, vol. 3150, pp. 311–319. Springer, Heidelberg (2004)
9. 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)
10. Tobii technology. User Manual (2003), http://www.tobii.se
11. Shi, J., Tomasi, C.: Good features to track. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 593–600 (1994)
12. Togami, H.: Effects on visual search performance of individual differences in fixation time and number of fixations. Ergonomics 27, 789–799 (1984)
13. Tsai, R.: A versatile camera calibration technique for high accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses. IEEE J. Robotics Automation 3(4), 244–323 (1987)