# Gaze-Contingent Soft Tissue Deformation Tracking for Minimally Invasive Robotic Surgery

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

Royal Society/Wolfson Medical Image Computing Laboratory,
Imperial College London, London, United Kingdom
{george.mylonas, danail.stoyanov, fani.deligianni, a.darzi,
g.z.yang}@imperial.ac.uk

**Abstract.** The introduction of surgical robots in Minimally Invasive Surgery (MIS) has allowed enhanced manual dexterity through the use of microprocessor controlled mechanical wrists. Although fully autonomous robots are attractive, both ethical and legal barriers can prohibit their practical use in surgery. The purpose of this paper is to demonstrate that it is possible to use real-time binocular eye tracking for empowering robots with human vision by using knowledge acquired *in situ*. By utilizing the close relationship between the horizontal disparity and the depth perception varying with the viewing distance, it is possible to use ocular vergence for recovering 3D motion and deformation of the soft tissue during MIS procedures. Both phantom and in vivo experiments were carried out to assess the potential frequency limit of the system and its intrinsic depth recovery accuracy. The potential applications of the technique include motion stabilization and intra-operative planning in the presence of large tissue deformation.

## 1 Introduction

The field of Minimally Invasive Robotic Surgery (MIRS) is increasingly attracting considerable attention from the computer-assisted surgical research community. In robotic surgery, dexterity is enhanced by microprocessor controlled mechanical wrists, which allow motion scaling for reducing gross hand movements and the performance of micro-scale tasks that are otherwise not possible. The continuing evolution of the technology, including force feedback and virtual immobilization through real-time motion adaptation, will permit more complex procedures such as beating heart surgery to be carried out under a static frame-of-reference. The use of robotically assisted minimally invasive surgery provides an ideal environment for integrating patient specific pre-operative data for performing image guided surgery and active constraint control, which can all be conducted without the need of the surgeon to remove his/her eyes from the operating field of view. Under the dichotomy of autonomous and manipulator technologies in robotics, intelligence of the robot is typically pre-acquired through high-level abstraction and environment modelling. For systems that require robotic vision, this is known to create major difficulties. Additionally, the ethical and legal barriers imposed on interventional surgical robots

give rise to the need of a tightly integrated perceptual docking between the operator and the robot.

The purpose of this paper is to demonstrate that eye gaze derived from binocular eye-tracking can be effectively used to recover 3D motion and deformation of the soft tissue during MIRS. Compared to the use of other input channels, eye gaze is the only input modality that implicitly carries information about the focus of the user's attention at a specific point in time. This allows seamless *in vivo* registration of the motion and deformation fields within the area that is directly under foveation. This research extends our existing work in eyetracking and saccadic eye movement analysis, for gaze-contingent compensation of complex motion of the soft tissue.

## 2   Methods

### 2.1   Ocular Vergence Measurement for Gaze-Contingent 3D Deformation Recovery

One of the strongest depth cues available to human is the horizontal disparity that exists between the two retinal images. There is a close relationship between the horizontal disparity and the depth perception, varying with the viewing distance. More specifically, as the fixation point moves away from the observer, the horizontal disparity between the two retinal images is diminished and vice-versa. In order to extract quantitative information regarding the depth of the fixation point, ocular vergence needs to be measured, thus providing a veridical interpretation of stereoscopic depth [1]. One technique of achieving that is video-oculography. This is a non-intrusive, video based, approach used to measure the corneal reflection from a fixed IR light source in relation to the centre of the pupil [2]. The two centres define a vector, which can be mapped to a unique eye gaze direction. The combined tracking of both eyes provides the binocular vergence measure, which in turn determines the fixation point. In order to establish the relationship between pupil-glint vectors and points in 3D space, calibration is required prior to any eye tracking session. Different methods of binocular eye-tracking calibration exist. In [3] and [4], two of these calibration methods are described. Our eye-tracking calibration method however, takes also into consideration the intrinsic and extrinsic robot camera characteristics, including their relative pose.

### 2.2   Experimental Design and System Setup

In order to assess the basic characteristics of using ocular vergence for 3D depth recovery, a replication of the DaVinci (Intuitive Inc, CA) robotic surgical environment was created. The system consists of a stereoscopic console and an industrial robot equipped with a stereo-camera rig. The stereo console allows the user to look at 3D video captured by the two cameras held by the robot. Inside the stereoscope, a video based binocular eye tracking system was installed [4]. With binocular eyetracking, it is possible to determine the 3D position of the fixation point of the user while observing the stereo images. The robot used in this experiment is a Stäubli RX60 robotic arm with six degrees of freedom (DOF) and a repeatability accuracy of ±0.02mm at high speed and acceleration. For accurate positioning of the

camera tips, a Polaris (Northern Digital Inc, Ontario) 6-DOF tracker was used. The Polaris tracker is able to simultaneously track a number of passive, active, wired and wireless IR tools in real time. Data interfacing is achieved through RS-232/RS-422 and the provided tracking accuracy is 0.35 mm RMS at a sampling rate of 60Hz.
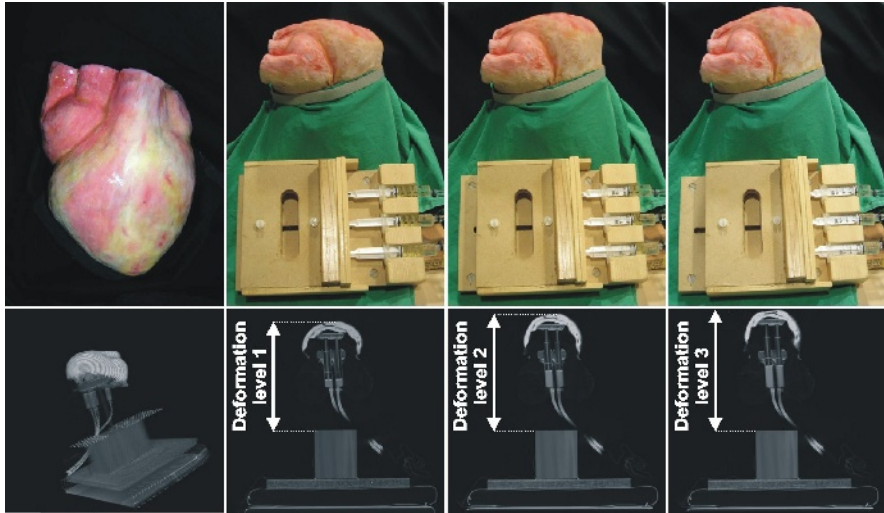


**Fig. 1.** *Top:* The phantom heart at different deformation levels. The oil filled pistons (only 3 shown here) allow for reproducible deformation control through numbered injection levels. *Bottom:* The reconstructed phantom heart from a series of CT slices. On the right, CT slices at exactly the same position of the phantom and at three deformation levels.
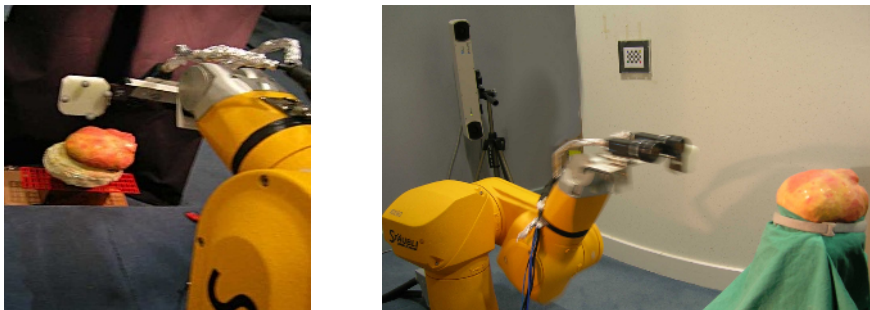


**Fig. 2.** On the left, the robot with the mounted optical-tracker retroreflectors and the stereo-camera rig. On the right, the polaris optical tracker is visible on the background.

For assessing the accuracy of the proposed 3D depth recovery framework through ocular vergence, a phantom heart model was created by using thixotropic silicone mould rubber and pre-vulcanised natural rubber latex with rubber mask grease paint to achieve a specular appearance and high visual fidelity. The phantom is deformable by means of a system of 4 oil filled pistons with controllable injection levels. In this

way, the amount of deformation can be accurately controlled and reproduced at any time as this can be seen in Fig. 1. The same figure also depicts the reconstructed phantom heart from a series of CT slices at different deformation levels. In Fig. 2, the entire setup can be seen. For determining the exact position of the epicardial surface, an Aurora (Northern Digital Inc, Ontario) 5-DOF electromagnetic catheter-tip tracker was used. This device has an accuracy of 0.9-1.3mm RMS, depending on the distance of the tool from the magnetic field generator. The maximum sampling rate is 45Hz for up to 5 tracked tools, which drops down to 22Hz when more than 6 tools are tracked.

## 2.3  Autoregressive Moving Average Modelling (ARMA) for Visual System Identification subjected to Oscillatory Stimulus

In order to study the oculomotor response, it is necessary to obtain a model that can closely describe the system. To this end, the first part of our experiment is to record a set of data for deriving the input response of the eye-tracking system. By collecting the stimulus and the response data sets, the Steiglitz-McBride method was used [5] to derive the rational transfer function of the proposed ocular vergence 3D depth recovery system. This method is commonly referred to as ARMA modelling, which attempts to find the coefficients of the rational transfer function that approximates a given time-domain linear system. This method is based on identifying the unknown system from given input and output sequences that describe the system's behaviour. The method solves for the numerator and denominator coefficients simultaneously in an attempt to minimize the signal error between the output and the given output signal. To include the effect of exogenous inputs to the system in a time series model, the basic ARMA model can be further extended to a number of more elaborate variants, like ARMAX (Autoregressive Moving Average with Exogenous Input). A general input-output linear model for a single output system with input $u$ and output $y$ can be written [7]

$$A(q)y(t) = \sum_{i=1}^{nu} \left[ B_i(q)/F_i(q) \right] u_i(t - nk_i) + \left[ C(q)/D(q) \right] e(t) \tag{1}$$

In the above equation, $u_i$ is the input $i$ and $A$, $B_i$, $C$, $D$ and $F_i$ are polynomials in the shift operator (z or q). The general model is defined by setting the time delays $nk$ and the orders of the above polynomials. According to the variance in use, the above equation can be reduced down to a special case, *e.g.*, for ARMAX Eq. (1) can be confined to

$$A(q)y(t) = B(q)u(t - nk) + C(q)e(t) \tag{2}$$

## 2.4  Gaze-Contingent Decoupling of Cardiac and Respiratory Motion

One of the basic problems in intraoperative registration between 3D tomographic data and cardiac tissues is the large deformations that can occur during the MIS procedures. If deformation is significant, convergence of any iterative registration algorithm is not guaranteed. One of the largest components of cardiac deformation is the respiratory motion, which is typically coupled with the higher frequency cardiac motion. We show in this paper that gaze-contingent motion recovery of the composite

cardiac motion is practical and *in vivo* decoupling of the superimposed components is possible. Given that the dominant variation of the gaze recovered signal is due to cardiac motion and it has a constant direction during a given time window, then Wavelet and Principal Component Analysis can be employed to decouple the motion components [6]. The principal vector is expected to lie on the same plane over time, thus projecting the signal on this plane can isolate any variations due to respiration.

## 3   Results

### 3.1   Binocular System Frequency Response

To assess the binocular system frequency response, experiments were carried out involving six subjects. The first set of experiments investigates the oscillatory response of the binocular visual system over a frequency range. The subjects involved, were asked to keep fixating on a feature of their preference on the surface of the heart model. While eyetracking is performed, the robot is set to oscillation of gradually increasing frequency in the z-axis (depth). While the 3D fixation point of a subject is tracked, the position of the robot cameras is also recorded by the Polaris optical tracker. After data collection, ARMAX modelling is used to derive the coefficients of the parametric system that describes the transfer function of the system. Table 1 summarises the response of the visual system in oscillation on the z-axis, which indicates that the response of the system is good up to frequencies of about 1.8Hz. Beyond this limit, there is considerable attenuation and noticeable phase shift.

**Table 1.** Error analysis comparing the oculomotor responce of the 6 subjects over a range of frequencies

| Frequency | Amplitude | | Phase (rad) | |
|---|---|---|---|---|
| (rad/s) | Mean | std | Mean | std |
| 0.1 | 0.950 | 0.118 | -0.031 | 0.016 |
| 1.0 | 0.918 | 0.132 | -0.307 | 0.120 |
| 3.0 | 0.822 | 0.204 | -0.871 | 0.098 |
| 6.0 | 0.650 | 0.156 | -1.494 | 0.150 |
| 10.0 | 0.577 | 0.125 | -2.394 | 0.278 |
| 12.0 | 0.520 | 0.126 | -2.955 | 0.316 |

### 3.2   Gaze-Contingent Depth Extraction and Tissue Deformation Tracking

The second part of our experiment was to extract gaze-contingent depth information from the phantom. For this experiment, an Aurora catheter tip was positioned on the epicardial surface of the phantom heart. Though slightly less accurate than the Polaris system, the Aurora tools were used because of their small size. These tools can be easily embedded over or under the epicardial surface and not influence the fixation point position as this would be the case with the much larger Polaris tools. The subjects were asked to keep fixating on a surface feature close to the position sensor. While the phantom was subjected to different levels of deformation, both the fixation

point and the position of the electromagnetic sensor were tracked. In Fig. 3, the gaze-contingent recovered deformation is compared with the actual deformation levels.
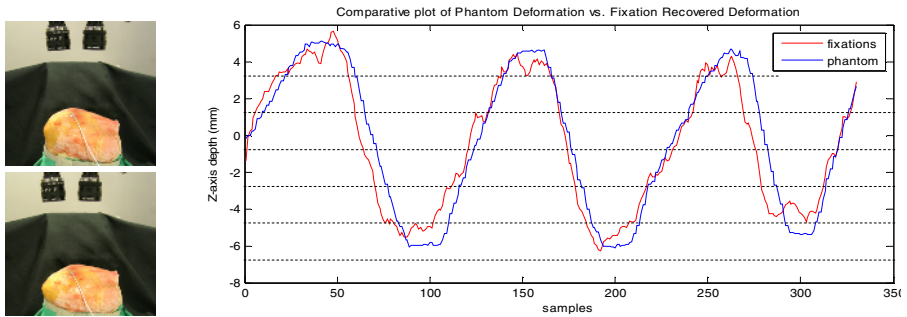


**Fig. 3.** The recovered deformation from ocular vergence. On the right, actual deformation of the phantom heart surface vs. the gaze-contingent reconstructed deformation. On the left, the setup for the experiment.

### 3.3   Gaze-Contingent In-Vivo Cardiac and Respiratory Motion Decoupling

To assess the *in vivo* accuracy of the proposed ocular-vergence 3D depth recovery system, data from a real robotic assisted operation where used. The operation was a totally endoscopic coronary artery bypass graft (TECAB) performed with the DaVinci robot. The stereo footage was played back in the stereoscopic viewer while a subject was eyetracked. The purpose of the experiment was to demonstrate how cardiac and respiratory motion can be recovered and decoupled using eyetracking. Fig. 4 shows a snapshot of the two views from a 40sec long footage. A large portion of the view was filled with the robotic endowrist, grasping the cardiac stabilizer just before positioned in place. During the entire footage, the laparoscope was stationary. The deformed tissue under foveation is at the bottom left portion of the image. What appears on the video sequence is a deforming epicardial surface with the respiratory motion principally manifested along the horizontal axis.
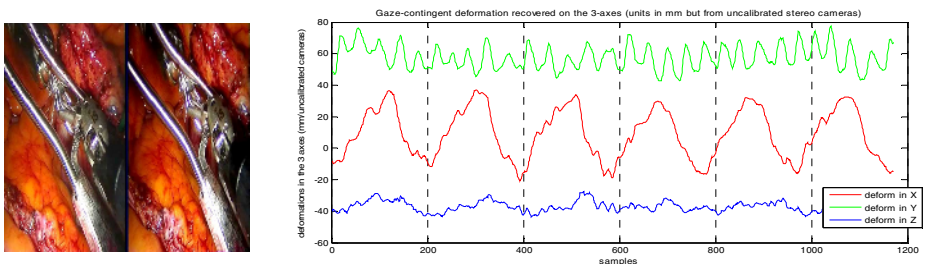


**Fig. 4.** The recovered epicardial surface deformation for a TECAB procedure. On the left, a snapshot of the binocular views used on the experiment is shown. On the right, the eyetracking acquired tissue deformations on the x, y and z-axes are plotted.

The cardiac motion is superimposed on top of this motion throughout the respiratory periods. The subject was asked to keep fixating a landmark on the tissue. Fig. 4 on the right, shows the collected eyetracking data on the x, y and z-axis (depth). Wavelet and PCA analysis is used to successfully decouple respiratory from cardiac motion as it can be seen on Fig. 5.
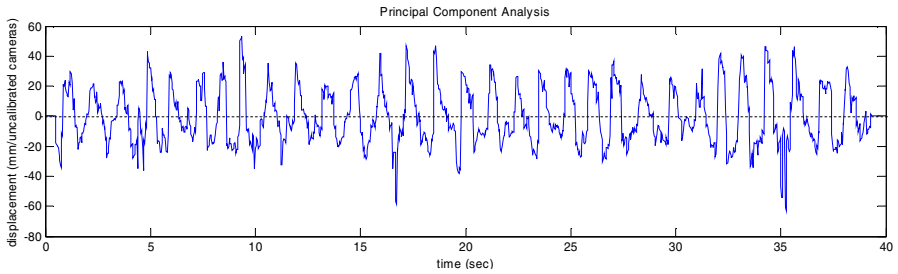


**Fig. 5.** The extracted cardiac motion by using Wavelet and PCA analysis of the gaze data

## 4   Discussion and Conclusions

In this paper, we have demonstrated the potential of gaze-contingent soft tissue motion and deformation recovery in MIRS. In this study, it is shown that the cut-off of 3D depth recovery with the proposed method reached at frequencies of about 1.5-1.8Hz when oscillations occur on the z-axis (depth). This corresponds to a heartbeat rate in the range of 100bps. It is expected that the response of the visual system in oscillations on the x and y axes will be better. In this case, and when depth recovery is not required, it would be possible to decouple the two eyes during eyetracking and use just the saccadic information of one of them preferably by using the dominant eye. In this way, it is possible to eliminate a large amount of noise caused by the intrinsic behaviour of the visual system during binocular coordination. It is also possible to get a better overall response by re-positioning the robotic endoscope in such a way that the principal oscillation component is shifted from z to the other two axes (*x* or *y*).

It is worth noting that there are also other important factors that can adversely affect the dynamic response of the visual system. These factors are accumulated through the video relay chain from the endoscope cameras up to the monitors of the stereoscope. It is worth noting that the human visual system is highly dynamic, none of the fitted models could accurately describe the response of the visual system to the oscillatory stimulus. We have demonstrated in this paper that it is practically possible to determine the 3D deformation of the surface at an oscillation frequency of about 100bps. Potential applications of the proposed approach for MIRS include motion stabilization, active constraints and intra-operative image registration under large tissue deformation.

## Acknowledgements

# References

1. Mon-Williams, M., Tresilian, JR, Roberts, A., Vergence provides veridical depth perception from horizontal retinal image disparities. Exp Brain Res **133** (2000) 407-413
2. Yang, G.-Z., Dempere-Marco, L., Hu, X.-P., Rowe,A., Visual search: psychophysical models and practical applications. Image and Vision Computing **20** (2002) 291-305
3. Andrew T. Duchowski , Eric Medlin, Nathan Cournia , Anand Gramopadhye , Brian Melloy , Santosh Nair, 3D eye movement analysis for VR visual inspection training, Proceedings of the symposium on ETRA 2002: eye tracking research & applications symposium, March 25-27, 2002, New Orleans, Louisiana 103-110
4. Mylonas, P.G., Darzi, A., Yang, G.-Z., Gaze Contingent Depth Recovery and Motion Stabilisation for Minimally Invasive Robotic Surgery. MIAR (2004), **LNCS 3150**, 311-319
5. Steiglitz, K., McBride, L.E., "A Technique for the Identification of Linear Systems." *IEEE Trans. Automatic Control*. Vol. **AC-10** (1965) 461-464
6. Deligianni, F., Chung, A., Yang, G.-Z., Decoupling of Respiratory Motion with Wavelet and Principal Component Analysis. Proceedings of the eighth Annual MIUA, (2004) 13-16
7. Ljung, L., System Identification Toolbox for use with MATLAB, User's Guide, www.mathworks.com