Soft-Tissue Motion Tracking and Structure Estimation for Robotic Assisted MIS Procedures

Danail Stoyanov[1], George P. Mylonas[1], Fani Deligianni[1],
Ara Darzi[2], and Guang Zhong Yang[1,2]

[1] Royal Society/Wolfson Foundation Medical Image Computing Laboratory
[2] Department of Surgical Oncology and Technology,
Imperial College of Science, Technology and Medicine, London SW7 2BZ, UK
{danail.stoyanov, george.mylonas, fani.deligianni,
a.darzi, g.z.yang}@imperial.ac.uk
http://vip.doc.ic.ac.uk

Abstract. In robotically assisted laparoscopic surgery, soft-tissue motion tracking and structure recovery are important for intraoperative surgical guidance, motion compensation and delivering active constraints. In this paper, we present a novel method for feature based motion tracking of deformable soft-tissue surfaces in totally endoscopic coronary artery bypass graft (TECAB) surgery. We combine two feature detectors to recover distinct regions on the epicardial surface for which the sparse 3D surface geometry may be computed using a pre-calibrated stereo laparoscope. The movement of the 3D points is then tracked in the stereo images with stereo-temporal constrains by using an iterative registration algorithm. The practical value of the technique is demonstrated on both a deformable phantom model with tomographically derived surface geometry and *in vivo* robotic assisted minimally invasive surgery (MIS) image sequences.

1 Introduction

Recent advances in robotic assisted Minimally Invasive Surgery (MIS) for performing micro-scale tasks using motion scaling and miniaturized mechanical wrists have made it possible to perform closed-chest cardiothoracic surgery on a beating heart. This approach minimizes patient trauma and avoids certain adverse effects associated with cardiopulmonary bypass. In practice, deformation of the epicardial surface due to cardiac and respiratory motion can impose significant challenges to delicate tasks such vessel anastomosis. The use of mechanical stabilizers can effectively remove most of the bulk motion, but residual tissue deformation remains significant in most cases. For intraoperative guidance and applying image guided active constraints to avoid critical anatomical structures such as nerves and blood vessels, it is necessary to develop complementary techniques for accurate 3D surface structure reconstruction and motion estimation *in situ* [1].

The determination of tissue deformation can be approached with a number of approaches that involve intraoperative imaging such as endoscopic ultrasound, or motion sensors such as mechanically or optically based accelerometers [2,3]. Marker based techniques have been proposed, but they involve suturing or projecting fiducals

J. Duncan and G. Gerig (Eds.): MICCAI 2005, LNCS 3750, pp. 139 – 146, 2005.

onto the heart surface [4,5]. Region based tracking of natural epicardial regions has also been investigated using monocular video sequences [6], but only for recovering 2D image motion. Since robotic assisted MIS procedures typically involve a pair of miniaturized stereo cameras, detailed 3D motion and structure recovery from the stereo laparoscope with image registration was recently proposed [7,8]. The major advantage of these methods is that they do not necessitate additional modification to the existing MIS hardware, but computationally they require complex computer vision algorithms inferring dense 3D correspondence which is often an ill-posed problem. Existing research has shown that sparse sets of well known feature correspondences can be used as ground control points to enforce additional constraints and increase the inherent accuracy and robustness of dense stereo techniques [9]. Furthermore, the integration of other visual cues such as shading and specular reflectance and their temporal characteristics in response to soft-tissue deformation can further improve the practical value of optically based methods.

The purpose of this paper is to introduce a method for inferring precise 3D structure and motion for a set of sparse salient features on the soft-tissue surfaces during robotic assisted MIS procedures. With a calibrated stereo laparoscope, a combination of landmarks is used to provide robust performance in the presence of specular reflections. The temporal behavior of each landmark is then derived by using constraints in the stereo video sequence. Detailed validation of the proposed method was performed on both a phantom model with known geometry and *in vivo* robotic assisted MIS data.

2 Methods

2.1 Salient Landmarks on the Epicardial Surface

Traditionally, the identification of salient landmarks is usually achieved with edge or corner features for sparse stereo matching and motion tracking. For robotically assisted MIS, these features can be unstable and prone to errors due to the homogeneity of surface texture and the presence of specular highlights, which can cause clustering of high frequency features on the specular boundary. In the context of wide-baseline stereo matching, Matas *et al.* [10] defined maximally stable extremal regions (MSER) based on thresholding the image intensity to create connected components with local minima. The use of MSER landmarks has a number of desirable properties as they are invariant to monotonic changes in illumination. Furthermore, if they are detected by starting from the lowest intensity (MSER-), they can implicitly avoid specular reflections. It can also be shown that on MIS cardiac surfaces, MSER- generally corresponds to physically meaningful texture details such as superficial blood vessels or small tissue bruising.

In this paper, a combination of MSER- regions and the traditional gradient based image features [11] is used for salient landmark selection. The use of different feature descriptors can provide added robustness [12], which is necessary in the presence of occlusions and specular highlights as encountered in cardiac MIS procedures. We associate a measurement region (MR) around each landmark for computing the dissimilarity metrics. The MR is a rectangular window for corner features and an ellipse that bounds the convex hull of the component for MSER- regions.

2.2 Stereo Feature Matching

To recover 3D measurements from a stereoscopic laparoscope, the correspondence of landmarks of the stereo pair needs to be determined. There are many algorithms for matching sparse feature sets, and in this work we used the method proposed by Pilu [13] for combining proximity and similarity measures to discriminate between potential matches. For each feature type, we build a cost matrix with row entries corresponding to features in the left image and columns for the right image. Each entry of the cost matrix depicts how well respective features correspond to each other by using the following dissimilarity measure:

$$Cost_{ij} = e^{-\frac{(C_{ij}-1)^2}{2\gamma^2}} e^{-\frac{r_{ij}^2}{2\mu^2}} \tag{1}$$

In Eq. (1), r is the Euclidian distance between features, γ and μ are sensitivity control parameters set to the values suggested in [13], and C_{ij} is the normalized cross correlation (NCC) between the measurement regions of features i and j. When matching MSER-, we used the largest MR for computing correlation. The algorithm makes use of the properties of Singular Value Decomposition (SVD) of the cost matrix to attenuate matrix values for poor matches. Once corresponding points are determined, 3D points on the epicardial surface can be inferred by using the centre of mass of the MSER- as a reference. By calibrating the camera before the MIS procedure, the epipolar constraint can be introduced to the proximity cost. Calibration is also important to the intrinsic accuracy of the proposed technique as otherwise the recovered 3D points will be ambiguous up to a projective transformation if just using the determined stereo correspondences for estimating the camera matrices.

2.3 Temporal Tracking Using Stereo Constraints

Once the stereo correspondence is established, we used temporal motion tracking of salient features to iteratively update their temporal positions in 3D space by using both stereo frames. This extends the Lucas-Kanade (LK) [15, 16] registration algorithm for incorporating the inter-stereo epipolar constraint. The goal of the LK tracker is to align a reference image template $T(\mathbf{x})$ with an image region subject to the squared pixel difference, given a warping function $W(\mathbf{x};\mathbf{p})$ of arbitrary complexity and $\mathbf{p}$ parameters. The algorithm starts with an initial estimate of the warping parameters $\mathbf{p}$ and iteratively computes an update term $\Delta\mathbf{p}$ until convergence below a predefined threshold $\Delta\mathbf{p} \leq \xi$. The error function e used for minimizing the modified stereo LK tracker is defined as:

$$\varepsilon = \sum_{\mathbf{x}} \left[\left[I\left(W\left(\mathbf{x};\mathbf{p}+\Delta\mathbf{p}\right)\right) - T\left(\mathbf{x}\right) \right]^2 + \left[J\left(W'\left(\mathbf{x}';\mathbf{p}+\Delta\mathbf{p}\right)\right) - T'\left(\mathbf{x}'\right) \right]^2 \right] \tag{2}$$

where $I(W(\mathbf{x};\mathbf{p}))$ and $J(W'(\mathbf{x};\mathbf{p}))$ are the images transformed by the respective warping function. The error function can be linearized by taking the first order Taylor expansion about $\mathbf{p}$, such that the partial derivative with respect to $\Delta\mathbf{p}$ can be determined by using the chain rule. Setting the partial derivative to zero and solving for $\Delta\mathbf{p}$ yields a least-squares solution (constants are ignored), we have:

$$\Delta\mathbf{p} \approx H^{-1}\sum_{\mathbf{x}}\left[\left[\nabla I\frac{\partial W}{\partial \mathbf{p}}\right]^{T}\left[I(W(\mathbf{x};\mathbf{p}))-T(\mathbf{x})\right]^{2}+\left[\nabla J\frac{\partial W'}{\partial \mathbf{p}}\right]^{T}\left[J(W'(\mathbf{x}';\mathbf{p}))-T'(\mathbf{x}')\right]^{2}\right] \tag{3}$$

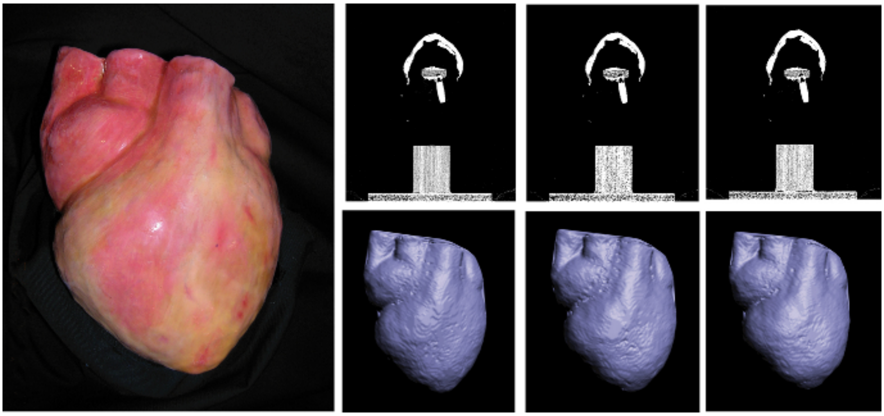
Where ∇I and ∇J are the warped image gradients of each stereo channel, $\partial W / \partial \mathbf{p}$ is the *Jacobian* of each warping function and H^{-1} is the inverse of the *Hessian* matrix:

$$H = \sum_{\mathbf{x}}\left[\nabla I\frac{\partial W}{\partial \mathbf{p}}\right]^{T}\left[\nabla I\frac{\partial W}{\partial \mathbf{p}}\right]+\left[\nabla J\frac{\partial W'}{\partial \mathbf{p}}\right]^{T}\left[\nabla J\frac{\partial W'}{\partial \mathbf{p}}\right] \tag{4}$$

The motion parameterization used in this is study is based on a pure translation model for each feature, which incorporates terms for the vertical and horizontal motion in the reference image, and an additional term for disparity changes in the stereo pair. More complex models can readily be incorporated into the proposed framework (for further details consult the study by Baker *et al* [16]) but the increased search space may have an adverse effect on the actual system performance [6]. Furthermore, it has been demonstrated that for small inter-frame motion, the use of translation tracking alone can be sufficient [11].

2.4 Experimental Design

The proposed method was implemented in C++ on a standard desktop PC with a Pentium IV 2.4 GHz CPU and 512 Mb RAM, running the Windows XP operating system. In the current implementation, initialization took 0.5s (mostly for MSER detection) after which the algorithm processed 320×288 images at 11 frames per second (fps). With further optimization real-time performance can be achieved.



(a) **(b)**

Fig. 1. The cardiac phantom model used for validating the proposed technique **(a)** image of the heart model showing the visual and geometrical fidelity of the model and **(b)** CT slice for three levels of deformation and 3D renditions of reconstructions from the respective CT series

To validate the proposed method, a scaled simulation environment was created with a phantom heart model shown in **Fig. 1 (a)**. A stereo rig mounted on a Stäubli RX60 robotic arm with six degrees of freedom (DOF) and repeatability accuracy of ±0.02mm. The phantom model was created using thixotropic silicone mould rubber and pre-vulcanized natural rubber latex with rubber mask grease paint to achieve a specular appearance and high visual fidelity. The deformable silicone surface was mounted onto a piston mechanism with controllable injection levels to simulate the heart beat motion in a reproducible manner. The precise phantom model geometry was recovered at seven discrete heart beat simulation levels using a Siemens Somatom Sensation 64 CT scanner with slice thickness of 0.6mm, an example CT slice and 3D rendition of a reconstruction are shown in **Fig. 1 (b)**.

For *in vivo* analysis, data from robotic assisted cardiac surgery carried out with a daVinci™ surgical system (Intuitive Surgical, CA) was used. The cameras of the stereoscopic endoscope were hardware synchronized by using a proprietary FPGA device designed by this institution. The stereo cameras were calibrated before the procedure using a planar calibration object [17]. The proposed method was used to detect and then track landmarks on the epicardial surface after the positioning of a mechanical stabilizer. Since, ground truth data for the 3D structure and motion of the soft-tissue cannot be easily obtained for robotic procedures, we used the motion of landmarks on the epicardial surface to determine the respiratory and cardiac motion as a means of qualitative analysis.

3 Results

The phantom heart model described above was used to generate an image sequence of 50 frames, with each frame showing consecutive deformation of the heart associated with the CT data. The setup was devised so that the resultant inter-frame pixel motion

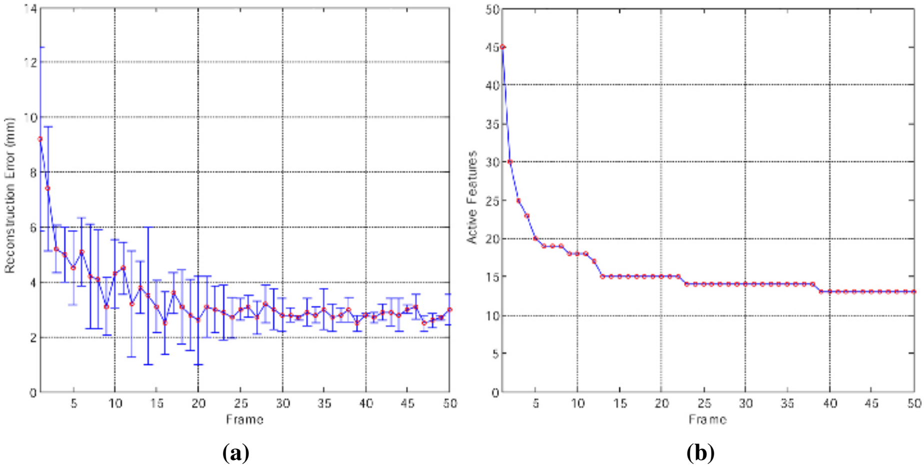


Fig. 2. Phantom model experiment for evaluating reconstruction accuracy of stereo feature tracking **(a)** the average and standard deviation of error in millimeters for feature correspondences in each experimental frame **(b)** the number of features actively tracked at each frame of the sequence

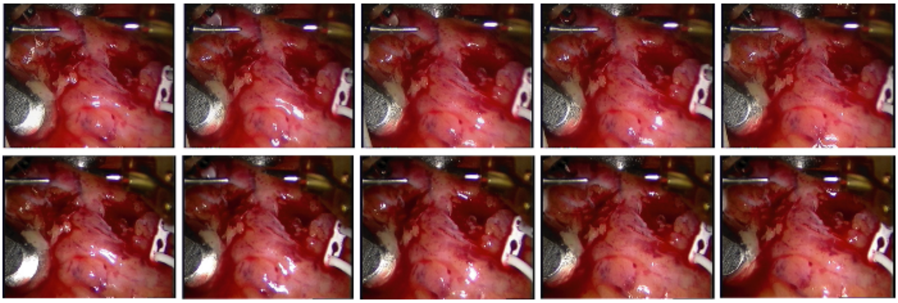
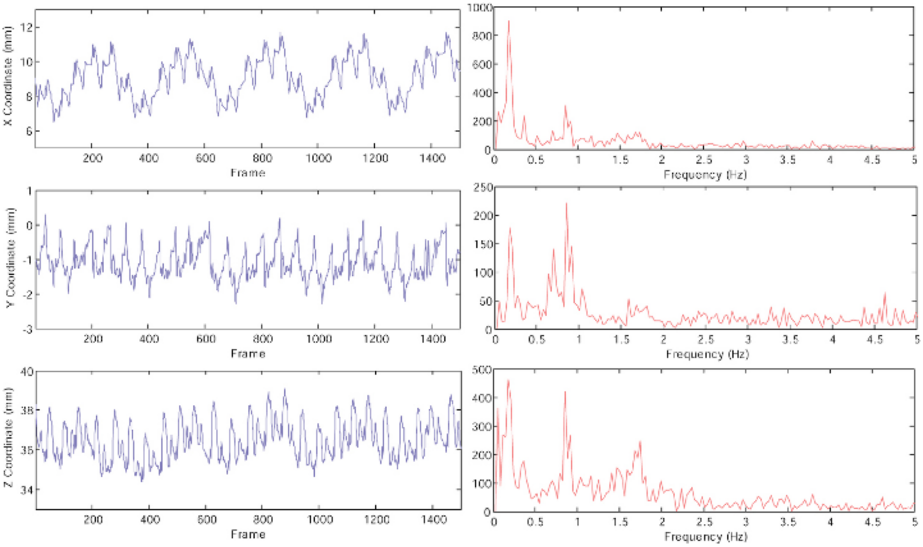


Fig. 3. Example stereoscopic image pairs of robotic assisted totally endoscopic coronary artery bypass surgery used for the *in vivo* analysis in this study



(a) **(b)**

Fig. 4. Results for *in vivo* robotic assisted MIS **(a)** the recovered 3D coordinates in the left camera reference system for a landmark tracked through 1500 video frames at 50 fps **(b)** power spectral analysis clearly identifies the heart beat and respiratory frequencies in the 3D motion

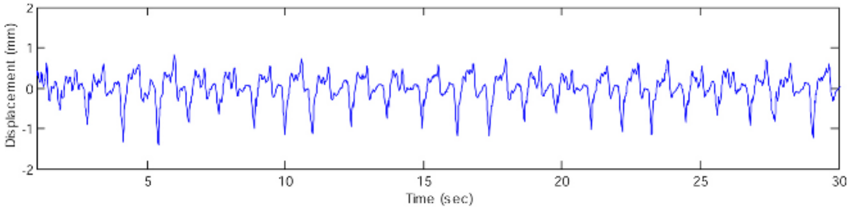


Fig. 5. Principal component analysis of the recovered motion signal indicating the decoupled cardiac motion component

did not exceed 15 pixels, which was consistent with observations from *in vivo* data for consecutive frames. Metric error was measured as the distance between the reconstructed 3D point and the point on the CT reconstructed surface along the ray back-projected from the left camera. In **Fig. 2**, we demonstrate the reconstruction accuracy of stereo correspondence obtained with the proposed technique. Not all features are suitable for temporal tracking, and initial outliers were rejected depending on the correlation threshold. This results in fewer features being tracked over the entire period but improves the overall accuracy by ensuring that only consistent landmarks are considered. With the proposed framework additional landmarks may be introduced at any stage.

The *in vivo* performance of the algorithm is assessed with a robotic assisted totally endoscopic coronary artery bypass graft (TECAB) as shown in **Fig. 3**. In **Fig. 4**, it is evident that the recovered motion clearly captures the coupled deformation of the epicardial surface due to cardiac as well as respiratory motion. It is worth noting that the graph shown in **Fig. 4** illustrates the surface motion as projected onto the x, y and z axes of the camera coordinate system. Within this figure, the power spectrum of each of the motion components is also provided, which illustrates the dominant frequencies derived from the proposed algorithm. In **Fig. 5**, we show the decoupled motion component indicating only the cardiac motion by using a localized principal component analysis (PCA) cardiac/respiratory decoupling technique.

4 Discussion and Conclusions

In this paper, we have proposed a practical method for determining soft-tissue deformation for robotic assisted MIS from a set of landmarks. We have used a combination of landmarks including MSER- regions and the traditional gradient-based image features for ensuring robust system performance. Results from the phantom model have demonstrated the accuracy of 3D reconstruction that can be achieved and analysis of *in vivo* robotic assisted MIS data has further demonstrated the clinical value of the proposed technique. With the current implementation, features occluded by the instruments or tissue effects such as bleeding are detected through correlation and epipolar geometry thresholds and set as outliers in the tracking process. The introduction of new features or labeling lost features as occluded and performing subsequent searches with statistical motion models can be used improve the tracking process.

Acknowledgements

The authors thank Eddy Kavanagh, Danny Sharpe and Bob Pullen for their valuable help and Roberto Casula for kindly providing the *in vivo* data used in this study.

References

1. Taylor, R. H., Stoianovici, D.: Medical Robotics in Computer-Integrated Surgery. IEEE Transactions on Robotics and Automation, (19):765-781, 2003.
2. Hoff, L.,Elle, O.J., Grimnes, M.J., Halvorsen, S., Alker, H.J., Fosse, E.: Measurements of heart motion using accelerometers. In: Proc. EMBC, 2049 – 2051, 2004.

3. Thrakal A, Wallace J, Tomlin D, Seth N, Thakor N. Surgical Motion Adaptive Robotic Technology (SMART): taking the motion out of physiological motion. In: Proc. MICCAI, 317-325, 2001.
4. Nakamura, Y., Kishi, K., Kawakami, H.: Heartbeat synchronization for robotic cardiac surgery. In: Proc. ICRA, 2014-2019, 2001.
5. Ginhoux, R., Gangloff, J. A., de Mathelin, M. F., Soler, L., Arenas Sanchez, M., Marescaux, J. : Beating Heart Tracking in Robotic Surgery Using 500 Hz Visual Servoing, Model Predictive Control and an Adaptive Observer. In: Proc. ICRA, 274-279, 2004.
6. Gröger, M., Ortmaier, T., Sepp, W., Hirzinger, G.: Tracking local motion on the beating heart. In: Proc. SPIE Medical Imaging Conference, 233-241, 2002.
7. Stoyanov, D., Darzi, A., Yang, G-.Z.: Dense 3D Depth Recovery for Soft Tissue Deformation During Robotically Assisted Laparoscopic Surgery. In: Proc. MICCAI, 41-48, 2004.
8. Lau, W., Ramey, N., Corso, J., Thakor, N., Hager, G.: Stereo-Based Endoscopic Tracking of Cardiac Surface Deformation. In: Proc. MICCAI, 494-501, 2004.
9. Bobick, A. F., Intille, S. S.: Large occlusion stereo. International Journal of Computer Vision, (33):181-200, 1999.
10. Matas, J., Chum, O., Martin, U., Pajdla, T.: Robust wide baseline stereo from maximally stable extremal regions. In: Proc. BMVC, 384-393, 2002.
11. Shi, J., Tomasi, C.: Good features to track, In: Proc. CVPR, 593 - 600, 1994.
12. Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, F., Van Gool, L.: A comparison of affine region detectors. International Journal of Computer Vision, *in press*.
13. Pilu, M.: A Direct Method for Stereo Correspondence based on Singular Value Decomposition. In: Proc. CVPR, 261-266, 1997.
14. Hager, G. D., Belhumeur, P. N.: Efficient Region Tracking With Parametric Models of Geometry and Illumination. IEEE Transactions on Pattern Analysis and Machine Intelligence, (20):1-15, 1998.
15. Lucas, B. D., Kanade, T.: An Iterative Image Registration Technique with an Application to Stereo Vision. In: Proc. IJCAI, 674-679, 1981.
16. Baker, S., Matthews, I.: Lucas-Kanade 20 Years On: A Unifying Framework. International Journal of Computer Vision, (56):221-255, 2004.
17. Zhang, Z.: A flexible new technique for camera calibration. IEEE Transactions on Pattern Analysis and Machine Intelligence. (22):1330-1334, 2000.