

3D Reconstruction of Internal Organ Surfaces for Minimal Invasive Surgery

Mingxing Hu¹, Graeme Penney¹, Philip Edwards², Michael Figl², and David Hawkes¹

¹ Centre for Medical Image Computing, University College London

² Department of Surgical Oncology and Technology, Imperial College London,
London, United Kingdom

{mingxing.hu,g.penney,d.hawkes}@ucl.ac.uk, {eddie.edwards,
m.figl}@imperial.ac.uk

Abstract. While Minimally Invasive Surgery (MIS) offers great benefits to patients compared with open surgery surgeons suffer from a restricted field-of-view and obstruction from instruments. We present a novel method for 3D reconstruction of soft tissue, which can provide a wider field-of-view with 3D information for surgeons, including restoration of missing data. The paper focuses on the use of Structure from Motion (SFM) techniques to solve the missing data problem and application of competitive evolutionary agents to improve the robustness to missing data and outliers. The method has been evaluated with synthetic data, images from a phantom heart model, and *in vivo* MIS image sequences using the da Vinci telerobotic surgical system.

1 Introduction

The past decade has witnessed significant advances on robotic assisted Minimally Invasive Surgery (MIS) evolving from early laboratory experiments to an indispensable tool for many surgeries. MIS offers great benefits to patients: the incisions and trauma are reduced and hospitalisation time is shorter. Robotic assisted techniques further enhance the manual dexterity of the surgeon and enable him to concentrate on the surgical procedure. Despite of all these advantages, MIS using an endoscope still suffers from a fundamental problem: the narrow field-of-view. As a result, the restricted vision impedes the surgeon's ability to collect visual information (locating landmarks, etc.) from the scenes and his/her awareness of peripheral sites.

A large amount of research work has been carried out which attempts to provide more 2D or 3D preoperative or intraoperative information to the surgeon. For example, Seshamani et al. presented an endoscopic mosaicing technique to display a broader field-of-view of the scene by stitching together images in a video sequence [1]. Unfortunately, the mosaicing results are only 2D images with larger scene areas, which lack any 3D information. In order to recover the 3D depth of a scene, Devernay proposed a five-step strategy to achieve 3D reconstruction of the operating field based on vision techniques [2]. More recently stereo-based techniques were proposed to reconstruct soft tissue structure in MIS, which can also track the temporal motion of the deformable surface [3, 4]. However, a drawback of these methods is that the images captured for 3D reconstruction are mainly from a fixed field-of-view, so they

cannot offer 3D information of peripheral scenes around the operating field. Exceptions to this are [5, 6], where a 3D map of the scene is built and camera movement is recovered based on the Simultaneous Localization and Mapping technique. However, they require long term repeatable landmarks to estimate the camera parameters, which are hard to obtain in dynamic scenes with soft tissue.

This paper tries to solve the 3D reconstruction problem of internal organ surface for MIS and makes the following contributions: (i) 3D structure is recovered from a moving vision system using Structure from Motion (SFM) techniques [7]. Feature points are detected and tracked when the camera system moves around the operating area. The 3D positions of these features are reconstructed to provide a broader field-of-view. Therefore, we can offer a wider operating scene with 3D information to the surgeon, which can be used for 3D-3D registration of the anatomy to the pre-operative data. We also want to emphasize that the proposed method is able to reconstruct 3D organ surface without any laser scanner or fiducial markers but only with endoscopic cameras. (ii) A Competitive Evolutionary Agent-based (CEA) algorithm is proposed to solve the missing data problem and improve robustness to outliers [8]. Features may move out of the field-of-view during the feature tracking so the missing data needs to be filled before estimating the 3D structure. Outliers obtained from bad location and false tracking, on the other hand, can distort the estimation to such a degree that the fitted parameters become arbitrary. The CEA algorithm is applied to remove outliers from the feature dataset, so as to minimize their effect on the whole estimation. The proposed technique is validated using a heart phantom with known 3D geometry and using *in vivo* data from totally endoscopic coronary artery bypass (TECAB) surgery with the da VinciTM system.

2 Material and Methods

2.1 Stereo Feature Tracking

The first step to reconstruct 3D structures during a MIS endoscopic procedure is to track image features as the cameras move. One of the well-known tracking methods is the Lucas-Kanade (LK) tracker [9], which we extended to a stereo framework. The LK tracker minimizes the sum of squared errors between two images I_k and I_{k+1} by altering the warping parameters \mathbf{p} which are used to warp I_{k+1} to the coordinate frame of I_k . For a general motion model with transformation function $W(\mathbf{x}, \mathbf{p})$, the objective function is

$$\min_{\mathbf{x}} \sum [I_{k+1}(W(\mathbf{x}; \mathbf{p} + \Delta \mathbf{p})) - I_k(\mathbf{x})]^2 \quad (1)$$

This expression is linearized by a first order Taylor expansion on $I_{k+1}(W(\mathbf{x}; \mathbf{p} + \Delta \mathbf{p}))$

$$\min_{\mathbf{x}} \sum \left[I_{k+1}(W(\mathbf{x}; \mathbf{p})) + \nabla I_{k+1} \frac{\partial W}{\partial \mathbf{p}} - I_k(\mathbf{x}) \right]^2$$

Where ∇I_{k+1} is the image gradient vector and $\partial W / \partial \mathbf{p}$ is the Jacobian of the transformation function. In our stereo framework, the tracking process is extended to four directions, namely, two temporally consecutive directions, $Tr(I_k, I_{k+1})$ (where

$Tr(I_k, I_{k+1})$ denotes tracking from I_k to I_{k+1}) and $Tr(J_k, J_{k+1})$, and two corresponding (stereo) directions, $Tr(I_k, J_k)$ and $Tr(J_k, I_k)$, as shown in Fig. 1. Since the relative position of the stereoscopic cameras is fixed and remains unchanged during MIS surgery, we can use this to adapt the window size and initial position for feature tracking in the corresponding directions. Then features $\mathbf{x}_{I(k)}$ and $\mathbf{x}_{J(k)}$ are considered to be a stereo pair (two projections of a 3D point \mathbf{X} from two viewpoints), if they satisfy

$$\varepsilon = \|\mathbf{x}_{I(k)} - \mathbf{x}'_{J(k)}\|_F + \|\mathbf{x}_{J(k)} - \mathbf{x}'_{I(k)}\|_F \leq \delta \quad (2)$$

where \mathbf{x}'_I (or \mathbf{x}'_J) is the corresponding point of \mathbf{x}_I (or \mathbf{x}_J) in image J (or I) and they are added to the feature dataset if the error ε is less than the threshold δ .

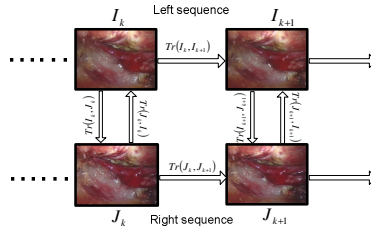


Fig. 1. Framework of stereo tracking

2.2 Structure from Motion with Missing Data

In the feature tracking process, it is possible that some features are only detected and tracked for few frames. These features cannot be used for reconstruction using conventional vision techniques [7, 11]. The missing data needs to be recovered before 3D reconstruction can be carried out.

Given a set of tracked feature points, their 2D and 3D locations can be related using the following matrix:

$$\underbrace{\begin{bmatrix} \lambda_1^1 \mathbf{x}_1^1 & \cdots & \lambda_n^1 \mathbf{x}_n^1 \\ \vdots & \ddots & \vdots \\ \lambda_1^m \mathbf{x}_1^m & \cdots & \lambda_n^m \mathbf{x}_n^m \end{bmatrix}}_{\mathbf{M}} = \underbrace{\begin{bmatrix} \mathbf{P}^1 \\ \vdots \\ \mathbf{P}^m \end{bmatrix}}_{\mathbf{P}} \underbrace{\begin{bmatrix} \mathbf{X}_1 & \cdots & \mathbf{X}_n \end{bmatrix}}_{\mathbf{X}} \quad (3)$$

where $\mathbf{x}_i^j = (x_i^j, y_i^j, 1)^T$ is the projection of 3D point $\bar{\mathbf{X}}_i = (X_i, Y_i, Z_i, 1)^T$ ($i=1, \dots, n$) onto the j -th ($j=1, \dots, m$) image plane. \mathbf{P}^j is a 3×4 projective matrix of frame j , and λ_i^j is the projective depth, which can be computed using epipolar geometry [11]. The measurement matrix can be further factorized into shape and motion components, \mathbf{P} and \mathbf{X} , by enforcing the rank four constraint. In the absence of missing data, the solution to this linearized problem is found as the \mathbf{P} and \mathbf{X} ($\text{rank}(\mathbf{P}) = \text{rank}(\mathbf{X}) = 4$) that minimize:

$$\min_{\mathbf{P}, \mathbf{X}} \mathbf{R}_e = \min_{\mathbf{P}, \mathbf{X}} \|\mathbf{M} - \mathbf{P}\mathbf{X}\|_F^2 \quad (4)$$

where \mathbf{R}_e is the residual between the model $\mathbf{P}\mathbf{X}$ and the data \mathbf{M} .

However, in the presence of missing data, this is not the case. Recently Martinec and Pajdla proposed a structure recovery method [10], which modified the projective reconstruction in [11] with the Jacobs' linear fitting method [12]. The central idea is to apply a linear algebra operation to fit an unknown matrix \mathbf{M}_r of a certain rank $r=4$ to the incomplete noisy measurement matrix \mathbf{M} by using a number of incomplete submatrices \mathbf{S} (quadruplet of linear independent columns of \mathbf{M}). Since each submatrix \mathbf{S} can provide some constraints, putting all the constraints together will generate much stronger constraints to the whole fitting problem. Readers are referred to [12] for further information about data filling techniques.

Ideally, each submatrix (quadruplet of the columns) should be considered, but the vast number of candidates makes the computation almost infeasible. In [10] a random selection of submatrices is used instead. This is a major drawback as it does not maximize the contributions of the submatrix samples to the whole problem. A system which allows more information to be obtained from each submatrix would be able to solve more difficult fitting problems given the limited amount of information. Moreover, if some outliers are included in \mathbf{S} , it may not contribute any useful information to the estimation, or even divert the whole estimation to some incorrect result.

2.3 Competitive Evolutionary Agents for 3D Reconstruction

In order to more effectively choose a selection of submatrices we employ a new approach based on competitive evolutionary agents to solve the 3D reconstruction problem with missing data and outliers. The agents represent a subset of the columns used to construct the submatrices. They carry out evolutionary behavior, such as reproduction and diffusion, to search for good samples from a large number of candidates. The quality of the submatrix is also supervised using a fitness function, which measures the amount of missing data and perturbation levels of noise and outliers.

Agent definition

Suppose that Ψ is the dataset representing all the columns of \mathbf{M} . Ψ may be viewed as a one-dimensional index to measurements of the feature points, and also as an environment in which the agents inhabit and evolve. Then the evolutionary agent can be defined as

$$Agent = \langle \mathbf{V}, a, F_{fitness}, fml, Rep, Diff, Die \rangle$$

where \mathbf{V} is an r -dimensional position vector which denotes positions in Ψ and the value of its component V_k , $k=1, \dots, r$, is an index number within Ψ . a , $F_{fitness}$ and fml represents its internal behavior, i.e., agent age, fitness and family index respectively. Rep , $Diff$ and Die describe its external behavior, i.e., reproduction, diffusion and vanishing, respectively.

Fitness function

The value of the fitness function is mainly determined by the number and position of missing data and unknown depths. For example, for each agent we can construct a

submatrix $S_t = [s_t^1 \ s_t^2 \ s_t^3 \ s_t^4]$ using the selected quadruplet from \mathbf{M} , and $t \in T$ (T is a random set of indices). Then we calculate the number of rows without missing data n_{fill}^4 , and the number of rows with one missing entry n_{fill}^3 . Rows with more than one missing element are seldom considered, because they cannot provide good constraints for the relevant viewpoints. In this way a concise matrix \hat{S}_k is constructed with the rows of S_k containing either zero or only one missing element. Then the number of unknown depths in \hat{S}_k , n_{depth} , is calculated. Matrix \hat{S}_k can be used for decomposition to provide constraints for matrix factorization. Therefore, the fitness function can be written as

$$F_{fitness} = n_{fill}^4 + \alpha_1 n_{fill}^3 - \alpha_2 n_{depth} \quad (5)$$

where $0 < \alpha_1, \alpha_2 \leq 1$ are the constants used to weight the data-missing level and depth-unknown level terms. Generally, the larger the fitness value, the fewer missing entries and unknown depths in the selected submatrix.

Evolutionary behavior

Evolutionary agents adapt to their environment by way of behavioral responses, namely, reproduction and diffusion. Letting $\Phi^{(g)}$ denote the set of active agents in generation g .

(1) Reproduction: Each active agent $\alpha^{(g)}$ will breed a finite number m_{br} of offspring agents $\alpha^{(g+1)}$. The differences between $\alpha^{(g)}$ and $\alpha^{(g+1)}$ primarily lie in the position vectors $\mathbf{V}^{(g)}$ and $\mathbf{V}^{(g+1)}$. Some elements of $\mathbf{V}^{(g)}$ are chosen and changed to a new index number selected randomly from Ψ by a random number generator. Then the fitness of the new generated agent is computed using equation (5), and compared with that of the siblings of the same family, $\alpha_{(l)}^{(g+1)}$ (where l denotes the family number). Only the agent with largest fitness value will survive and move on to the diffusion process.

(2) Diffusion: After the reproduction process, the successful agent $\alpha^{(g+1)}$ will compare its fitness with that of its parent $\alpha^{(g)}$. If the offspring has a fitness advantage, it will be appended to the agent set $A^{(g+1)}$, and its parent will become inactive and be removed from the evolutionary environment. If not, the offspring will be deleted and the age of its parent will be increased by one. In addition, if the age of an agent exceeds its life span, it will be removed from the environment.

2.4 Experimental Design

In order to assess and quantify the performance of the proposed method, a heart phantom made of silicone rubber (The Chamberlain Group, Great Barrington, MA, USA) was used, dimensions 6" (atria / apex) \times 4" wide \times 3.5" (anterior / posterior). Then the 3D surface of the phantom was reconstructed using the AlignRT (VisionRT, London, UK) system to provide the ground truth surface data. This enabled us to choose points from surface as ground truth data to evaluate the accuracy of the proposed method, or to register the reconstructed point cloud directly to the surface. In addition a CT scan of the heart phantom was also used (Philips MX8000 IDT 16

slice CT scanner: voxel size $1 \times 1 \times 2 \text{mm}$ with $768 \times 768 \times 145$ voxels). Fig. 2 shows the heart phantom, the 3D surface obtained from the AlignRT system and the 3D surface reconstructed from CT scan using the marching cubes algorithm implemented in VTK.

We have carried out experiments on three sets of data: synthetic test data, heart phantom data and *in vivo* data from endoscopic surgery. These sets of data and a description of the experiments are described in the following paragraphs.

The synthetic data was produced using 3D points \underline{X} randomly selected from the 3D surface captured using the AlignRT system. These points were projected to twenty image planes with different rotation and translation parameters to generate sets of 2D image points. Two sets of experiments were then carried out:

- (G1): Ten different ranges of Gaussian noise are added to the image measurements, with a zero mean and standard deviation varying from 0.5 to 5.0. The percentage of missing data was fixed at 20%.
- (G2): The standard deviation of Gaussian noise was fixed at 2.0, and the percentage of missing data was varied from 10% to 60%.

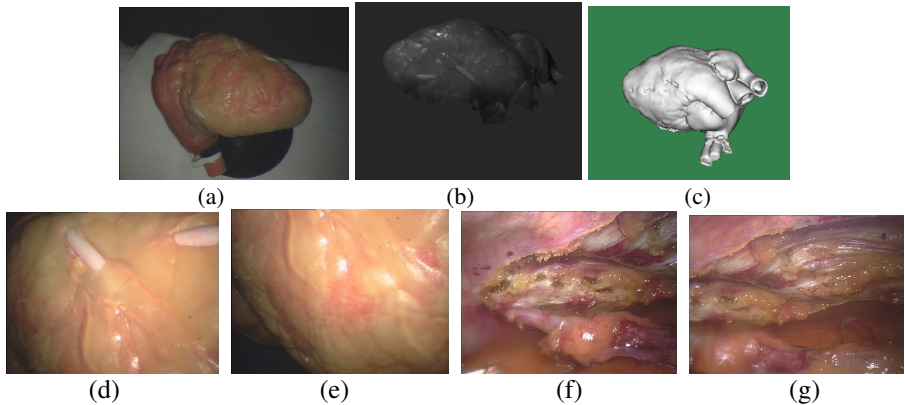


Fig. 2. Examples of experimental data: (a) Image of the heart model; (b) 3D surface reconstructed from AlignRT scanning; (c) 3D model reconstructed from CT scanning; (d) and (e) show the first and last frames of the image sequence of phantom heart; (f) and (g) show the first and last frames of the image sequence from TECAB surgery;

The da VinciTM robotic surgical system (Intuitive Surgical, Inc., Sunnyvale, CA, USA) was used to obtain images of the heart phantom. The heart phantom was placed on a bed and the endoscopic cameras were positioned approximately 40mm from the heart surface. Twenty images were selected from this set of data (every fifth frame) and the first and last frames captured from left camera are shown in Fig. 2 (d) and (e), respectively. Using these data our algorithm detected and tracked feature points over the image sequence. In total 401 features were detected and on average 48.8% of the data within the matrix \mathbf{M} was missing.

Our *in vivo* data set consisted of endoscopic images from a TECAB surgical procedure using the da Vinci system. Fig. 2 (f) and (g) shows the first and last frames

from the left sequence respectively. A surface was reconstructed from this data set. In total 397 features were detected and tracked and on average 45.8% of data within the matrix \mathbf{M} was missing. In order to further improve the accuracy of the proposed method, a bundle adjustment technique [13] was also applied after data filling. This technique minimizes the RMS error, which is the mean squared distance between the observed image points and the projected image points $\hat{\mathbf{x}} = \mathbf{P}\mathbf{X}$.

3 Experimental Results

3.1 Synthetic Data

In the noise testing experiment (G1), we ran 100 trials for each noise level. A graph of the result is shown in Fig. 3 (a). It can be seen that the CEA-based method performs better than the *M&P* method [10] over all the experiments. In particular, as the noise level increases ($\sigma = 4, 5$), the performance of our approach is affected less than the *M&P* method. Fig. 3 (b) displays the results of the missing-data testing (G2). We can see that the average error of CEA increases when the amount of missing data increases, but not as significant as that of the *M&P* method. This suggests that evolutionary agents can effectively select good samples from a vast pool of candidates, even when a large percentage of the elements are missing.

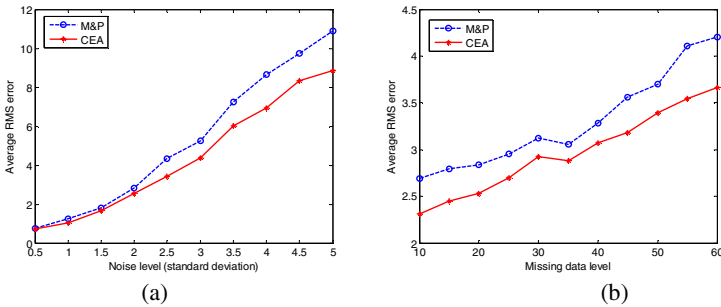


Fig. 3. Experimental results of synthetic data testing. (a) Average RMS error under different noise levels; (b) Average RMS error under different missing-data levels.

3.2 Phantom Model

Fig. 4 (a) and (b) illustrate the results of our data filling method. The measured features are marked with blue “o” and the reprojected points to those image planes are marked with red “x”. In Fig. 4 (a), it can be seen that LK tracker lost the feature since the 7th frame, but using the proposed data filling method we can still “track” the feature in frames 8 and 9 and the recovered positions appear visually reasonable. Fig. 4 (c) displays the reconstructed 3D point cloud using the proposed technique. Since we did not have a gold standard transformation between the 3D points and our CT image, we used the Iterative Closet Point (ICP) [14] algorithm to register the 3D

points to the surface derived from CT to obtain a measure of 3D reconstruction accuracy. Fig. 4 (d) shows the final position after ICP registration. Most of the points lie on or are close to the surface and the mean residual is around 1.87 mm. Moreover, using visual inspection we can see that the positions of 3D points are overlaid on the CT surface after ICP registration in roughly the correct position.

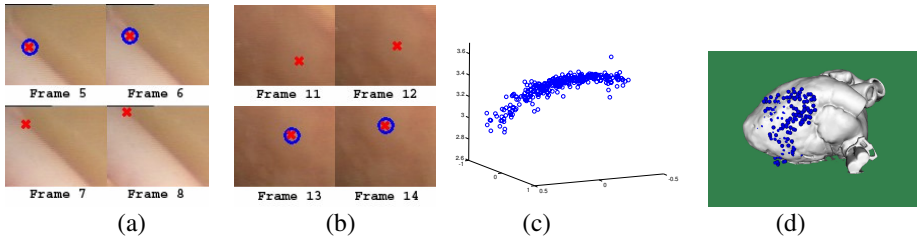


Fig. 4. Results using the heart phantom. (a) and (b) show the data filling results of two features; (c) and (d) display the reconstructed point cloud and its registration result to CT scan model.

3.3 *In vivo* Data

Fig. 5(a) shows the results of our missing data filling algorithm on the *in vivo* data, the feature point shown was not detected or tracked until the 9th frame, however our data filling method was able to accurately recover its position in frames 7 and 8.

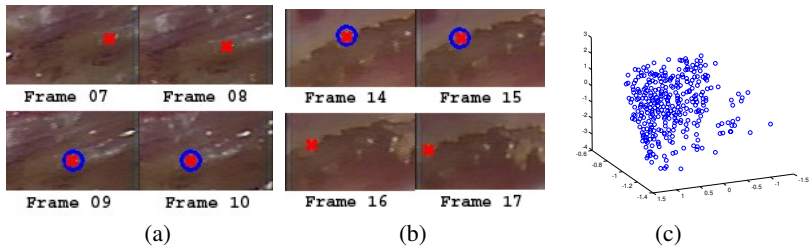


Fig. 5. Experimental results of *in vivo* data testing. (a) and (b) illustrate the data filling results of two features, respectively; (c) displays the reconstructed point cloud using the proposed method.

Unfortunately no ground truth 3D data was available for this patient. Instead we report the RMS surface reconstruction error. Before removing the outliers, the average RMS error for all the visible features was 3.977. This reduced to 1.535 after 65 outliers were detected and removed from the feature dataset. Finally the bundle adjustment technique was applied, after initialization using the output from the CEA method, and the error was further reduced to 1.040. Fig. 5 (c) shows the point cloud reconstructed from the endoscopic image sequence.

4 Discussion and Conclusions

In this paper we have introduced a new method to reconstruct a 3D surface of soft tissue structure with missing data for robotic assisted MIS surgery, which can provide a wider field-of-view of the operating area with 3D information. 3D structure of operating scenes is recovered from a moving stereoscopic vision system and a competitive evolutionary agent optimisation strategy is used to improve the robustness of the proposed technique to missing data and outliers. Our approach has been evaluated with both a heart phantom of known geometric structure and *in vivo* data from TECAB surgery. The results indicate the proposed method can handle both missing data and outliers, even when a large amount of missing data is involved in the estimation.

Efforts in the near future will focus on assessing reconstructed surface accuracy of the phantom heart with known gold standard position (for example, using some manually selected 3D points or attached fiducial markers). Our long term goal is to register the 3D information from optical images with the pre-operative data (CT or MRI) which will also be used as a ground truth for further *in vivo* testing of our method.

Acknowledgments. The authors would like to thank Dr. Fernando Bello for his help in obtaining the VisonRT data of the heart phantom. Special thanks also go to Tim Carter for data display and Dr. Gang Gao for ICP registration.

References

1. Seshamani, S., Lau, W., Hager, G.: Real-Time Endoscopic Mosaicking. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4190, pp. 355–363. Springer, Heidelberg (2006)
2. Devernay, F., Mourgues, F., Coste-Maniere, E.: Towards Endoscopic Augmented Reality for Robotically Assisted Minimally Invasive Cardiac Surgery. In: Proc. International Workshop on Medical Imaging and Augmented Reality, pp. 16–20 (2001)
3. Stoyanov, D., Darzi, F., Yang, A., Dense, G.-Z.: 3D Depth Recovery for Soft Tissue Deformation During Robotically Assisted Laparoscopic Surgery. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) MICCAI 2004. LNCS, vol. 3217, pp. 41–48. Springer, Heidelberg (2004)
4. Lau, W., Ramey, N., Corso, J., Thakor, N., Hager, G.: Stereo-Based Endoscopic Tracking of Cardiac Surface Deformation. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) MICCAI 2004. LNCS, vol. 3217, pp. 494–501. Springer, Heidelberg (2004)
5. Burschka, D., Li, M., Taylor, R., Hager, G.: Scale-Invariant Registration of Monocular Endoscopic Images to CT-Scans for Sinus Surgery. In: Proc. MICCAI. In: Barillot, C., Haynor, D.R., Hellier, P. (eds.) MICCAI 2004. LNCS, vol. 3217, pp. 413–421. Springer, Heidelberg (2004)
6. Mountney, P., Stoyanov, D., Davison, A., Yang, G.-Z.: Simultaneous Stereoscope Location and Soft-Tissue Mapping for Minimal Invasive Surgery. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4190, pp. 347–354. Springer, Heidelberg (2006)

7. Tomasi, C., Kanade, T.: Shape and Motion from Image Streams under Orthography: a Factorization Method. *Int. J. Computer Vision* 9(2), 137–154 (1992)
8. Hu, M.X., McMenemy, K., Ferguson, S., Dodds, G., Yuan, B.Z.: Epipolar geometry estimation based on evolutionary agents, *Pattern Recognition* (to be published)
9. Lucas, B., Kanade, T.: An Iterative Image Registration Technique with an Application to Stereo Vision. In: *Proc. IJCAI*, pp. 674–679 (1981)
10. Martinec, D., Pajdla, T.: Structure from Many Perspective Images with Occlusions. In: Heyden, A., Sparr, G., Nielsen, M., Johansen, P. (eds.) *ECCV 2002*. LNCS, vol. 2351, pp. 355–369. Springer, Heidelberg (2002)
11. Sturm, P., Triggs, B.: A Factorization based Algorithm for Multi-Image Projective Structure and Motion. In: Buxton, B.F., Cipolla, R. (eds.) *ECCV 1996*. LNCS, vol. 1065, pp. 709–720. Springer, Heidelberg (1996)
12. Jacobs, W.: Linear Fitting with Missing Data for Structure-from-Motion. *Computer Vision and Image Understanding* 82(1), 57–81 (2001)
13. Triggs, B., McLauchlan, P., Hartley, R., Fitzgibbon, A.: *Bundle Adjustment: A Modern Synthesis*. *Vision Algorithm: Theory and Practice*, 298–375 (2000)
14. Besl, P., McKay, N.: A Method for Registration of 3-D Shapes. *IEEE Trans. on PAMI* 14(2), 239–256 (1992)