

Dense 3D Depth Recovery for Soft Tissue Deformation During Robotically Assisted Laparoscopic Surgery

Danail Stoyanov¹, Ara Darzi², and Guang Zhong Yang^{1,2}

¹ Royal Society/Wolfson Foundation Medical Image Computing Laboratory,
²Department of Surgical Oncology and Technology
Imperial College of Science, Technology and Medicine, London SW7 2BZ, UK
{d.v.stoyanov,a.darzi,g.z.yang}@imperial.co.uk
<http://vip.doc.imperial.ac.uk>

Abstract. Recovering tissue deformation during robotic assisted minimally invasive surgery is an important step towards motion compensation and stabilization. This paper presents a practical strategy for dense 3D depth recovery and temporal motion tracking for deformable surfaces. The method combines image rectification with constrained disparity registration for reliable depth estimation. The accuracy and practical value of the technique is validated with a tissue phantom with known 3D geometry and motion characteristics. It has been shown that the performance of the proposed approach compares favorably against existing methods. Example results of the technique applied to *in vivo* robotic assisted minimally invasive surgery data are also provided.

1 Introduction

With recent advances in robotic assisted Minimally Invasive Surgery (MIS), it is now possible to perform closed-chest cardiothoracic surgery on a beating heart to minimize patient trauma and certain side effects of cardiopulmonary bypass. For robotic assisted MIS, 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. So far, two commercially available master-slave manipulator devices are specifically designed for MIS cardiac surgery [1]. Both systems improve the ergonomics of laparoscopic surgery and provide high dexterity, precision and 3D visualization of the operating field. One of the significant challenges of beating heart surgery is the destabilization introduced by cardiac and respiratory motion, thus severely affecting precise instrument-tissue interactions and the execution of complex grafts. Mechanical stabilizers [2] permit off-pump procedures by locally stabilizing the target area while the rest of the heart supports blood circulation. Despite this, residual motion remains, which complicates delicate tasks such as small vessel anastomosis.

Thus far, a number of techniques have been proposed for resolving intraoperative tissue deformation. Intraoperative 3D tomographic techniques offer precise information about soft tissue morphology and structure, but they introduce significant challenges to instrument design, integration and computational cost. A more practical al-

ternative is to use optical based techniques to infer surface deformation in real-time. In animal experiments, Nakamura *et al* [3] used a high-speed camera to track a fiducial marker on the epicardial surface. The trajectory changes of the markers were used to identify the frequencies due to cardiac and respiration motion by using an autoregressive model. A region based, reduced affine tracking model was used by Gröger *et al* [4] in robotic assisted MIS heart surgery for computing the local motion of the epicardial surface. Thakral *et al* [2] used a fiber optic displacement sensor to measure the motion of a rat’s chest for motion modeling with weighted time series. While these techniques demonstrate the feasibility of providing motion compensation, they generally do not consider detailed 3D deformation. With the use of a stereoscopic laparoscope for robotic assisted MIS, the feasibility of recovering the 3D structure of the operating field based on computer vision techniques has also been investigated [5]. Previously, monocular shading was used to infer surface shape in less interactive endoscope diagnostic procedures [6]. Although the recovery of the depth of a 3D scene based on different visual cues is one of the classic problems of computer vision, dense disparity measurement for deformable structure with high specularly is a difficult task. The purpose of this paper is to present a robust dense 3D depth recovery method with a stereoscopic laparoscope for motion stabilization. The method combines image rectification with constrained disparity registration for reliable depth recovery. The accuracy and practical value of the technique is validated with a tissue phantom with known 3D geometry and motion characteristics. Example results of the technique applied to *in vivo* robotic assisted MIS data are also provided.

2 Methods

2.1 Stereo Camera Model and Calibration

One of the first steps towards depth recovery is to compute both the intrinsic and extrinsic camera parameters of the stereoscopic laparoscope. In this study, the standard pinhole model is assumed and an upper triangular matrix \mathbf{K} is used to describe the internal optics of the camera. Denoting the camera’s position and orientation with respect to a world coordinate system by a rotation matrix \mathbf{R} and translation vector \mathbf{t} , the camera matrix can thus be defined as:

$$\mathbf{P}^k = \mathbf{K}^k \left[\mathbf{R}^k \mid -\mathbf{R}^k \mathbf{t}^k \right] \quad (1)$$

Without loss of generality, the camera matrices for the stereoscopic laparoscope can be represented by the following equation by taking the left camera as the reference:

$$\mathbf{P}^L = \mathbf{K}^L \left[\mathbf{I} \mid \mathbf{0} \right] \text{ and } \mathbf{P}^R = \mathbf{K}^R \left[\mathbf{R} \mid -\mathbf{R}\mathbf{t} \right] \quad (2)$$

In practice, laparoscope cameras deviate from ideal perspective projection and induce a high level of distortion. We consider henceforth the first three terms of the radial distortion, k_1^k, k_2^k, k_3^k and two tangential distortion terms, p_1^k and p_2^k [7].

For MIS, the stereo cameras are usually pre-calibrated before the surgical procedure and then remain unchanged during the operation. Off-line calibration by using objects with known geometry is therefore sufficient [8]. In this study, the intrinsic and extrinsic parameters of the cameras were derived by using a closed form solution as proposed in [9]. Following the initial estimate, the parameters were refined subject to the mean squared error between the measured image points \mathbf{m} and the re-projected world points \mathbf{M} . By parameterizing the rotation matrix \mathbf{R} as a vector of three parameters \mathbf{r} , the minimization criteria for a set of n images with m grid points can be written as:

$$\sum_{i=1}^n \sum_{j=1}^m \left\| \mathbf{m}_{ij} - \mathbf{m}(\mathbf{K}, \mathbf{r}_i, \mathbf{t}_i, k_1, k_2, k_3, k_1, p_2, \mathbf{M}_j) \right\|^2 \quad (3)$$

The optimization problem formulated above is non-linear and the Levenberg-Marquardt algorithm was used to derive the above parameters iteratively. After each camera has been calibrated, the relative pose of the two cameras is then introduced such that the following equation is minimized. This allows the use of the solution derived for each individual camera as the initial solution for the Levenberg-Marquardt algorithm.

$$\sum_{i=1}^n \sum_{j=1}^m \left[\left\| \mathbf{m}_{ij}^L - \mathbf{m}(\mathbf{K}^L, \mathbf{r}_i^L, \mathbf{t}_i^L, k_1^L, k_2^L, k_3^L, p_1^L, p_2^L, \mathbf{M}_j) \right\|^2 + \left\| \mathbf{m}_{ij}^R - \mathbf{m}(\mathbf{K}^R, \mathbf{r}_i^R, \mathbf{t}_i^R, k_1^R, k_2^R, k_3^R, p_1^R, p_2^R, \mathbf{M}_j, \mathbf{r}, \mathbf{t}) \right\|^2 \right] \quad (4)$$

2.2 Image Rectification

For reliable dense depth recovery, image rectification based on epipolar geometry is an important step for enhancing the robustness of the algorithm as this effectively restricts the search space for disparity to 1D. For common robotic MIS settings, the cameras are slightly verged to permit both positive and negative disparities so as to enhance the overall 3D depth perception. To ease the fusion of the stereo images for the observer, the stereo cameras are generally in near vertical alignment. However, this arrangement may not be perfect in practice and therefore a planar rectification process is applied to the stereo image pairs before dense correspondence is sought [10]. By definition, the intrinsic matrices of the two rectified images must be the same. Without changing the centers of the cameras, the new projection matrices can be defined through the same rotation matrix such that

$$\mathbf{P}_r^L = \mathbf{K}_r [\mathbf{R}_r | \mathbf{0}] \text{ and } \mathbf{P}_r^R = \mathbf{K}_r [\mathbf{R}_r | -\mathbf{R}_r \mathbf{t}] \quad (5)$$

In the above equation, \mathbf{R}_r may be computed by assuming that the new image planes are parallel to the baseline. As the camera centers remain unchanged so does the optical ray through each image point, the original and rectified camera matrices can therefore be written as:

$$\mathbf{P}^k = [\mathbf{Q}^k | \mathbf{q}^k] \text{ and } \mathbf{P}_r^k = [\mathbf{Q}_r^k | \mathbf{q}_r^k] \quad (6)$$

Subsequently, the rectifying transformations can be computed from the original and rectified camera matrices through the following pair of equations:

$$\mathbf{T}^L = \lambda^L \mathbf{Q}_r^L (\mathbf{Q}^L)^{-1} \text{ and } \mathbf{T}^R = \lambda^R \mathbf{Q}_r^R (\mathbf{Q}^R)^{-1} \quad (7)$$

Although this method does not directly minimize the distortion or resampling effects caused by the transformations [11], in the context of the current work the warping introduced above is inherently small due to the general settings of the stereoscopic laparoscope cameras.

2.3 Stereo Correspondence with Constrained Disparity Registration

Traditional computer vision techniques for dense stereo correspondence are mainly concerned with rigid objects and much emphasis is placed on issues related to occlusion and discontinuity [12]. Occlusion and object boundaries make stereo matching a difficult optimization problem, as disparity is not globally continuous and smooth. Existing techniques include winner-takes-all, graph-cuts, and dynamic programming approaches [13]. For soft tissue as observed in MIS, the surface is generally smooth and continuous and the difficulty of dense depth recovery is usually due to the paucity of identifiable landmarks. Explicit geometrical constraints of the deformation model is therefore required for ensuring the overall reliability of the algorithm. For this study, the free-form registration framework proposed by Veseer *et al* [14] was used as it provides a robust, fully encapsulated multi-resolution approach based on piece wise bilinear maps (PBM). The lattice of PBM permits non-linear transitions, which is suitable for temporally deforming surfaces and it easily lends itself to a hierarchical implementation. With image rectification, the search space for each iteration is constrained on scan lines and the number of PBM forming the image transformation is increased, refining the registration of finer structures. Within this framework, the disparity obtained at low-resolution levels are propagated to higher levels and used as starting points for the optimization process. To cater for surfaces in laparoscope images that have reflectance properties dependent on the viewing position, normalized cross correlation (NCC) was used as a similarity measure. The NCC of two image regions I^L and I^R of dimensions (u, v) is defined as:

$$NCC(I^L, I^R) = \frac{\sum_{u,v} (I^L(u, v) - \bar{I}^L)(I^R(u, v) - \bar{I}^R)}{(uv)^2 \sqrt{\sum_{u,v} (I^L(u, v) - \bar{I}^L)^2 (I^R(u, v) - \bar{I}^R)^2}} \quad (8)$$

For deriving disparity values of the soft tissue, the gradient of the given metric can be computed directly which permits the use of fast optimization algorithms. For this study, the Broyden-Fletcher-Goldberg-Shano (BFGS) method can be used. This is a quasi-Newton technique, which uses an estimate of the Hessian to speed up the iterative process [15].

2.4 Experimental Design and Validation

To model the real stereoscopic laparoscope, a stereo camera rig was built by using a pair of miniature NTSC cameras. Each camera has a physical diameter of just over 5mm and therefore it is possible to setup a configuration with a small baseline of just over 5mm. The described calibration procedure was employed by using a 5×7 square grid with a checked black and white pattern. Corners were detected through a semi-automated procedure, where the user indicates guidance positions and sub-pixel refinement is performed automatically. The pixel re-projection error after calibration was measured at less than half a pixel. We also measured the reconstruction error after triangulation at an average magnitude of 1mm. This can be improved by taking into account errors in the measured coordinates and refining the measurements [16].

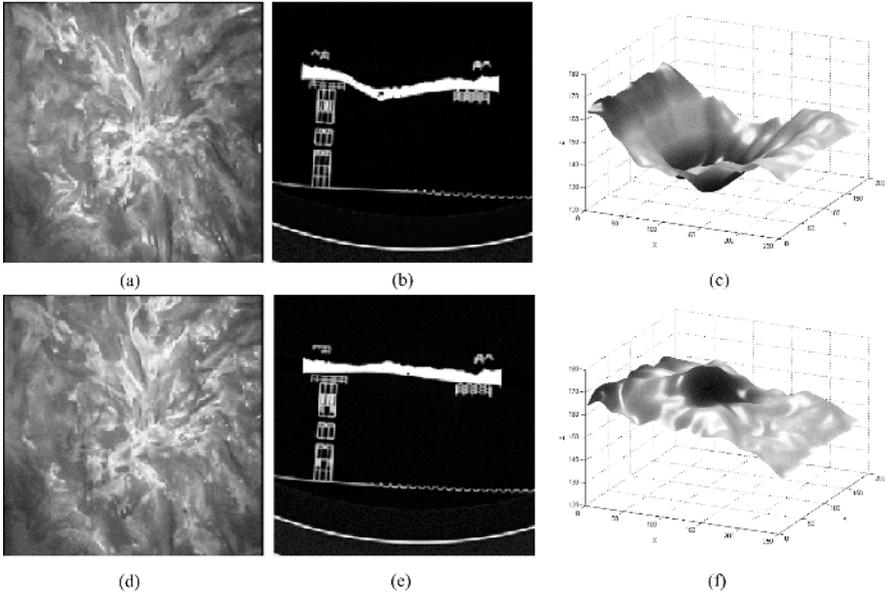


Fig. 1. A stereo image pair of the tissue phantom captured by the stereo rig. (b, e) CT cross sections of the phantom corresponding to two different phases of the deformation and their respective 3D surface representation (c, f).

In order to assess the accuracy of the proposed algorithm, a tissue phantom made of silicone rubber and painted with acrylic was constructed. The surface was coated with silicone rubber mixed with acrylic to give it a specular finish that looks similar to wet tissue. The tomographic model of the phantom was scanned with a Siemens Somatom Volume Zoom four-channel multi-detector CT scanner with a slice thickness of 0.5 mm and in-plane resolution of 1mm. To allow the evaluation of temporal surface deformation, the model was scanned at four discrete and reproducible deformation levels. Fig. 1 illustrates a pair of images captured by the stereo cameras and cross sectional images of two different time frames of the tissue phantom captured by CT scanning. The corresponding 3D surface plots are shown in Figs. 1 (c) and (f), respectively.

3 Results

Fig. 2 represents the reconstructed surfaces at four different levels of deformation as captured by CT and the proposed algorithm for dense 3D depth recovery. Figs. 2(c)-(e) demonstrate the regression of relative depth change over time as extracted by the two techniques. It is evident that the overall quality of the stereo reconstruction is good, but the scatter plots also show a certain level of deviation. This was largely due to the specular highlights, which were not explicitly modeled in the proposed method.

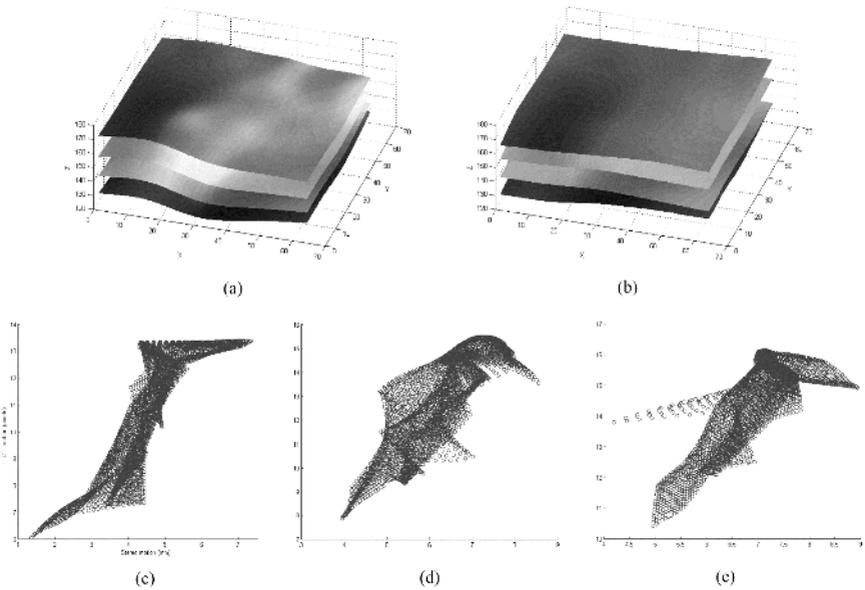


Fig. 2. The reconstructed 3D surface for four different levels of deformation as captured by 3D CT (a) and the proposed depth recovery method based on combined image rectification and constrained disparity registration (b). Scatter plots (c)-(e) illustrate the correlation of the recovered depth change between different levels of deformation by the two techniques.

To demonstrate the potential clinical value of the proposed technique, Fig. 3 illustrates three of the reconstructed depth maps from an *in vivo* stereoscopic laparoscope sequence. Both the depth maps and their associated 3D renditions illustrate the quality of the reconstruction technique. However, it is also evident that specular highlight represent a major problem to the proposed algorithm, as evident from the reconstruction errors indicated by the arrows.

In this study, we also compared the relative performance of the proposed method against existing depth recovery techniques. Fig. 4 shows a comparison of the results obtained from several popular stereo algorithms on an image pair of the phantom model. It is clear that due to the lack of texture these techniques perform poorly in comparison to the proposed technique.

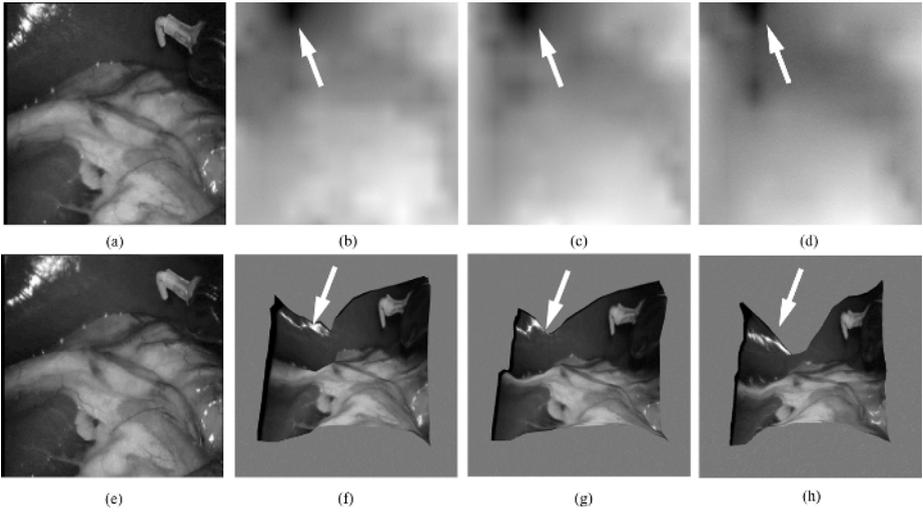


Fig. 3. A pair of stereo images (a, e) from an *in vivo* stereoscopic laparoscope sequence, and three temporal frames of the reconstructed depth map (b-d) and their corresponding 3D rendering results (f-h).

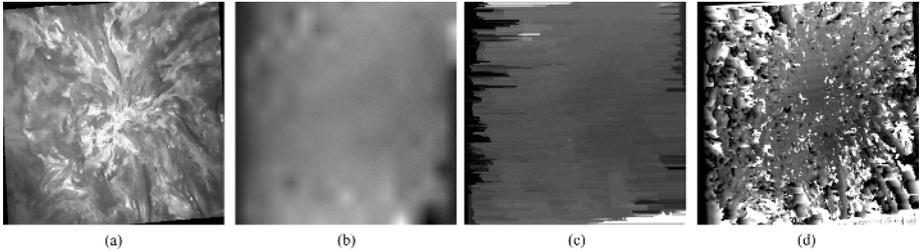


Fig. 4. Comparison of disparity results with popular stereo algorithms (a) rectified left image of a stereo pair of the phantom model (b) the proposed algorithm (c) dynamic programming [17] (d) SSD [13]

4 Discussion and Conclusions

In conclusion, we have developed a practical strategy for dense 3D structure recovery and temporal motion tracking for deformable surfaces. The purpose of the study is to capture real-time surface deformation during robotic assisted MIS procedures such that effective motion stabilization can be deployed. The method uses image rectification to simplify the subsequent free-form disparity registration procedure. Both phantom validation and *in vivo* results demonstrate the potential clinical value of the technique. It has been shown that the performance of the proposed method compares favorably against existing techniques, but the result also indicates the importance of handling specular highlights before depth reconstruction. For robotic assisted MIS procedures, it is possible to exploit the restricted lighting configuration imposed by the laparoscope to filter out these artifacts.

References

1. Ballantyne, G.: Robotic surgery, telerobotic surgery, telepresence, and telementoring. *Surgical Endoscopy*. 2 Springer (2002) 1389-402.
2. Thakral, A., Wallace, J., Tomlin, D., Seth, N., Thakor, N.: Surgical motion adaptive robotic technology (S.M.A.R.T): taking the motion out of physiological motion. In: Proc. of MICCAI. Volume 2208 of LNCS Springer (2001) 317-325.
3. Nakamura, Y., Kishi, K., Kawakami, H.: Heartbeat synchronization for robotic cardiac surgery. In: Proceedings of the 2001 IEEE International Conference on Robotics and Automation. (2001) 2014-2019.
4. Gröger, M., Ortmaier, T., Sepp, W., Hirzinger, G.: Tracking local motion on the beating heart. In: Proc. of SPIE Medical Imaging Conference. Volume 4681. (2002) 233-241
5. Mourgues, F., Devernay, F., Coste- Manière, E.: 3D reconstruction of the operating field for image overlay in 3D-endoscopic surgery. In: Proceedings of International Symposium on Augmented Reality. (2001).
6. Okatani, T., Deguchi, K.: Shape reconstruction from an endoscope image by shape from shading technique for a point light source at the projection centre. *Computer Vision and Image Understanding*. 66 (1997) 119-131.
7. http://www.vision.caltech.edu/bouguetj/calib_doc/
8. Morgues, F., Coste-Manière, È.: Flexible calibration of actuated stereoscopic endoscope for overlay in robot assisted surgery. In: Proceedings of MICCAI. Volume 2488 of LNCS. Springer (2002) 24-34.
9. Zhang, Z.: A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 22 (2000) 1330-1334.
10. Fusiello, A., Trucco, E., Verri, A.: A compact algorithm for rectification of stereo pairs. *Machine Vision and Applications*. 12 (2000) 16-22.
11. Loop, C., Zhang, Z.: Computing rectifying homographies for stereo vision. In: Proceedings of Computer Vision and Pattern Recognition. (1999) 125–131.
12. Brown, M., Burschka, D., Hager, G.: Advances in computational stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 25 (2003) 993-1008.
13. Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *The International Journal of Computer Vision*. 47 (2002) 7-42.
14. Veese, S., Dunn, M., Yang, G.-Z.: Multiresolution image registration for two-dimensional gel electrophoresis. *Proteomics*. 1 (2001) 856-870.
15. Nocedal, J., Wright, S.: Numerical optimization. Springer Verlag. (1999).
16. Hartley, R., Zisserman, A.: Multiple view geometry in computer vision. Cambridge University Press. (2000).
17. Birchfield, S., Tomasi, C.: Depth discontinuities by pixel-to-pixel stereo. In: Proceedings of The 6th International Conference on Computer Vision. IEEE Computer Society, IEEE Computer Society Press. (1998) 1073–1080.