

Belief Propagation for Depth Cue Fusion in Minimally Invasive Surgery

Benny Lo¹, Marco Visentini Scarzanella², Danail Stoyanov¹,
and Guang-Zhong Yang^{1,2}

¹ Institute of Biomedical Engineering

² Royal Society/Wolfson Foundation MIC Laboratory,
Imperial College London, United Kingdom

{benny.lo, marco.visentini-scarzanella02,
danail.stoyanov, g.z.yang}@imperial.ac.uk

Abstract. In minimally invasive surgery, dense 3D surface reconstruction is important for surgical navigation and integrating pre- and intra-operative data. Despite recent developments in 3D tissue deformation techniques, their general applicability is limited by specific constraints and underlying assumptions. The need for accurate and robust tissue deformation recovery has motivated research into fusing multiple visual cues for depth recovery. In this paper, a Markov Random Field (MRF) based Bayesian belief propagation framework has been proposed for the fusion of different depth cues. By using the underlying MRF structure to ensure spatial continuity in an image, the proposed method offers the possibility of inferring surface depth by fusing the posterior node probabilities in a node's Markov blanket together with the monocular and stereo depth maps. Detailed phantom validation and in vivo results are provided to demonstrate the accuracy, robustness, and practical value of the technique.

Keywords: Robotic Assisted Surgery, Image Guided Intervention, Intra-operative Navigation, Markov Random Fields, Belief Networks.

1 Introduction

In image guided Minimally Invasive Surgery (MIS), accurate 3D deformation recovery is essential for intra-operative image registration, motion stabilisation and prescribing dynamic active constraints. Despite significant research effort in this area, current progress towards reliable, real-time, intra-operative 3D tissue deformation recovery remains limited. This is largely due to the complexity of the MIS field-of-view, in which inter-lumen reflection, specular highlights, coupled with the general lack of stable distinctive visual features, constitute significant challenges [1]. In such situations, imposing strong geometrical constraints, as often used in early work of tissue deformation recovery, is not appropriate.

The need for accurate and robust tissue deformation recovery has motivated research into fusing multiple visual cues for depth recovery. In computer vision, the value of this approach has long been recognised for compensating for the weaknesses

of individual cues and to better reflect the nature of the human vision system. Currently, integrated depth recovery methods are based on a range of different techniques including probabilistic approaches using Bayesian formulations [2], game theory [3], and calculus of variations [4]. These techniques use either explicit or implicit data exchange and the chosen combination of different visual cues is determined by the suitability of each reconstruction method for separate regions.

Thus far, many of the methods developed are application specific and limited work has been directed towards deformable object reconstruction, which remains a challenging task in computer vision. In MIS, both monocular and stereo based depth recovery techniques have been proposed [5, 6]. Strong regularisation terms tend to be used, which limit the ability of the existing techniques in dealing with large tissue deformation and discontinuities. The purpose of this paper is to propose a more general Bayesian fusion framework for integrating multiple visual cues based on Markov Random Fields (MRFs). Although a Bayesian MRF-based solution for dense 3D depth reconstruction has been attempted for natural scenes [7], its use in MIS tissue deformation recovery is faced with largely textureless views, extensive changes in scene context, illumination and tissue surface appearance, as well as the difficulty of obtaining sufficiently large sets of ground truth training data.

In this paper, the depth cues employed include feature based stereo laparoscopic correspondence and surface shading based depth reconstruction techniques. A MRF based Bayesian Network (BN) has been proposed for the fusion framework. By using the underlying MRF structure to ensure spatial continuity in an image, the MRF-BN approach offers the possibility of inferring the surface depth by fusing the posterior node probabilities in a node's Markov blanket together with the shading and stereo information. In addition, a belief propagation scheme is designed to utilise the evidence provided by the sparse stereo points to infer the depth of the surrounding surface patches. Detailed phantom validation and *in vivo* results are provided to demonstrate the accuracy, robustness, and practical value of the technique.

2 Method

2.1 Stereoscopic Depth Recovery

Stereo methods for 3D reconstruction exploit the geometrical relationship between camera views in order to triangulate the 3D position of matching points from the two 2D views. In this paper, we follow the calibration procedure presented in [8] and the features used for stereo-correspondence include gradient-based features selected by the Shi and Tomasi detector [9], and Maximally Stable Extremal Regions (MSER) [10]. The latter exhibits desirable properties of invariance to monotonic illumination changes as well as affine transformations. The matching of the detected features is carried out according to the measure proposed in by Pilu in [11], *i.e.*,

$$\mathbf{C}(m, n) = e^{-\frac{(C_{mm-1})^2}{2\mu^2}} e^{-\frac{r_{mm}^2}{2\nu^2}} \quad (1)$$

where C_{mn} is the normalised cross-correlation (NCC) between features m, n from the two stereo channels, r_{mn} is their Euclidean distance and μ, ν are control parameters determining the influence of feature similarity and proximity respectively.

While this is ideal for natural scenes involving large quantities of highly distinctive high frequency features such as edges and corners, MIS data is generally lack of distinctive geometric features. This would result in obtaining only a semi-dense 3D representation of the target surface. The presence of highly saturated specular highlights commonly found in MIS images can also lead to errors in stereo calculations. For these reasons, additional depth cues must be employed.

2.2 Monocular Depth Recovery

To estimate the depth of homogenous regions of the tissue, monocular depth cues based on the surface shading information is used. As the specular reflection model cannot be measured from the laparoscopic scene, a near Lambertian reflection model is assumed. The technique used consists of two steps, where the depth of the surface is first estimated by using a minimisation technique followed by B-spline interpolation to approximate the depth in regions with specular highlights. To make use of the unique optical arrangement of the laparoscopic camera with coincidental axes of the cameras and light source, the general form of the depth estimation equations based on the shading information can be simplified as [12]:

$$Z(x, y) = \frac{I(x, y) + \beta^2 \nabla^2 Z(x, y)}{g(x, y)} \quad (2)$$

where $\nabla^2 Z(x, y)$ represents the Laplacian of the function $Z(x, y)$, β is a weighting factor determining the influence of the smoothness constraint employed and $g(x, y)$ is a normalising function needed because of the unevenness of the light distribution due to the proximity and point-like nature of the light source.

To minimise the impact of specular highlights, a specular highlight detector is used to identify the potentially problematic areas. In this paper, specular highlights are detected by assessing the intensity and chromaticity information of the tissue surface. This is achieved by simply thresholding based on the following equations:

$$S(x, y) = \begin{cases} 1 & \text{if } I(x, y) > \gamma \text{ and } Sat(x, y) > \kappa \\ 0, & \text{otherwise} \end{cases} \quad (3)$$

where $I(x, y)$ is the pixel intensity and $Sat(x, y)$ represents the pixel colour saturation defined as $Sat(x, y) = 1 - 3 \min(r, g, b) / I(x, y)$; γ, κ are empirically determined, and are here set as 160 and 60 respectively. The detected regions are then subjected to cubic B-spline interpolation.

2.3 MRF-BN Belief Propagation and Depth Fusion

The core of the proposed fusion technique lies in the MRF representation of the dense depth map as shown in Fig. 1. In this figure, each node $x_{i,j}$ represents the depth $d_{i,j}$ as a random variable dependent on the depth of its neighbouring nodes $x_{N_{i,j}}$, the depth obtained with monocular shading cue, $SFS_{i,j}$ and stereo correspondence, $SS_{i,j}$, if present.

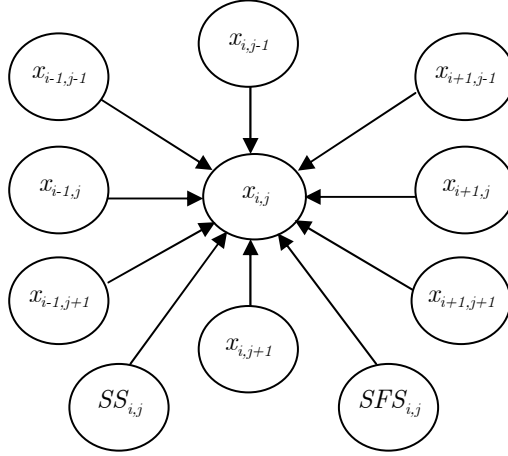


Fig. 1. Local neighborhood structure of the proposed Markov random field

In Fig. 1, the neighbouring nodes $x_{N_{i,j}}$ are defined as the Markov blanket of node $x_{i,j}$ and can be expressed as: $N_{i,j} = (i + a, j + b)$, where $a = -1, \dots, 1$, $b = -1, \dots, 1$, and $(a, b) \neq (0, 0)$. Given this Bayesian network, the belief of depth $d_{i,j}$ can be then defined as:

$$BEL(x_{i,j}) \triangleq P(x_{i,j} \mid x_{N_{i,j}}, SS_{i,j}, SFS_{i,j}) \quad (4)$$

Since the measurements obtained with the two techniques are inherently independent, the expression above can be simplified as:

$$\begin{aligned} BEL(x_{i,j}) &= \alpha P(x_{i,j}) P(x_{N_{i,j}} \mid x_{i,j}) P(SS_{i,j} \mid x_{i,j}) P(SFS_{i,j} \mid x_{i,j}) \\ &= \alpha \pi(x_{i,j}) \lambda(x_{i,j}) \end{aligned} \quad (5)$$

where α is a normalising constant such that the integral of the probability distribution function adds up to one, $\pi(x_{i,j})$ represents the prior probability of depth at node $x_{i,j}$ and $\lambda(x_{i,j})$ its likelihood from the evidence provided by the sparse stereo and photometric depth recovery.

The scheme presented shows how nodes are directly dependent on the posterior probability of neighbouring nodes, and its approximate solution can be found through a belief propagation algorithm. Since not all nodes are supplied with stereo depth estimates due to the semi-dense output of the technique, it makes sense to start the propagation seeds at the nodes with stereo depth information for the first iteration of the propagation algorithm. The posterior probability at the nodes with available stereo information can be initially calculated and propagated outwards by:

$$BEL(x_{i,j}) = \alpha P(x_{i,j})P(SS_{i,j} | x_{i,j})P(SFS_{i,j} | x_{i,j}) \quad (6)$$

neighbouring nodes can then rely on this information to calculate their belief:

$$BEL(x_{i+a,j+b}) = \alpha P(x_{i+a,j+b})P(x_{N_{i+a,j+b}} | x_{i+a,j+b})P(SFS_{i+a,j+b} | x_{i+a,j+b}) \quad (7)$$

With this scheme, it can be observed how the independence between processes resulting from the weak fusion structure adopted. It allows for a great degree of flexibility as additional modules providing depth information obtained using different techniques can be readily integrated without any substantial design modification.

3 Experimental Setup and Results

3.1 Phantom Reconstruction Results

In order to demonstrate the performance of the proposed algorithm, both phantom and *in vivo* data from a daVinci surgical robot is used. The first sequence consists of a beating phantom heart and in order to provide ground truth data to evaluate the accuracy of the algorithm, fiducial markers are placed on the epicardial surface with their actual 3D position being tracked with a NDI Optotrak Certus tracker. Multiple views of the output 3D reconstruction of the shading module alone compared with the output from the proposed algorithm are shown below in Fig. 2 (a) and (b) respectively, whereas numerical results from the evaluation of the algorithm with the ground truth data from the Optotrak sensor are shown in Fig. 3. Both sequences are normalised and low-pass filtered to remove high frequency noise and facilitate the comparison.

At a first glance, the output of the shading algorithm in Fig. 2 (a) may look smoother and more regular than the output from the proposed algorithm in Fig. 2 (b). However, a closer inspection reveals that the proposed system yields results closer to the actual object structure: the irregularities present on the reconstructed surface in Fig. 2 (b) can indeed be found on the epicardial surface of the model in the form of artificial adipose deposits. On the other hand, the shading based depth cue is unable to correctly interpret these local irregularities because of its smoothness constraint and a lack of high-frequency information. This information is provided instead from the stereo module.

The results in Fig. 3 show the correctness of the proposed algorithm when applied to tracking the problems, where Fig. 3 (a) depicts the actual and reconstructed depth

of one vertex of the triangular reference marker against time for the beating phantom heart. It can be observed that the phase and frequency information of the reconstructed signal do correspond to the ground truth data sequence, with a frequency of approximately 1.22Hz as shown in Fig. 3 (b). It should be noted that the reconstructed data is correct up to a scale factor, and this is reflected in the slope of the linear regression shown in Fig. 3 (c), where the accuracy of linear relationship is evident from the relatively tight regression ($\sigma = \pm 0.114$) of the plot.

3.2 *In vivo* 3D Surface Reconstruction

The suitability of the proposed algorithm for actual surgical procedures is qualitatively evaluated by applying it to an *in vivo* video sequence of a Totally Endoscopic Coronary Artery Bypass (TECAB) procedure. Fig. 4 highlights the points tracked during the procedure. As ground truth data is not available for the TECAB procedure, Fig. 5 only illustrates the reconstructed 3D position of the tracked points for qualitative analysis.

The reconstructed result shown in Fig. 4 presents a relatively smooth heart surface and a depression corresponding to the point of contact with the instrument. Also, the instrument is correctly reconstructed as a separate volume from the heart. In Fig. 5, the reconstructed motion of the same tracked points is shown. The recovered motion appears to be more irregular than the results for the phantom model, due not only to the presence of multiple frequencies in the heart surface motion, but also to the interferences introduced by the pressure exerted on the surface by the instrument. Nevertheless, the results for each tracked point do show the presence of a compound

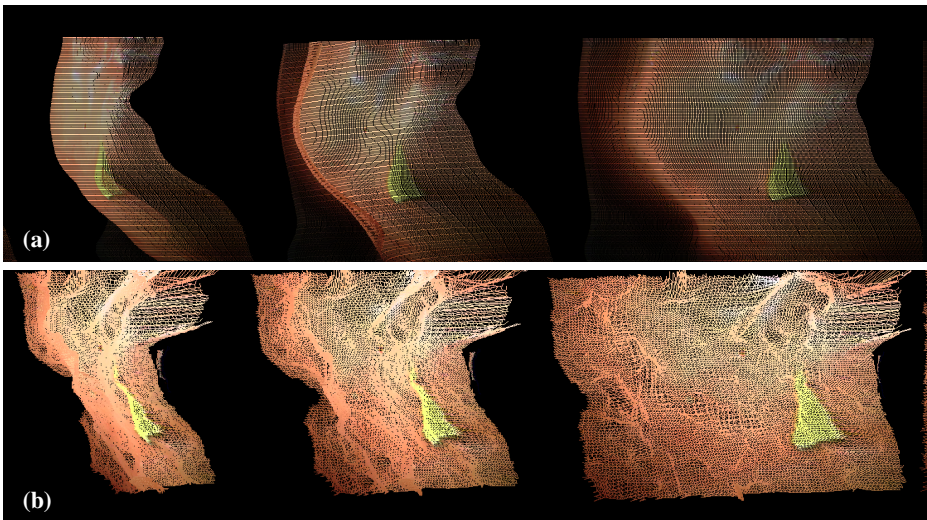


Fig. 2. Dense 3D surface reconstruction using (a) shape-from-shading and (b) the proposed MRF-based algorithm

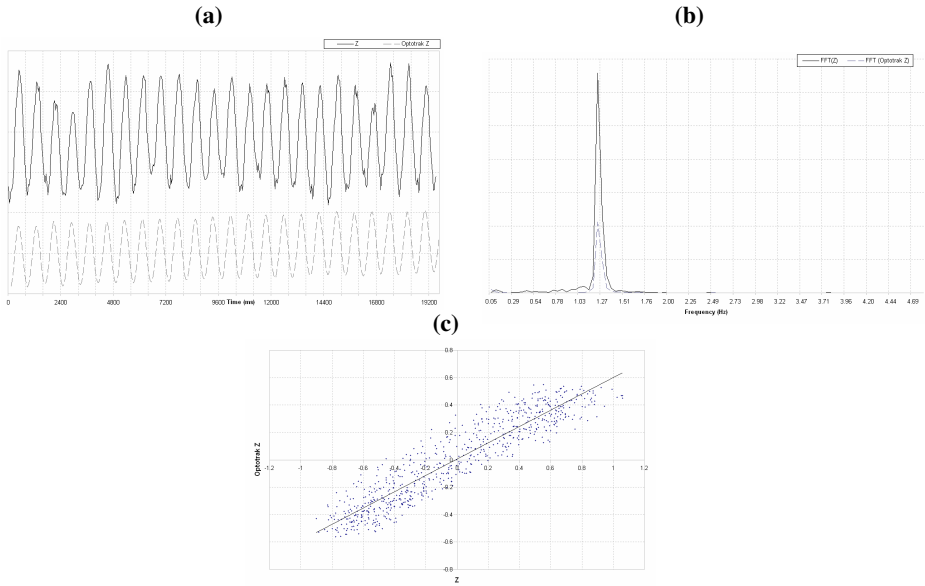


Fig. 3. (a) Time/position plot of the reconstructed motion (above) and ground truth data (below) for the tracked reference marker. (b) Frequency power spectrum of (a), highlighting the scale factor. (c) Linear regression results between the two datasets.

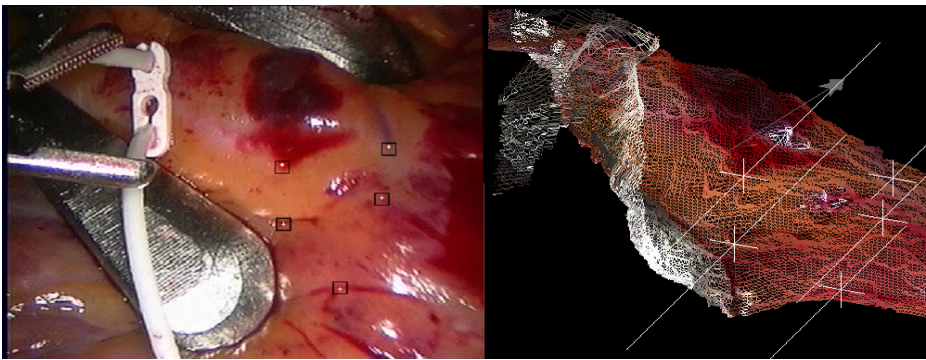


Fig. 4. Frame from the TECAB procedure and its corresponding 3D reconstruction with the five tracked points highlighted

of regular component frequencies in each point’s motion, as well as a constant reconstructed depth interval signifying constant bounds for the recovered 3D coordinates. The lack of objective ground truth data makes a quantitative evaluation of the proposed algorithm’s performance difficult; however, the observations on the results discussed above suggest a regular, constant performance of the algorithm through time, with a qualitatively accurate *in vivo* performance.

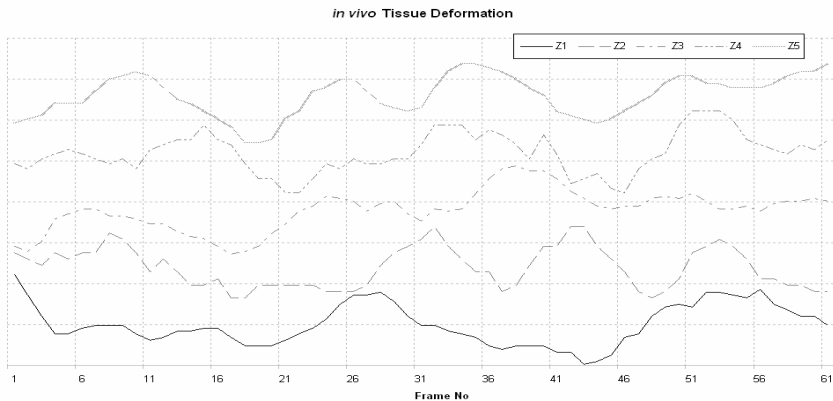


Fig. 5. Reconstructed motion of the five tracked points from Fig. 3

4 Discussion and Conclusions

In this paper, we have proposed a MRF-BN framework for dense 3D surface reconstruction in MIS. The validity of the proposed method has been demonstrated with applications on both phantom models and *in vivo* data. The results obtained illustrate the importance of fusion complementary depth cues for reliable tissue deformation recovery. Although the results presented only employ monocular and stereo depth maps, the proposed belief propagation framework can be readily extended to incorporate other depth measures. The strength of the proposed method is that it allows seamless integration of additional cues without significant modifications being required, thus greatly simplifies its practical application.

References

1. Delingette, H., et al.: Computational Models for Image-Guided Robot-Assisted and Simulated Medical Interventions. *Proceedings of the IEEE* 94(9), 1678–1688 (2006)
2. Pankanti, S., Jain, A.K.: Integrating vision modules: stereo, shading, grouping, and line labeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 17(9), 831–842 (1995)
3. Bozma, H.I., Duncan, J.S.: Integration of vision modules: a game-theoretic framework. In: *Proc. CVPR* (1991)
4. Shah, J., Pien, H.H., Gauch, J.M.: Recovery of surfaces with discontinuities by fusing shading and range data within a variational framework. *IEEE Transactions on Image Processing* 5(8), 1243–1251 (1996)
5. Deguchi, K., Okatani, T.: Shape reconstruction from an endoscope image by shape-from-shading technique for a point light source at the projection center. In: *Proceedings of the Workshop on Mathematical Methods in Biomedical Image Analysis* (1996)
6. Stoyanov, D., Darzi, A., Yang, G.Z.: Dense 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)

7. Saxena, A., Schulte, J., Ng, A.: Depth Estimation using Monocular and Stereo Cues. Proc. IJCAI (2007)
8. Stoyanov, D., Darzi, A., Yang, G.-Z.: Laparoscope Self-calibration for Robotic Assisted Minimally Invasive Surgery. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3750, pp. 114–121. Springer, Heidelberg (2005)
9. Shi, J., Tomasi, C.: Good Features to Track, Cornell University (1993)
10. Matas, J., et al.: Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing* 22(10), 761–767 (2004)
11. Pilu, M.: A direct method for stereo correspondence based on singular value decomposition. In: Proc. CVPR. IEEE Computer Society, Los Alamitos (1997)
12. Ikeuchi, K., Horn, B.K.P.: Numerical shape from shading and occluding boundaries. In: *Shape from shading*, pp. 245–299. MIT Press, Cambridge (1989)