Real-Time Stereo Reconstruction in Robotically Assisted Minimally Invasive Surgery

Abstract. The recovery of tissue structure and morphology during robotic assisted surgery is an important step towards accurate deployment of surgical guidance and control techniques in minimally invasive surgery. In this article, we present a novel stereo reconstruction algorithm that propagates disparity information around a set of candidate feature matches. This has the advantage of avoiding problems with specular highlights, occlusions from instruments and view dependent illumination bias. Furthermore, the algorithm can be used with any feature matching strategy allowing the propagation of depth in very disparate views. Validation is provided for a phantom model with known geometry and this data will be made available online in order to establish a structured validation scheme in the field. The practical value of the proposed method is further demonstrated by reconstructions on various *in vivo* images of robotic assisted procedures, which will also be made available to the community.

1 Introduction

In robotically assisted Minimally Invasive Surgery (MIS), recovering the underlying 3D structure of the operating field *in vivo* is important for registering pre-operative data to the surgical field-of-view for providing dynamic active constraints and motion compensation [1]. Tomographic intra-operative imaging modalities can potentially provide anatomically co-registered information about the 3D shape and morphology of the soft tissues but their deployment in operating theatres is a significant challenge [2]. Currently, the most practical method of recovering the 3D structure of the operating site *in situ* is through optical techniques using a stereo laparoscope. This information can subsequently be used to align multimodal information within a global reference 3D coordinate system and enhance robotic instrument control. However, the recovery of 3D geometry from stereo in real-time during robotic procedures is difficult due to tissue deformation, partial occlusion due to instrument movement, and specular inter-reflections.

The recovery of 3D information from stereo images is one of the classic problems in computer vision. Given a calibrated stereo rig, the task is to identify the unique correspondence of image primitives across the stereo image pair. Recent review articles provide a good summary of progress in the field [3, 4] and also establish a benchmarking framework with ground truth data [3]. However, these methods, as well as the data used, are not always well suited to MIS settings where the scene is complicated by the large disparity discontinuities and occlusions arising from surgical instruments. The presence of view-dependent reflectance characteristics and a lack of

fronto-parallel surfaces with unique colors and texture further complicate the issue. Thus far, for robotically assisted MIS, a number of stereoscopic techniques for recovering 3D shape and morphology have been proposed [5-9]. Many methods assume a geometric surface model of the tissue and use this constraint to track tissue 3D structure and morphology with a particular focus on cardiac procedures [7, 8, 10]. Such techniques have demonstrated the feasibility and potential of stereo vision in MIS but they are constrained by smooth surface parameterizations without adequate handling of instrument occlusions. Furthermore, there has been no extensive validation and comparative assessment of the existing methods.

In this study, we propose a technique for building a semi-dense reconstruction of the operating field in MIS that can operate in real-time. The method initially starts with a sparse 3D reconstruction based on feature matching across the stereo pair and subsequently propagates structure into neighboring image regions. The method is validated by experiments on phantom data with known ground truth. Qualitative validation is also provided for *in vivo* robotic MIS images. All the data used in this study will be available for access online to facilitate the community to establish a structured validation framework for stereo vision in robotic surgery.

2 Method

2.1 Feature Matching

The first step of the proposed method is to recover a sparse set of matches across the stereo-laparoscopic image pair using a feature based technique. The most commonly used image features are based on points where the image intensity gradient is high in both the vertical and horizontal directions [11]. Such salient points can be reliably matched across the stereo pair to recover a sparse set of 3D points in the surgical site as shown by Stoyanov *et al* [12]. The method works effectively with stereo-laparoscopic images because the vergence of the cameras creates a zero disparity region that assists convergence. Furthermore, the technique can be adapted to compensate for linear illumination changes and to incorporate the expected disparity as a starting solution and prior information about the 3D surface. Efficient implementations of this method have been reported to operate at very high frame-rates on GPU accelerated hardware [13].

It is important to note that for the proposed stereo propagation method, any feature matching strategy can be deployed. As the technique can be used to match images temporally and recover structure with a moving imaging device more complex and robust feature detectors and descriptors can be used to accommodate wider variations in perspective distortion and tissue deformation [14].

2.1 Structure Propagation

With a sparse set of 3D points established in the surgical field-of-view, it is possible to propagate 3D information to cover a semi-dense portion of operating field domain.

This is necessary since the sparse 3D structure is usually not sufficient to describe the tissue geometry in detail. During the first stage of the proposed propagation algorithm, all features correspondences are used as seed matches. They are sorted subject to the correlation score between their respective templates and stored using a priority queue structure. The algorithm then proceeds to propagate structure around the matches with highest correlation scores on a best-first basis by popping the priority queue as proposed by Lhuillier *et al* [15]. New matches are simply added to the queue as the algorithm iterates until no more matches can be popped.

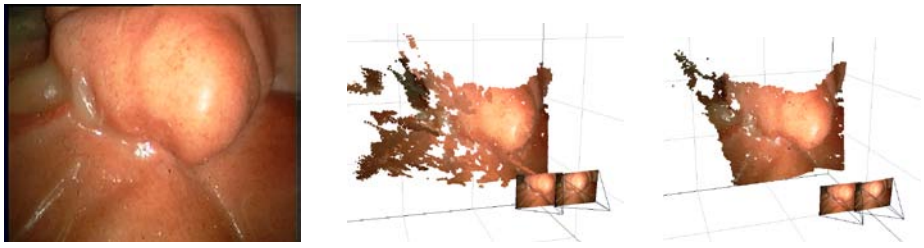

Fig. 1. Example image from a stereo-laparopscope and the corresponding stereo reconstructions using the proposed method with a sum of squared difference metric and with the ZNCC, which clearly performs more effectively.

The spatial neighborhood of a seed match $(\mathbf{x},\mathbf{x}')$, where $\mathbf{x}$ denotes image position, is defined as $N(\mathbf{x},\mathbf{x}')$ and it is used to enforce a 2D disparity gradient limit as a smoothness constraint. Rather than the 1D regions typically enforced by the epipolar geometry $N(\mathbf{x},\mathbf{x}')$. For each image, the spatial neighborhoods around a respective seed match are defined by $N(\mathbf{x}) = \{\mathbf{u}, \mathbf{u}-\mathbf{x} \in [-N,N]^2\}$ and $N(\mathbf{x}') = \{\mathbf{u}', \mathbf{u}'-\mathbf{x}' \in [-N,N]^2\}$ where $(\mathbf{u},\mathbf{u}')$ denotes a candidate pair of pixels. The formulation means that the points considered for propagation are within a spatial window of $(2N+1)\times(2N+1)$ pixels centered at the respective seed locations. The full match propagation neighborhood is then denoted as:

$$N(\mathbf{x},\mathbf{x}') = \left\{(\mathbf{u},\mathbf{u}'), \mathbf{u} \in N(\mathbf{x}), \mathbf{u}' \in N(\mathbf{x}'), \left\|(\mathbf{u}-\mathbf{u}')-(\mathbf{x}-\mathbf{x}')\right\| \le \gamma\right\} \qquad \textbf{(1)}$$

This defines all the possible candidate matches around the seed correspondence $(\mathbf{x},\mathbf{x}')$ and the strategy can be easily adapted to the 1D search space of rectified images [15]. The term γ is used to control the smoothness of the disparity map and we determine this adaptively depending on the color similarity and proximity between the seed and candidate pixels. Such a scheme was shown to be effective by Yoon *et al* [16] for locally adapting the weight of support windows in stereo aggregation. To avoid the heavy computational load of the CIELab color space the method was adapted to use the RBG space [17]. In this study, we only propagate information into only regions of similar color weighed by the proximity of the spatial neighborhood. By defining $\gamma = \beta^{-1}\Sigma \mid I_i(\mathbf{x}) - I_i(\mathbf{u}) \mid + \lambda^{-1} \mid\mid \mathbf{x}-\mathbf{u} \mid\mid$ and stopping the propagation if the value of γ we enforce a consistency and crude segmentation to the propagation process. The dissimilarity measure used during propagation is the zero mean normalized cross correlation (ZNCC), which is less prone to illumination bias in

homogeneous regions while it is also more indicative in regions with discriminative texture. This observation is illustrated in Fig 1 where the proposed algorithm was applied using the traditional sum of squared differences metric and the ZNCC.

2.1 Parallelization of Propagation

To improve the computational performance of the proposed method it is possible to exploit modern GPGPU technology to concurrently calculate multiple correlation windows and propagate structure over multiple pixels. The simplest parallelization strategy is to execute the correlation searches during propagation in parallel. This can be implemented to exploit the large number of concurrent threads that run on modern graphics hardware. In order to maximize the throughput on the graphics hardware, we implemented in this study the propagation of each pixel as a kernel in the NVIDIA language CUDA, including left-right consistency checking. Integral images were used to keep the running time invariant to the correlation window size as for example shown in the stereo algorithm by Veskler [18].

3 Experiments and Results

The proposed method was implemented using C++ and the NVIDIA® CUDA language for GPGPU performance enhancement. The CPU implementation of our propagation technique was able to operate at ~10fps for images of resolution 360 x 288 on a Hewlett-Packard P® xw4600 desktop workstation with an Intel® Pentium® 2.5 GHz Dual-Core Processor and 4Gb of RAM. On the same workstation our CUDA implementation using an NVIDIA® Quadro® FX 5800 card was able to operate at 15fps. We believe this can be significantly improved given optimization of our CUDA code to maximize the use of the GPU cores.

For validation a phantom model of the heart (Chamberlain Group, MA, USA) was embedded with high contrast CT markers and scanned using a Siemens SOMATOM CT scanner. Registration between the camera and CT coordinate systems was performed using the contrast fiducials in the CT coordinate frame and their observed 3D reconstructions from stereo images captured using the daVinci® surgical system. The absolute orientation algorithm by Horn [19] followed by non-linear refinement was used to compute the aligning transformation.

2.1 Phantom Experiment

To validate the stereo reconstruction accuracy of the proposed method we recorded two datasets using the previously described experimental setup. Ground truth information was obtained using the CT data to generate disparity maps for each laparoscopic image as shown in Fig 2. While this method for generating the ground truth has embedded registration error associated with it, it is representative of real clinical scenario.

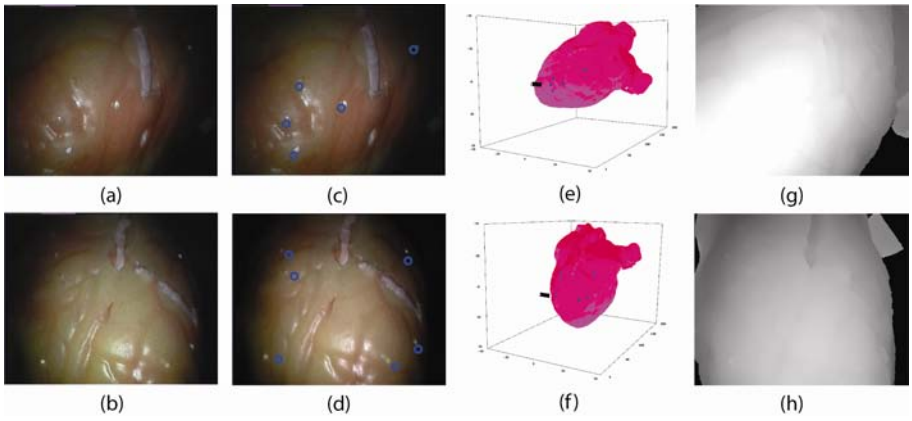

Fig. 2. (a-b) Images of the phantom heart model used in this study taken with the daVinci surgical system; **(c-d)** fiducial points located in the image space; **(e-f)** 3D reconstruction of the heart model from CT data registered in the calibrated stereo-laparoscope coordinate system; **(g-h)** ground truth disparity maps for the images generated by projecting the CT model into the stereo laparoscope images.

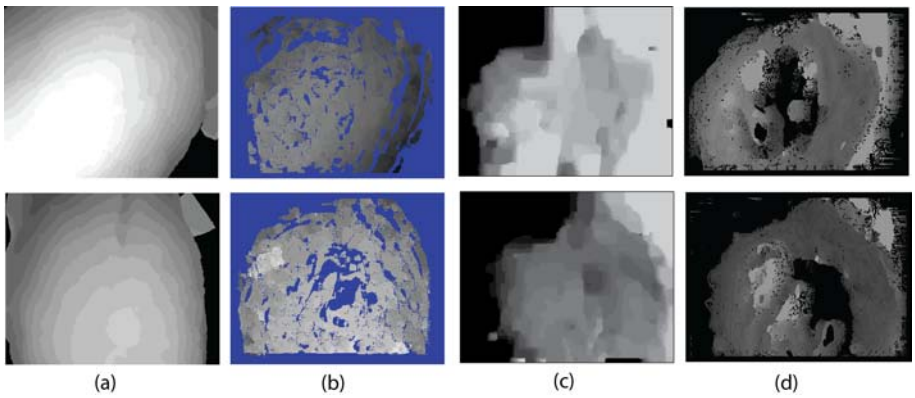

Fig. 3. (a) Ground truth disparity images, the same as shown in Fig 2 but discretized to integer disparity levels; **(b)** disparity map generated with the proposed technique; **(c)** disparity map generated with a global belief propagation (BP) algorithm [20]; (d) disparity map generated with a real-time technique [22].

The proposed technique was used to derive the 3D structure and disparity map of the scene and this was measured against the known ground truth information. The results for this experiment are shown in Fig 3 where the performance of our technique is compared to several dense computational stereo algorithms. The selection of comparison algorithms was based on the availability of the source code for the techniques and their suitability for efficient real-time implementation. It is clear that the disparity map generated by our approach yields a more consistent result than the other methods. A quantitative summary of the results is provided in Table 1 where it is evident that the proposed method outperforms the other approaches measured. It is important to note that the proposed technique does not recover as dense a result as the

other approaches. Thereby perhaps by discarding matches of low reliability their results could be improved to closer match the performance of our technique.

Table 1. Summary of the disparity reconstruction error for the phantom model datasets used in this study (currently not available to preserve anonymity). Different stereo techniques are compared with the approach proposed in this work and the mean disparity error and standard deviations are reported.

Method	Heart 1 Disparity	Heart 2 Disparity
Proposed	0.89 [± 1.13]	1.22 [± 1.71]
BP [20]	9.95 [± 5.22]	9.59 [± 2.77]
RT [22]	12.58 [± 4.57]	9.32 [± 2.80]
CUDA [21]	10.12 [± 4.21]	9.92 [± 3.43]

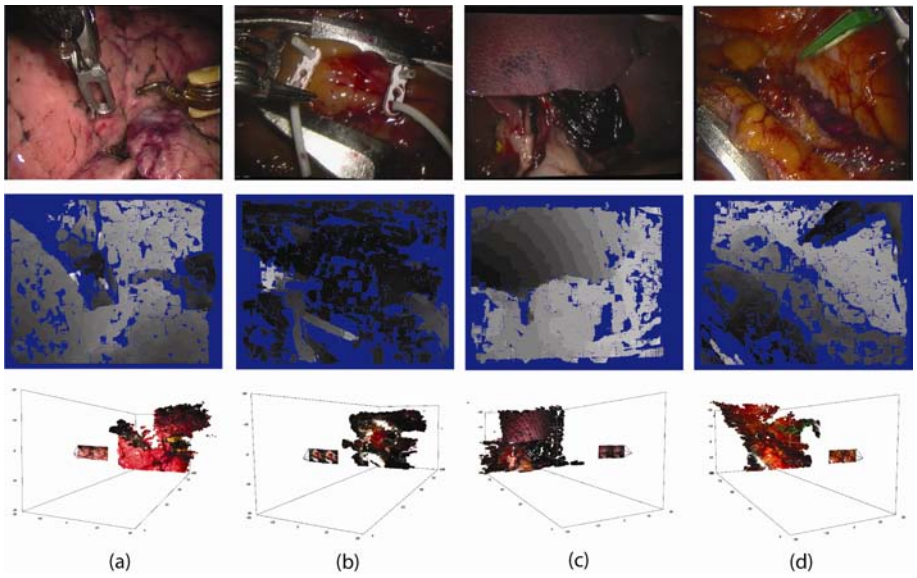

Fig. 4. (top row) Examples of *in vivo* robotic assisted MIS images taken with the daVinci® surgical system; (middle row) the corresponding disparity maps for each image computed with the method proposed in this study, where light colors indicate further away from the camera; (bottom row) 3D renditions of the corresponding reconstruction of the surgical field of view in the camera coordinate space.

2.1 Qualitative *In Vivo* Data Experiments

To qualitatively evaluate the performance of the proposed method on *in vivo* images, we have used several datasets taken from different surgical procedures using the

daVinci® surgical system. The results for the disparity map and the corresponding 3D renditions of reconstructions are shown in Fig 4. It is evident that the proposed technique effectively captures the 3D geometry of the surgical site. It copes well with the presence of large disparity discontinuities due to the surgical instruments and with large specular reflections. However, there are errors, particularly at occlusion boundaries, which need to be addresses further. It is important to note that we do not explicitly cater for specular reflections or model occlusion, for example by using detection, and we do not employ any surgical instrument tracking. By incorporating such strategies in our method we believe that results can be improved significantly.

4 Discussion

In this article, we have presented a real-time stereo reconstruction framework for robotic assisted MIS. The proposed technique relies on propagating a sparse set stereo correspondences into a semi-dense 3D structure by using a best-first principle growing scheme. By incorporating constraints into the propagation framework to consider uniqueness, consistency and disparity smoothness the algorithm produces robust, semi-dense 3D reconstructions of the operating field. We have validated the effectiveness of the approach using phantom data with known ground truth. This data will be made available online together with an executable of the approach to help the development and benchmarking of future work in the field. In our future work, we hope to extend the validation database to more complex phantom models with known ground truth and to include *in vivo* images with manually labeled disparity.

References

[1] D. Stoyanov, M. Lerotic, G. Mylonas, A. J. Chun, and G.-Z. Yang, "Intra-operative Visualizations: Perceptual Fidelity and Human Factors," *IEEE/OSA Journal of Display Technologies*, vol. 4, pp. 491-501, 2008

[2] R. H. Taylor and D. Stoianovici, "Medical Robotics in Computer-Integrated Surgery," *IEEE Transactions on Robotics and Automation*, vol. 19, pp. 765-781, 2003

[3] D. Scharstein and R. Szeliski, "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms," *International Journal of Computer Vision*, vol. 47, pp. 7-42, 2002

[4] M. Z. Brown, D. Burschka, and G. D. Hager, "Advances in Computational Stereo," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, pp. 993-1008, 2003

[5] F. Devernay, F. Mourgues, and E. Coste-Maniere, "Towards endoscopic augmented reality for robotically assisted minimally invasive cardiac surgery," in *Medical Imaging and Augmented Reality*, 2001

[6] F. Mourgues, F. Devernay, G. Malandain, and È. Coste-Manière, "3D reconstruction of the operating field for image overlay in 3D-endoscopic surgery," in *International Symposium on Augmented Reality*, 2001

[7] W. W. Lau, N. A. Ramey, J. Corso, N. V. Thakor, and G. D. Hager, "Stereo-Based Endoscopic Tracking of Cardiac Surface Deformation," in *International Conference on Medical Image Computing and Computer Assisted Intervention*, pp. 494-501, 2004

[8] R. Richa, P. Poignet, and C. Liu, "Efficient 3D Tracking for Motion Compensation in Beating Heart Surgery," in *International Conference on Medical Image Computing and Computer Assisted Intervention*, vol. II, pp. 684-691, 2008

[9] G. Hager, B. Vagvolgyi, and D. Yuh, "Stereoscopic Video Overlay with Deformable Registration," in *Medicine Meets Virtual Reality*, 2007

[10] D. Stoyanov, A. Darzi, and G.-Z. Yang, "A Practical Approach Towards Accurate Dense 3D Depth Recovery for Robotic Laparoscopic Surgery," *Computer Aided Surgery*, vol. 10, pp. 199-208, 2005

[11] J. Shi and C. Tomasi, "Good features to track," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 593-600. 1994

[12] D. Stoyanov, G. P. Mylonas, F. Deligianni, A. Darzi, and G.-Z. Yang, "Soft-tissue Motion Tracking and Structure Estimation for Robotic Assisted MIS Procedures," in *International Conference on Medical Image Computing and Computer Assisted Intervention*, pp. 139-146, 2005

[13] S. N. Sinha, J.-M. Frahm, M. Pollefeys, and Y. Genc, "Feature Tracking and Matching in Video Using Programmable Graphics Hardware," *Machine Vision and Applications*, 2007

[14] P. Mountney, B. P. L. Lo, S. Thiemjarus, D. Stoyanov, and G.-Z. Yang, "A Probabilistic Framework for Tracking Deformable Soft Tissue in Minimally Invasive Surgery," in *International Conference on Medical Image Computing and Computer Assisted Intervention*, pp. 34-41, 2007

[15] M. Lhuillier and L. Quan, "Robust dense matching using local and global geometric constraints " in *International Conference on Pattern Recognition*. vol. 1, pp. 968-972, 2000

[16] K.-J. Yoon and I. S. Kweon, "Adaptive Support-Weight Approach for Correspondence Search," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, pp. 650-656, 2006

[17] Q. Yang, L. Wang, R. Yang, H. Stewenius, and D. Nister, "Stereo Matching with Color-weighted Correlation, Hierarchical Belief Propagation and Occlusion Handling," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, pp. 492-504, 2009

[18] O. Veksler, "Fast Variable Window for Stereo Correspondence using Integral Images," in *International Conference on Computer Vision and Pattern Recognition*. vol. 1, pp. 556-564, 2003

[19] B. K. P. Horn, "Closed-form solution of absolute orientation using unit quaternions," *Journal of the Optical Society of America*, vol. 4, pp. 629-642, 1987

[20] P. F. Felzenszwalb and D. P. Huttenlocher, "Efficient Belief Propagation for Early Vision," *International Journal of Computer Vision*, vol. 70, 2006.

[21] J. Fung, S. Mann, and C. Aimone, "OpenVIDIA: Parallel GPU Computer Vision," in *ACM Multimedia*, pp. 849-852, 2005

[22] D. Demerdjian http://people.csail.mit.edu/demirdji/download/index.html