Dynamic View Expansion for Minimally Invasive Surgery using Simultaneous Localization And Mapping

Peter Mountney and Guang-Zhong Yang

Abstract— **Navigation during Minimally Invasive Surgery (MIS) has recognized difficulties due to limited field-of-view, off-axis visualization and loss of direct 3D vision. This can cause visual-spatial disorientation when exploring complex *in vivo* structures. In this paper, we present an approach to dynamic view expansion which builds a 3D textured model of the MIS environment to facilitate *in vivo* navigation. With the proposed technique, no prior knowledge of the environment is required and the model is built sequentially while the laparoscope is moved. The method is validated on simulated data with known ground truth. Its potential clinical value is also demonstrated with *in vivo* experiments.**

I. Introduction

MINIMALLY invasive surgery is becoming a preferred choice of operation for many surgical procedures due to reduced hospitalisation and patient trauma. Technically, the method has introduced significant challenges to surgeons as they are required to perform the procedures in confined space with elongated tools without direct 3D vision. This is exacerbated by a lack of tactile feedback and the fulcrum effect. Due to the limited distance between the laparoscope camera and the target anatomy, the field-of-view of the surgical site is usually limited. This results in restricted vision which can affect the visual-spatial orientation of the surgeon and the awareness of peripheral events.

A potential solution to expanding the surgeon's field-of-view is dynamic view expansion [1]. To this end, a fisheye lens is a simple hardware solution but this can distort the image and change the appearance of *in vivo* structures. Although distortion can be limited by using rectilinear lenses, the capabilities of an optical solution are limited by the physical confines of the workspace and instrument design. In [2], the authors have used ultrasound to aid navigation through augmented reality. Preoperative CT is registered to the ultrasound data, which is tracked using an electromagnetic tracker. Without introducing additional hardware, image based dynamic view expansion is proposed in [1], where optical flow is used to expand the view of a monocular endoscope for Natural Orifice Translumenal Endoscopic Surgery (NOTES). This approach does not require the scene to be static; however, it relies on the use of a brightness constraint, which can be problematic in homogeneous regions. Direct image mosaicing by creating a large image from multiple local views has been used on images of the bladder [3], brain [4], and retina [5]. These techniques work well for planar geometries but the underlying assumption does not hold for many surgical sites. In [6], an image mosaicing approach is proposed which assumes a tubular structure. This approach requires prior knowledge of the structure of the scene. Despite these problems, image based approaches are attractive because they require no additional hardware. However, issues related to specular reflections, tissue deformation and the paucity of feature landmarks must be addressed.

In this paper, we propose an image based approach for dynamic view expansion that requires no prior knowledge of the 3D structure. The method is based on using Simultaneous Localization And Mapping (SLAM) to generate a sparse probabilistic 3D map of the surgical site while tracking the position of the laparoscope relative to the map. The proposed method is sequential, real-time and capable of working with a limited number of features. Validation has been performed on both simulated data and clinical value is demonstrated with *in-vivo* procine experiments.

II. METHODS

A. Dynamic View Expansion

The proposed method expands the effective field-of-view by incrementally creating a 3D model of the tissue as the laparoscope is navigated around the surgical scene. The 3D model is used to augment the current view from the laparoscope with information from outside the current field–of-view. The model is projected onto a virtual camera situated in the same location as the laparoscope, thus creating a larger field-of-view than the physical laparoscope itself. The method builds a 3D model of the tissue *in situ* and recovers the position of the laparoscope relative to the tissue. It has been shown that SLAM can be used to simultaneously recover the geometrical map of tissue and the position of the laparoscope in MIS[7, 8].

Manuscript received April 7, 2009. This work was supported in part by the Department of Trade and Industry.

P. Mountney is with the Department of Computing and Institute of Biomedical Engineering, Imperial College London. South Kensington Campus, London SW7 2AZ, (peter.mountney@imperial.ac.uk).

G-Z. Yang is with the Department of Computing and Institute of Biomedical Engineering, Imperial College London. South Kensington Campus, London SW7 2AZ,

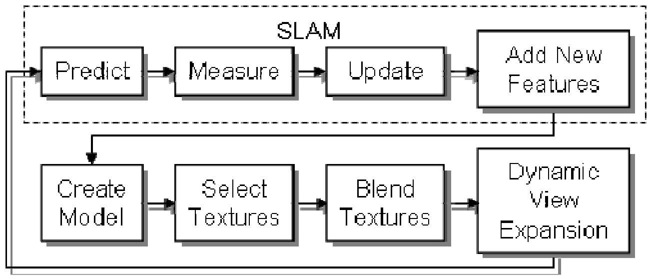

Fig. 1 A schematic illustration of the SLAM system used for dynamic view expansion, in which the key steps involved are highlighted.

B. Simultaneous Localization And Mapping (SLAM)

The SLAM approach used in this paper is based on [7-9], the major steps are outlined in Fig. 1. The 3D map and the position and orientation of the camera are held in a state vector. This vector, along with a covariance matrix, forms the components of an Extended Kalman Filter (EKF) which is used to model the non linear dynamic system. The SLAM EKF consists of four major steps:

Predict: In the prediction step, the position of the camera is estimated for the new frame based on a constant velocity, constant acceleration motion model with added Gaussian noise assuming smooth motion trajectory. The positions of the features are estimated in the camera coordinate system relative to the newly estimated camera position.

Measure: The positions of the features are projected onto the image plane based on the prediction step. The features are measured by actively searching for the best feature match.

Update: By using the predicted state and the measured state, the Kalman gain is computed. Subsequently, the state and covariance estimation are updated.

New features: New features are added to the map when the total number of visible features falls below a predetermined threshold. New features are added by matching salient regions in the left and right stereo images and epi-polar geometry is used to compute their 3D position relative to the camera. Specular highlights are detected with thresholding and features close to specular highlights are ignored. Features close to the edge of the image are added first to create a larger map of the environment.

C. Model Generation

The map created by using the SLAM algorithm is a sparse set of 3D points. These points are interpolated to create a 3D model of the tissue. The solid tissue surface model is generated by performing Delaunay triangulation on the SLAM map. This meshing approach provides an estimate for every 3D point within the observed and mapped environment. An example of a surface model is shown in Fig. 2.

D. Texture Selection

After model generation, it is textured with images taken from the laparoscopic camera to recreate a realistic representation of the environment. A single texture is selected for each triangular face on the model. In MIS, a point light source is attached to the laparoscope, which leads to changes in the surface brightness of the tissue as the laparoscope is navigated. This effect is inevitable when creating a composite image as both spatial and temporal information is combined. In practice, visually inconsistent texturing and the artifacts can be reduced by using a small set of images to texture the model. This set is chosen by searching for images that can texture the largest number of faces. Areas close to the edge of the image are ignored as vignetting can cause poor visual quality. In this study, image rectification is performed before the textures are applied to the mesh in order to remove possible distortions. This texture selection process effectively reduces visual artifacts in the model appearance; however seams are visible where adjacent faces in the model are textured with different images as shown in Fig. 2(c). Seams are removed by blending adjacent textures described below.

E. Poisson Blending

For planar surfaces, seams between adjacent textures can be removed by simple blending. In this study, however, it is not applicable as the surfaces are not planar and may assume arbitrary topology. Approaches such as multiband blending are not straight forward to apply either as images may not have overlap faces.

Previous studies have shown that the Poisson image editing [1, 10] can be used to reduce the effects of seams by locally adjusting the brightness values. It has been successfully used for dynamic view extension [1] in NOTES. With this approach, the new image is mapped onto the existing image by formulating it as a partial differential equation. The border on the new image is constrained to be equal to the intensities on the existing image by enforcing the Dirichlet boundary conditions [10]. The new texture is added to the existing texture with a one pixel overlap $\delta\Omega$. A large sparse positive definite system of linear equations is solved iteratively in the RGB channels with a conjugate gradient method with successive over relaxation as a pre-condition:

$$|N_i| f_i^{'} - \sum_{j \in N_i} f_j^{'} = \sum_{j \in \delta\Omega \cap N_i} g_j^{'} + |N_i| f - \sum_{j \in N_i - \delta\Omega} f_j \qquad \textbf{(1)}$$

where N_i is the set of pixels neighbouring pixel i, f_i are the pixel values of the mosaic before updating, $f_i^{'}$ are the unknown pixel values of the updated mosaic and $g_i^{'}$ are the pixel values of the new texture. Pairwise face blending is performed. The existing image is taken to be the image which covers the larger number of faces. An affine transformation is applied to the new image to align and scale it into the same coordinate system as the existing image before blending is performed. This generates a seamless textured model of the tissue.

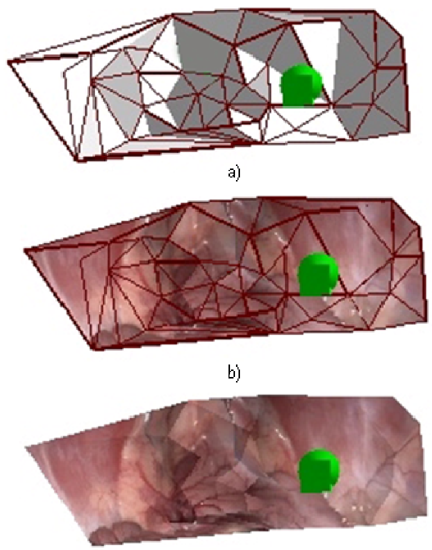

Fig. 2 **(a)** Delaunay triangulation of the points in a SLAM map with current camera position shown in green. **(b)** Selected textures for each triangle **(c)** the textured 3D tissue model before seam removal.

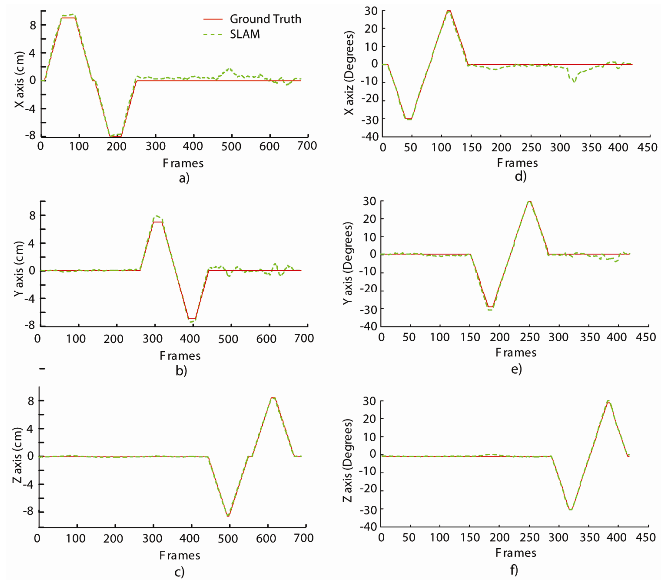

Fig. 3 Quantitative analysis of the camera motion. The SLAM estimated position is shown in green and the ground truth is shown in red for the X (top), Y (middle) and Z (bottom) axis.

F. *Dynamic View Expansion*

In essence, view expansion is the augmentation of the video images from the laparoscope with a textured model. The SLAM algorithm provides the position of the laparoscope, enabling the model to be projected into the image plane. The expanded view is achieved by extending this image plane, and thus the effective field-of-view. Direct combination of the model and laparoscopic images can cause a seam around the edge of the image as shown in Fig. 4(c). The Poisson image editing operation as described above is thus applied to remove this seam by blending the model into the image from the laparoscope. The image from the laparoscope is not adjusted or changed in any way. The extended field-of-view is an approximation and can be considered as a navigational aid.

III. Results

The proposed system has been quantitatively validated on simulated data with known ground truth. An image from a laparoscopic procedure showing the liver and part of the stomach was used to texture a 3D surface model. A virtual stereo laparoscope was navigated around the model to capture images of the texture mapped surface. This provided known ground truth for the camera's position in the environment, enabling quantitative validation of the SLAM system. Results are shown in Fig. 3(a-c) for translation where the laparoscope is navigated left and right, up and down and finally away and towards the 3D surface. In this study, the average error was 0.4, 0.22 and 0.1 cm in the X, Y and Z axes. The corresponding standard deviation was 0.28, 0.22, 0.09, respectively. This represents average errors of 2.3%, 1.5% and 0.5% of the total movement in the X, Y and Z axes. These results demonstrate the accuracy of the system in recovering the laparoscope's position, which is fundamental to dynamic view expansion. In general, the errors involved are relatively small. The largest errors are found when the laparoscope is changing direction, in which case the constant velocity motion model used in EKF no longer holds and motion is modeled with a Gaussian distribution. Small errors are introduced into the X and Y estimation when navigating along the Z axis away and towards the tissue model. These errors are due to the higher uncertainty in the Z position of the features resulting from the small baseline of the stereo cameras used.

Figs. 3(d-f) show the laparoscope rotating around the X, Y and Z axes with average error of 1.34°, 0.8° and 0.295° and standard deviation of 1.57°, 0.75° and 0.32° respectively. This represents 2.23%, 1.33% and 0.49% of the total rotation in the X, Y and Z axes. Small errors occur when rotating around the optical axis (Z axis). This rotation causes significant changes in the appearance of the features.

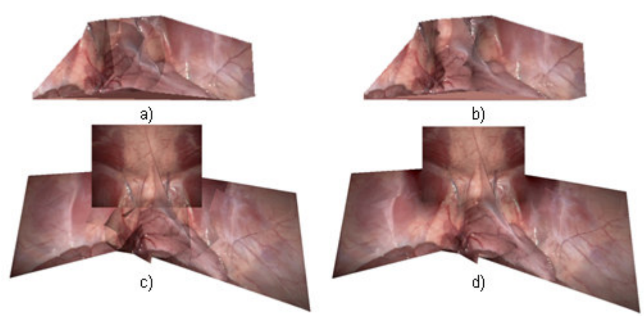

Fig. 4 A visual comparison of the effect of Poisson texture blending on *in vivo* data. **(a)** Textured model without blending. **(b)** Textured model with blending. **(c)** Current view augmented with textured model without blending. **(d)** Current view augmented with textured model with blending.

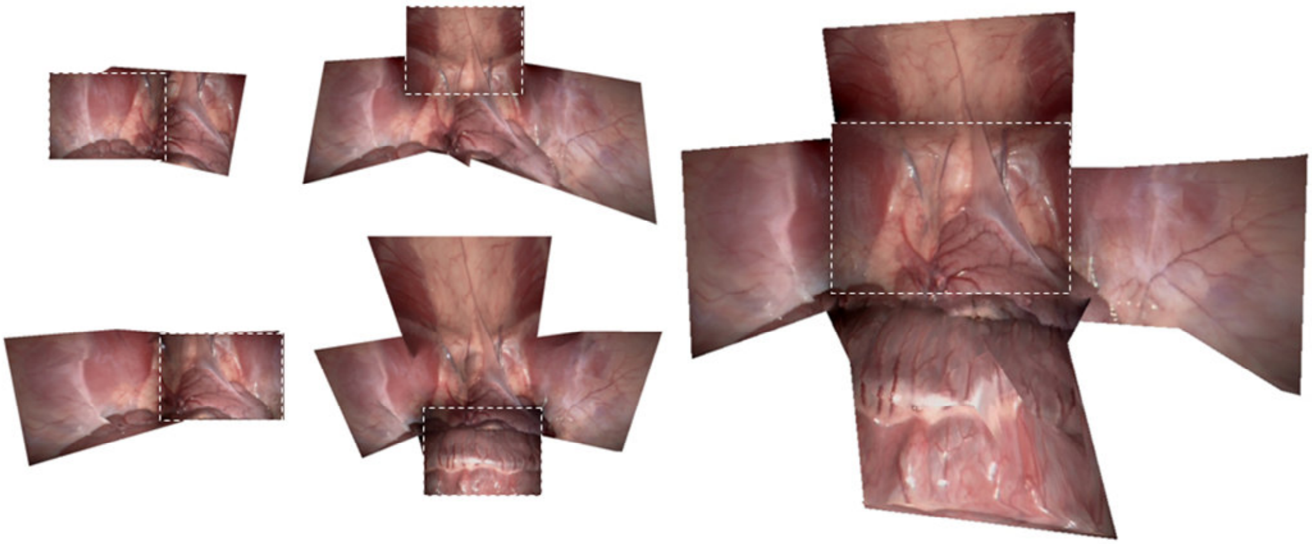

Fig. 5 Five *in-vivo* examples of dynamic view expansion performed during an exploration of the abdomen. The current image from the endoscope is highlighted with a white dashed border.

The proposed technique has also been applied to *in-vivo* porcine data. Since the ground truth data for the *in-vivo* procedure is not available, only qualitative results are provided in Fig. 4, which shows a textured surface model before and after texture selection and blending. It is evident from Fig. 4(a) that seams are clearly visible on the model between textured facets from difference images. Fig. 4(b) shows the same surface model after texture selection and blending, illustrating the benefit of this approach in seam removal. Figs 4 (c) and (d) illustrate the augmentation of the model to the current image from the laparoscope and the resulting view expansion without and with blending.

Further results are presented in Fig. 5 to demonstrate the clinical value of dynamic view expansion for *in-vivo* abdominal exploration during MIS. The current image from the laparoscope is highlighted in a white box. It is evident that the views derived from the proposed method are seamless.

IV. Conclusion

In this work, we have developed a technique for dynamic view expansion during MIS. It simultaneously builds a 3D model of the environment and tracks the position of the laparoscope. The method has been validated and applied to both simulated and *in vivo* data sets. The results have shown that the proposed technique achieves improved visual appearance. The application of Poisson blending further enhances the visual fidelity of the results. It is worth noting that accuracy can be further improved by directly incorporating tissue deformation in the SLAM framework. Future work will also include more sophisticated meshing of the SLAM map to cater for more complex environments.

Acknowledgment

The authors would like to thank Danial Stoyanov, Mirna Lerotic, Adrian Chung and Andrew Davison for their help and advice. Mono-SLAM [9] is available as an open source project under the LGPL license from the Scene homepage at http://www.doc.ic.ac.uk/~ajd/Scene/.

References

[1] M. Lerotic, A. Chung, J. Clark , S. Valibeik, and G.-Z. Yang, "Dynamic View Expansion for Enhanced Navigation in Natural Orifice Transluminal Endoscopic Surgery," presented at MICCAI, 2008.

[2] R. San José Estepar, N. Stylopoulos, R. Ellis, E. Samset, C. Westin, C. Thompson, and K. Vosburgh, "Towards scarless surgery: An endoscopic ultrasound navigation system for transgastric access procedures.," *Computer Aided Surgery.*, vol. 12, pp. 311-24, 2007

[3] A. Behrens, "Creating Panoramic Images for Bladder Fluorescence Endoscopy," *Acta Polytechnica Journal of Advanced Engineering*, vol. 48, pp. 50-54, 2008.

[4] W. Konen, M. Naderi, and M. Scholz, "Endoscopic image mosaics for real-time color video sequences," presented at Computer Assisted Radiology and Surgery, Berlin, 2007.

[5] P. Cattin, C, , H. Bay, L. Van Gool, and G. Székely, "Retina Mosaicing Using Local Features," presented at MICCAI, 2006.

[6] S. Seshamani, W. Lau, and G. Hager, "Real-time endoscopic mosaicking," presented at MICCAI, 2006.

[7] Mountney P, Stoyanov D, J. Davison A, and Yang G-Z, "Simultaneous Stereoscope Localization and Soft-Tissue Mapping for Minimal Invasive Surgery," presented at MICCAI, 2006.

[8] D. Noonan, P. Mountney, D. Elson, A. Darzi, and G.-Z. Yang, "A Stereoscopic Fibroscope for Camera Motion and 3D Depth Recovery During Minimally Invasive Surgery.," presented at ICRA 2009.

[9] Davison A.J., Reid I., Molton N., and S. O., "MonoSLAM: Real-Time Single Camera SLAM," *to appear in IEEE PAMI*, 2007.

[10] P. Perez, M. Gangnet, and A. Blake, "Poisson image editing," *ACM Trans. Graph* vol. 3, pp. 313–318, 2003.