Dense Surface Reconstruction for Enhanced Navigation in MIS

Johannes Totz, Peter Mountney, Danail Stoyanov, and Guang-Zhong Yang

The Hamlyn Centre for Robotic Surgery
Imperial College London, UK

Abstract. Recent introduction of dynamic view expansion has led to the development of computer vision methods for minimally invasive surgery to artificially expand the intra-operative field-of-view of the laparoscope. This provides improved awareness of the surrounding anatomical structures and minimises the effect of disorientation during surgical navigation. It permits the augmentation of live laparoscope images with information from previously captured views. Current approaches, however, can only represent the tissue geometry as planar surfaces or sparse 3D models, thus introducing noticeable visual artefacts in the final rendering results. This paper proposes a high-fidelity tissue geometry mapping by combining a sparse SLAM map with semi-dense surface reconstruction. The method is validated on phantom data with known ground truth, as well as in-vivo data captured during a robotic assisted MIS procedure. The derived results have shown that the method is able to effectively increase the coverage of the expanded surgical view without compromising mapping accuracy.

1 Introduction

In Minimally Invasive Surgery (MIS), disorientation due to restricted field-of-view of the endoscopic camera has a recognised effect on 3D visuomotor control and navigation. Dynamic View Expansion (DVE) combines previously captured images with the current view of the camera, thus revealing the surrounding anatomical structure to minimise these effects. In practise, DVE can be performed using image mosaicing algorithms. Such algorithms register two or more images together based on the assumption of the 3D structure of the scene. The most common assumption is that the tissue or organ is planar and the registration is performed by using image- or feature-based techniques. The former generally requires a large overlap between the images, whereas feature-based techniques [1] are based on the detection and matching of salient regions in the images. In reality, tissue structure and organ geometry in MIS tend to not satisfy the planar assumptions. To alleviate this effect, parallax motion correction has been proposed [2] and prior knowledge of the organ geometry has been used to model intraluminal structures such as the colon [3] and oesophagus [4] as cylinders. However, if the assumed tissue model does not represent the tissue geometry, significant artefacts can be introduced into the final result with expanded field-of-view. In recent work based on Simultaneous Localisation and

G. Fichtinger, A. Martel, and T. Peters (Eds.): MICCAI 2011, Part I, LNCS 6891, pp. 89–96, 2011.

Mapping (SLAM), methods are proposed to build a sparse model of the tissue structure sequentially and incrementally without relying on prior knowledge or explicit assumption of the tissue structure [5,6]. The sparse model is meshed and textured with images from the endoscope. This enables the model to be projected onto the estimated camera position to create an expanded view. With this approach, two problems have been identified due to the sparse representation of the tissue. Firstly, artefacts are introduced into the visualisation due to the coarse model. Secondly, the expanded view is limited to the size of the modelled region. Any part of the tissue that does not fall within the model cannot be visualised with DVE. In [7], an approach for dense surface reconstruction is used to create 3D tissue models from stereo laparoscopic images. The method matches a sparse set of salient regions using stereo Lucas-Kanade and propagates the disparity around each matched region. The method is used to reconstruct deforming tissue from a static camera. Dense surface reconstruction can also be achieved using monocular images combined with [8] or without [9] the use of a robotic arm. However, limited work has been performed in dealing with a dense model of the tissue from a moving stereo laparoscopic camera.

The purpose of this paper is to explore the use of sparse SLAM with dense surface reconstruction for DVE. Under the proposed framework, the disc-homeomorphic 3D model is parameterised into a planar texture domain into which video images and surface details are projected. The accuracy of the proposed method is validated on phantom data and the potential clinical application is demonstrated on in-vivo data.

2 Methods

Fig. 1 outlines the main steps involved in creating the expanded visualisation. SLAM is used to estimate the position of the camera and incrementally build a sparse map or model of the tissue. This model is extended using a "skydome" and a dense stereo algorithm is used to reconstruct the surface of the tissue, which is registered to the sparse model. The dense tissue model is then textured and reprojected onto the live laparoscopic video stream to create an augmented visualisation. It should be noted that the original live video stream is never modified to ensure the visual fidelity of the current surgical view.

2.1 Dynamic View Expansion with Sparse Stereo SLAM

For sequential scene reconstruction, SLAM is used to estimate the position and orientation of the stereo camera and a sparse set of features in 3D which represent the map or tissue model. This information comprises the state of the system which is modelled probabilistically using an Extended Kalman Filter (EKF), enabling uncertainty and noise in the system to be modelled. SLAM detects and matches regions of interest in the stereo images, which are triangulated to estimate a 3D map feature. In subsequent video frames, the motion of the camera is predicted using a constant velocity, constant acceleration motion model. The

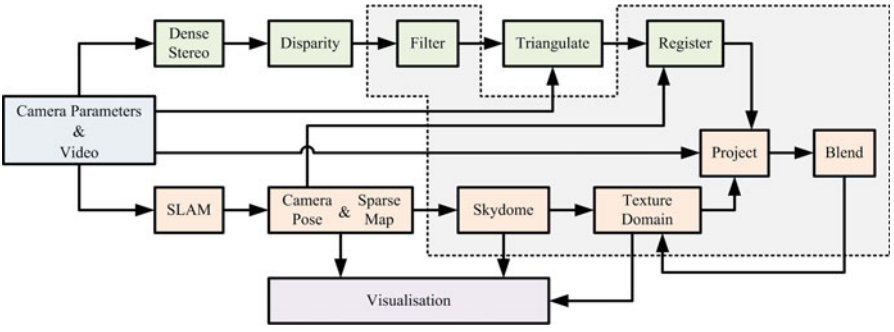

Fig. 1. Processing pipeline of the proposed method. Video images are processed independently by SLAM and dense-disparity until they are registered to each other. Projection blends surface detail and video colour into texture maps used for visualisation. This paper addresses mainly components enclosed in the dashed box.

position of each 3D map feature is estimated relative to the predicted camera position and the map features are measured by matching the regions of interest in the image space. The prediction and measurement are used to perform the update step of the EKF and estimate a new state (camera position and 3D map). A photorealistic model of the tissue is generated by meshing the sparse 3D map using a Delaunay triangulation and texturing using the approach outlined in [6].

Using the SLAM-map alone for DVE can result in a loss of visual information because parts of the video frame may not be covered with trackable features and map triangles. Previously, this was avoided by introducing virtual features into the map. For the proposed method, the SLAM-map triangulation Y is padded with additional triangles around the boundary to produce Y', which is referred to as "skydome" in this paper. The 3D position of vertices is computed by $y'_i = r \cdot (y_i - c) + c$, with c being the centroid of all n map vertices, y_i being a vertex on the boundary of the map and y'_i the position of the new skydome vertex. The factor r controls by how much the triangles are enlarged; suitable values are between 2 and 4, trading coverage for texture map resolution. Fig. 2a illustrates this aspect of the algorithm. The position of y' in the texture domain is derived in the same way, using 2D texture coordinates instead of 3D positions. In Figs. 2b and c, the green area marks video content covered by the sparse map only, while the red area marks additional video pixels covered by the skydome. Thus, the skydome provides a larger target for video frames to be mapped on; in fact, its exact geometry is not important because dense 3D positions mapped on it (Sec. 2.2) override its sparse vertex positions (Fig. 3d and e).

2.2 Dense Surface Reconstruction

Using the stereo images of the observed tissue, the 3D position of each pixel is computed by first establishing a disparity map that sets up correspondences between pixels in the left and the right video frame. The algorithm used to

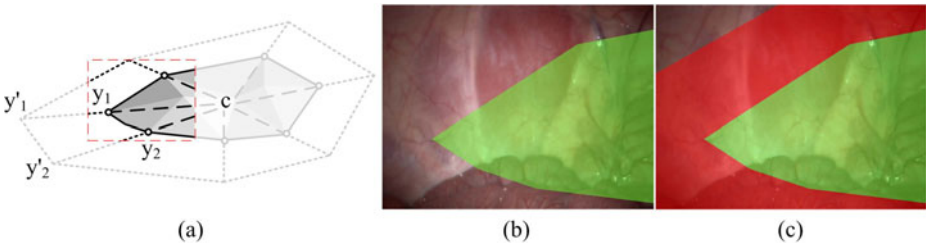

Fig. 2. Padding the SLAM-map with additional triangles. b) and c) illustrate the effect in the video image with green being the original sparse SLAM-map and red the additional area covered by the skydome.

compute the disparity between the left and right frames is the one proposed in [7]. Given a set of sparse correspondences between left and right images used as seeds, the algorithm starts by propagating for any neighbouring pixels at which a match in the other image can be found. The output is the disparity map D (Fig. 3b): for every pixel in the left image it contains a coordinate into the right image for the corresponding pixel, or an invalid coordinate if no such correspondences can be established. The outliers and regions without stereo-disparity correspondence are then filled by using Telea's in-painting method [10] followed by a box-blur. Figs. 3a-c show an example of the left and right video frame, the unprocessed disparity with holes and outliers, followed by in-painting. Each pixel in the disparity map is triangulated into 3D space. The reconstructed surface is mapped from the camera space to the SLAM coordinate system by transforming it with the SLAM-estimated camera pose. The remaining difference is due to SLAM's noise modelling, but mainly limited to the viewing direction. Therefore, for every SLAM feature visible in the current frame, the offset between its current position in 3D and the dense estimated surface is computed, subject to outlier removal. The resulting 3D position map P in the SLAM-coordinate system is rendered into the texture domain T already used for the video pixels [6], i.e. every video pixel that has been observed previously has a position in 3D.

2.3 Texture Domain Update and Visualisation

Within the proposed framework, updating texture domain T consists of three steps: reprojection, position update and colour update. There are three images attached to it: colour from video T_C, age-map T_A (storing a pixel's age, see [6]) and 3D position T_{3D} encoded in RGBA. The contents of these images is first reprojected from the previous frame to accommodate the possibly updated and retriangulated SLAM-map. Colour and position are normalised by the accumulated value in the alpha channel: $g_{rgba} = g_{rgba}/g_a$, where g is a pixel's RGBA colour value representing either colour or position.

Following the registration described above, the 3D position image is updated by rendering the sparse map Y' textured with P into T_{3D} with additive blending,

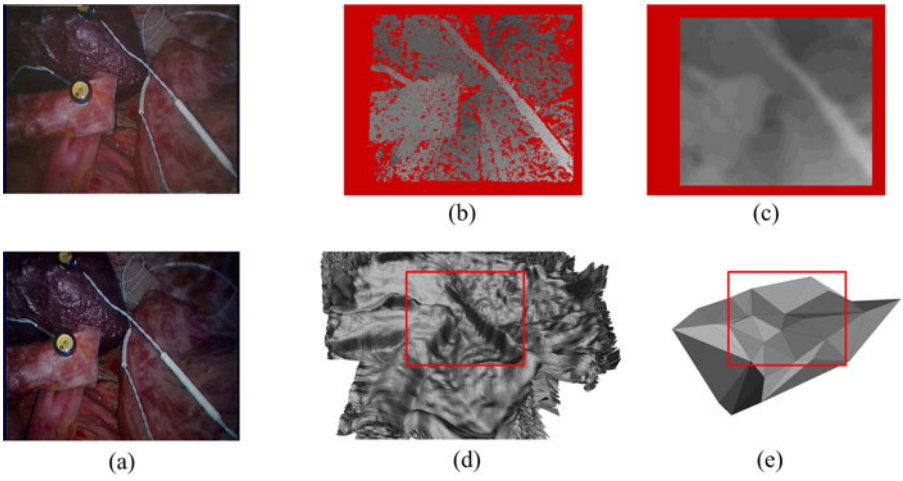

Fig. 3. Stereo disparity and estimated dense surface. a) shows left and right video frame of the phantom sequence; b) shows raw stereo disparity; c) filtered and smoothed disparity; d) shows the flat-shaded highly tesselated Y' with the red rectangle marking the part that would be covered by the video frame. For comparison, the sparse SLAM map surface Y is shown in e).

combining new and existing data (in effect dense temporal surface reconstruction). The colour and age-map are updated by taking the up-to-date detailed surface position from T_{3D} into account and each pixel is warped to account for perspective distortion. This two-pass approach is necessary to avoid an unstable feedback loop between the current and the previous frame's positions.Without normalisation during texture domain reprojection, newly arriving data would have minimal effect as many more frames have already been accumulated. Instead, a running average is computed. In addition, video and position projection is radially attenuated to account for spot-light illumination effects by applying a linear fall-off function towards the edges of the video frame.

The final result for each frame is visualised by using the fade-to-grey method of [6] but additionally rendering Y' with high tessellation and displacement mapping. The colour and position information are texture-mapped onto the triangles and then subdivided by at least eight levels using the graphics hardware. Each vertex is positioned in 3D space using the estimated dense surface position. The surface is never explicitly meshed; Fig. 3d serves for illustration purposes only. Pixels are rejected for which attenuation has caused $g_a < 1$ accumulation to avoid noisy edge cases.

3 Results

The method has been evaluated on a phantom sequence and an in-vivo porcine experiment, both representing an exploration of the abdomen. The stereo cameras were calibrated at the start and remain unchanged during the experiments.

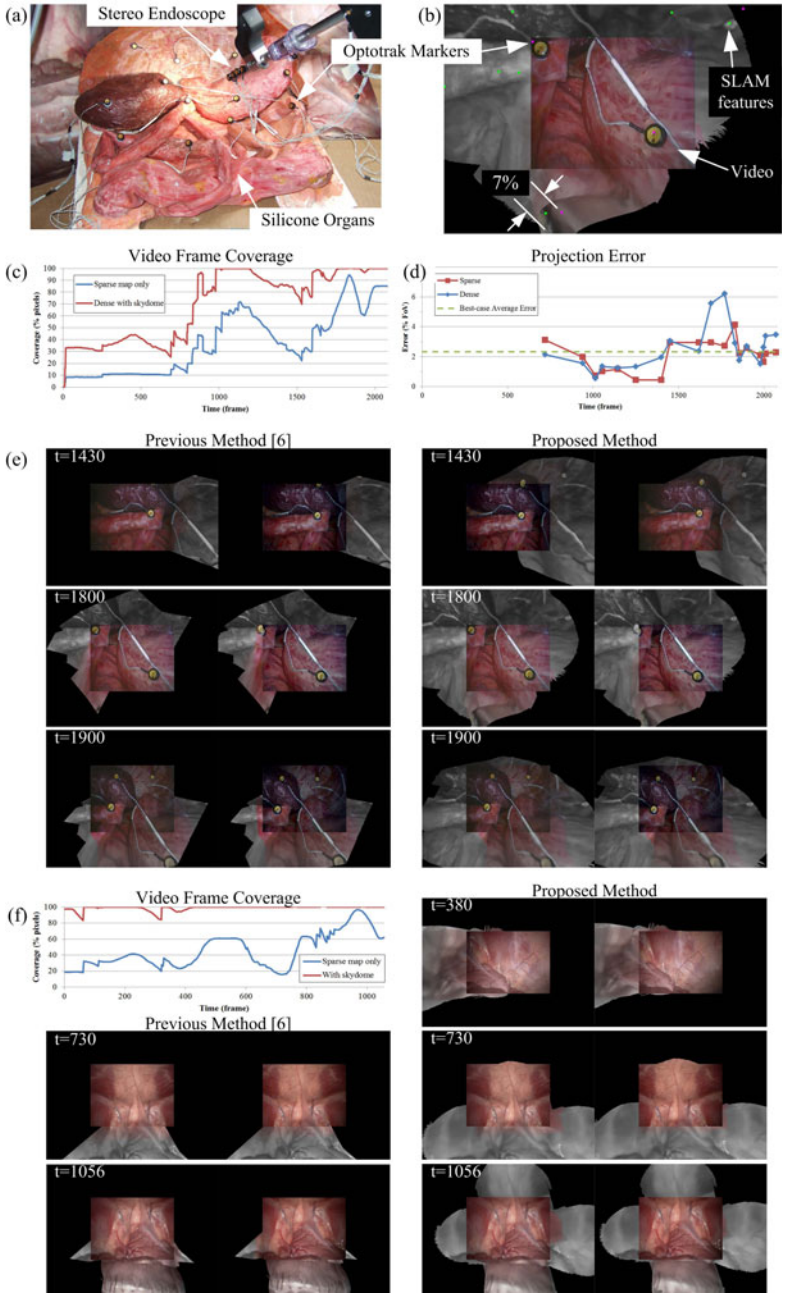

Fig. 4. a) Experimental setup with silicone organs, OptoTrak markers and stereo laparoscope. b) Projected markers (purple) and SLAM features (green) in the DVE. c) Video frame coverage of SLAM-map with and without skydome. d) Projection error of markers into image plane. e) Stereoscopic DVEs for phantom sequence at various points in time, comparing sparse with dense surface. f) Coverage plot and DVEs for in-vivo sequence.

The abdominal phantom used for validation comprised a full set of silicone organs painted with acrylic paint to provide realistic texture. Optical markers were placed on the surface of the phantom (see Fig. 4a) to provide the ground truth data represented as a sparse 3D model of the phantom. A further eight markers were placed on a rigid body attached to the laparoscope. These were used, in conjunction with hand-eye calibration, to determine the position of the camera. The quality of DVE was validated by projecting the position of the markers on the model into the expanded image plane of the camera at a given camera position. This provided a set of ground truth points in the 2D image plane for comparison with the reprojected dense model obtained using the proposed method. Due to the limited accuracy of the optical tracker and noise in the calibration the validation framework was evaluated to be accurate to within 2.3% of the FoV. The graph in Fig. 4d shows the average reprojection error for the sparse (red) and dense (blue) approaches. The average visual angel error for the sparse is 2.2% and for the dense 2.7% of the FoV, making the accuracy of the two results comparable. The accuracy of the dense method decreases around frame 1700 due to noisy estimation of the camera position causing inaccurate registration of the dense map but recovers afterwards. Qualitative evaluation of the sparse and dense approaches are provided in Fig. 4e. The effect of ageing the expanded view is clearly visible on these sequences with older regions appearing desaturated.

Quantitative evaluation is also performed with respect to coverage measured by how much of the observed tissue is represented in the expanded view. It is computed as the number of video frame pixels covered by a projection of the sparse map Y'. This correlates directly with how much information in the video is actually used for DVE: only video pixels that map to a position in the sparse map (with or without skydome) can be used. Fig. 4c depicts the effect of extending sparse map coverage via the skydome with $r = 2$. The phantom sequence starts with small rotational motion that only covers a relatively small area and does not introduce new SLAM features. Only after frame 700 does the camera move enough for new features to be introduced. The average video pixel coverage without skydome is 37% and with skydome 70%. The in-vivo sequence [11] shows the peritoneal wall, large intestines and stomach. The sequence is quantitatively evaluated in Fig. 4e and the average coverage without and with skydome is 45% and 99%, respectively, for $r = 3.5$ (chosen due to initial SLAM feature distribution). Easily visible in the coverage graph are the introduction of new SLAM features and the resulting extension of the skydome: sudden spikes in increased coverage. Fig. 4e shows the video frames with the corresponding DVE. The current single-threaded implementation of the proposed method runs at about 1 frame per second.

4 Conclusions

In this paper, a method has been proposed for DVE by exploiting the spatial-temporal correspondence of intra-operative laparoscopic imaging. The novelty of the proposed approach is in the creation of a dense tissue model from a moving stereoscopic camera and the method is validated on both silicon phantom

with know ground truth and in-vivo data. The proposed method has shown to increase the coverage of the expanded view without compromising accuracy. Another feature of the proposed system is that the method is sequential and well suited for real-time operation. In addition the system has been designed to make it compatible with non-static SLAM algorithms and dense non-rigid tissue models. It is worth noting that noise modelling in SLAM takes into account unreliable feature tracking thus increasing the robustness of the method. Further improvements of the proposed method could include estimation of the surface's BRDF for perspective correct relighting.

References

1. Hu, M., Penney, G., Rueckert, D., Edwards, P., Bello, F., Figl, M., Casula, R., Cen, Y., Liu, J., Miao, Z., Hawkes, Z.M.D.: A robust mosaicing method with super-resolution for optical medical images. In: Liao, H., Edwards, P.J., Pan, X., Fan, Y., Yang, G.-Z. (eds.) MIAR 2010. LNCS, vol. 6326, pp. 373–382. Springer, Heidelberg (2010)
2. Lerotic, M., Chung, A.J., Clark, J., Valibeik, S., Yang, G.-Z.: Dynamic view expansion for enhanced navigation in natural orifice transluminal endoscopic surgery. In: Metaxas, D., Axel, L., Fichtinger, G., Székely, G. (eds.) MICCAI 2008, Part II. LNCS, vol. 5242, pp. 467–475. Springer, Heidelberg (2008)
3. Seshamani, S., Lau, W., Hager, G.D.: Real-time endoscopic mosaicking. In: Larsen, R., Nielsen, M., Sporring, J. (eds.) MICCAI 2006. LNCS, vol. 4190, pp. 355–363. Springer, Heidelberg (2006)
4. Carroll, R.E., Seitz, S.M.: Rectified surface mosaics. In: IEEE 11th International Conference on Computer Vision, pp. 1–8 (October 2007)
5. Mountney, P., Yang, G.Z.: Dynamic view expansion for minimally invasive surgery using simultaneous localization and mapping. In: 31st Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 1184–1187 (September 2009)
6. Totz, J., Fujii, K., Mountney, P., Yang, G.Z.: Enhanced visualisation for minimally invasive surgery. International Journal of Computer Assisted Radiology and Surgery (2011), doi:10.1007/s11548-011-0631-z
7. Stoyanov, D., Scarzanella, M.V., Pratt, P., Yang, G.-Z.: Real-time stereo reconstruction in robotically assisted minimally invasive surgery. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010. LNCS, vol. 6361, pp. 275–282. Springer, Heidelberg (2010)
8. Moll, M., Tang, H.-W., Van Gool, L.: GPU-accelerated robotic intra-operative laparoscopic 3D reconstruction. In: Navab, N., Jannin, P. (eds.) IPCAI 2010. LNCS, vol. 6135, pp. 91–101. Springer, Heidelberg (2010)
9. Newcombe, R.A., Davison, A.J.: Live dense reconstruction with a single moving camera. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1498–1505 (2010)
10. Telea, A.: An image inpainting technique based on the fast marching method. Journal of Graphics Tools 9(1) (2004)
11. Mountney, P., Stoyanov, D., Yang, G.Z.: Three-dimensional tissue deformation recovery and tracking: Introducing techniques based on laparoscopic or endoscopic images. IEEE Signal Processing Magazine 27(4), 14–24 (2010)