Multi-view Stereo and Advanced Navigation for Transanal Endoscopic Microsurgery

Christos Bergeles, Philip Pratt, Robert Merrifield,
Ara Darzi, and Guang-Zhong Yang

The Hamlyn Centre,
Imperial College London, London SW7 2AZ, UK
{c.bergeles,p.pratt,rdm99,a.darzi,g.z.yang}@imperial.ac.uk

Abstract. Transanal endoscopic microsurgery (TEM), *i.e.*, the local excision of rectal carcinomas by way of a bimanual operating system with magnified binocular vision, is gaining acceptance in lieu of more radical total interventions. A major issue with this approach is the lack of information on submucosal anatomical structures. This paper presents an advanced navigation system, wherein the intraoperative 3D structure is stably estimated from multiple stereoscopic views. It is registered to a preoperatively acquired anatomical volume based on subject-specific priors. The endoscope motion is tracked based on the 3D scene and its field-of-view is visualised jointly with the preoperative information. Based on *in vivo* data, this paper demonstrates how the proposed navigation system provides intraoperative navigation for TEM[1].

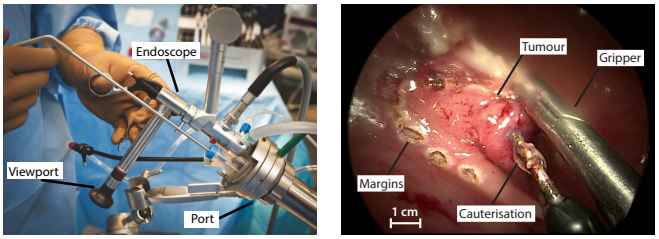
1 Introduction

Colorectal cancer is a significant health problem in most countries. Traditionally treated via radical excisions, advances in laparoscopy and the miniaturisation of endoscopes are supporting more localised interventions via Transanal Endoscopic Microsurgery (TEM) [1]. In TEM, elongated instruments and a stereo endoscope are inserted through a hermetically sealed port that allows pneumorectum, *i.e.* rectal insuflation (see Fig. 1). With dexterous manoeuvres, skilled surgeons excise and remove the tumour. One of the constraining factors, however, is the limited anatomical information that can be inferred by only the endoscopic view. Intraoperative registration of preoperatively acquired anatomical volumes is considered an elegant solution for information augmentation.

Existing approaches in surgical domains outside of TEM use preoperatively acquired Magnetic Resonance Imaging (MRI) and Computed Tomography (CT) volumes to augment endoscopic views. For example, Dey *et al.* used MRI volumes and texture-mapped images acquired from a neuroendoscope intraoperatively [2]. Pratt *et al.* and Vemuri *et al.*, in [3] and [4], respectively, used segmented

[1] With support from the Department of Health and Wellcome Trust through the Health Innovation Challenge (HIC) Fund. Source code available at: http://christos.bergeles.net/software/code/miccai2014.zip

P. Golland et al. (Eds.): MICCAI 2014, Part II, LNCS 8674, pp. 332–339, 2014.



(a) (b)

Fig. 1. (a) Illustration of TEM procedure. (b) Endoscopic images provide a limited field-of-view with no ability to visualise tissue layers and adjacent anatomy.

manually registered anatomical volumes acquired from MRI images to augment endoscopic views during partial nephrectomy and adrenalectomy. In addition to existing challenges, TEM presents the complexity that the intervention is carried out within a lumen, and information such as tissue thickness, metastatic lymph nodes, neighbouring organs (*e.g.* prostate, vagina) can be retrieved only from memory. This lack of information limits registration landmarks and may also lead to perforation or partial tumour excision and unfavourable histologies.

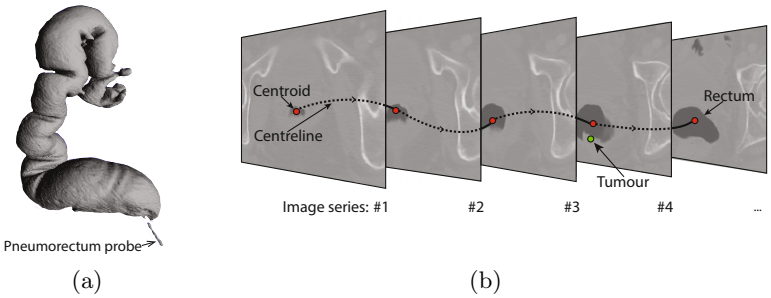
A prerequisite of information augmentation is 3D reconstruction of the operating field-of-view. Despite a number of algorithms being recently proposed for minimally invasive surgery (MIS), *e.g.* the CPU/GPU algorithm of [5], and the variational algorithm of [6], TEM presents important limitations since the rectum is largely textureless. This leads to unstable and noisy estimations.

This paper demonstrates that multi-view stereo fusion, inspired from [7], takes advantage of even slight endoscope motions to provide stable high fidelity reconstructions. Moreover, contrary to information augmentation in bronchoscopy [8] or sinus skull-base surgery [9], two procedures that also involve intraluminal interventions, in TEM the endoscope does not exhibit a large travel range and does not allow the use of bifurcations as landmarks between the endoscopic view and the preoperative volume. It lends itself, however, to two important patient-specific priors: a) the endoscope is inserted close to the centreline of the rectum, and b) the tumour to be excised is identifiable in the endoscopic view and in the MRI/CT. These two priors can be exploited for registration. In the sections that follow, the entire algorithmic workflow and results based on *in vivo* acquired data justify the applicability of the proposed navigation framework.

2 Materials and Methods

2.1 Preoperative Imaging and Segmentation

CT or MRI scans of the patient are acquired several days before the procedure to evaluate the extent of tumour and the possibility of local excision via TEM.



(a) (b)

Fig. 2. (a) Segmented colon after pneumocolon (CT data). (b) Illustration of centreline extraction procedure based on CT images of the patient of case 3 (see Sec. 3).

Pneumorectum is applied in order to visualise the rectum and colon, as in Fig. 2(a). The presence of a clear air/tissue interface facilitates segmentation. The entry angle and potential path of the endoscope during TEM is aligned with the centreline that passes through the rectum. By identifying the rectum in each image, which is possible via brightness thresholding, its centroid and the centreline are extracted, similar to [10], and the first registration prior is established (see Fig. 2(b)). Subsequently, the tumour extent is manually segmented, *e.g.* using ITK-Snap, and its centroid, the second prior, is calculated.

2.2 Endoscopic Image Preprocessing

The stereo endoscope used in this study is a Wolf Stereo Endoscope (Richard Wolf, Germany), equipped with two Storz HD camera heads (1920x1080i, Karl Storz, Germany). Access to the rectum is provided through a Wolf TEM Port. Image capture is performed with an NVIDIA Quadro SDI card (NVIDIA Corp., California), at a rate of 25 Hz. The tilted lens of the endoscope observes the intraoperative scene at 50°. In the preprocessing step, the images are rectified and specular reflections are detected via simple thresholding. The regions of specularity, together with the black regions denoting the limits of the field of view, form a mask of pixels that are disregarded in the algorithmic workflow.

2.3 Stereo Reconstruction

The smoothness of the disparity maps generated by the variational algorithm of Chang *et al.* [6] makes that algorithm an appropriate choice for surgical scenarios such as TEM where limited textural information is available. However, even for a relatively small camera motion, the reconstructed scenes demonstrate inconsistencies. As Fig. 3 shows, the camera motion is negligible (zero/black image difference) but the reconstructed point clouds exhibit large variations.

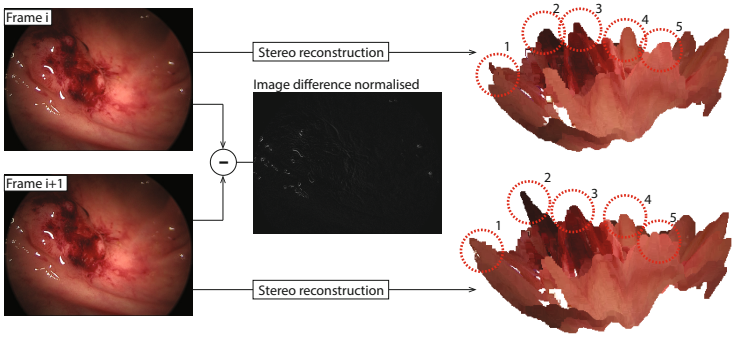


Fig. 3. Example of reconstruction noise for two subsequent frames. The reconstructions demonstrate a large reconstruction variability, examples of which are encircled.

Hence, before the reconstructed views can be used for metrology, view augmentation, or registration, stable reconstruction of the intraoperative view should be performed, which can be handled by camera motion tracking and depth fusion.

2.4 Camera Motion Tracking

Camera tracking can be achieved through Structure-from-Motion or Simultaneous Localisation and Mapping (SLAM). The former does not rely on the 3D information of the stereo endoscope, while the latter requires large travel ranges and features that span the entire field-of-view, which is counter to the narrow field-of-view of TEM images. Alternatively, when dense 3D representations are available, ICP can be used to relate subsequent views via a homogeneous transformation $H \in SE(3)$. ICP, however, cannot cope well with flat surfaces and misestimated correspondences/outliers, both of which are encountered in rectal-surface images. The proposed workflow implements a two-step approach wherein RANSAC provides an estimate of the transformation H and the inliers, and, subsequently, ICP - based on the inlying points - estimates the precise camera transformation between two subsequent frames. These are multiplied along the sequence's timeline to relate to the original coordinate frame.

2.5 Muti-view Fusion

The proposed algorithm achieves both dynamic field-of-view expansion and reconstruction stabilisation. Scene information is recorded in accumulators:

- $\mathcal{X}_{\mathrm{3D}} = \{\boldsymbol{X}^i_{\mathrm{3D}}\}$ stores the coordinates of the scene;
- $\mathcal{C}_{\mathrm{rgb}} = \{\boldsymbol{c}^i_{\mathrm{rgb}}\}$ stores the colour triplet for each reconstructed point;
- $\mathcal{F}_{\mathrm{fa}}$ stores the ID of the first frame of detection of each point;
- $\mathcal{A}_{\mathrm{noa}}$ stores the number of appearances of each point;
- $\mathcal{Z}_{\mathrm{zncc}} = \{z^i_{\mathrm{zncc}}\}$ stores the confidence of the reconstruction from ZNCC.

where $i = 1, \cdots, n$ corresponds to the index of each point in entire scene. Accumulators $\mathcal{F}_{\text{fa}}$ and $\mathcal{A}_{\text{noa}}$ identify inconsistently detected points.

In SLAM-based approaches for dynamic field of view expansion in MIS, *e.g.*, [11], reconstructed scenes are combined based on moving averages, which is inadequate since image regions do not present the same degree of fidelity. Instead, using $\mathcal{Z}_{\text{zncc}}$ leads to information fusion based on matching confidence. For every X^i_{3D} in $\mathcal{X}_{\text{3D}}$, when a new pair of frames is processed:

$$\begin{aligned} z'^i_{\text{zncc}} &= z^i_{\text{zncc}} + \sum z^j_{\text{zncc}} \\ X'^i_{\text{3D}} &= (z^i_{\text{zncc}} X^i_{\text{3D}} + \sum z^j_{\text{zncc}} X^j_{\text{3D}}) / z'^i_{\text{zncc}} \\ c'^i_{\text{rgb}} &= c^i_{\text{rgb}} + \sum z^j_{\text{zncc}} c^j_{\text{rgb}} \\ f'^i_{\text{fa}} &= \min(f^j_{\text{fa}}, \mathcal{N}(f^i_{\text{fa}})) \\ a'^i_{\text{noa}} &= \max(a^j_{\text{noa}}, \mathcal{N}(a^i_{\text{noa}})) + 1, \forall j \in \mathcal{N}(\mathcal{X}^{\text{i}}_{\text{3D}}) \end{aligned} \tag{1}$$

where a $\mathcal{N}(X)$ denotes the neighbours of X based on the latest stereo image pair, where the "neighbour-ness" is calculated both based on euclidean distance and point-to-camera-centre ray orientation. In order words, points that are close both in 3D space and in image space are fused together and their appearance counter is increased. New points that do not match these criteria are added to the accumulators, and points that have not been selected for fusion (*i.e.,* have not reappeared) after a preselected number of frames are rejected as outliers. Fusing both 3D information and colour (c_{rgb}) provides a smoother texture.

2.6 Registration for Intraoperative Information Augmentation

The priors extracted in Sec. 2.1 are used to register the intraluminal views to the global patient anatomy. Aligning the endoscope axis to the centreline of the rectum constrains the endoscope orientation. The viewport orientation, however, needs to be accounted for if the endoscope possesses an inclined field-of-view, as is common in TEM. The remaining degrees-of-freedom of the endoscope are constrained by matching the tumour location in the coordinate frame of the camera with the tumour location in the patient anatomy, which requires the selection of the tumour on an endoscopic image. These constraints initialise the camera coordinate frame within the anatomy. Subsequently, based on the motion tracking, the endoscopic views are related to the anatomy for information augmentation. This information can be presented either overlaid in the endoscopic views, or as a separate entity that the surgeon can navigate through. The former option presents the risk that registration errors may affect the surgeon's perception since he/she is not be able to separate the augmented information from the endoscopic view. Hence, we opt for the latter option which shows the preoperative information based on the endoscope registration as a separate view.

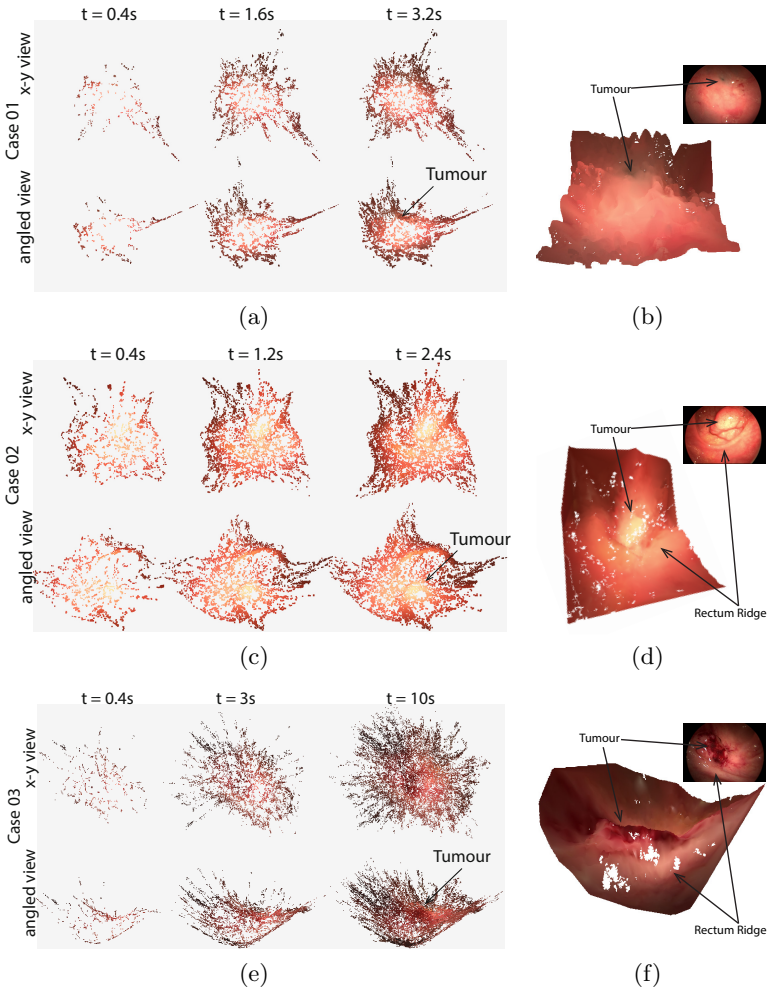
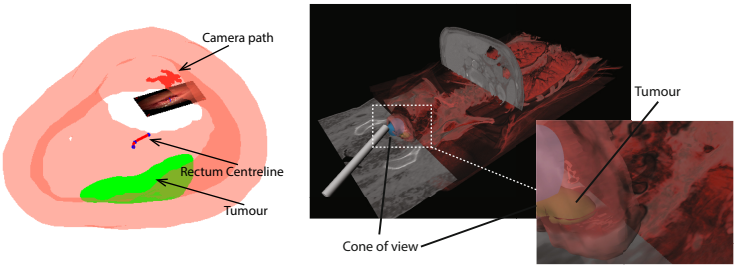


Fig. 4. *In vivo* evaluation of multiview stereo reconstruction: (a), (c), (e) Evolution of the reconstructed point cloud. (b), (d), (f) Interpolated reconstruction and meshing.

3 Results

Results are presented for *in vivo* data acquired from three patients that were admitted for TEM. In the first case, the tumour had been tattooed during a previous excision attempt, and no three dimensional structure was visible. In the second and third cases, a tumour on the order of several centimetres was excised. Prior to data processing, which, at this stage, occurs offline, endoscope calibration was performed using Zhang's technique [12]. First, we demonstrate multiview stereo fusion in TEM for all cases, and, subsequently, the registration and intraoperative view augmentation are described for the third case. In all cases, camera motion was tracked via ICP. Results are quantified in Table 1.



(a) (b)

Fig. 5. (a) Centreline and endoscopic view registration. The red line indicates the tracked camera motion. (b) Augmented views acting as a "Surgical GPS".

Table 1. Quantification of algorithm performance. Values in [mm].

	Case 1	**Case 2**	**Case 3**
Tumour size (real vs virtual)	N/A	32×20 vs. 37×21	34×21 vs. 43×21
ICP convergence	1.1 ± 0.6	2.3 ± 1.3	1.4 ± 0.8
Camera distance	139	32	25

The results of multiview fusion for the first case are shown in Fig. 4(a),(b). The endoscopic images suffer from limited texture, also resulting from the absence of a canonical tumour structure. Within seconds, the proposed algorithm extracts a stable 3D point cloud that captures the surface of the rectum and its tubular structure. In the second case, shown in Fig. 4(c),(d), there is a clear tumour structure. Details of the tumour and the rectum ridge and walls appear in the 3D reconstruction. Similarly, the third case, shown in Fig. 4(e), (f), also exhibits a prominent tumour, and, additionally, substantial vascularisation. Our algorithm allowed dynamic field of view expansion and a good degree of spatial resolution. Fig. 4(f) showcases capturing both the tumour and the rectum ridge in great detail, which contrasts the noisy and varying reconstruction of Fig. 3.

The remaining workflow is demonstrated for the third case. The priors are used to register the endoscope to the preoperative volume (see Fig. 5(a)). The endoscope has a 50° angle-of-view, which is accounted for. The entry angle and camera path are consistent with our observations during surgery. The endoscope, with its conical field-of-view, is rendered on the anatomy (see Fig. 5(b)). Separating the augmented view from endoscopy allows its independent manipulation, and minimises clinical risk arising from registration errors and preop-vs-intraop insuflation. Similarly, the 3D reconstruction is currently not overlaid on the preoperative volume. To conclude, the presented system serves as a "Surgical GPS", indicating points of action without obstructing the clinician's field-of-view.

4 Conclusions and Discussion

This paper presented a novel method for information augmentation and robust 3D reconstruction in TEM. After establishing unique priors that allow easy semi-automated registration of preoperative images to the endoscopic view, it was demonstrated how multi-view stereo fusion can lead to significantly better estimation of the 3D structure of the operative view. Finally, registration of the endoscope to the anatomy allows visualisation of submucosal information.

References

1. Cataldo, P., Buess, G.: Transanal endoscopic microsurery: principles and techniques. Springer, Heidelberg (2009)
2. Dey, D., Gobbi, D.G., Slomka, P.J., Surry, K.J.M., Peters, T.M.: Automatic fusion of freehand endoscopic brain images to three-dimensional surfaces: creating stereoscopic panoramas. IEEE Trans. Medical Imaging 21(1), 23–30 (2002)
3. Pratt, P., Mayer, R., Vale, J., Cohen, D., Edwards, E., Darzi, A., Yang, G.Z.: An effective visualisation and registration system for image-guided robotic partial nephrectomy. J. Robotic Surgery 6(1), 23–31 (2012)
4. Vemuri, A.S., Wu, J.C.H., Liu, K.C., Wu, H.S.: Deformable three-dimensional model architecture for interactive augmented reality in minimally invasive surgery. Surgical Endoscopy 26(12), 3655–3662 (2012)
5. Rohl, S., Bodenstedt, S., Suwelack, S., Kenngott, H., Muller-Stich, B.P., Dillmann, R., Speidel, S.: Dense GPU-enhanced surface reconstruction from stereo endoscopic images from intraoperative registration. Medical Physics 39(3), 1632–1645 (2012)
6. Chang, P.L., Stoyanov, D., Davison, A.J., Edwards, P.E.: Real-time dense stereo reconstruction using convex optimisation with a cost-volume for image-guided robotic surgery. In: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (eds.) MICCAI 2013, Part I. LNCS, vol. 8149, pp. 42–49. Springer, Heidelberg (2013)
7. Goesele, M., Brian, C., Seitz, S.M.: Multi-view stereo revisited. In: IEEE Int. Conf. Computer Vision and Pattern Recognition, pp. 2402–2409 (2006)
8. Merritt, S.A., Khare, R., Bascom, R., Higgins, W.E.: Interactive CT-video registration for the continuous guidance of bronchoscopy. IEEE Trans. Medical Imaging 32(8), 1376–1396 (2013)
9. Mirota, D.J., Uneri, A., Schafer, S., Nithiananthan, S., Reh, D.D., Ishii, M., Gallia, G.L., Taylor, R.H., Hager, G.D., Siewerdsen, J.H.: Evaluation of a system for high-accuracy 3D image-based registration of endoscopic video to C-arm cone-beam CT for image-guided skull base surgery. IEEE Trans. Medical Imaging 32(7), 1215–1226 (2013)
10. Samara, Y., Fiebich, M., Dachman, A.H., Doi, K., Hoffmann, K.R.: Automated centreline tracking of the human colon. In: Medical Imaging, pp. 740–746 (1998)
11. Totz, J., Mountney, P., Stoyanov, D., Yang, G.-Z.: Dense surface reconstruction for enhanced navigation in MIS. In: Fichtinger, G., Martel, A., Peters, T. (eds.) MICCAI 2011, Part I. LNCS, vol. 6891, pp. 89–96. Springer, Heidelberg (2011)
12. Zhang, Z.: A flexible new technique for camera calibration. IEEE Trans. Pattern Analysis and Machine Intelligence 22(11), 1330–1334 (2000)