RECOVERING TISSUE DEFORMATION AND LAPAROSCOPE MOTION FOR MINIMALLY INVASIVE SURGERY

Peter Mountney, Danail Stoyanov and Guang-Zhong Yang

Institute of Biomedical Engineering, Imperial College London, United Kingdom

ABSTRACT

Recent advances in surgical robotics have provided a platform for extending the current capabilities of minimally invasive surgery by incorporating both pre-operative and intra-operative imaging data. In this tutorial paper, we introduce techniques for *in vivo* 3D tissue deformation recovery and tracking, based on laparoscopic or endoscopic images. These optically based techniques provide a unique opportunity for recovering surface deformation of the soft-tissue without the need of additional instrumentation. They can therefore be easily incorporated into the existing surgical workflow. Technically, the problem formulation is challenging due to non-rigid deformation of the tissue and instrument interaction. Current approaches and future research directions in terms of intra-operative planning and adaptive surgical navigation are explained in detail.

Index Terms— image-guidance, robotic assisted minimally invasive surgery, 3D deformation recovery.

1. INTRODUCTION

Over the past two decades, technological innovations have played a major role in reshaping the general practice of surgery. Solid state cameras and fibre optic devices have made Minimally Invasive Surgery (MIS) a reality. In MIS, specialized instruments are inserted into the anatomy through small access ports and operated under remote video guidance. By avoiding large incisions, MIS greatly reduces patient trauma, post-operative recovery period and the risk of comorbidity. These advantages have made MIS a viable treatment option for a wider range of patients [1, 2].

Recently, robotic technologies have been used to overcome the limitations of traditional MIS tools and provide the control and maneuverability required for precise microsurgical tasks. Robotic devices represent one of the most promising enhancements in modern operating theatres for MIS. They facilitate the performance of dexterity demanding procedures with improved repeatability and precision through the use of microprocessor controlled mechanical wrists. By using master-slave set-ups, surgical robots have been shown to significantly improve the ergonomics in the operating theatre, enable the use of motion scaling and provide a unique platform for real-time navigation based on multi-modal patient specific imaging and sensing data.

For performing complex procedures using robotic assisted MIS, medical image computing plays an important role for improving the surgeon's operating capabilities. Despite the advantages of robotic assisted MIS instruments, performing micro-surgical tasks in a highly dynamic environment is challenging. This is reflected in complex procedures such as robotic assisted, beating heart Totally Endoscopic Coronary Artery Bypass (TECAB) surgery, for which, despite the apparent patient benefits, clinical uptake has been slow [3]. While imaging modalities such as intra-operative MRI and CT can provide accurate information about the tissue morphology, they are constrained by the operating environment mainly due to their size and accessibility. Optical techniques based on laparoscopic or endoscopic cameras provide a unique opportunity for recovering the morphology, as well as the structure of the soft-tissue *in situ*. In MIS, recovering tissue deformation is essential for co-registering intra-operative and pre-operative data. It is also important for providing intra-operative guidance and accurately fusing multimodality intra-operative information. With robotic assistance, the recovered tissue deformation can further be used for providing motion stabilization and prescribing dynamic active constraints to avoid critical anatomical structures such as nerves and blood vessels as illustrated in Fig. 1.

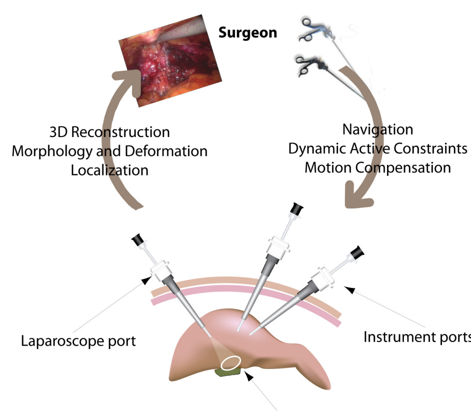

Fig. 1. A schematic diagram showing the information flow in robotic assisted MIS. By using information from the laparoscopic cameras, it is possible to recover tissue deformation in 3D, which permits intra-operative navigation, motion compensation and dynamic active constraints.

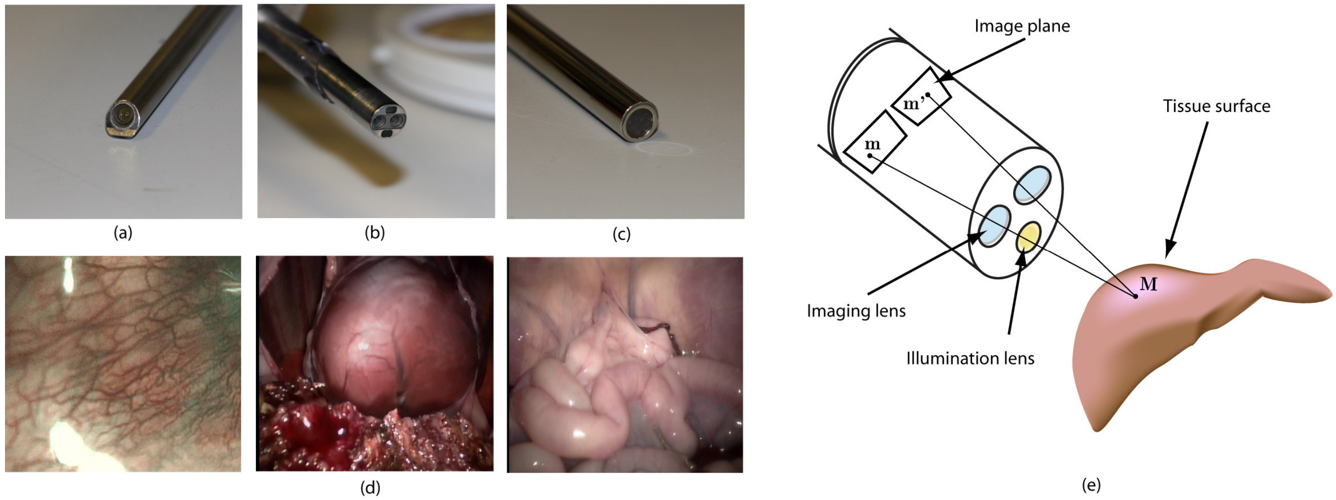

Fig. 2. **(a)** A 30 degree laparoscope, **(b)** a stereo laparoscope with two point light sources, **(c)** a zero degree laparoscope with circular light source **(d)** example images acquired during MIS. **(e)** Schematic of a laparoscope with imaging optics observing a sample of tissue in 3D.

In this tutorial paper, we provide an explanation of the physical configuration of the optical imaging environment in MIS with a geometric camera model and camera calibration. This serves as the basis of techniques for recovering 3D soft tissue deformation and relative pose of the laparoscopic cameras. We describe how these techniques can be used for tissue deformation tracking and 3D reconstruction, with specific focus on the use of a moving camera model for structure recovery. Quantitative validation is discussed to highlight the practical challenges involved for *in vivo* applications. To summarize we discuss the major challenges and future research directions, particularly in dealing with deformable tissue structures.

2. OPTICAL SET-UP

The laparoscope camera used in MIS is typically inserted into the patient via a small incision or natural orifice. The surgeon maneuvers the external, proximal end of the laparoscope to navigate through the body via a video displays. The MIS environment is illuminated with a light source embedded in the laparoscope. Fig. 2 shows the optical configuration of several laparoscopes and example images displayed to the surgeon. Quantitative measurements can be made from laparoscopic images only if the instrument has been accurately modeled and calibrated.

The camera of a laparoscope can be modeled by its optical characteristics called intrinsic parameters and its position and orientation in a world coordinate system called extrinsic parameters [4]. Typically, the pinhole projection model is used to describe the mapping of a 3D point $\mathbf{M} = [X \quad Y \quad Z \quad 1]^T$ in homogeneous coordinates onto the image point $\mathbf{m} = [x \quad y \quad 1]^T$ as a matrix multiplication:

$$\mathbf{m} = \mathbf{K}\begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0} & 1 \end{bmatrix}\mathbf{M} = \mathbf{PM} \qquad (1)$$

where $\mathbf{K}$ is a matrix of the intrinsic camera parameters and $\mathbf{R}$ and $\mathbf{t}$ describe the extrinsic orientation of the device in the world coordinate system. Fig. 2(e) shows a schematic illustration of this model in a stereo configuration. Lens distortion can be effectively modeled using radial and tangential distortion coefficients [5, 6].

In general, the unknown parameters of the laparoscope model are estimated by a preoperative calibration process. To obtain these unknown parameters, certain constraints are usually imposed on the projection of points, with known coordinates in the 3D world, onto the image plane. There are several well established algorithms for this procedure from the computer vision communities and implementations of these methods are available online [7].

After calibration, the metric 3D structure of the surgical scene can be recovered given the correspondence of image primitives ($\mathbf{m}$ and $\mathbf{m}'$ in Fig. 2(e)) among multiple views of the surgical site. This process is called triangulation [4], which is also illustrated in Fig. 2(e).

3. RECOVERING SOFT-TISSUE 3D SHAPE

Recovering 3D information from images is a long standing problem in computer vision. Typical solutions are stimulated by our basic understanding of biological vision systems and the intrinsic relationship of how 2D images are acquired from 3D space. The early work of Marr [8] led to the establishment of Shape-From-X where different visual cues can be used to infer information about the shape and position of objects with respect to the camera. The wealth of research in this area has resulted in many publications [9]. In this section, we will only summarize those approaches reported in MIS.

Approaches to 3D tissue surface reconstruction are summised in Table I and and an example is shown in Fig. 3(d). They can be broadly divided into passive and active techniques. Passive techniques do not introduce additional light or sensing devices into the MIS environment and are

TABLE I SUMMARY OF METHODS USED FOR 3D RECONSTRUCTION FROM IMAGES IN MIS

SFS Assumptions	Stereo Approaches	Active Technique
Orthographic [10]	Computational [11, 12]	Fiducial [13, 14]
Perspective [6, 15, 16]	Surface Priors [17, 18, 19, 20, 21]	One-shot [22]
Illumination [23]	Cue Fusion [6, 24]	Progressive [2, 25, 26]

purely based on the existing images as observed by the operating surgeon. The two main visual cues that have been exploited are shading and stereo.

For Shape-From-Shading (SFS), laparoscopic images do not obey many of the traditional assumptions used to simplify the Bidirectional Reflectance Distribution Function (BRDF). Lambertian reflectance is not compatible with specular reflections, which are common due to the mucus layer of the soft-tissue and the relatively high intensity of the laparoscopic light source. Furthermore, the assumption of a light source located at infinity is not satisfied due to the co-positioning of the light source at the tip of the laparoscope. In addition, the camera cannot be assumed to perform orthographic projection as perspective effects and lens distortions are significant in laparoscopic images.

Therefore, the special optical arrangement between the scope, illumination source and the surgical scene must be used to simplify the image irradiance equation. This was first proposed by Rashid and Berger in 1992 [10] where the light source and the optical centre of the camera were considered to be coincident. This approach was subsequently combined with the assumption that the BRDF is a monotonically decreasing function with respect to the viewing angle [21]. More recent work has expanded the camera projection model to incorporate lens distortion [15] and perspective projection [6, 16]. The assumption of coincident camera and light source positions has also been relaxed [23], although this requires the calibration of the relative positions [27].

One of the main drawbacks of SFS approaches in MIS is that the information recovered is not in a metric coordinate space and only relative surface orientation information can be measured. Passive stereo techniques and SFS are complementary and can be combined to overcome this limitation [24].

Early work on stereo methods in MIS used a simple normalized cross correlation algorithm [11]. Subsequently this was adapted to incorporate hierarchical solutions with a geometric surface prior and to recover the 3D shape of the heart [17, 18, 19]. The use of explicit assumptions (*e.g.* smoothness) about the observed soft-tissue surfaces enables the reconstruction of homogenous tissue regions but does not handle discontinuities arising at instrument-tissue boundaries. To address this issue, methods based on a sparse set of salient features have been used to first recover a sparse 3D reconstruction of the surgical site and then propagate this information to achieve a semi-dense 3D map [28].

It is worth noting that an important feature of MIS images is the abundance of specular reflections. They are view-dependent and can cause significant errors in recovering 3D structure and tracking deformation. It is therefore necessary to correctly identify these regions prior to stereo-correspondence [19, 29]. Alternatively they can be used as constraints when the illumination direction is known or as a starting point in SFS algorithms [24, 27].

The main limitation of passive reconstruction techniques is that they have limited robustness when dealing with the dynamic environment of MIS. For this reason, methods based on fiducial markers or the use of structured lighting have been proposed. Fiducial markers are predominantely used for temporal tissue tracking, which is discussed in more detail in the following section. In terms of structured lighting, an overview of the general techinques is provided in [30]. In surgery, the use of light projection for 3D measurements has attracted extensive attention [25]. For augmented reality, a structured light system was developed to recover the shape of the surgical site [22]. Subsequently, methods based on a laser plane sweeping over the surgical scene have been developed [2, 26]. All of these systems require an additional instrument port, which has not been clinically popular.

More recently, the use of projected coded patterns has been investigated [20] and methods based on time-of-flight technologies have been explored. They have been shown to produce promising results, albeit at limited resolution and frame rates with the current technologies [31].

4. SOFT-TISSUE TRACKING AND MORPHOLOGY ESTIMATION

4.1. Tissue Tracking

The 2D/3D morphology and dynamic motion of soft tissue can be recovered by temporally tracking regions of interest in the image. This approach is illustrated in Fig. 3(a-b) and has been used to recover 3D tissue morphology and deformation in a variety of anatomical regions as summarized in Table II. The problem of locating a region of interest in one image and finding the corresponding region in another is difficult in MIS. This is because MIS images can be low in contrast, noisy and poorly illuminated. The appearance of tissue also varies greatly from homogenous, to highly textured and many regions contain view dependent specular reflections. It is also necessary to deal with occlusion by surgical instruments, image artifacts and dynamic effects such as bleeding and cauterisation smoke. The performance of a region tracking algorithm is largely influenced by how distinguishable the region is from its surroundings. This is affected by what regions are detected for tracking, how the region is represented in image or

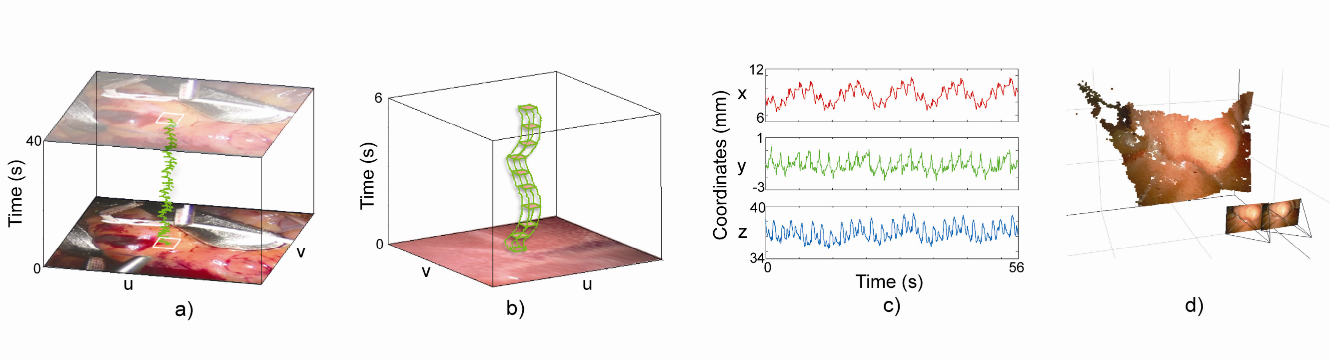

Fig. 3. **(a)** A region tracked on the cardiac surface illustrating motion from the cardiac and respiratory cycles, **(b)** a region tracked on the liver illustrating motion resulting from respiration, **(c)** the tracked 3D motion of a region on the surface of the heart, **(d)** a dense stereo reconstruction of the tissue surface.

feature space, and the matching strategy used to locate the corresponding region in a new image or video frame.

Region detection is the process of identifying salient regions in the image which are distinguished from their surroundings. Passive techniques that detect naturally occurring features such as vessels, corners or blobs [32, 33, 34, 35, 36, 37] are preferred as they do not interfere with the surgeon's view or require user interaction. A comparison of region detectors in MIS is provided in [38]. Tissue can appear homogenous, making region detection challenging. This can be overcome by manually selecting regions [39, 40], using fiducial markers [13, 14] or by marking the tissue of interest (*e.g.* with diathermy) [39, 41]. These active approaches limit the number of tracked regions.

In general, the region can be represented in **image space or feature space**. In image space, the region is simply represented by pixels as an image patch or template [33, 39]. The main problems with this approach are that the representation is not invariant to large image transformation and the image information may not be sufficient to distinguish a region from its surroundings. Alternatively, descriptors can be used to represent the region in feature space. Feature descriptors select what information from the image will be used (*e.g.* grayscale, color, gradient) and how this information will be represented (*e.g.* energy in the co-occurrence matrix [40], non-uniformity of the run length matrix [40], probability density histograms [41], histograms of gradients [34], contours, active appearance models).

Descriptors can be made invariant to image transformation such as scale and rotation through explicit modeling. However, *ad hoc* modeling of non-linear deformation is not trivial. Selecting a feature descriptor is context specific and the performance of descriptors can be affected by low contrast images, changes in illumination and specular highlights, making the selection of a robust descriptor challenging. In [13, 34, 40], machine learning techniques are used to select and combine discriminant descriptors.

For tracking purposes, the region representation can be created on the first frame and remain constant or updated at each frame. Updating enables temporal persistency to be assumed but can lead to error propagation.

Matching strategies can be categorized as recursive methods or 'tracking by detection' [42]. Recursive methods such as Lucas Kanade (LK) attempt to minimize the difference between the region representation and a region in the new image. LK operates in image space and uses the previous location of the region to search for a match locally. This minimization approach works well on small deformations and has been successfully applied to MIS [17, 29, 32, 33, 36, 39, 43]. However, recursive approaches using image space can be sensitive to changes in illumination and specular highlights. They are not well suited to dealing with occlusion and require frame-to-frame updates, leading to error propagation.

In 'tracking by detection,' the region detector is applied to each new video frame to extract a set of potential matches. This set is searched to find a match by comparing feature descriptors. Matching strategies can be one to one (*e.g.* nearest neighbor), one to many (*e.g.* nearest neighbor ratio) or many to many (*e.g.* RANSAC [44]). Detectors and descriptors can be complementary such as SIFT [45] and SURF [46]. 'Tracking by detection' is well suited to dealing with occlusion as no temporal information about the region's location is required. The main problem with the application of these techniques in MIS is related to the *ad hoc* assumptions they make about what image features to use and the expected image transformations. In addition, this approach is dependent on the region detector to correctly locate the region in each new video frame and the global uniqueness of the region as represented in the feature space. 'Tracking by detection' has been applied in MIS [34], and in [35], an approach is proposed which exploits a recursive technique (which requires no prior knowledge) to learn a feature descriptor online.

4.2. Tissue Morphology Modeling

Extracting and modeling the 2D/3D motion of dynamic tissue is an important prerequisite of image guided surgery. The 3D position of tissue, shown in Fig. 3 (c) can be estimated with a stereo laparoscope as described earlier or with a monocular laparoscope based on fiducial markers with known geometry [13]. In practice, tissue deformation can be caused by the cardiac and respiratory cycles, tissue tool interaction or muscular contraction.

Deformation resulting from cardiac and respiratory cycles can be modeled as quasi-periodic or periodic signals [47]. Respiration during MIS is usually regulated by a ventilator, creating an asymmetric periodic signal with an extended exhale phase. For example, the effect of respiration on the liver is modeled in [41] by a Prototype Repetitive Controller and using a weighted-frequency Fourier linear combiner in [48]. The motion of the cardiac surface, however, is more complex as it contains deformations caused by both the cardiac and respiratory cycle. The deformations can be decoupled [33, 35] into their intrinsic components or considered together. A number of approaches have been suggested for modeling cardiac motion, which include Fourier series [49], vector autoregressive models [49], Taken's theorem [36] and Linear Parameter Variant Finite Impulse Response Models [50]. Information from the ventilator and ECG has also been incorporated to increase accuracy [51]. Modeling large scale, non-periodic tissue deformation caused by tissue-tool interaction or muscular contraction is more challenging. It is likely to require the application of statistical shape, finite element and biomechanical models such as those used in needle steering and surgical simulators [52].

5. STRUCTURE AND CAMERA MOTION ESTIMATION

The methods described in the previous sections are based on the assumption that the laparoscopic camera is static. This is not true in practice, particularly with the recent emergence of Natural Orifice Transluminal Endoscopic Surgery (NOTES) or Single Port Access (SPA) techniques. In this section, we will describe two approaches for recovering the structure of the MIS environment, as well as the camera position: Structure from Motion and Simultaneous Localization And Mapping (SLAM). These competing techniques are compared schematically in Fig. 4. Both approaches are based on the assumption that the structure of the environment is relatively stable. It is worth noting that this is a strong assumption for MIS. Nevertheless, these methods have been applied to various parts of the anatomy (Table II) where tissue motion or deformation is minimal. The extension of these techniques to non static environments will be discussed.

5.1. Structure from Motion

Structure from Motion [4] is a computer vision technique developed to recover the structure of a scene and the motion of the camera. A wide variety of approaches exist. However, the basic framework contains three components as illustrated in Fig. 4; 1) image registration and frame-to-frame camera motion estimation; 2) global optimization or bundle adjustment where multiple images are registered; and 3) scene reconstruction.

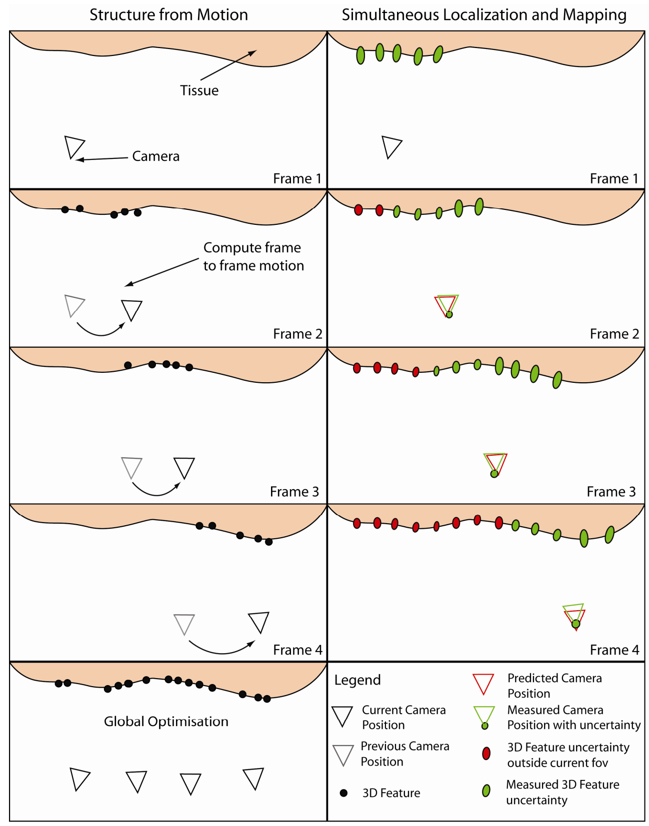

Fig. 4. Illustration of structure and camera motion estimation. Structure from Motion (left) with frame to frame estimation and global optimization. Simultaneous Localization And Mapping (right) with sequential incremental long term mapping, uncertainty estimates, motion prediction and state updates.

Image registration and frame-to-frame motion estimation can be performed in the image space by using direct alignment [53, 54, 55, 56] or in feature space using region matching [57, 58, 59, 60]. Direct alignment uses every pixel in the image and is well suited to environments with sparse regions of interest. However, it requires a large image overlap, suffers from the aperture problem and can be affected by specular highlights. Operating in feature space enables registration with smaller image overlap and non sequential matching. Camera motion is estimated by minimizing the equation [4] defined by the motion model.

The motion model describes the assumptions made about the structure and geometry of the environment. It defines the mathematical relationship between pixels in images captured from different locations. In MIS, planar models have been used on a variety of organs [59, 61, 62, 63, 64] (see Table II), cylindrical models for the esophagus and colon [53, 54, 55, 56] and full projective models for the abdomen, colon [57], heart [58] and bladder [60]. The main problem with Structure from Motion is error propagation

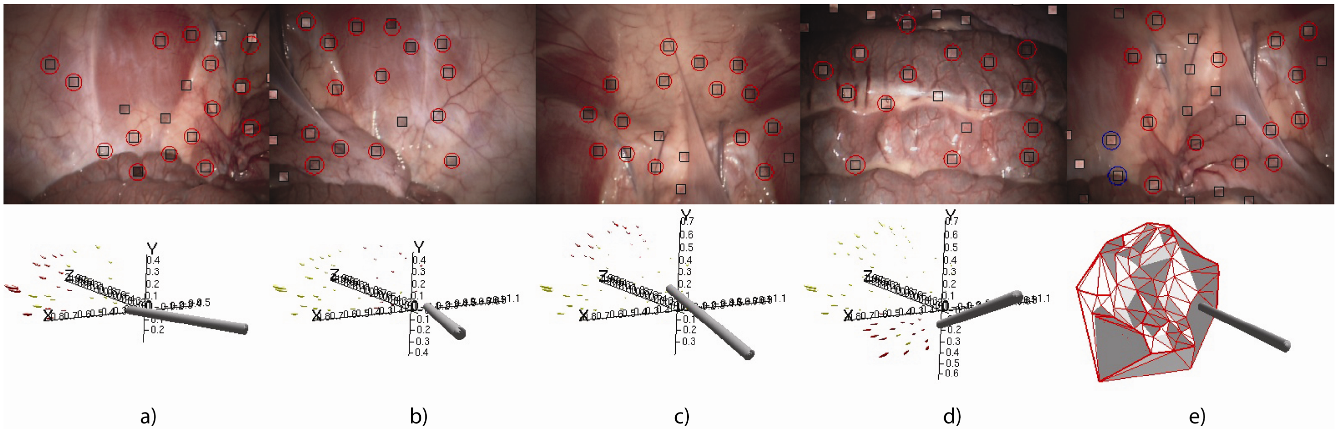

Fig. 5. Laparoscopic SLAM as applied to the abdominal MIS. Top – Laparoscopic video with tracked regions (squares) and projected uncertainly (circles). Bottom – Laparoscope position and **(a-d)** 3D sparse map of tissue with position uncertainties and **(e)** 3D surface mesh of tissue.

caused by the frame-to-frame camera motion estimation. Small errors propagate over time and can cause inaccuracies in the camera and structure estimations. This problem can be addressed using global optimization.

Global optimization is the use of batch operations or bundle adjustment to register multiple images together and find the optimal set of transformations which minimizes error and removes drift. Global optimization with multiple images can be computationally expensive, making it inappropriate for online *in vivo, in situ* applications. Nevertheless, it is suited to offline applications [56, 59, 61].

Scene reconstruction is the process of generating a model of the tissue structure. Given the estimated positions of the camera, scene reconstruction can be performed by matching regions of interest between images. The matched regions are triangulated to estimate 3D points relative to the camera. These points can be meshed or interpolated to create a model of the tissue structure.

The work described above is based on the assumption that the MIS environment is static. Non rigid Structure from Motion has been proposed for tracking faces [65] and clothing [66]. These techniques are based on the factorization method and shape basis representation. They are not suitable for real-time applications as the deformation is dealt with in an offline, global optimization step. Non-rigid Structure from Motion has been applied to the heart [67]. However, it is used to deal with residual motion when constructing a static cardiac surface at a preselected point in the cardiac cycle and not to generate a deforming surface model.

5.2. Simultaneous Localization And Mapping (SLAM)

SLAM has its origin in autonomous robotic navigation. It is designed to solve the problems of consistent incremental environment mapping and localization of a robot within the map. Previously, these had been treated as separate problems where either the map or robot location is assumed to be known. This approach was unsuccessful as neither can be known for certain due to noise in sensor measurements. The solution is to formulate mapping and localization into a single state estimation problem within a probabilistic framework. Originally developed for laser ranger finders and sonar, SLAM has been reformulated for cameras [68].

In MIS, SLAM has been applied to the abdomen [69, 70, 71], esophagus [72] and sinus [73] (in conjunction with pre-operative data) where deformation and tissue motion is minimal. Fig. 5 shows the results of SLAM when applied to laparoscopic surgery, illustrating the 3D map and camera position. The fulcrum effect of the laparoscope is clearly visible. In MIS, the goal is to localize the laparoscope camera and build a map of the tissue surface. A typical feature based SLAM system is illustrated in Fig. 4. The SLAM system alternates between a prediction step, where the motion of the camera is blindly predicted, and an update step, where the map is measured relative to the camera. A vision SLAM system consists of a state vector, a probabilistic framework, feature initialization, a prediction model and a measurement model.

The **state vector** contains a map and the position of the laparoscope camera. The map contains the 3D xyz position of a set of features or points in the environment. The camera's position is represented by xyz position and *roll, pitch, yaw* rotations. In addition, this state vector contains the velocity and angular velocity of the camera. Real-time performance has been demonstrated [68] on sparse maps containing 100 features with full covariance.

The **probabilistic framework** in SLAM enables uncertainty or noise in the system to be modeled. The framework represents the joint probability between the position of the camera and the features in the map at a given point in time. It therefore corresponds to the current estimate of the state vector and the uncertainty in the state estimation. In MIS, the Extended Kalman Filter (EKF), which assumes Gaussian noise, has been employed [69, 70, 71, 72, 74]. The uncertainty in the state estimate is represented in a covariance matrix, which describes the variance from the estimate. In the wider SLAM community, a variety of approaches have been implemented included Unscented Kalman Filters and Rao-Blackwellised particle filters (FastSLAM) [75].

Features initialization is dependent on the optical set-up. In stereoscopic systems [69, 71, 74], features are matched in the left and right images and the 3D position is triangulated relative to the camera. In monocular systems, the 3D position is estimated by matching features temporally and requires the camera motion to be estimated. This is estimated using inverse depth [70, 72] or structure from motion [73]. SLAM uses a full covariance matrix between all features in the map to enable map convergence. For real-time performance, the size of the map is restricted and feature initialization should be carefully managed.

The **prediction model** or motion model describes how the camera is expected to move. This model contains two elements; 1) deterministic element - the motion is estimated based on a sensor (*e.g.* odometry) or an assumption. In [69, 70, 72], a constant velocity constant acceleration model is assumed. 2) Stochastic element – this is a probability distribution represented by a Gaussian or collection of particles. It represents the unknown motion which cannot be easily modeled. A constant velocity, constant acceleration motion model assumes the camera motion will be smooth. This assumption can be violated in both handheld MIS and robotic assisted MIS, thus leading to system failure. The motion of a rigid laparoscope is limited by the fulcrum effect which may help to constrain the problem.

In the **update step**, the predicted state is compared to the measured state. The **measurement model** provides a means of measuring the current state of the system. SLAM measures the location of features in the map relative to the camera. In stereo SLAM, visible features are compared in 3D by stereo region matching and triangulation, while in monocular SLAM visible features are projected onto the camera image plane and regions are matched using measurements in the 2D image plane.

SLAM is a recent success story in mobile robotics largely due to its probabilistic foundations and real-time capabilities. Unlike Structure from Motion, it is naturally suited to returning to previously visited areas and does not require a batch process to converge to an accurate estimation of the environment structure. Practical future work in the application of SLAM to MIS will be focused on creating denser maps covering larger areas, identifying more robust long term features, developing motion models better suited to rapid motion, and recovering from failure. However, the main challenge in the application of SLAM to MIS is the theoretical treatment of deformation.

SLAM has been widely applied to non static civil environments where motion is caused by people and cars. Non static motions are treated as outliers. Outliers can be identified using approaches such as RANSAC [44]. This assumes a global rigidity model and identifies outliers as features which do not fit to the model. This approach relies on parts of the environment being static which may not be the case in MIS. In [77], however, moving objects (cars) are identified and incorporated into the probabilistic framework of SLAM. This work demonstrates that SLAM can be applied without the static assumption by explicitly creating motion models for moving objects. As we have seen in section 4.2, it is possible to estimate motion models representing the morphology of deforming tissue. Future work on deforming SLAM will investigate the incorporation of morphological models into the probabilistic framework.

The output SLAM is generally a sparse set of 3D points representing the structure of the environment. These points can be meshed to create a solid model shown in Fig. 5 (e). Textures from the laparoscopic video can be applied to make the model visually accurate.

5.3 Monocular and Stereo Systems

Structure from Motion and SLAM can be used with either monocular or stereo cameras. Monocular systems are commonly used in operating theatre. However, the number of stereoscopic systems is steadily increasing particularly for robotic assisted MIS. Ideally, the integration of computer vision into the surgical theatre will operate with existing monocular laparoscopes, however, the significant drawback of monocular vision is that acquiring depth information requires camera motion or fiducials of a known size. Therefore, the application of monocular vision in MIS is more limited than stereo.

TABLE II SUMMARY OF TISSUE MORPHOLOGY AND STRUCTURE ESTIMATION METHODS APPLIED IN MIS

Organ	Recovered Scene Geometry	
	Static	**Deforming**
Heart	[11, 58, 67]	[13, 14, 17, 18, 29, 32, 33, 35, 36, 40, 43, 49, 51, 76]
Abdomen / Liver / Gall Bladder / Kidney	[37, 59, 64, 69, 70, 71, 74]	[34, 35, 39, 41]
Colon	[53, 55, 57]	-
Bladder	[60, 61, 62, 63]	-
Esophagus	[54, 56, 72]	-
Sinus	[73]	-

6. VALIDATION AND VERIFICATION

Validation is a crucial step in the evaluation of the discussed methods. Practically, the validation process is challenging due to a lack of ground truth for *in vivo* cases. Experiments are usually performed on numerically simulated data or on phantom models. The ideal metric for measuring error should be Euclidian distances in metric 3D space or in the projected 2D image plane. However, for algorithms where rotations need to be evaluated, as with mosaicing, the exact method for measurement is less well defined [59]. Qualitative evaluations using physiological signal frequency comparisons have been used in the literature [14, 36].

Computer simulations are used to test the numerical stability of algorithms under different levels of modeled noise to establish the baseline performance [41, 58, 60, 69]. However, simulations are not capable of modeling all noise sources and the complexity of the MIS set-up. Therefore, more realistic phantom experiments with known ground truth geometry and motion characteristics are used [41, 58, 60, 74, 76].

In practice, the ground truth for phantom models can be obtained using tomographic scanning and reconstruction techniques or surface scanning using range finders [58, 60, 73, 74, 76]. A practical challenge is to ensure the structural integrity of the model during ground truth acquisition. This is particularly difficult for dynamic models, where the model morphology must be consistently repeatable and synchronized between modalities [28]. Repeatable dynamic motions can be achieved by a combination of mechanical devices and signal generators [19, 28, 41]. High contrast fiducial markers are typically embedded in the phantom enabling registration between the experimental and ground truth coordinate systems. The quality of the resulting alignment is of crucial importance to the values obtained during validation and controlling the error in the ground truth to measurement registration is an important consideration.

Ground truth for the camera or surgical tools can be obtained using optical trackers or electromagnetic tracking devices [73, 74]. They require hand-eye calibration to relate the tracking device and the camera coordinate systems. In addition, controlling the error propagation between the optical, camera and phantom model coordinate systems can often be a practical challenge that needs to be handled with care.

For better visual fidelity, a cadaver can be used in experiments, however, the ground truth for this is difficult to obtain and maintain due to gradual changes in tissue property [19, 74, 76]. The same problems arise during *in vivo* and wet lab experiments with animal studies. In these cases, structural and morphological ground truth is not available and results are usually presented to qualitatively demonstrate practical feasibility rather than metric measurements. Some experimental analysis may be performed by obtaining user feedback [34, 35, 39] and by comparisons with other physiological sensing equipment such as ECG signals [13, 18, 32, 35, 67].

Currently, there is no quality data repository providing a series of datasets for algorithm benchmarking, evaluation and comparison. This makes it particularly difficult for research centers without established MIS infrastructure to work in this area. To address this problem, we have introduced a database containing video data, calibration information and ground truth data http://vip.doc.ic.ac.uk.

7. TECHNICAL CHALLENGES AND CLINICAL APPLICATIONS

The future of navigation and control in robotic assisted MIS is in the intelligent use of pre-operative and intra-operative patient specific data. For intra-operative guidance and applying image guided, dynamic active constraints to avoid critical anatomical structures, it is necessary to develop fast and accurate techniques for 3D surface reconstruction and motion estimation *in situ*. However, the development of computer vision techniques for these dynamic and non-rigid surgical scenes remains challenging.

The robustness of computer vision in MIS is affected by a number of factors including the paucity of features, specular highlights, rapid camera motion, small baseline, tissue deformation, surgical smoke and occlusion. One of the major challenges is the theoretical treatment of tissue deformation, in particular, when combined with camera motion. New methods are required to adapt to the changing environment and to understand the dynamics of the structural morphology in order to anticipate risks and apply motion prediction.

Tissue motion caused by the respiratory and cardiac cycles can be modeled using periodic and quasi periodic models. This is particularly important for highly dynamic procedures around the beating heart where motions arising from the cardiac and respiratory cycles affect the stability of the operating field. In these cases, an important control issue to consider is motion compensation, where the robotic tools are synchronized with the physiological motion to cancel out its rhythmic components. In cardiothoracic surgery, despite the use of mechanical stabilizers the anastomosis site can be unstable and motion compensation is required facilitate less invasive procedures such as TECAB [36, 50]. For the effective deployment of motion compensation, the operating frequency of the robotic device control must be determined to avoid redundancy and signal aliasing. Some preliminary studies indicate that this may be in the region of 100Hz, which requires fast intra-operative processing. In fact the frequency of operation required by the computer-integrated surgical system to update information or robotic control needs to be identified and accuracy requirements clearly defined for different applications [78].

Non-periodic tissue deformation is likely to require the fusion of optical information with prior biomechanical or statistical anatomical models and patient specific information. The problem is complicated further by tissue-tool interaction and topological changes of the tissue due to dissection. There is a critical need for a synergy between the robotic instruments' interactions with tissue and the surgeon. For systems directed at orthopaedic surgery, for example, this can be achieved by imposing active constraints on the tool's motion by using the pre-operatively acquired, segmented and modeled patient data [79].

For soft-tissue procedures, the problem is significantly more complex, largely due to the deformation and dynamics of the anatomy during surgery. In order to impose control constraints on the robotic instruments and to establish "no go" zones for protecting delicate parts of the anatomy, patient specific data must be updated *in vivo* to reflect the current location and changes in anatomical structure. This requires 3D surface recovery in real-time and the subsequent augmentation of geometric and biomechanical models that are physically accurate. By incorporating biomechanical tissue properties, it may be possible, to accurately delineate critical anatomical structures and deliver tactile sensing to reflect the dynamic active constraints imposed. However, a major challenge of physical based modeling such as Finite Element Modeling (FEM) is how to obtain the model parameters using information from medical images to conform to the appearance and behavior of real tissue. By considering the tissue deformation in real-time, the model parameters may be updated to improve the most up-to-date anatomical representation. The modeled tissue can then be used for intra-operative simulations, establishing dynamic active constraints and delivering tactile feedback through the surgical console.

Information regarding the computer-integrated system must be effectively presented to the surgeon with considerations for error and uncertainty in the data visualization. In image-guided surgery, Augmented Reality (AR) is the most common form of data fusion. The clinical benefit of image guidance has been well recognized in neuro and orthopaedic surgeries where the operating field is stable and undergoes only limited deformation [80].

The main problem with implementing AR for surgical navigation in robotically assisted MIS is in the accurate alignment of the computer generated images with the real world. Accurate alignment of the real and virtual objects depends on the accurate tracking of the position and orientation of the viewing source with respect to the anatomy of interest. The complexity of tissue deformation during surgery imposes significant challenges to the AR display and it is a major factor that limits the more widespread use of AR for surgical guidance in soft-tissue procedures. In particular, deformation inhibits two important aspects of navigation: 1) recovery of the motion and the location of the imaging device with respect to the tissue; 2) the computation of the relationship between the pre-operative model of the anatomy and its intra-operative status. The incorporation of 3D shape recovery from stereo video sequences provides the possibility of AR being used for robotic assisted laparoscopic surgeries. An important area of work is how to extend the current state-of-the art in localization techniques to handle deformable environments.

Human computer/robot interaction is another important part of future MIS platforms. Developing interfaces for the surgical theatre is challenging as the surgeons use their hands to perform surgery, making traditional interfaces such as keyboards and mice inappropriate. Foot peddles offer an additional source of input however they are limited in their range of input and in [81], it has been shown that voice control can be employed to position the endoscope. Eye-gaze tracking and brain machine interfaces are elegant solution to the interface problem and have the potential to provide more information than traditional techniques, such as the focus and attention of the surgeon. This information have been exploited for visual servoing in [82] and for motion compensation in [83]. Developing intuitive interfaces for surgery can be challenging as surgical workflow can vary greatly between surgeons. It is envisaged that for complex image guided procedures, a new profession of surgical analysts may be created in future.

8. SUMMARY

Advanced surgical techniques such as image guided navigation with intra-operative motion stabilization and dynamic active constraints have the potential to change the current functional capabilities of MIS. For these techniques to be successful in complex MIS procedures, accurate recovery of 3D tissue structure and morphology, as well as, camera motion estimation are important prerequisites. In this tutorial, we have outlined the current approaches to estimating this information using laparoscopic cameras. We have reviewed optical methods from camera models to tissue morphology recovery techniques for robotic guidance. This is an active research area which has witnessed a significant amount of research output in recent years. It is anticipated that with its maturity, the information derived will play a pivotal role in the future of image guided or robotic assisted MIS.

9. AUTHORS

Peter Mountney (peter.mountney05@imperial.ac.uk) received his MEng (4 years Hons) in Computer Science from the University of Bristol in 2003. He is currently a PhD candidate at Imperial College. His research interests are in the use of laparoscopic cameras as a means of providing measurement of tissue structure during MIS. His current work focuses on image guided surgery, soft tissue tracking, 3D estimation, tissue modeling and SLAM.

Danail Stoyanov (danail.stoyanov@imperial.ac.uk) received the PhD degree from Imperial College London in 2006 for his work on soft-tissue 3D structure and motion recovery in robotic assisted MIS. He is currently a Royal Academy of Engineering/EPSRC Research Fellow at the

Institute of Biomedical Engineering, Imperial College London. His main research interests are in computer vision, image-guided surgery and surgical robotics.

Guang-Zhong Yang (g.z.yang@imperial.ac.uk) received PhD in Computer Science from Imperial College London, London, UK. He is the Founding Director of the Royal Society/Wolfson Medical Image Computing Laboratory, research director of the Institute of Biomedical Engineering, and co-founder of the Hamlyn Centre for Robotic Surgery and Wolfson Surgical Technology Laboratory at Imperial College London. His main research interests include medical imaging, sensing and robotics.

.

REFERENCES

[1] Mack, M. J., "Minimally Invasive and Robotic Surgery," *The Journal of the American Medical Association,* vol. 285, pp. 568-572, 2001.

[2] McKinlay, R., Shaw, M., and Park, A., "A Technique for Real-Time Digital Measurements in Laparoscopic Surgery," *Surgical Endoscopy,* vol. 18, pp. 709–712, 2004.

[3] Damiano, R. J., "Robotics in Cardiac Surgery: The Emperor's New Clothes," *The Journal of Thoracic and Cardiovascular Surgery,* vol. 134, pp. 559-561, 2007.

[4] Hartley, R. and Zisserman, A., *Multiple View Geometry in Computer Vision*: Cambridge Press, 2000.

[5] Heikkila, J. and Silven, O., "A Four-Step Camera Calibration Procedure with Implicit Image Correction," in *Computer Vision and Pattern Recognition*, San Juan, Puerto Rico, pp. 1106-1112, 1997.

[6] Prados, E. and Faugeras, O., "Shape from Shading: A Well Posed Problem?," in *Computer Vision and Pattern Recognition*, San Diego, USA, vol. 2, pp. 870-877, 2005.

[7] Bouguet, J.-Y., "Camera Calibration Toolbox for Matlab," 2004. Available: http://www.vision.caltech.edu/bouguetj/calib_doc/.

[8] Marr, J. L., *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. San Francisco: W. H. Freeman and Company, 1982.

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

[10] Rashid, H. U. and Burger, P., "Differential Algorithm for the Determination of Shape from Shading Using a Point Light Source," *Image and Vision Computing,* vol. 10, pp. 119-127, 1992.

[11] Mourgues, F., Devernay, F., Malandain, G., and Coste-Manière, È., "3d Reconstruction of the Operating Field for Image Overlay in 3d-Endoscopic Surgery," in *International Symposium on Augmented Reality*, New York, USA, 2001.

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

[13] Sauvée, M., Noce, A., Poignet, P., Triboulet, J., and Dombre, E., "Three-Dimensional Heart Motion Estimation Using Endoscopic Monocular Vision System: From Artificial Landmarks to Texture Analysis " *Biomedical Signal Processing and Control,* vol. 2, pp. 199-207 2007.

[14] Ginhoux, R., Gangloff, J. A., Mathelin, M. d., Soler, L., Sanchez, M. M. A., and Marescaux, J., "Beating Heart Tracking in Robotic Surgery Using 500 Hz Visual Servoing, Model Predictive Control and an Adaptive Observer," in *International Conference on Robotics and Automation*, pp. 274-279, 2004.

[15] Forster, C. H. Q. and Tozzi, C., "Towards 3d Reconstruction of Endoscope Images Using Shape from Shading," in *Brazilian Symposium on Computer Graphics and Image Processing*, pp. 90-96, 2000.

[16] Tankus, A., Sochen, N., and Yeshurun, Y., "Perspective Shape-from-Shading by Fast Marching," in *Computer Vision and Pattern Recognition*, pp. 43-49, 2004.

[17] Stoyanov, D., Darzi, A., and Yang, G.-Z., "Dense 3d Depth Recovery for Soft Tissue Deformation During Robotically Assisted Laparoscopic Surgery," in *Medical Image Computing and Computer Assisted Intervention*, pp. 41-48, 2004.

[18] Lau, W. W., Ramey, N. A., Corso, J., Thakor, N. V., and Hager, G. D., "Stereo-Based Endoscopic Tracking of Cardiac Surface Deformation," in *Medical Image Computing and Computer Assisted Intervention*, St. Malo, France, vol. 3217, pp. 494-501, 2004.

[19] Richa, R., Poignet, P., and Liu, C., "Deformable Motion Tracking of the Heart Surface," in *International Conference on Intelligent Robots and Systems*, pp. 3997-4003, 2008.

[20] Albitar, I. C., Graebling, P., and Doignon, C., "Robust Structured Light Coding for 3d Reconstruction," in *International Conference on Computer Vision*, pp. 1-6, 2007.

[21] Okatani, T., . Deguchi, K,. "Shape Reconstruction from an Endoscope Image by Shape from Shading Technique for a Point Light Source at the Projection Center," *Computer Vision and Image Understanding,* vol. 66, pp. 119–131, 1997.

[22] Fuchs, H., Livingston, M. A., Raskar, R., Colucci, D., Keller, K., State, A., Crawford, J. R., Rademacher, P., Drake, S. H., and Meyer, A. A., "Augmented Reality Visualization for Laparoscopic Surgery," in *Medical Image Computing and Computer-Assisted Intervention*, vol. 1496, pp. 934-943, 1998.

[23] Wu, C., Narasimhan, S. G., and Jaramaz, B., "A Multi-Image Shape-from-Shading Framework for near-Lighting Perspective Endoscopes," *International Journal of Computer Vision,* 2009.

[24] Lo, B. P. L., Visentini-Scarzanella, M., Stoyanov, D., and Yang, G.-Z., "Belief Propagation for Depth Cue Fusion in Minimally Invasive Surgery," in *Medical Image Computing and Computer Assisted Intervention*, pp. 104-112, 2008.

[25] Haneishi, H., Ogura, T., and Miyake, Y., "Profilometry of a Gastrointestinal Surface by an Endoscope with Laser Beam Projection," *Optics Letters,* vol. 19, pp. 601-603, 1994.

[26] Hayashibe, M., Suzuki, N., and Nakamura, Y., "Laser-Scan Endoscope System for Intraoperative Geometry Acquisition and Surgical Robot Safety Management," *Medical Image Analysis,* vol. 10, pp. 509-519, 2006.

[27] Stoyanov, D., Elson, D., and Yang, G.-Z., "Illumination Position Estimation for 3d Soft-Tissue Reconstruction in Robotic Minimally Invasive Surgery," in *Intelligent Robots and Systems*, pp. 2628-2633, 2009.

[28] Visentini-Scarzanella, M., Mylonas, G. P., Stoyanov, D., and Yang, G.-Z., "I-Brush: A Gaze-Contingent Virtual Paintbrush for Dense 3d Reconstruction in Robotic Assisted Surgery " in *Medical Image Computing and Computer-Assisted Intervention*, pp. 353-360, 2009.

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

[30] Koninckx, T. P. and Van-Gool, L., "Real-Time Range Acquisition by Adaptive Structured Light," *Pattern Analysis and Machine Intelligence,* vol. 28, pp. 432-445, 2006.

[31] Penne, J., Höller, K., Stürmer, M., Schrauder, T., Schneider, A., Engelbrecht, R., Feußner, H., Schmauss, B., and Hornegger, J., "Time-of-Flight 3-D Endoscopy," in *Medical Image Computing and Computer Assisted Interventions*, pp. 467-474, 2009.

[32] Gröger, M., Ortmaier, T., Sepp, W., and Hirzinger, G., "Tracking Local Motion on the Beating Heart," in *SPIE Medical Imaging Conference*, San Diego, USA, vol. 4681, pp. 233-241, 2002.

[33] Stoyanov, D., Mylonas, G. P., Deligianni, F., Darzi, A., and Yang, G.-Z., "Soft-Tissue Motion Tracking and Structure Estimation for Robotic Assisted Mis Procedures," in *Medical Image Computing and Computer Assisted Intervention*, vol. 3750, pp. 139-146, 2005.

[34] Mountney, P., Lo, B. P. L., Thiemjarus, S., Stoyanov, D., and Yang, G.-Z., "A Probabilistic Framework for Tracking Deformable Soft Tissue in Minimally Invasive Surgery," in *Medical Image Computing and Computer Assisted Intervention*, vol. 2, pp. 34-41, 2007.

[35] Mountney, P. and Yang, G.-Z., "Soft Tissue Tracking for Minimally Invasive Surgery: Learning Local Deformation Online," in *Medical Image Computing and Computer Assisted Intervention*, pp. 364-372, 2008.

[36] Ortmaier, T., Gröger, M., Boehm, D. H., Falk, V., and Hirzinger, G., "Motion Estimation in Beating Heart Surgery," *IEEE Transactions on Biomedical Engineering,* vol. 52, pp. 1729-1740, 2005.

[37] Wengert, C., Bossard, L., Häberling, A., Baur, C., Székely, G., and Cattin, P. C., "Endoscopic Navigation for Minimally Invasive Suturing," in *Medical Image Computing and Computer Assisted Intervention*, vol. 792, pp. 620–627, 2007.

[38] Giannarou, S., Visentini-Scarzanella, M., and Yang, G.-Z., "Affine-Invariant Anisotropic Detector for Soft Tissue Tracking in Minimally Invasive Surgery," in *International Symposium on Biomedical Imaging*, pp. 1059–1062, 2009.

[39] Masson, N., Nageotte, F., Zanne, P., Mathelin, M. d., and Marescaux, J., "Comparison of Visual Tracking Algorithms on in Vivo Sequences for Robot-Assisted Flexible Endoscopic Surgery," in *Engineering in Medicine and Biology Conference*, pp. 5571-5576, 2009.

[40] Noce, A., Triboulet, J., Poignet, P., and Dombre, E., "Texture Features Selection for Visual Servoing of the Beating Heart," in *BioRob* pp. 335 – 340, 2006.

[41] Ott, L., Zanne, P., Nageotte, F., Mathelin, M. d., and Gangloff, J., "Physiological Motion Rejection in Flexible Endoscopy Using Visual Servoing," in *International Conference on Robotics and Automation*, pp. 2928-2933, 2008.

[42] Lepetit, V. and Fua, P., "Monocular Model-Based 3d Tracking of Rigid Objects: A Survey," *Foundations and Trends in Computer Graphics and Vision,* vol. 1, pp. 1–89, 2005.

[43] Gröger, M. and Hirzinger, G., "Optical Flow to Analyse Stabilised Images of the Beating Heart," in *International Conference on Computer Vision Theory and Applications* pp. 237 - 244, 2006.

[44] Torr, P. H. S. and Murray, D. W., "The Development and Comparison of Robust Methods for Estimating the Fundamental Matrix," *Internation Journal of Computer Vision,* vol. 24, pp. 271-300, 1997.

[45] Lowe, D. G., "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision,* vol. 60, pp. 91-110, 2004.

[46] Bay, H., Tuytelaars, T., and van-Gool, L., "Surf: Speeded up Robust Features," in *European Conference on Computer Vision*, vol. 3951, pp. 404-417, 2006.

[47] Riviere, C. N., Gangloff, J., and Mathelin, M. d., "Robotic Compensation of Biological Motion to Enhance Surgical Accuracy," *Proceedings of the IEEE,* vol. 94, pp. 1705-1716, 2006.

[48] Riviere, C., Thakral, A., Iordachita, I. I., Mitroi, G., and Stoianovici, D., "Predicting Respiratory Motion for Active Canceling During Percutaneous Needle Insertion," in *Engineering in Medicine and Biology Conference*, pp. 3477-3480, 2001.

[49] Richa, R., Bo, A. P. L., and Poignet, P., "Motion Prediction for Tracking the Beating Heart," in *Engineering in Medicine and Biology Conference*, pp. 3261-3264, 2008.

[50] Ginhoux, R., Gangloff, J., Mathelin, M. d., Soler, L., Sanchez, M. M. A., and Marescaux, J., "Active Filtering of Physiological Motion in Robotized Surgery Using Predictive Control," *IEEE Transactions on Robotics,* vol. 21, pp. 67-79, 2005.

[51] Cuvillon, L., Gangloff, J., Mathelin, M. d., and Forgione, A., "Toward Robotized Beating Heart Tecabg: Assessment of the Heart Dynamics Using High-Speed Vision," in *Medical Image Computing and Computer Assisted Intervention*, vol. 2, pp. 551-558, 2005.

[52] Misra, S., Ramesh, K. T., and Okamura, A. M., "Modeling of Tool-Tissue Interactions for Computer-Based Surgical Simulation: A Literature Review," *Presence: Teleoperators and Virtual Environments,* vol. 17, pp. 463-491, 2008.

[53] Zhou, J., Das, A., Li, F., and Li, B., "Circular Generalized Cylinder Fitting for 3d Reconstruction in Endoscopic Image Based Mrf," in *Computer Vision and Pattern Recognition Workshops*, pp. 1-8, 2008.

[54] Seibel, E. J., Carroll, R. E., Dominitz, J. A., Johnston, R. S., Melville, C. D., Lee, C. M., Seitz, S. M., and Kimmey, M. B., "Tethered Capsule Endoscopy, a Low-Cost and High-Performance Alternative Technology for the Screening of Esophageal Cancer and Barrett's Esophagus," *IEEE Transactions on Biomedical Engineering,* vol. 55, pp. 1032-1042, 2008.

[55] Seshamani, S., Lau, W., and Hager, G., "Real-Time Endoscopic Mosaicking," in *Medical Image Computing and Computer Assisted Intervention*, vol. I, pp. 355-363, 2006.

[56] Carroll, R. E. and Seitz, S. M., "Rectified Surface Mosaics," in *International Conference on Computer Vision*, pp. 1-8, 2007.

[57] Koppel, D., Chen, C.-I., Wang, Y.-F., Lee, H., Gu, J., Poirson, A., and Wolters, R., "Toward Automated Model Building from Video in Computer-Assisted Diagnoses in Colonoscopy," in *In Proc SPIE*, vol. 6509, 2007.

[58] Hu, M., Penney, G. P., Edwards, P. J., Figl, M., and Hawkes, D. J., "3d Reconstruction of Internal Organ Surfaces for Minimal Invasive Surgery," in *Medical Image Computing and Computer Assisted Intervention*, vol. 1, pp. 68-77, 2007.

[59] Atasoy, S., Noonan, D. P., Benhimane, S., Navab, N., and Yang, G.-Z., "A Global Approach for Automatic Fibroscopic Video Mosaicing in Minimally Invasive Diagnosis.," in *Medical Image Computing and Computer Assisted Intervention*, vol. I, pp. 850-857, 2008.

[60] Wu, C.-H., Sun, Y.-N., and Chang, C.-C., "Three-Dimensional Modeling from Endoscopic Video Using Geometric Constraints Via Feature Positioning," *IEEE Transactions on Biomedical Engineering,* vol. 54, pp. 1199-1211, 2007.

[61] Miranda-Luna, R., Daul, C., Blondel, W. C. P. M., Hernandez-Mier, Y., Wolf, D., and Guillemin, F., "Mosaicing of Bladder Endoscopic Image Sequences: Distortion Calibration and Registration Algorithm," *IEEE Transactions on Biomedical Engineering,* vol. 55, pp. 541-553, 2008.

[62] Olijnyk, S., Mier, Y. H., Blondel, W. C. P. M., Daul, C., Wolf, D., and Bourg-Heckly, G., "Combination of Panoramic and Fluorescence Endoscopic Images to Obtain Tumor Spatial Distribution Information Useful for Bladder Cancer Detection," *Progress in Biomedical Optics and Imaging,* vol. 8, 2007.

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

[64] Igarashi, T., Suzuki, H., and Naya, Y., "Computer Based Endoscopic Image Processing Technology for Endourology and Laparoscopic Surgery," *International Journal of Urology,* pp. 533–543, 2009.

[65] Brand, M., "Morphable 3d Models from Video," in *Computer Vision and Pattern Recognition*, vol. 2, pp. 456-463, 2001.

[66] Torresani, L., Yang, D. B., Alexander, E. J., and Bregler, C., "Tracking and Modeling Non-Rigid Objects with Rank Constraints," in *Computer Vision and Pattern Recognition*, vol. 1, pp. 493-500, 2001.

[67] Hu, M., Penney, G. P., Rueckert, D., Edwards, P. J., Bello, R., Casula, R., Figl, M., and Hawkes, D. J., "Non-Rigid Reconstruction of the Beating Heart Surface for Minimally Invasive Cardiac Surgery," in *Medical Image Computing and Computer Assisted Intervention*, pp. 34-42, 2009.

[68] Davison, A. J., Reid, I., Molton, N., and Stasse, O., "Monoslam: Real-Time Single Camera Slam," *Pattern Analysis and Machine Intelligence,* vol. 29, pp. 1052–1067, 2007.

[69] Mountney, P., Stoyanov, D., Davison, A. J., and Yang, G.-Z., "Simultaneous Stereoscope Localization and Soft-Tissue Mapping for Minimal Invasive Surgery," in *Medical Image Computing And Computer Assisted Intervention*, vol. 1, pp. 347-354., 2006.

[70] Garcıa, O., Civera, J., Gueme, A., Munoz, V., and Montiel, J. M. M., "Real-Time 3d Modeling from Endoscope Image Sequences," in *International Conference on Robotics and Automation Workshop on Advanced Sensing and Sensor Integration in Medical Robotics*, 2009

[71] Mountney, P. and Yang, G.-Z., "Dynamic View Expansion for Minimally Invasive Surgery Using Simultaneous Localization and Mapping," in *Engineering in Medicine and Biology Conference*, pp. 1184-1187, 2009.

[72] Castaneda, V., Atasoy, S., Mateus, D., Navab, N., and Meining, A., "Reconstructing the Esophagus Surface from Endoscopic Image Sequences," in *Russian Bavarian Conference on Bio-Medical Engineering*, 2009.

[73] Burschka, D., Li, M., Ishii, M., Taylor, R., and Hager, G. D., "Scale-Invariant Registration of Monocular Endoscopic Images to Ct-Scans for Sinus Surgery," *Medical Image Analysis,* vol. 9, pp. 413 - 426, 2005.

[74] Noonan, D., Mountney, P., Elson, D., Darzi, A., and Yang, G.-Z., "A Stereoscopic Fibroscope for Camera Motion and 3d Depth Recovery During Minimally Invasive Surgery.," in *International Conference on Robotics and Automation*, pp. 4463-4468, 2009.

[75] Thrun, S., Burgard, W., and Fox, D., *Probabilistic Robotics (Intelligent Robotics and Autonomous Agents)*: The MIT Press, 2005.

[76] Richa, R., Poignet, P., and Liu, C., "Efficient 3d Tracking for Motion Compensation in Beating Heart Surgery.," in *Medical Image Computing and Computer Assisted Intervention*, pp. 684-691, 2008.

[77] Wang, C. C., Thorpe, C., and Thrun, S., "Online Simultaneous Localization and Mapping with Detection and Tracking of Moving Objects: Theory and Results from a Ground Vehicle in Crowded Urban Areas," in *International Conference on Robotics and Automation*, vol. 1, pp. 842-849, 2003.

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

[79] Davies, B., Jakopec, M., Harris, S. J., Rodriguez, Y., Baena, F., Barrett, A., Evangelidis, A., Gomes, P., Henckel, J., and Cobb, J., "Active-Constraint Robotics for Surgery," *Proceedings of the IEEE,* vol. 94, pp. 1696-1704, 2006.

[80] Peters, T. M., "Image-Guidance for Surgical Procedures," *Physics in Medicine and Biology,* vol. 51, pp. R505-R540, 2006.

[81] Jacobs, L. K., Shayani, V., and Sackier, J. M., "Determination of the Learning Curve of the Aesop Robot," *Surgical Endoscopy,* pp. 54-55, 1997.

[82] Noonan, D. P., Mylonas, G. P., Darzi, A., and Yang, G.-Z., "Gaze Contingent Articulated Robot Control for Robot Assisted Minimally Invasive Surgery," in *Intelligent Robots and Systems*, pp. 1186-1191, 2008.

[83] Mylonas, G. P., Stoyanov, D., Deligianni, F., Darzi, A., and Yang, G.-Z., "Gaze-Contingent Soft Tissue Deformation Tracking for Minimally Invasive Robotic Surgery," in *Medical Image Computing and Computer Assisted Intervention*, vol. 3749, pp. 843-850, 2005.