Optical techniques for 3D surface reconstruction in computer-assisted laparoscopic surgery

MAIER-HEIN, L., MOUNTNEY, P., BARTOLI, A., ELHAWARY, H., ELSON, D., GROCH, A., KOLB, A., RODRIGUES, Marcos <http://orcid.org/0000-0002-6083-1303>, SORGER, J., SPEIDEL, Suzanne and STOYANOV, D.

This document is the Submitted Version

Citation:

Optical Techniques for 3D Surface Reconstruction in Computer-assisted Laparoscopic Surgery

L. Maier-Hein[a,*], P. Mountney[b], A. Bartoli[c], H. Elhawary[d], D. Elson[e], A. Groch[a], A. Kolb[f],
M. Rodrigues[g], J. Sorger[h], S. Speidel[i], D. Stoyanov[j,*]

[a]*Div. of Medical and Biological Informatics, German Cancer Research Center (DKFZ), Germany*
[b]*Siemens Corporate Research & Technology, Princeton, NJ, USA*
[c]*Advanced Laparoscopy and Computer Vision group, Université d'Auvergne and CNRS, France*
[d]*Philips Research North America, New York, NY, USA*
[e]*Department of Surgery and Cancer, Imperial College London, UK*
[f]*Institute for Vision and Graphics, University of Siegen, Germany*
[g]*Geometric Modelling and Pattern Recognition Research Group, Sheffield Hallam University, UK*
[h]*Intuitive Surgical, Inc., Sunnyvale, CA, USA*
[i]*Institute for Anthropomatics, Karlsruhe Institute of Technology, Germany*
[j]*Centre for Medical Image Computing (CMIC) and Dept. of Computer Science, University College London, UK*

Abstract

One of the main challenges in computer-assisted surgery (CAS) is the intra-operative imaging of soft-tissue morphology and motion. This information is a prerequisite for the registration of multi-modal patient-specific data to enhance the surgeon's navigation capabilites by observing beyond the exposed tissue surface and to provide intelligent control of roboticized instruments. In minimally invasive surgery (MIS), optical techniques are an increasingly attractive approach for *in vivo* 3D reconstruction of the soft-tissue surface geometry. This paper reviews the various state-of-the-art methods for optical intra-operative 3D reconstruction in laparoscopic surgery and discusses the technical challenges and future perspectives towards clinical translation. With the recent paradigm shift of interventional healthcare delivery towards MIS and rapid developments in 3D optical imaging, this is a timely discussion of technologies that could facilitate complex CAS procedures in dynamic and deformable anatomical regions.

Keywords: Surface Reconstruction, Surgical Vision, Stereoscopy, Shape-from-Motion, Shape-from-Shading, Simultaneous Localization and Mapping, Structured Light, Time-of-Flight, Laparoscopy, Computer-assisted Surgery, Computer-assisted Interventions, Intra-operative Registration, View Enhancement, Biophotonics

1. Introduction

Today, numerous diseases are diagnosed or treated using interventional techniques to access the internal anatomy of the patient. While open surgery involves cutting the skin and dividing

---

*Corresponding authors: Lena Maier-Hein (l.maier-hein@dkfz-heidelberg.de) and Danail Stoyanov (danail.stoyanov@ucl.ac.uk)

the underlying tissues to gain direct access to the surgical target, minimally invasive surgery (MIS) is performed through small incisions in order to reduce surgical trauma and morbidity. The term *laparoscopic* surgery refers to MIS performed in the abdominal or pelvic cavities. The abdomen is usually insufflated with gas to create a working volume (*pneumoperitoneum*) into which surgical instruments can be inserted via ports. As direct viewing of the surgical target is not possible, an endoscopic camera (*laparoscope*) generates a 2D view of the anatomical structures and of the surgical instruments. In contrast to open surgical procedures, MIS provides the surgeon with a restricted, smaller view of the surgical field, which can be difficult to navigate for surgeons only trained in open surgery techniques. To compound the visual complexity of MIS, laparoscopic instruments are operated under difficult hand-eye ergonomics and usually provide only four degrees of freedom (DOF) which severely inhibits the dexterity of tissue manipulation.

To improve the visualization capabilities during MIS, recent developments in medical imaging and image processing have opened the way for computer-assisted surgery (CAS) (Yaniv and Cleary, 2006; Cleary and Peters, 2010) in which computer systems support the physician by providing highly precise localization information about the patient anatomy relative to the interventional instruments. The ergonomics of the MIS operating room mean that there is a natural interface between the surgeon and the patient as the surgical site is inherently displayed on a digital screen. This alleviates the difficuty of providing overlay with specialized hardware to visualize the computed anatomical information over the surgical site as for example in percutaneous procedures (Fichtinger et al., 2005). One of the main difficulties to be addressed in soft-tissue CAS is the fast, accurate and robust acquisition of the patient anatomy during surgery. For Augmented Reality (AR) visualization of subsurface anatomical details overlayed on the laparoscopic video, intra-operative 3D data has to be registered non-rigidly to 3D pre-procedural planning images and models. Tomographic intra-operative imaging modalities, such as ultrasound (US), intra-operative computed tomography (CT) and interventional magnetic resonance imaging (iMRI) have been investigated for acquiring detailed information about the anatomic morphology. However, there are significant technological challenges, costs and risks associated with real-time image acquisition in a surgical theatre or interventional radiology suite with traditional instrumentation and providing images with acceptable signal-to-noise ratio (SNR).

In MIS, an increasingly attractive approach involves 3D reconstruction of soft-tissue surfaces using the endoscope itself by interpreting the properties or geometry of light reflecting off the surfaces at the surgical site. Optical techniques for 3D surface reconstruction can roughly be divided into two categories (Mirota et al., 2011): *passive* methods that only require images, and *active* methods that require controlled light to be projected into the environment. Passive methods include *stereoscopy*, monocular *Shape-from-X (SfX)* and *Simultaneous Localization and Mapping (SLAM)* while the most well-known active methods are based on *structured light* and *Time-of-Flight (ToF)*. Both active and passive technologies have found successful applications in a wide spectrum of fields including domestic and industrial robotics, and the film and games industries. Reconstruction of the patient anatomy for MIS, however, poses several specific challenges that have not yet been solved. While many applications focus on the 3D reconstruction of static scenes (Mountney et al., 2010), the methods applied in MIS must be able to cope with a *dynamic* and *deformable* environment. Furthermore, tissue may have *homogeneous texture* making automatic salient feature detection and matching difficult. The criticial nature of surgery

means that techniques must have *high accuracy and robustness* in order to ensure patient safety. This is particularly challenging in the presence of specular highlights, smoke, and blood, all of which occur frequently in laparoscopic interventions. New technologies in the operaring room also require *seamless integration* into the clinical workflow with minimum setup and calibration times. Finally, *miniaturization* is a challenging issue considering that methods relying on *triangulation* (cf. sec. 2), such as stereoscopy and structured light, require a certain distance (*baseline*) between the optical centers of the two cameras or the camera and the projector. Because reconstruction accuracy increases with the length of the baseline, there is a tradeoff between compactness and reconstruction quality. Table 1 summarizes and compares the most well-known 3D surface reconstruction techniques in MIS.

In this article we report a comprehensive review of the literature for optical 3D reconstruction in MIS that summarizes the state-of-the-art for the different techniques and identifies the main technical challenges as well as future perspectives for the field. Recent reviews have discussed the role of computer vision for MIS (Mountney et al., 2010; Mirota et al., 2011; Stoyanov et al., 2012) as well as for surgical navigation (Baumhauer et al., 2008; Nicolau et al., 2011; Cleary and Peters, 2010). Mirota et al. (Mirota et al., 2011) include an excellent review on the general working principles of surface reconstruction. With this paper, which evolved from a tutorial on 3D surface reconstruction organized at the 14th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI) 2011, we address a technical audience that is already familiar with most of the fundamental concepts in surgical navigation and computer vision. In contrast to previous papers that focus on the underlying *basic* principles of surface reconstruction, we aim to provide a deeper insight into the state-of-the-art in this field. In sections 2-6, we review the relevant literature on laparoscopic 3D surface reconstruction based both on passive (sec. 2-4) and active (sec. 5-6) illumination techniques. Each reconstruction method is introduced by a brief description of the working principle, a general review of the state-of-the-art of the technique as well as a summary and discussion in the context of laparoscopic surgery. To relate the underlying technology to CAS, we also present possible clinical applications for surface reconstruction (sec. 7) and conclude with a discussion of the technical challenges and future perspectives toward clinical translation (sec. 8).

2. Stereoscopy

2.1. Introduction

The 3D spatial understanding capabilities of the human visual system are heavily reliant on binocular vision and stereopsis (Marr and Poggio, 1979; Marr, 1983). By observing a scene from two distinct viewpoints, parallax between the observations provides a strong cue about the distance between the world and the observer. Objects close to the eye have larger binocular parallax than those further away and this can be used to infer metric distances by using the geometry of triangles and a process called *triangulation*. In MIS, stereo laparoscopes, as shown in Fig. 1(a-b), have been introduced to provide the surgeon with a 3D view of the operating site because various studies have shown that the loss of depth cues on 2D monitors impairs the surgeon's control of the instruments and that microsurgical tasks can be performed more easily with depth perception (Taffinder et al., 1999). When a stereo laparoscope is used for surgery,
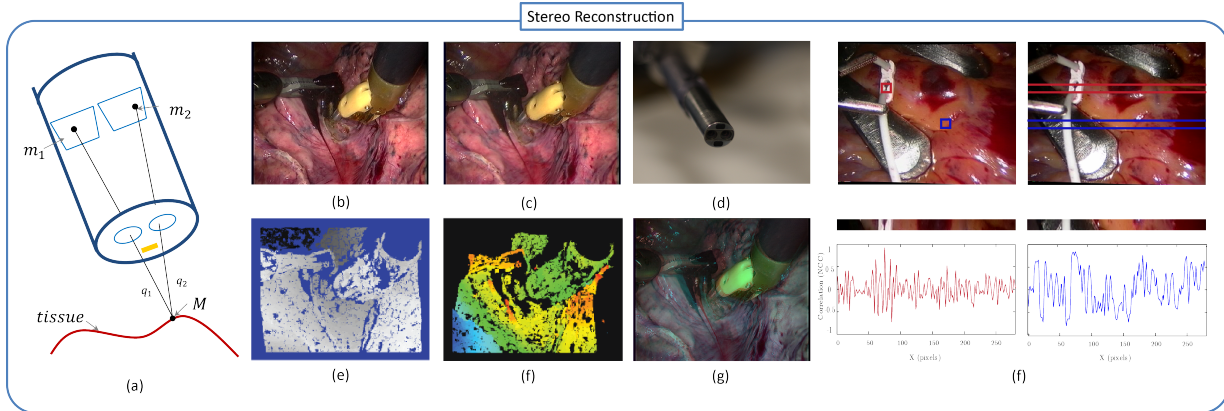
Figure 1: (a) Principle of 3D surface reconstruction based on stereo vision. A point on the soft-tissue surface, M, is projected to image points m_1 and m_2 on the two image planes by the line-of-sight rays q_1 and q_2 respectively. With a geometrically calibrated device the line-of-sight ray for each image pixel is known. For two corresponding pixels in the stereoscopic image pair, the metric position of surface points can therefore be recovered by computing the ray intersection. (b-d) Two images obtained using a stereo laparoscope during robotic assisted surgery on the lung and an image of the device. (e) Disparity image obtained using the images in (b) and (c) with the algorithm reported in (Stoyanov et al., 2010) where lighter colours are closer to the camera. (f) 3D motion of the surface in images (b) and (c) obtained between two time frames using stereoscopic scene flow where warmer colours represent larger motion (Stoyanov, 2012a). (g) An overlay of the stereoscopic image pair illustrating the parallax between the stereo views. (h) Rectified stereo pair obtained during robotic beating heart surgery with two plots illustrating the correspondence problem. The two strips represent the image search space in 1D for a correlation window, and the plots show the correlation metric for a window highlighted in the images above sliding along the search space. The left plot shows a unique correct match in the similarity measure and the one on the right shows an incorrect one with multiple candidate matches.

computer vision techniques can be used to obtain the tissue surface geometry from the pair of images as a passive technique requiring no additional light or hardware to enter the patient.

The basic principle of stereo reconstruction is illustrated in Fig. 1(a) and can be broken down into the following steps: calibrating the cameras; acquiring multiple images of the scene; establishing stereo correspondences of points in the images; structure triangulation using the known geometric properties of the cameras; and structure refinement using filtering or priors on the shape of objects. Stereo relies on the parallax between different observation points and the geometry of projection to intersect line-of-sight rays of corresponding image pixels and compute the 3D position of the source point on the tissue surface. The process requires knowledge of the intrinsic camera parameters describing the geometry of pinhole projection such that image pixels can be back-projected to line-of-sight rays (Faugeras, 1993; Hartley and Zisserman, 2003). The projection process can be expressed as a matrix multiplication in each camera:

$$m_1 = P_1 M = K_1[I|0]M \quad and \quad m_2 = P_2 M = K_2[R|t]M$$

where points m_1 and m_2 are homogeneous vectors in pixel coordinates and world point M is a homogeneous point in the metric coordinate system. The intrinsic camera parameters are encapsulated by matrix K_i and the camera pose in a reference coordinate system is given by

4

the rotation matrix R and the camera center $t = -RC$. Note that one camera is usually chosen as the reference coordinate system as above and hence has $R = I$ and $C = 0$. All projection parameters can be combined in one matrix P and are usually obtained in a pre-operative calibration procedure using calibration objects with precisely known geometry (Zhang, 2000). Several toolboxes are available online for performing this step (Bouguet, 2012; Stoyanov et al., 2012; Sepp and Fuchs, 2012). Once calibration is known, the projection equations can be rearranged to describe the direction of the rays q_1 and q_2 shown in Fig. 1. The rays are unlikely to intersect at M due to noise but an estimate of the 3D point can be obtained by finding the midpoint of the shortest line between the two rays. Because structure triangulation is straightforward with calibrated stereo cameras the main problem of stereo reconstruction is establishing correspondence between the images. The correspondence problem involves determining image primitives (pixels or higher level features) across image pairs that are projections of the same point in 3D space. It is typically approached in either a sparse manner where a number of salient regions (features) are detected and matched using some strategy (Ullman, 1979) or as a dense problem where correspondence is computed for every image pixel (Barnard and Fischler, 1982). For calibrated cameras, it is common to consider only dense techniques as stereo images can be rectified to align vertically according to their epipolar geometry[1] and hence reduce the search space for correspondence to 1D. The dense correspondence problem is often referred to as computational stereo. With rectified images its solution can be shown as a disparity map (shown in Fig. 1(f)), which represents the parallax motion of pixels.

2.2. State-of-the-art

To compute the disparity map, computational stereo algorithms can typically be broken down into four stages as described by (Scharstein and Szeliski, 2002): cost computation, cost aggregation, disparity computation and optimization, and disparity refinement. Early work in the field focused on local techniques with simple winner-takes-all (WTA) strategies for selecting the disparity at image pixels based on the cost function measuring the similarity between pixels or windows around pixels in the stereo pair (Barnard and Fischler, 1982). While the limitation of local approaches is that they fail to exploit regional constraints, recent works have shown that even WTA can produce compelling results when combined with biologically inspired cost aggregation (Nalpantidis and Gasteratos, 2010; Yoon and Kweon, 2006). Another approach to enforcing more global constraints is through optimization of the disparity cost function over image scan lines using techniques such as dynamic programming (DP) (Criminisi et al., 2007) or over the entire image using belief propagation (Tappen and Freeman, 2003), graph cuts (Kolmogorov et al., 2008) or variational methods (Swirski et al., 2011). More recently, the disparity optimization and refinement stages usually combine explicit processing for object recognition (Bleyer et al., 2011), region segmentation (Bleyer et al., 2010) and fitting for structures such as planes (Yang et al., 2008) which are common in man-made environments. While most algorithms assume Lambertian reflectance to simplify similarity calculations, methods for handling view dependent

---

[1]The epipolar geometry describes the projective geometry relationships between two views of a scene. Rectification is the process of transforming image pixel coordinates to display the epipolar relationship as vertical alignment between the two views (Faugeras, 1993; Hartley and Zisserman, 2003).

specular highlights have also been reported (Zhou and Kambhamettu, 2006). With numerious algorithms reported each year, computational stereo is a very mature method for 3D reconstruction from images, and the above paragraphs have only outlined some of the main lines of thought in the field as a full review is impractical and beyond the scope of this article. The main remaining challenge is to build robust systems that perform reliably in practical applications where changes in the environment or imaging setup do not influence the quality of reconstruction. For a detailed perspective on the state-of-the art in computational stereo techniques outside surgery, the reader is referred to two comprehensive reviews of the field up to 2003 (Scharstein and Szeliski, 2002; Brown et al., 2003). For the most recent advances we refer the reader to the *Middlebury Stereo Vision*[2] repository of data with ground truth and evaluation metrics which has served the community as a baseline for algorithm performance over the past decade.

2.3. Application to Laparoscopy

The first work reporting stereoscopic depth reconstruction in MIS used a hardware implementation of a dense computational stereo algorithm (Devernay et al., 2001). Using normalized cross-correlation as a similarity metric for cost computation and a WTA strategy without any global optimization the algorithm was reported to operate at 1 Hz using field programmable gate arrays (FPGA). Semi-global optimization using DP was developed in (Hager et al., 2007) and used to register the depth map acquired during surgery to pre-operative models in robotic partial nephrectomy (Su et al., 2009). The run-time of the algorithm was around 10 Hz with a central processing unit (CPU) implementation. Faster run-times have recently been reported at around 30 Hz with a global or semi-global optimization strategy in addition to bilateral disparity filtering and meshing (Röhl et al., 2012). This method has been used for registering endoscopic video images to biomechanical models of the liver. Another technique also operating at near real-time frame rates that uses local optimization to propagate disparity information around correspondence seeds obtained using feature matching as developed in (Stoyanov et al., 2005b) has also been shown to perform well for endoscopic images (Stoyanov et al., 2010). This method has the desirable property of readily ignoring regions with highlights, occlusions or high uncertainty and has also recently been extended to recover temporal motion as well as 3D shape (Stoyanov, 2012a). An extension of the algorithm showing reduced noise artefacts has also been presented (Bernhardt et al., 2012). To boost the computational performance of reconstruction, algorithms using the graphics processing unit (GPU) have recently been reported (Kowalczuk et al., 2012; Richa et al., 2010; Röhl et al., 2012). These approaches rely on executing computationally expensive elements of the algorithm, such as cost computation simultaneously on multiple cores of the GPU.

In some MIS applications, a large part of the surgical field of view (FoV) is composed of a single tissue surface, such as in endoscopic beating heart surgery, for example. For this application, several stereoscopic techniques have been proposed for reconstructing the cardiac surface shape and subsequently tracking its motion and deformation using a geometric surface model as a smoothness constraint (Lau et al., 2004; Stoyanov et al., 2004). These approaches do not

---

[2]http://vision.middlebury.edu/stereo

solve the correspondence problem like conventional computational stereo algorithms, but they do inherently reconstruct the tissue surface shape from stereoscopic cues and can incorporate motion models and constraints for handling specular reflections and occlusions (Richa et al., 2008b, 2011). The methods can be readily implemented using GPU parallelization to achieve fast processing (Richa et al., 2011).

Table 1 summarizes some of the properties of stereo reconstruction applied to laparoscopy. In general, the computational performance of algorithms is dependent on image resolution. Reconstruction accuracy is also dependent on the image resolution as well as on the extrinsic and intrinsic camera parameters which heavily influence the recovered depth resolution (Kyto et al., 2011). As already mentioned above, the accuracy of methods based on triangulation decreases with distance away from the cameras as the relative baseline between the camera centers and the reconstructed points becomes smaller. In stereoscopy, the baseline is inherently bounded by the diameter of the laparoscope. The majority of stereo laparoscopes have a 10 mm diameter and a resulting baseline of about 5 mm though the most recent da Vinci® systems use an 8 mm scope. Experimentally, we have experienced instability and errors to grow once the target tissues are further than a multiple of 10-15 of the baseline. For images with Video Graphics Array (VGA) resolution, recent quantitative results on silicone phantom models indicate that stereo can achieve an accuracy of 1.5 mm (Röhl et al., 2012). This is likely to improve with increased image resolution, although high definition (HD), with its greater computational complexity presents a greater challenge than standard definition (SD) images, for which algorithms reaching video frame rates can easily be implemented. Nevertheless, frame rates for quarter-resolution HD images running at 13 frames per second (fps) have been reported with GPU support (Röhl et al., 2012).

2.4. Discussion

Computational stereo has been demonstrated as a feasible technique for 3D reconstruction from laparoscopic images on *in vivo* clinical data. It is currently perhaps the most well-placed technique for translation into clinical practice because stereoscopic hardware for both imaging and display is already used in the operating room, for example with the da Vinci® surgical system (Intuitive Surgical, Inc; Sunnyvale, CA, USA). Other stereo systems are either already available or in development by Karl Storz Gmbh, Richard Wolf GmbH and Viking Systems, Inc. with more manufacturers looking at entering the market in the near future. It is important to note that stereo laparoscope systems developed in the early days of MIS were not adopted into widespread clinical use, most likely due to the poor ergonimics of headmounted display systems. However, with the recent developments in display technologies and the popularity of 3D video in the entertainment industry, it is likely that the re-emergence of stereo laparoscope systems will have a wider impact in surgical practice.

When a stereo laparoscope is used, computational stereo approaches do not require any amendment to the existing operating room setup and do not introduce pertubations to the surgical environment or the clinical workflow. The major challenge for 3D reconstruction from a pair of laparoscopic images is the visual complexity of scenes in MIS, which means that algorithms developed in the computer vision community do not always perform well without customization, despite good performance on non-surgical data. Particular difficulties in the clinical setting include specular highlights, which are view-dependent and can be caused by interreflections as well

as multiple light sources. In many cases highlights can be detected by simple filtering techniques but the variability in equipment, anatomical appearance and procedure-dependent setup all complicate the performance of such approaches (Gröger et al., 2005; Stoyanov and Yang, 2005). Another challenge is calibration and maintaining this during surgery. Often assumed to be fixed, calibration parameters may change due to focusing on different depths during the procedure or due to the coupling between camera heads and laparoscope, which may not be entirely fixed in practice. Methods to cope with focus changes have been reported but have limited capabilities in practical use (Stoyanov et al., 2005a). Additionally, the robustness of computational stereo techniques is yet to be determined, especially when smoke or blood are present in the scene (cf. Fig. 10) or there are large occlusions due to the surgical instruments. Indeed, a quantitative measure of algorithm robustness is needed to evaluate the practical value of different approaches and to identify problem cases when a system dependent on the recovered 3D information would be compromised and potentially unstable or inaccurate.

While stereo is a practical approach it is also important to note that the majority of MIS procedures are currently performed with monocular scopes. This may change in the future with the development of display technologies and systems that are ergonomic, do not cause fatigue and can be observed comfortably from different viewpoints. However, not all stereoscopes can be used for reconstruction, as some techniques for producing stereoscopic images rely on beam splitters and do not have a natural baseline between the two cameras. Currently, such devices are not common due to the limited working volume over which depth perception is comfortable for the operator. Another consideration is that the baseline between the cameras reduces with the miniaturization of instruments, which limits the capability of the system to compute 3D information. Nevertheless, stereo techniques are promising and can be combined with additional visual cues to recover 3D tissue surfaces *in vivo* (Lo et al., 2008).

3. Monocular Shape-from-X

3.1. Introduction

Computing the 3D shape of an environment observed by a single moving camera has been a focus of research in computer vision for decades. The main advantage of passive monocular techniques is that they do not require hardware modification to standard laparoscopes: these techniques take as inputs images directly acquired by the laparoscope and output an estimate of the observed 3D shape. Two types of passive monocular techniques are especially important in the context of laparoscopy: Shape-from-Motion (SfM) and Shape-from-Shading (SfS). They are illustrated in Fig. 2(a-e) and 2(g-i), respectively.

Deformable Shape-from-Motion. SfM uses the apparent image motion as a cue to recover depth. It is somewhat similar to stereoscopy, but with increased difficulty: in SfM, the camera displacement is unknown and the observed surface could potentially deform between the input images. Here we present template-based Deformable SfM (DSfM), which, unlike 'classical' rigid SfM (Hartley and Zisserman, 2003), handles surface deformations. DSfM uses a 3D template and a single input image for 3D reconstruction. Here the camera is geometrically calibrated and modeled by a projection operator $\Pi : \mathbb{R}^3 \to \mathbb{R}^2$. The basic principle of template-based DSfM can
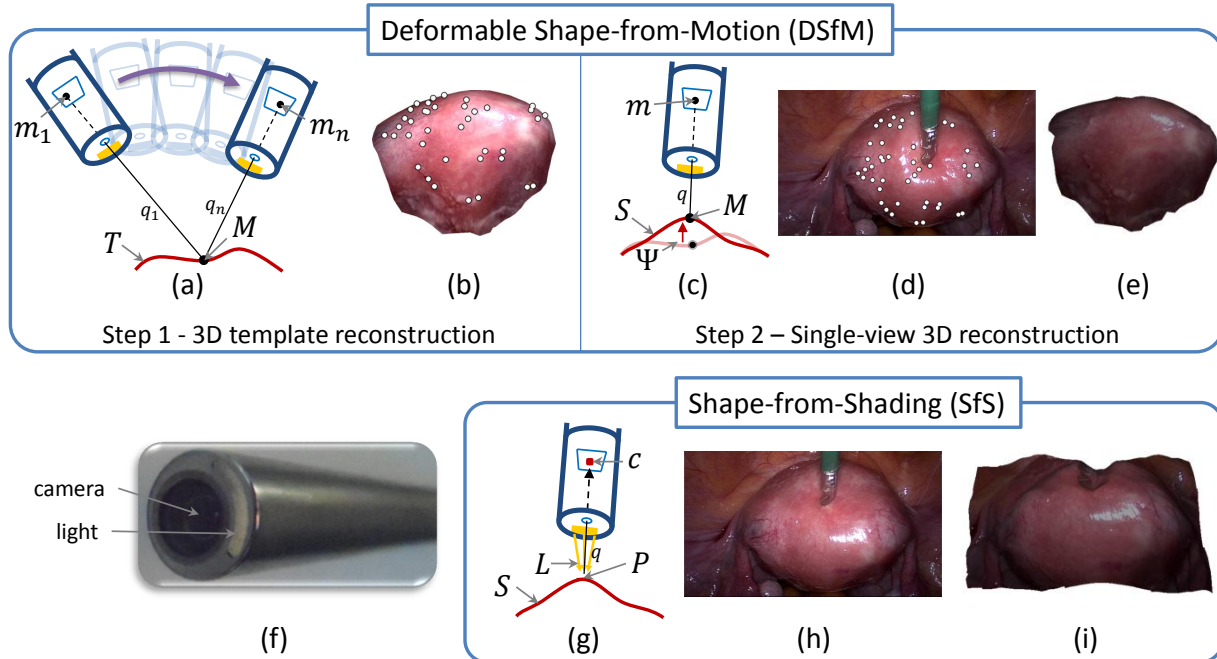
Figure 2: Passive monocular techniques: Deformable Shape-from-Motion (DSfM, a-e) and Shape-from-Shading (SfS, g-i). (a) illustrates the principle of reconstructing the 3D template shape T using rigid SfM, as the first step of DSfM. A point M projects along the line-of-sight rays $q_1, \ldots, q_n$ while the laparoscope moves, and gives rise to the image points $m_1, \ldots, m_n$, related by multiple-view constraints. (b) shows an example of a reconstructed template. (c) illustrates the principle of template-based deformation computation: notice how the observed surface S is deformed with respect to the template shape T by the unknown 3D deformation Ψ. Point M, projecting along the line-of-sight ray q is observed at position m in the image. It is matched to the template, and thus constrains the 3D deformation relating the template shape T to the deformed 3D shape S that is sought. (d,e) show an example input image, overlaid with 37 keypoint matches, with the template T and the reconstructed 3D model. (f) shows a monocular laparoscope. (g) illustrates the principle of SfS: the light L emitted by the laparoscope is reflected by a surface patch P. Traveling along the line-of-sight ray q, it is then imaged at a pixel with colour c in the image. SfS simply uses the fact that the brighter the pixel colour c, the more fronto-parallel the surface patch P. (h,i) show an example input image and the obtained depth map rendered from a new viewpoint.

9

be broken down into three main steps. First, reconstructing a *template*: a 3D model of the observed shape in some configuration. This can be done by any means, such as rigid SfM. Second, matching the template to the input image, for example by finding keypoint correspondences. Third, finding a 3D deformation $\Psi : \mathbb{R}^3 \to \mathbb{R}^3$ that will fit the template to the input image, as illustrated in Fig. 2(c). Step one is done only once, but steps two and three are repeated for every input image. Step three is the most important step; it combines the camera projection with the unknown 3D deformation to predict the location $m \in \mathbb{R}^2$ in the input image of a point $M \in \mathbb{R}^3$ on the template as:

$$\Pi(\Psi(M)) = m.$$

By applying this equation to the template to input image correspondences, the 3D deformation is then estimated under some additional physical constraints, such as isometry[3]. Because it uses a single image, DSfM does not require a baseline or motion.

Shape-from-Shading. SfS follows a different route. It uses only one image, and models the relationship between the observed pixel intensity and the surface normal; roughly speaking, the brightness of a pixel is used to infer to what extent the surface is tilted at this pixel. The camera is radiometrically calibrated here; a pixel's colour $c \in \mathbb{R}^3$ can thus be converted to an image irradiance $\iota \in \mathbb{R}$. The light is also calibrated here; the light direction and intensity are encapsulated in a known vector $L \in \mathbb{R}^3$. The basic principle of SfS can be broken down in two main steps. First, constraints on the surface's normal N are formed for the input image. Second, these constraints are 'integrated' to compute the depth at each pixel. Under the Lambertian reflectance model the fundamental equation of SfS relates the normal N of surface S to ι, as illustrated in Fig. 2(g), via:

$$\iota = \rho L^\top N.$$

In this equation ρ is the surface's reflective power, referred to as *albedo*. Step two recovers the depth by introducing it in the expression of N, and solving the above equation for all pixels simultaneously.

3.2. State-of-the-art

Deformable Shape-from-Motion. SfM has been solved over the last few decades under the hypothesis that the observed scene was rigid (Hartley and Zisserman, 2003). It was established that both the camera displacement and the scene shape are recoverable from image observations. However, if the camera displacement can be computed using an external sensor, rigid SfM is then equivalent to stereoscopy (with varying baseline). Rigid SfM has numerous applications, for instance in robotics and the film industry, and is also applicable in some laparoscopic procedures (Grasa et al., 2011). However, the laparoscopic environment is generally deformable, as Fig. 2(d,h) illustrates using a simple example showing a tool exerting pressure on the surface of a uterus. Because SfM here relies on a single camera, the multiple input images will show the observed shape in different states of deformation, and have to be handled with DSfM (or 'non-rigid' SfM). DSfM has been studied for about a decade (Bregler et al., 2000) and is an extremely

---

[3]Isometric deformations preserve the geodesic distance on the deformed surface.

challenging problem, which has not yet received a stable general solution in the literature. To alleviate the difficulty, some approaches use a template. This is a lot easier than the *template-free* problem, which we do not describe here. The template-based approach has initially been proposed for objects with simple material properties such as sheets of paper (Perriollat et al., 2011; Salzmann and Fua, 2011). At run-time, the input image is first matched to the template robustly using keypoints (Pilet et al., 2008; Pizarro and Bartoli, 2012). Pixel colours (Garg et al., 2011; Pizarro and Bartoli, 2012) have also been successfully used to increase accuracy. The template 3D shape is then deformed by combining two types of energies: a data term forcing the 3D shape's reprojection to match the input image and a prior forcing the 3D shape's deformation to match physical constraints. SfM typically achieves reliable global shape recovery, but misses shape details in textureless areas (Cryer et al., 1995).

Shape-from-Shading. SfS has been studied since the early seventies (Horn, 1970). The vast majority of SfS methods only use a single input image (Zhang et al., 1999). The colour of a pixel is a camera measurement of the interaction between the light, the scene surface and its albedo. This interaction is modeled by the scene reflectance function. As illustrated in Fig. 2(g), the shading cue exploits the scene reflectance function to relate the normal of the shape and the observed intensity in the image. Most methods make four fundamental assumptions: *(1)* the scene contains only a single light source, whether directional or proximal, *(2)* the scene's reflectance is Lambertian (light is reflected equally in all directions), *(3)* the shape's albedo is constant or known and *(4)* the shape is continuously differentiable, and its projection does not create discontinuities. Using assumptions *(1)-(3)*, it is simple to derive an equation relating the brightness of each image pixel to the angle between the shape normal at this pixel and the camera's depth axis. Because the shape normal has two degrees of freedom (it is a direction in the 3D space), this single per-pixel constraint is not sufficient to recover the 3D shape uniquely. Assumption *(4)* is therefore used to relate the normals of neighboring pixels, which provide the additional constraints to recover the shape up to a few discrete convex/concave ambiguities. Most SfS methods also assume that the camera projection model is known, whether affine or perspective. They also assume that the light and camera's radiometric calibration was carried out properly (Rai and Higgins, 2008). The light calibration can be done once for all. It requires one to move a flat piece of white paper in front of the laparoscope, preferably with little ambient light. The camera transfer function is calibrated automatically by showing a flat color calibration checkerboard to the laparoscope. In principle, recalibration is necessary when the intrinsic camera parameters are changed due to, for instance, changing white balance. In practice, it happens at most once or twice per use. SfS generally recovers shape details accurately, but estimates shapes that may be globally distorted (Cryer et al., 1995).

Combining multiple cues. Combining SfM and SfS has proven useful in the rigid case (Cryer et al., 1995). They have only been combined in the deformable case recently, however (Moreno-Noguer et al., 2010). Specularities are very complimentary to shading. The common shading constraints cannot be applied at specular points, because the camera generally saturates. However, the shape's normal can be set to be colinear with the light direction for most materials.

11

3.3. Application to Laparoscopy

We present the application of first DSfM and then SfS to laparoscopy. Current DSfM and SfS techniques used in laparoscopy require the intrinsic parameters of the laparoscope to be calibrated. Both techniques have a close depth range (depending on the light power and camera sensitivity for SfS) and lateral resolution of up to image resolution. Their accuracy has been empirically shown to be of the order of a few to a dozen millimeters. SfS runs in real time on the GPU (for instance 23 fps for 720×576 images was achieved (Collins and Bartoli, 2012)), and DSfM is expected to reach real-time in the near future.

Deformable Shape-from-Motion. In the context of MIS, rigid SfM has primarily been applied in SLAM-based algorithms, as shown in sec. 4. The current research in applying DSfM to laparoscopy mainly follows the template-based paradigm (Malti et al., 2011), although the template-free low-rank shape model has also been tested (Hu et al., 2012; Wu et al., 2007). The template-based workflow has two main phases. The first phase is to reconstruct the template's 3D shape using rigid SfM while the surgeon explores the abdominal cavity. This is illustrated in Fig. 2(a,b). At this phase, the surgery tools have not yet been introduced into the patient, and the rigidity assumption holds for rigid and semi-rigid organs such as the uterus shown in Fig. 2. Note that this template reconstruction method cannot be used for organs which naturally deform, such as a beating heart. The second phase is to reconstruct the current shape while the surgeon starts deforming the observed tissues. This is the 'active' surgery phase. The current frame is registered to the template, and the observed 3D shape is reconstructed. Fig. 2(b,d) shows images overlaid with keypoint matches. The unpredictable elastic deformation of the tissues does not allow one to learn a prior global deformation model, and one has to resort to local models, which are more flexible. While the isometric deformation assumption does not hold, convincing results were obtained under the conformal deformation model, which preserves angles and allows for local isotropic scaling (Malti et al., 2011). An analytical solution exists to this problem (Bartoli et al., 2012). The surface hidden by the tool is recovered by deforming the template. Figure 2(d,e) shows a sample input image and the corresponding estimated 3D shape.

Shape-from-Shading. SfS can be applied to laparoscopy well, mainly because the light source has a constant relative pose with respect to the camera of the laparoscope (Wu et al., 2010; Yeung et al., 1999; Okatani and Deguchi, 1997). Knowing this relative pose is a key assumption in many SfS methods (assumption *(1)* in sec. 3.2). This specificity can be exploited by calibrating the light source, and the transfer function of the camera (Rai and Higgins, 2008). However, the shape of the light source differs between laparoscopes. Figure 2(a) shows an example of a light source with a partially circular shape. A step forward has been recently made by using a non-parametric light model (Collins and Bartoli, 2012) which adapts to any light source shape.

3.4. Discussion

The state-of-the-art in SfM and SfS as used in the context of laparoscopy demonstrates that they have complementary strengths and weaknesses. Their advantage compared to other techniques is that they can be used with almost any hardware (they do not generally require modification of a standard laparoscope's hardware). Both methods may fail in the presence of unmodeled

phenomena such as bleeding, smoke and occlusions. However, they both recover from temporary failure since they do not rely on the previously estimated 3D shape. In SfM, one of the main open problems is correspondence establishment. Indeed, laparoscopic images tend to have poor or repeated structures, and thus defeat current matching methods. In SfS, a strong limitation is caused by the importance and amount of assumptions made on the imaging process. One of the unsolved problems therefore uses a more advanced model of the laparoscopic environment's Bidirectional Reflectance Distribution Function (BRDF). An interesting avenue for future research is applying template-free DSfM in the context of laparoscopy (Taylor et al., 2010; Russell et al., 2011; Collins and Bartoli, 2010; Varol et al., 2009) and combining it with SfS (Moreno-Noguer et al., 2010).

4. Simultaneous Localization and Mapping

4.1. Introduction

The process of estimating the 3D structure of an environment from a moving camera and estimating the pose of the camera in the environment is well studied. The previous section introduced two techniques - SfM and DSfM - which use the frame-to-frame motion of a camera to recover shape. These techniques typically have offline or batch processing components which makes their application to live surgical navigation challenging. SLAM, sometimes referred to as online SfM, is a sequential and real-time technique for simultaneously estimating 3D structure (*mapping*) and camera pose (*localization*). Unlike DSfM, it requires *multiple* images as input and is based on the rigid body assumption. It is a general framework that can be used with a variety of input sensors (monocular and stereo cameras, laser range finders, structured light etc.) and has a particular focus on uncertainty handling.

SLAM systems have a state $x_t = (c_t, M_1, ..., M_n)$ which holds the camera pose $c_t = (t, R)$, consisting of a translation vector t and a rotation matrix R and a set of n 3D landmarks $M_i = (x, y, z)$ which describe the 3D structure of the environment at time t. The live camera images are processed individually to update the state of the system. At each frame, a new camera pose is estimated, existing landmarks are re-observed and new 3D landmarks are added to the state. The computational complexity of SLAM is dictated by the size of the state (i.e. the number of landmarks) and not the number of images (as in SfM). This formulation of the problem makes it computationally feasible to sequentially estimate the camera pose and 3D structure in real time.

Modeling image noise and uncertainty is a fundamental component of SLAM. Sequentially updating the state with noisy observations of landmarks would lead to error propagation and an inconsistent state. Uncertainty in the state is modeled by a full covariance matrix. The state and covariance matrix are managed and updated using a probabilistic framework where the joint posterior density of the 3D landmarks and the camera pose is described by the probability distribution $P(x_t|Z_{0:t}, U_{0:t}, x_0)$ given the observations Z_i of visible landmarks and any control inputs U_i from position sensors on the camera (e.g. accelerometer). The Bayesian formulation of the problem give rise to a recursive framework comprised of prediction, measurement and update steps illustrated in Fig. 3.

In the prediction step, the pose of the camera is estimated using a motion model. Motion models comprise a deterministic and a stochastic element. The deterministic part is a prediction

13

based on a sensor measurement (e.g. Inertial Measurement Unit (IMU)) or on previous history of camera motion. The stochastic part is a probabilistic model of the uncertainty in the predicted motion, which may be derived experimentally. Given the predicted new pose of the camera it is possible to project the 3D landmarks into the image in preparation for the measurement and update steps.

The measurement or observation step solves the association problem by establishing the correspondence between 3D landmarks and features in the image space. In vision SLAM systems, the 3D landmarks may be associated to an image patch or template. Matching the template in the image provides new measurements of the location of the 3D landmarks relative to the camera. The measurement can be made in the image space for monocular cameras or in 3D for stereo cameras. A measurement model is defined which relates the measurement to the state. Finally, the state is updated using the predicted model, measurement model and the observed measurements of the 3D landmarks. A wide variety of solutions to the SLAM problem have been proposed.
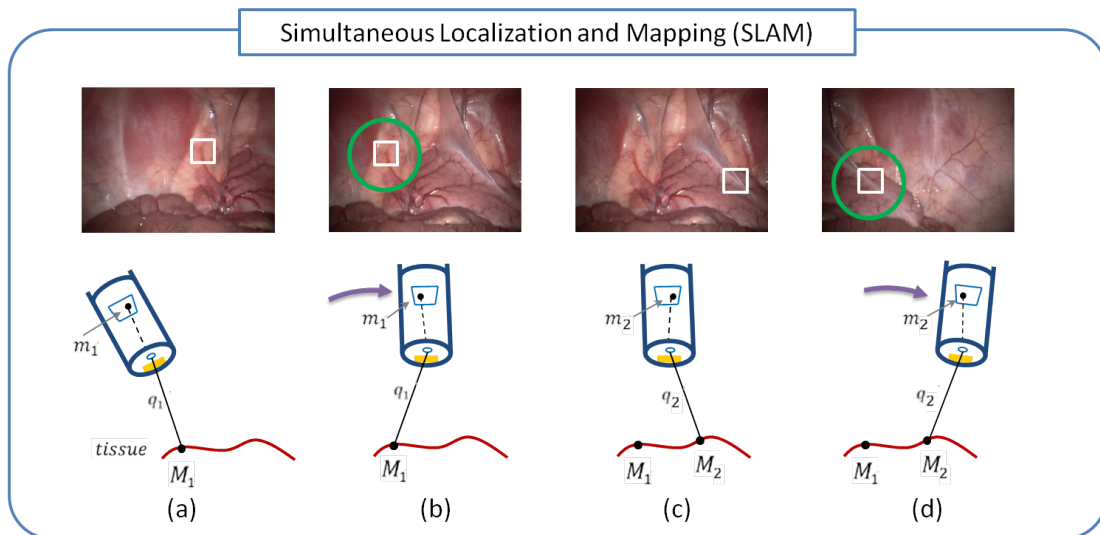

Figure 3: Simultaneous Localization and Mapping (SLAM): (a-d) A passive monocular SLAM system incrementally building a model of the tissue surface and estimating the pose of the camera. A point on the soft-tissue surface, M_i, is projected onto the image planes to the point m_i by the line-of-sight rays q_i. (a) A point m_1 in the image is detected (white box) and added to the map. (b) the camera moves and a new image is captured. The SLAM algorithm predicts the camera's motion. According to the predicted motion, the points in the map are projected into the image and a local search (green circle) is performed to match or measure the points. The SLAM state containing the pose of the camera and the map are updated. (c) A new point m_2 is detected and added to the map. (d) The camera moves and the SLAM algorithm repeats the *predict*, *measure*, *update* loop to incrementally build the map.

4.2. State-of-the-art

SLAM was developed largely by the robotics community for autonomous navigation with laser and sonar range finders, but it has found camera (monocular and stereo) based applications in computer vision and augmented reality. Extensive research has been undertaken over the past two decades, creating a large variety of solutions. A comprehensive review can be found in (Durrant-Whyte and Bailey, 2006; Bailey and Durrant-Whyte, 2006). The most common

probabilistic frameworks are the Extended Kalman Filter (EKF), the Particle Filter and the Rao-Blackwellised Filter. Readers are directed to (Thrun et al., 2005) for detailed mathematical descriptions. These frameworks provide the necessary tools to solve the SLAM problem, however the practical application of SLAM remains challenging and it is here where recent research has been focused.

The computation complexity of EKF SLAM is largely governed by the number of landmarks in the state. Thrun et al. (Thrun et al., 2004) proposes a sparsification method which represents the probability density in information form. Components close to zero in the normalized information matrix are ignored, leading to a sparse representation which can be efficiently updated with little compromise in performance. Larger-scale maps can be built and updated efficiently. An alternative to large-scale mapping is global or local submapping. Submapping creates small computationally manageable maps which are either linked to each other in a common global coordinate system or via a local relative transformation. In (Galvez-Lopez and Tardos, 2011) the authors propose a method to create consistent local and global submaps by addressing the loop closing problem.

The Parallel Tracking And Matching (PTAM) (Klein and Murray, 2007) algorithm creates high-quality submaps by separating localization and mapping and processing them on parallel threads on a dual-core computer. Uncertainty in submaps is reduced by exploiting limited batch processing without affecting real-time localization. The PTAM system includes relocalization using randomized lists. This enables it to recover from lost tracking caused by image blurring, rapid camera motion or occlusion. The Dense Tracking And Matching (DTAM) (Newcombe et al., 2011) builds on this approach to produce dense maps and robust camera localization. The approach builds dense photorealistic depth maps at selected key frames. Camera localization is performed using whole image alignment with the dense models and does not rely on feature matching. Additional research has focused on loop closing, dynamic environments and long-term mapping.

4.3. Application to Laparoscopy

In MIS, SLAM approaches can be used to localize the pose of the endoscopic camera and build a 3D model of an organ *in vivo* while the endoscope is navigated by the surgeon. The *in vivo* organ model can be used for registration to a pre-operative model. A fundamental component of AR or image guidance is knowing the camera's pose relative to the object or organ of interest. Real-time SLAM provides two fundamental components of CAS; 3D *in vivo* tissue model and camera pose estimation while allowing camera movement.

Burschka et al. (Burschka et al., 2005) proposes using an approach called V-GPS to create long-term SLAM-style maps/reconstructions for sinus surgery using a monocular endoscope. A method is proposed for estimating the scale of the 3D reconstruction which cannot be recovered from a monocular camera. The scaled 3D reconstruction of the rigid sinus is registered to a pre-operative CT to enable AR overlay of critical subsurface anatomy. The system was reported to run at 10 Hz with sub-millimeter registration accuracy on phantom data.

An EKF SLAM approach was proposed (Mountney et al., 2006) to build sparse 3D reconstructions of the abdomen and recover the motion of a stereo laparoscope. The system was has been adapted to low-resolution stereo fibre image guides (10,000 fibres) (Noonan et al., 2009)

and demonstrated reconstruction accuracy of less than 3 mm on phantom data. Monocular EKF SLAM has also been proposed for MIS (Grasa et al., 2011), combining 1 point RANSAC and randomized list relocalization for recovering from tracking failure. The system reports run times of around 12 Hz. It increases the number of actively tracked landmarks, creating a denser reconstruction which can be used for relocalization.

In EFK SLAM the reconstructed surface of the tissue is represented by the set of 3D landmarks. These landmarks can be meshed and textured with images from the endoscope to create visually more realistic tissue models (Mountney et al., 2009; Totz et al., 2011a). Such models are an approximation of the organ's surface and may contain inaccuracies. Combining sparse SLAM with dense stereo techniques (Totz et al., 2011b) creates more comprehensive 3D reconstructions without increasing the computational complexity of SLAM.

The models discussed so far are based on the assumption that the physical world is static. In anatomical environments such as the nasal passage this assumption is held, however, in the abdomen, respiration causes tissue motion. In (Mountney and Yang, 2010) dynamic mapping is proposed where the tissue model deforms with periodic motion caused by respiration. The error in the estimated camera position was less than 2 mm for *ex vivo* data and the system demonstrated accurate recovery of respiration models.

Tougher evaluation of SLAM systems for MIS remains a challenge for the community. Optical tracking systems have been used to obtain ground truth for camera motion, however these are still subject to errors from tracking, camera calibration and hand-eye calibration. Validation of the 3D reconstruction can use CT/MRI phantom or *ex vivo* data for rigid environments and synthetic data for non-rigid environments. No solutions have been proposed for validation of *in vivo* non-rigid tissue. The SLAM systems described above are sequential and capable of running in real time at up to 25 Hz, however the increased complexity of non-rigid modeling, dense surface reconstruction and recovery from failure introduce additional computational burdens.

4.4. Discussion

SLAM is a maturing technology and its use in MIS is attractive due to its real-time capabilities and integration with existing laparoscopic imaging equipment. The feasibility of SLAM has been demonstrated for the MIS environment but there remain a number of theoretical and practical research challenges in transferring this technology to the operating room. A fundamental assumption in SLAM is a rigid environment. Although this holds for some anatomy, fully non-rigid tissue motion is regularly observed in cardiac and abdominal soft-tissue surgery. A theoretical framework must be established for dealing with deformation caused by respiration, cardiac motion, organ shift and tissue tool interaction. Periodic biological signals (respiration, cardiac motion) have been well modeled in the medical imaging community and such models can be incorporated into SLAM (Mountney and Yang, 2010). However, complex tissue tool interaction and organ shift are likely to require complex biomechanical modeling. Tissue cutting and removal is an additional complication which remains an open research question. SLAM's real-time capabilities rely on establishing a set of 3D landmarks which can be repeatably matched in the image over long periods of time. Correct matching directly affects robustness and reconstruction accuracy. In well illuminated, well textured MIS environments SLAM has been shown to work well. The MIS environment can be challenging and procedure-long tracking is challenging due

to repetitive textures, large changes in lighting conditions, specular reflections and deformation. Partial occlusion due to tools, blood and smoke can generally be dealt with by using outlier removal. Tissue surfaces without texture or detectable features will require additional information from alternative approaches such as structured light or SfS algorithms. A significant challenge for the research community is *in vivo* validation for non-rigid tissue. Currently, there exists no simple, practical, highly accurate method of estimating the surface of a 3D organ as it deforms during surgery. Current methods for estimating the surface of tissue are either not real-time (CT, MRI, laser range scanners) and fail to capture tissue deformation, are not practical to introduce into the operating room, or are experimental systems under evaluation (e.g. structured light).

5. Structured Light

5.1. Introduction

Structured light techniques aim to recover the 3D surface information of an object in a similar way to stereoscopy but using an artificial pattern of light. The principle is again based on parallax and the use of the geometry of triangles and triangulation, using either stereo or monocular detection. In the case of stereo camera detection, the use of artificial features projected onto the surface of the tissue relaxes the requirement in the stereo endoscopy methods described above to detect intrinsic tissue features and properties to the tissue surface. If the artificial features can be uniquely detected then the parallax between the feature locations in the left and right stereo view may be used to find the intersection point of the line-of-sight rays, which must lie on the tissue surface. However, stereo detection using two cameras is no longer a requirement with structured lighting methods, and a simple trigonometric relationship can be established between the projection system and a single camera. If the pattern is known a priori together with the geometrical relationship between the light source and the imaging sensor (Robinson et al., 2004), (Brink et al., 2008), then the object's surface position may be accurately calculated based on the measurement of the deformation of the light pattern. In this case the line-of-sight rays projected from the camera and the structured light source intersect at the tissue surface. Similar to stereoscopy, this technique relies on the knowledge of the intrinsic camera properties and the projection is a matrix multiplication for each camera or for the projection device. Many different implementations of the structured lighting method have been proposed and Salvi et al. (Salvi et al., 2004) have suggested a classification for structured light patterns based on their coding strategy, e.g. time multiplexing, direct codification and neighbourhood codification.

5.2. State-of-the-art

While the stereo vision problem is to detect image correspondences, the structured light problem is to accurately detect (and index) the projected patterns in the presence of surface discontinuities or spatially isolated surfaces. In the case of a stereo camera detection system, the calibration and reconstruction issues are very similar to the stereoscopy case described above, whereby the intrinsic parameters and geometry of the camera system are typically determined prior to operation using standard calibration targets. In this case the resilience of the reconstruction then depends on the ability to uniquely determine the location within each image of the structured lighting pattern, usually a more straightforward problem than relying on intrinsic tissue surface
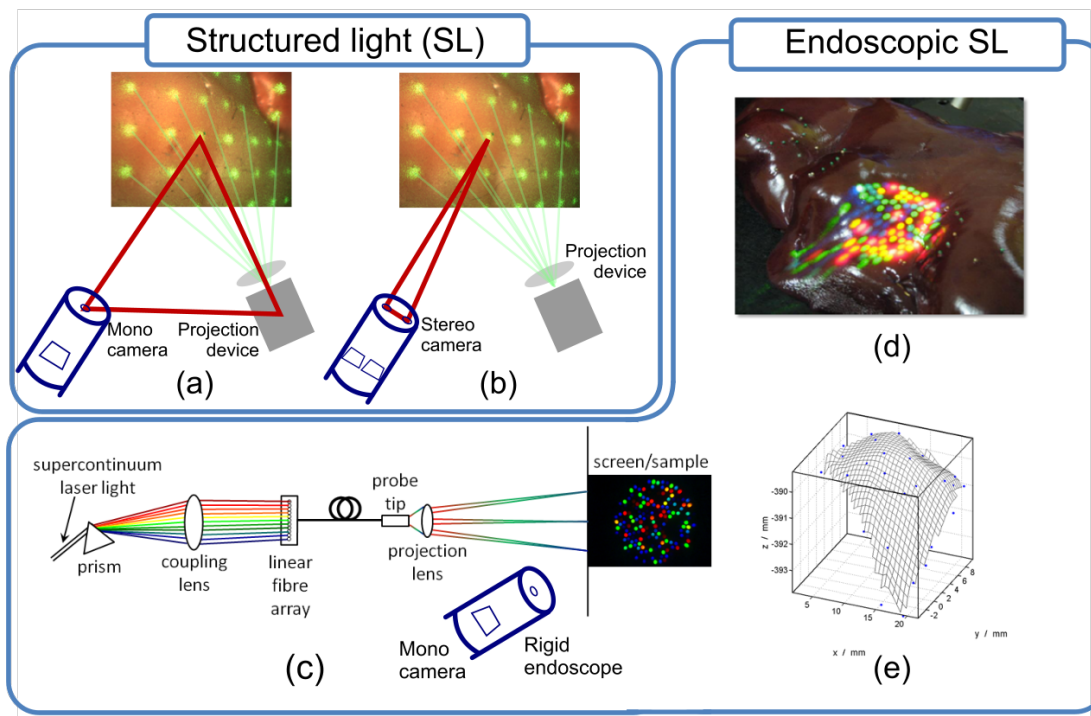
Figure 4: Schematic for structured lighting detection using (a) a stereo, and (b) a mono camera, where the principal lines used for triangulation are highlighted in dark red. (c) an experimental setup Clancy et al. (2011b) for generating a unique colour coded pattern of high brightness through a narrow diameter probe as well as allowing simultaneous white light imaging; (d) image of the structured pattern generated by this instrument and (e) the reconstructed tissue surface that is generated in this case.

features. If a striped pattern is used, it is necessary to find each stripe index in relation to a reference position. However, there are a number of methods that have been proposed to facilitate pattern detection through the use of more complex coded patterns, and a number of coding algorithms have been proposed in the literature (e.g. (Albitar et al., 2007; Pavlidis et al., 2007; Chen and Li, 2008; Kawasaki et al., 2008; Gorthi and Rastogi, 2010)); processing and display may be achieved in real time. In the case of a mono camera detection, the challenge is to establish the relationship between the pattern generation and the camera parameters, a process that is again generally achieved prior to operation using a calibration target. The remaining challenges are to apply these techniques in endoscopic formats, as detailed below and in the discussion section.

5.3. Application to Laparoscopy

In open surgery a number of structured lighting instruments have been proposed, including pattern projection systems using data projectors for determining the skin surface profiles of patients (Keller and Ackerman, 2000; Tardif et al., 2003) and for guiding percutaneous punctures in interventional radiology (Nicolau et al., 2008). The main challenge in structured lighting endoscopy is the transmission of a sufficiently bright illumination pattern to the tissue surface through the limited size of typical laparoscope optical or working channels. Many endoscopic instrumentation developers are exploring alternatives such as the use of laser technology (Clancy et al., 2011a), since optical coherence allows light to be focused more easily into narrow delivery channels and diffractive and holographic optics can be used for pattern generation. One laser-based approach used a laser spot scanning system that was able to steer a small laser beam through a laparoscope imaging channel such that a small point of light was translated across the endoscopic image field using scanning mirrors (Hayashibe and Nakamura, 2001; Knaus et al., 2006). The spot's position was detected by a separate laparoscope and millimetre depth accuracy, although one disadvantage is that many images must be acquired to build up a dense matrix of projected spots (Knaus et al., 2006), (Hayashibe et al., 2006). For faster image acquisition a miniature Liquid Crystal Display (LCD) projector may be mounted within a custom made laparoscope and a second laparoscope rigidly attached with a high speed camera for imaging at up to 180 Hz (Ackerman et al., 2002). A similar concept was adapted for coloscopic use by using a laser diode and a grating to produce a grid pattern on the cervix. This projector was rigidly fixed to a registered camera system recorded the reflected images and recover the three dimensional tissue profile (Wu and Qu, 2007). Systems that are compatible with flexible endoscopy are less common. An early example using the spot scanning approach scanned a focused laser onto a flexible fibre image guide to transmit the scanned spot onto the tissue surface (Haneishi et al., 1994). An extension of this technique used a laser to project line patterns on the near surface of a fibre image guide that was small enough to be inserted into a flexible endoscope (Hasegawa et al., 2002). Recently a spectrally encoded spot projection pattern system was proposed that uses a highly broadband laser source and a fibre bundle delivery system to create a coloured pattern of spots on the tissue surface (Clancy et al., 2011a). This approach has the advantage of producing a high brightness pattern as well as mitigating problems with tissue occlusion since spectral coding gives each spot a unique wavelength. A system for creating structured lighting within a lumen was proposed by Schmalz et al. who use a 3.6 mm probe to create a colour-coded stripe pattern to solve the correspondence problem (Schmalz et al., 2012). This instrument could also be used together with a flexible endoscope and has been tested on *ex vivo* biological tissue including real-time operation.

5.4. Discussion

The advantages of structured light systems are speed, accuracy and robust 3D reconstruction of featureless objects (e.g. objects with large smooth surfaces). Such systems have found widespread applications for macroscopic detection and depth profiling of objects (volumes with 0.5-5 m side length), but they have not been frequently applied during MIS due to a number of problems as described in this discussion. One advantage of the structured lighting techniques is that they do not rely on the automatic detection of intrinsic features on the tissue surface

19

as in stereo endoscopy, but instead artificial, and potentially unique, features can be artificially projected onto the surface. The size, density and dissimilarity of the projected patterns then determines the performance of the structured lighting system and its immunity to artefacts caused by occlusions and changes in tissue texture. The limit of the feature density for full spectroscopic projection and detection is the number of discrete measurements made, i.e. usually the number of pixels in the camera detection system. One disadvantage is the requirement for triangulation, which means that the projection axis must be offset from the imaging axis, or a stereo camera system must be used. If only one endoscopic device contains both the projection and imaging optics then this angle is necessarily small and is likely to reduce further with the trend towards smaller interventions. If two instruments are used then their alignment must be fixed or determined for accurate results or a method of live calibration must be proposed, which has not yet been demonstrated for endoscopic use. The lack of a standard single commercially available device for further testing of this approach is also a limit, as the systems described here all use different approaches and technologies with their own advantages and limitations. An ongoing issue in endoscopy in general, and structured lighting endoscopy in particular, is the challenge in transmitting enough light to the tissue as devices become ever smaller. Xenon lamp sources, light emitting diodes (LEDs) and lasers have all been investigated, and variously suffer from excess heating problems, low etendue or undesirable spectral features (Clancy et al., 2011a). The intensity of lighting is further compounded in structured lighting endoscopy because the devices that are used to create the pattern reduce the intensity of the light, for instance holographic devices or spectral filters both reject illumination photons or reduce the transmission efficiency of the system. Structured lighting optics are also often difficult to miniaturize. One final challenge is to create a unique and dense feature pattern, which can allow a dense surface reconstruction. Most of the methods described above are based on localization of distinct projected objects on the tissue surface, which necessarily limits the overall feature density to be at least nine pixels, but practically requires a larger number. One future goal is to use a continuous spectral pattern that can uniquely encode every pixel within the detector with a different spectrum. Initial steps have already been taken towards this goal (Schmalz et al., 2012). While most structured lighting approaches may be achieved in real time due to the limited number of artificial features introduced, a full multispectral analysis may require a longer image acquisition and analysis time.

6. Time-of-Flight

6.1. Introduction

The ToF technique is an active reconstruction method based on measuring the time that light emitted by an illumination unit requires to travel to an object and back to a detector. Even though the human visual system does not incorporate a comparable component, similar range detection can be found in nature; for example, bats use ToF-based range detection in the ultrasonic domain. Recently, the ToF principle has been the basis for the development of new range-sensing devices, so-called *ToF cameras*, which acquire dense range images with high update rates and without a scanning device. There are two main approaches currently employed for measuring the run-time of emitted light (Lange, 2000): *Pulsed modulation* is the most obvious method,

because it directly measures the ToF of an emitted pulse of light. It was first used for studio cameras (Iddan and Yahav, 2001) and was later developed for miniaturized cameras (Yahav et al., 2007). Continuous wave (CW) modulation, the most commonly applied method in the medical field, utilizes modulated, incoherent light, to measure the phase difference ϕ between emitted and reflected light (Hostica et al., 2006; Oggier et al., 2005; Xu et al., 1998). A range image is typically obtained by determining the distances to the object(s) under observation in all image pixels in parallel using so-called *smart pixels* realized in Complementary Metal Oxide Semiconductor (CMOS)/Charge-coupled Device (CCD) technique. The basic principle is depicted in Fig. 5. The scene is commonly illuminated with intensity-modulated near infrared (NIR) light emitted from one or more illumination units, and ϕ is determined by an on-chip correlation of the reflected signal with a reference signal. Based on the measured phase difference ϕ, the distance in a pixel is then obtained by:

$$ d \;=\; \frac{c}{4\pi f_m}\phi. $$

where $c \approx 3 \cdot 10^8 \frac{m}{s}$ is the speed of light and f_m is the modulation frequency of the emitted light. Commonly, $f_m \approx 20$ MHz, yielding an unambiguous distance measurement range of $0.5 \frac{1}{20MHz} c = 7.5\ m$.

Unlike most approaches described above, the ToF technique does not rely on some kind of correspondence search and does not require a baseline. Hence, ToF cameras are potentially very compact devices which deliver real-time range information at high frame rates (typically $20-40$ Hz). Currently, they feature moderate image resolution of up to 360×240px.

6.2. State-of-the-art

In principle, any application requiring range information at high update rates can benefit from ToF cameras. Examples include geometric reconstruction of static scenes (Huhle et al., 2008), building of 3D maps for mobile applications and robotics (May et al., 2009), interaction control, e.g. for touch-free navigation for 3D medical visualization (Soutschek et al., 2008) as well as image segmentation (Wang et al., 2010), e.g. used for AR applications (Koch et al., 2009). In the context of biomedical applications ToF cameras have been used as an imaging modality for respiratory motion gating (Schaller et al., 2008) and patient positioning (Placht et al., 2012; Schaller et al., 2009) in radiotherapy as well as for building patient-specific respiratory motion models (Fayad et al., 2011, 2009).

ToF cameras are still evolving, and a lot of work is currently devoted to understanding the sources of errors and to minimizing them, as well as to model their effect for camera simulation (Foix et al., 2011). Foix et al. (Foix et al., 2011) provides an overview of ToF related errors, classifying them into *systematic errors*, which can be compensated for by calibration, and *non-systematic errors*, which are typically reduced by filtering. The authors identified five different sources of systematic errors:

- *Wiggling error*: Due to irregularities in the modulation process, the infrared light emitted cannot be generated as theoretically planned (generally sinusoidal) in practice. This results
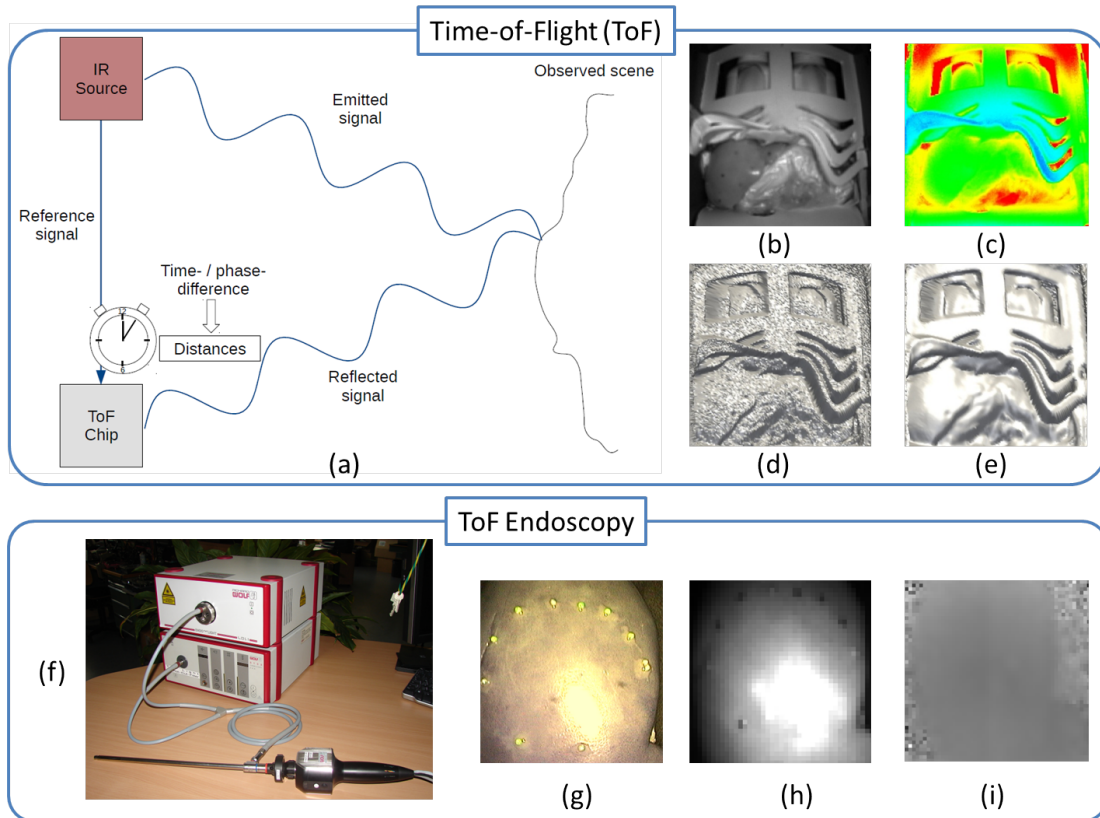
Figure 5: (a) Time-of-Flight (ToF) surface reconstruction based on continuous wave (CW) modulation: Intensity-modulated light is emitted by an incoherent light source, and the reflected light is correlated with a reference signal in every pixel, yielding (b) an intensity image and (c) a range image of the observed scene (here: human liver in a respiratory liver motion simulator). The range map can be converted to a surface based on the calibrated intrinsic camera parameters (d), preferably after applying a denoising filter (e). (f) First prototypical ToF endoscope developed by Richard Wolf GmbH. (g) RGB image of a kidney and corresponding ToF intensity (h) and range (i) image. (Images (a)-(e) courtesy of Alexander Seitel and Sven Mersmann (Div. Medical and Biological Informatics, DKFZ). Image (f) courtesy of Hubert Völlinger (Richard Wolf GmbH).)

in an error that only depends on the measured depth and typically follows a sinusoidal shape (Rapp, 2007). This error can, for example, be corrected on the basis of reference data using look-up tables (Kahlmann et al., 2007) or error correction functions such as B-splines (Lindner and Kolb, 2006).

- *Intensity/Amplitude-related error*: Depth measurements depend heavily on the intensity measured in a pixel. The error can be corrected in a similar manner to the wiggling error, i.e. using error correction functions. In (Lindner and Kolb, 2007), a bivariate correction B-spline function is used to simultaneously correct the intensity and the systematic error. As this approach requires a large number of reference ground truth measurements, Lindner et al. (Lindner et al., 2010) further proposed a decoupled calibration approach.

- *Integration time-related error*: Depth measurements also depend on the so-called *integration time*, i.e., the exposure time of the sensor for acquiring a single range image. As stated by Foix et al. (Foix et al., 2011), the main reason for this effect is still the subject of investigation. Some authors proposed performing the depth calibration with a set of integration times of interest (Foix et al., 2011), while others modelled the error as a constant offset (Kahlmann et al., 2006; Lindner and Kolb, 2007; Rapp, 2007).

- *Built-in pixel-related errors*: There are several pixel-related errors, resulting from different material properties and the readout mechanisms, that result in an offset per pixel that can be stored in a correction table (Foix et al., 2011).

- *Temperature-related errors*: As internal camera temperature affects depth processing, depth values suffer from a drift in the whole image until the camera temperature is stabilized. A common approach to compensate for this error is thus to allow a warm-up period (typically about 40 min. (Foix et al., 2011)).

Due to the above-mentioned systematic distance errors, ToF camera calibration not only requires a standard lateral calibration to determine the intrinsic camera parameters (Beder et al., 2007; Lindner et al., 2010), but also an additional calibration procedure to compensate for depth errors. Further challenges to be addressed include so-called *flying pixels* (i.e. pixels that observe regions with discontinuities in depth), *overexposed/saturated pixels*, *motion artefacts*, *multi-path reflexions* and *scattered light*, as well as *non-uniform illumination*. Foix et al. (Foix et al., 2011) provide a comprehensive overview of the majority of these errors as well as the methods proposed to compensate for them.

6.3. Application to Laparoscopy

The first ToF-based endoscope was proposed by Penne et al. (Penne et al., 2009). A commercial ToF camera (PMD[vision]3k-S, PMD Technologies, Siegen, Germany) with a lateral resolution of 48×64 was combined with a rigid standard endoscope optics. The standard LED-array illumination units of the ToF camera were replaced by a fiber-coupled high-power laser diode connected to the illumination fiber bundle of the endoscope. In a subsequent study, a new generation ToF camera (CamCube 2.0, PMD Technologies) featuring a higher resolution of 204

× 204 was used in a similar setup (Groch et al., 2011). The reconstruction error, determined with CT reference data, was of the order of magnitude of 4 mm. Recently, the company Richard Wolf GmbH (Knittlingen, Germany) introduced their first ToF endoscope, shown in Fig. 5(f). It features both a white light source as well as a ToF illumination unit and simultaneously generates range images, corresponding gray-scale amplitude images and standard definition RGB images at a framerate of about 30 frames/sec.

6.4. Discussion

The major advantages related to ToF are the registered depth and intensity data at a high frame rate as well as the compact design without scanning component or baseline. In contrast to open procedures, laparocopic interventions have a small working volume and thus require a relatively small depth range. Although the ToF endoscopes proposed so far operate at modulation frequencies similar to those offered by standard ToF cameras, which need to ensure unambiguous depth ranges of several meters, higher modulation frequencies could be applied to improve measurement precision (Lange, 2000). Another advantage is that errors caused by background illumination can be neglected due to the controlled environment.

On the negative side, the ToF technique requires additional hardware to be integrated in standard medical equipment and currently still suffers from severe systematic distance errors and noise. Due to the challenge in transmitting enough light to the tissue, the SNR in endoscopic ToF images and thus the measurement precision in camera direction (Lange, 2000) is very low. In theory, the systematic errors can be compensated for by a calibration procedure performed once before clinical use. However, this requires acquisition of a large amount of reference data in a high-dimensional space incorporating pixel ID, distance, amplitude, integration time and temperature. A practical approach involves assuming a constant temperature after a warm-up period, a constant integration time that can be chosen in an application-specific manner, and a pixel offset that is independent of distance, integration time and amplitude. The remaining error can then be compensated by determining the calibration parameters as a function of measured distance and amplitude, either in a coupled or in a decoupled approach (Lindner et al., 2010). Generation of the calibration parameters, however, is still cumbersome and the simplifying assumptions lead to larger errors. It has been shown, for example, that the temperature-related error does not remain constant after a warm-up period (Mersmann et al., 2012). Furthermore, a warm-up period of at least half an hour can lead to problems with respect to clinical workflow integration.

Major challenges to be addressed in the context of laparoscopic surgery further include scattered light, multi-path reflexions and tissue penetration. The light scattering effect is caused by multiple light reflexions between the camera lens and its sensor and causes a depth underestimation over the affected pixels (Foix et al., 2011). The amount of interference increases with a decreasing distance to the objects under observation, which makes this error important in the context of laparoscopic surgery. Multi-path errors result from the interference of multiple light reflections captured at each sensor's pixel. (Foix et al., 2011). These errors occur mainly with concave objects, which makes them highly relevant in the context of endoscopic applications. First attempts of compensating them have recently been published (Dorrington et al., 2011; Fuchs, 2010). Finally, infrared light may penetrate into the tissue, thus leading to an overestimation of depth.

As a consequence of all of these issues, the reconstruction accuracy of the first prototypical ToF endoscopes, which yield maximum errors > 1 cm (Seitel, 2012; Groch et al., 2011) is not yet sufficient for clinical application. Yet the increase of performance of ToF-devices in recent years and the growing number of applications in various areas clearly shows that upcoming ToF cameras will feature even more advanced characteristics. Due to the potential for realizing extremely compact ToF devices, it can be expected that the number of medical applications in the coming years will increase.

7. Clinical Applications

While every surgical procedure has its own specific requirements, common challenges exist across laparoscopic surgery as a whole. Surgeons wish to avoid damaging critical structures, such as nerves and blood vessels, that are beyond the exposed tissue surface. Real-time detection of tumor margins or tissue characterization *in situ* can have an immediate impact on surgical outcomes in terms of both oncologic control and quality of life by allowing all malignant tissue to be removed. Surface reconstruction techniques, while not offering a solution, are an enabling technology in order to achieve these goals through CAS and advanced imaging capabilities during surgery. This section gives an overview of some potential application areas of 3D surface reconstruction, namely *view enhancement* (sec. 7.1), *AR guidance* (sec. 7.2), and *biophotonics* (sec. 7.3).

7.1. View Enhancement

The limited FoV and the absence of 3D vision are the most important current limitations of MIS because vision is the primary sensory feedback from the surgical site. To avoid visual-spatial disorientation, the concept of *dynamic view expansion* has been proposed based on optical flow (Lerotic et al., 2008) and subsequently extended to incorporate SLAM for more robust performance (Mountney and Yang, 2009). It allows the exploration of complex anatomical structures by providing 3D textured models of the anatomy based on a sequence of endoscopic images, as shown in Fig. 6. This method has recently been extended to allow full 3D mapping of the extended view (Totz et al., 2012) , which can both enhance the visual appearance of the enlarged image and also be used to support orientation correction schemes, which can potentially be powerful aids in flexible endoscopic systems for reaching difficult anatomical sites (Warren et al., 2012). A similar system for observing wide-angle 3D on external monitors has also been reported with preliminary results (Bouma et al., 2012). For monocular endoscopes view enhancements by mosaicing has been recently demonstrated in bladder procedures where constraints on the shape of the bladder can be used to construct the expanded view (Soper et al., 2012).

A different ehancement is the use of 3D reconstruction to allow visual aids to be inserted into the FoV much alike to modern media sports reporting and commentary. Real-time annotation over the MIS video of the surgical site can support both surgical training as well as intra-operative guidance (Ali et al., 2008). A *telestrator* is a device that allows its operator to draw a freehand sketch over a video image. In telemedicine, it has recently been deployed with the latest da Vinci® Si surgical system and used to annotate anatomical details in medical images observed during surgery (cf. Fig. 7). An important application of telestration technology is the positioning

of visual guides for training in the skills suite and mentoring purposes in the operating room. The positioning of these visual annotations within the surgical FoV when observed with stereoscopic displays such as the da Vinci® console requires 3D information from the surgical site to ensure that the projected overlay markers within the stereoscopic display align correctly and appear at the right depth to the surgeon. With temporal tracking in 3D it is also potentially possible to synthesise a view of the operating field that stabilizes particular points of interest on the tissue surface. This concept has been investigated for motion compensation in robotic beating heart surgery where robotic instruments may potentially be synchronized with the computed surface motion (Stoyanov and Yang, 2007).
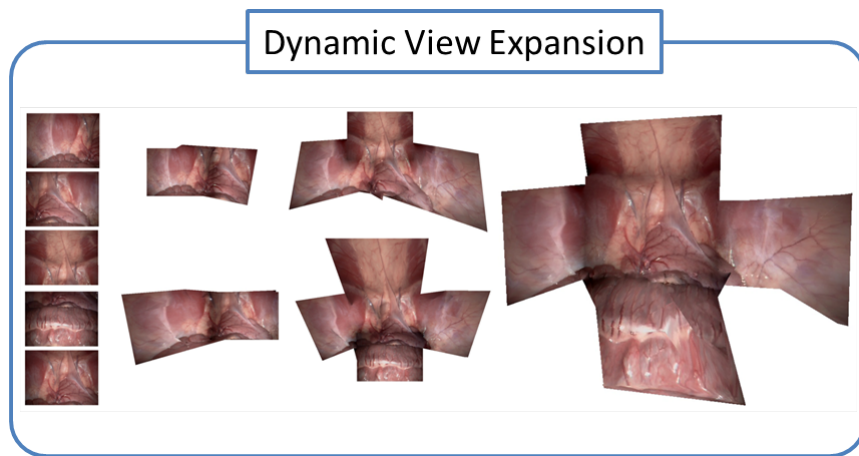

Figure 6: Concept of *dynamic view expansion* (Mountney and Yang, 2009).

7.2. Intra-operative Registration for Augmented Reality Guidance

In CAS, visual assistance to the physician is typically provided by displaying the spatial relationship between anatomical structures and medical instruments, located, for example, by a tracking system. The term *registration* refers to the alignment of pre-operative patient-specific models to intra-operatively acquired data. It may be used to augment the surgeon's view by visualization of structures below the tissue surface (Nicolau et al., 2011) (cf. Fig. 8). While computer-assisted *open* surgery generally requires the application of additional imaging modalities to acquire intra-operative anatomical information, the advantage of MIS is that the endoscope itself can be applied for this purpose.

While numerous methods have been proposed for multi-modal image registration in general (cf. e.g. (Pluim et al., 2003; Markelj et al., 2010; Glocker et al., 2011)), literature on registration in computer-assisted *laparoscopic* interventions is relatively sparse. In fact, most methods related to registration of endoscopic image data have been developed in the context of cardiac surgery (e.g. (Falk et al., 2005; Figl et al., 2008; Mourgues et al., 2003; Szpala et al., 2005)), skull base and sinus surgery (e.g. (Burschka et al., 2005; Mirota et al., 2009, 2011)), spine surgery (e.g. (Wengert et al., 2006)) and interventional radiology (e.g. (Deguchi et al., 2003; Deligianni et al., 2006)). In the interventions addressed, organ motion is generally rigid and/or periodic. In the context of laparoscopic surgery, several authors (cf. e.g. (Marescaux et al., 2004; Mutter
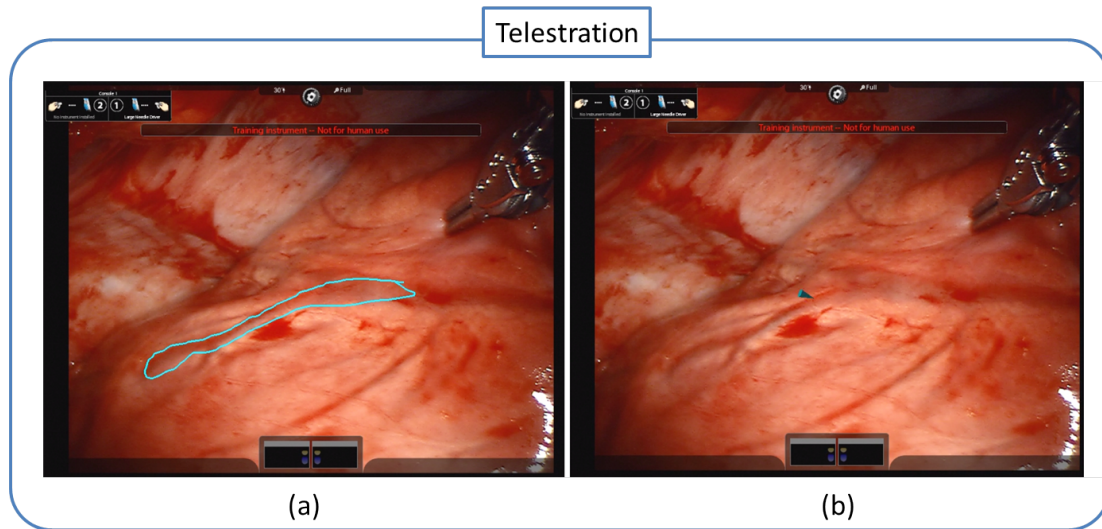
Figure 7: Example of 2D (a) and 3D (b) telestration. On the left image, the coronary artery has been outlined using the da Vinci® touch screen display and is displayed on either the right or left eye. 3D telestration can be performed using a dual console setup, where the robot master manipulators not involved in instrument control can utilize a 3D arrow to demonstrate anatomic features.

et al., 2010; Nozaki et al., 2012; Pratt et al., 2012)) proposed manual alignment of pre-operatively and intra-operatively acquired images. The majority of (semi-) automatic approaches for registering the endoscopic image data with 3D anatomical data acquired pre- or intra-operatively are either marker-based (Baumhauer et al., 2008; Falk et al., 2005; Ieiri et al., 2011; Marvik et al., 2004; Megali et al., 2008; Mourgues et al., 2003; Simpfendorfer et al., 2011; Suzuki et al., 2008) or use external tracking devices that are initially calibrated with respect to the imaging modality (Ukimura and Gill, 2008; Konishi et al., 2007; Shekhar et al., 2010; Feuerstein et al., 2008; Konishi et al., 2007; Feuerstein et al., 2007; Leven et al., 2005; Blackall et al., 2000)). In an alternative approach, reconstructed surface data may be used to perform the registration with pre-operative models (Audette et al., 2000). Comprehensive reviews on shape matching in general have been published by the computer vision community (cf. e.g. (van Kaick et al., 2011)). Regardless of the application, shape matching methods can be classified into two categories: *Fine* registration methods that assume a rough alignment of the input data, and *global* matching methods that establish correspondences between the input shapes without any prior knowledge on their poses relative to each other. In the medical domain, the surface matching methods proposed concentrate on fine registration, given a manually defined rough alignment of the data (cf. e.g. (Benincasa et al., 2008; Cash et al., 2007, 2005; Clements et al., 2008; Dumpuri et al., 2010; Maier-Hein et al., 2010, 2012; Rauth et al., 2007)).

To the authors' knowledge, all shape-based intra-operative registration methods presented for laparoscopic interventions so far are rigid and rely on the Iterative Closest Point (ICP) algorithm (Besl and McKay, 1992; Chen and Medioni, 1992) or one of its many variants (Lamata et al., 2009; Rauth et al., 2007; Su et al., 2009).
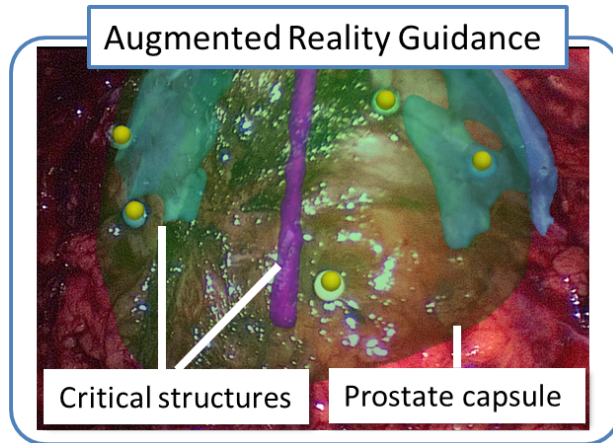
Figure 8: (a) Augmented Reality (AR) visualization during a prostatectomy provided by the marker-based computer-assisted surgery (CAS) system proposed in (Simpfendorfer et al., 2011).

7.3. Biophotonics

Optical imaging modalities, known as biophotonics, that interpret the interaction between light and tissue to acquire information about the tissue's structural and functional characteristics are emerging as very promising for *in vivo* acquisition during surgery (Iftimia et al., 2011) (cf. Fig. 9). Measuring 3D tissue surface shape and motion during surgery has important implications for emerging biophotonics modalities (Stoyanov, 2012b). Often biophotonics techniques have physical limitations that impede the practical FoV and the *in vivo* imaging of tissue that may be undergoing physiological motion. By acquiring 3D information about the tissue shape there are possibilities to overcome these limitations using registration algorithms for expanding the FoV, aligning images of moving tissues and using 3D information to support the interpretation of photometric tissue properties. Two biophotonic modalities that illustrate this potential and are used clinically are *multispectral imaging* and *confocal laser endomicroscopy* (CLE).

Multispectral imaging involves the acquisition of multiple images at different illumination wavelengths to build a complete spectrum of the tissue's response to light. Analysis of the spectral response has been shown to allow the identification of chromophores in the tissue and potentially provide tissue characterization capabilities. In imaging of the brain during neurosurgery, the modality can show functional information at the cortical surface that is equivalent to functional MRI (fMRI) (Chen et al., 2011a). However, a significant challenge in the acquisition of a stack of multispectral images is that both the camera and the target tissue can move during the image acquisition process. This can cause significant misalignment of the multispectral image stack and requires correction to enable effective spectral processing where the signal at different wavelengths is spatially aligned. Optical reconstruction and motion tracking in 3D using white light has been shown to be a promising approach to aligning the multispectral data for specific points on the tissue (Clancy et al., 2010, 2012). In intra-operative neurosurgery, non-rigid surface tracking has also been used to remove physiological motion induced by blood flow at the cortical surface (Stoyanov et al., 2012).

Probe-based CLE (pCLE) is an imaging modality that has been shown to enable *in vivo*
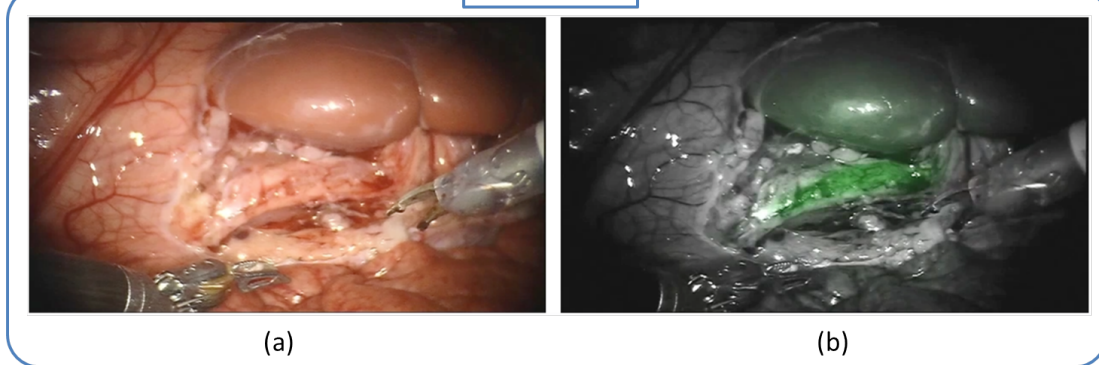
Figure 9: Biophotonics application: The use of endogenous or exogenous fluorescence can be used to provide additional anatomic information to the surgeon regarding critical structures to avoid, or pathologic anatomy to remove. In this example, the location of a porcine ureter is rather difficult to visualize in the white light image (a), while the fluorescing ureter is quite easy to discern (b). Images courtesy of Intuitive Surgical, Inc (Sunnyvale, CA, USA)

histopathology by showing the cellular structure at the tissue surface in real time during surgical or diagnostic interventions. Imaging with pCLE has been shown to be effective for *in situ* diagnosis for Barrett's Esophagus, colorectal lesions and for monitoring bilio-pancreatic structures[4]. While effective at acquiring microscopic images at the tissue surface, a difficulty with pCLE is that the area of imaging is restricted to the site in contact with the probe's tip. To overcome this FoV problem many methods for linking pCLE images using mosaicing techniques, which have similar aims to view enhancement methods, have been proposed (Vercauteren et al., 2005). Mosaiced images can potentially be mapped in the correct place in the surgeon's view of the operating field if the 3D surface of the tissue shape is known. Such methods can highlight locations where optical samples have been acquired as biopsy (Mountney and Yang, 2009). Another problem with pCLE images is that evidence linking pathologies to certain visible pCLE patterns is still limited and even manual examination of the images by endoscopists is not certain. This can be formulated as a recognition problem and 3D information at the biopsy site could potentially enhance automated recognition techniques.

Besides these exemplar techniques, there are many other applications of biophotonics methods that may also benefit from *in vivo* 3D surface reconstruction, including *fluorescence spectroscopy*, *optical coherence tomography*, *diffuse reflection spectroscopy*, *fluorescence lifetime imaging* and *photoacoustics imaging*. A more ambitious underlying consideration is that the 3D geometry of the tissue surface provides information about the way that incident light penetrates and scatters within the tissue and in the future this could be used to support interpretations of the light interaction in biophotonic modalities.

---

[4]http://www.maunakeatech.com/

8. Discussion

Intra-operative imaging techniques for obtaining the shape and morphology of soft-tissue surfaces *in vivo* are a key enabling technology for advanced surgical systems. In this review paper, we have discussed optical methods, which are currently an appealing modality for recovering 3D geometry from the surgical site without invasive instruments or exposure to harmful radiation. Table 1 provides an overview of the methods we have presented and compares their capabilities. The main advantages of passive methods, such as stereoscopy, SfX and SLAM is that they can be used with standard laparoscopic equipment and can therefore be tested in the clinical setting at the present time. However a current drawback is that passive methods require intensive processing and are thus computationally demanding. Furthermore, passive techniques, while based on the principles of biological vision system, cannot currently achieve their robustness, and failures when reconstructing homogeneous areas are common for those methods that rely on correspondence search. In contrast, active techniques require additional light to be introduced at the surgical site and as a result can deliver dense depth maps at high update rates because they do not rely on natural features. Their limitations are a result of the required hardware equipment adaptation, though these may be overcome more easily with the emergence of scopes housing multiple optical channels. To date, however, neither passive nor active reconstruction methods have found widespread use in the clinical setting. In the following paragraphs, we summarize some of the key areas of future development required to move the new technologies from the laboratory into the hospital with the ultimate goal of improving patient care:

Robustness. For translation into clinical practice, reconstruction methods must prove to be robust in the presence of dynamic, deformable and complexly illuminated environments, featuring specular highlights, smoke and blood as well as medical instruments that occlude the patient anatomy or even interact with the tissue. In this context, definition of a strategy to validate the performance of each method in challenging situations is required. So far, quantitative validation of the different reconstruction methods have typically been performed under (close to) ideal conditions using phantoms or explanted organs. To address this issue, some of the authors of this paper have performed a comprehensive evaluation study to assess and compare the robustness of different reconstruction techniques. Different hardware and algorithms were applied to acquire *in vitro* data from different organs with various shape and texture and in the presence of blood and smoke (cf. Fig. 10). The study, which will be published in the near future, concluded that none of the state-of-the-art reconstruction methods yielded accurate reconstruction results under all conditions. As achieving high accuracy that is invariant to the widely varying clinical challenges is practically not likely in the short term with any of the methods reviewed here, potential clinical applications should closely state performance requirements.

Sensor fusion. The different reconstruction methods reviewed here provide different, often complementary advantages, as already shown by Groch et al. (Groch et al., 2011) for laparoscopic interventions as well as by Beder et al. (Beder et al., 2007) and Ringbeck (Ringbeck, 2009) in the non-medical context. For example, stereo approaches perform best on textured objects, while structured light and ToF yield the best results on homogeneous objects. Consequently, sensor fusion could potentially combine the advantages of different sensor types and is therefore

Method	active/ passive	requires baseline	additional hardware	depth range	lateral resolution	frame rate
Stereo	passive	yes	no[a]	depends on baseline	$\approx$ image resolution	real-time on GPU
DSfM	passive	no[b]	no	close range	up to image resolution	not yet real-time
SfS	passive	no	no	depends on light power and camera sensitivity	same as image resolution	real-time on GPU
SLAM	passive	yes	no	depends on baseline	$\approx$ image resolution	real-time
SL	active	yes	yes	depends on light power and sensitivity	depends on complexity of patterning scheme (typically 0.1-1% of image resolution)	real-time
ToF	active	no	yes	depends on modulation frequency and light power	up to 360×240	real-time

Table 1: Overview of 3D surface reconstruction methods reviewed in this paper: Stereoscopy (*Stereo*), Deformable Shape-from-Motion (*DSfM*), Shape-from-shading (*SfS*), Simultaneous Localization and Mapping (*SLAM*), Structured Light (*SL*) and Time-of-Flight (*ToF*). The table shows whether the methods are *passive* (i.e. only require images for reconstruction) or *active* (i.e. require controlled light to be projected into the environment), whether they require modification of the hardware currently deployed in the clinical environment (*additional hardware*) and if a *baseline* is needed for reconstruction (cf. sec. 1). Furthermore, it provides general comments on depth range, lateral resolution and frame rate. Details can be found in the corresponding sections.

---

[a]but requires stereo laparoscope
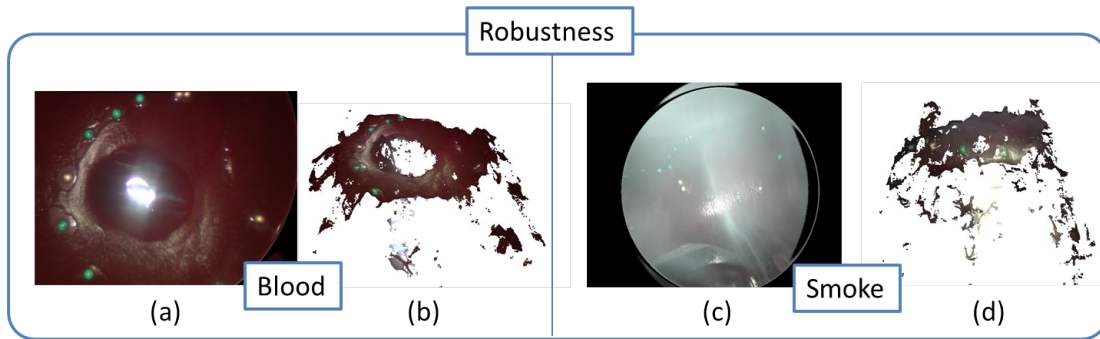
[b]just for template generation

Figure 10: Reconstruction challenges arising in MIS: In the presence of blood and smoke, stereo algorithms fail to establish correspondences between pairs of images due to homogeneous texture or occlusion. (a) Photograph of liver with blood and (b) corresponding surface obtained with a stereo reconstruction algorithm. (c) Endoscopic image of liver tissue with smoke and (d) corresponding surface obtained with a stereo reconstruction algorithm.

an important research field for improving the robustness of surface reconstruction. In the non-endoscopic context, the approaches proposed so far have focussed on combining ToF sensors with one or more RGB cameras with the purpose of increasing depth accuracy (Zhu et al., 2011; Fischer et al., 2011; Huhle et al., 2010; Gudmundsson et al., 2008) or resolution (Henry et al., 2010; Yang et al., 2007; Chan et al., 2008) or improving camera pose estimation (Castaneda et al., 2011; Henry et al., 2010; Streckel et al., 2007). Other approaches combine stereo and SfS (Blake et al., 1986; Cryer et al., 1995; Jin et al., 2008; Wöhler and D'Angelo, 2009; Wu et al., 2011) or active range scanning and SfS (Böhme et al., 2010; Herbort et al., 2011).

Fusion concepts for the endoscopic context have found less attention in the literature to date. From a practical point of view, the following issues must be considered in this context. Firstly, combination of different sensors requires methods for synchronizing and calibrating the modalities with each other. Furthermore, combining two of the methods that require additional hardware for surface reconstruction (e.g. stereo and ToF) would further increase the complexity and size of the setup. Also, active methods could potentially interfere with each other. As a consequence of these issues, all of the approaches to fusion of 3D surface reconstruction methods in the endoscopic context combine one of the reviewed methods with a method that requires only the RGB images as input (e.g, shape-from-X or SLAM). In this case, the hardware setup remains the same, and only the run-time of the algorithm is increased. Some approaches use several stereo image pairs, acquired over the time, in order to obtain a more accurate pose estimation than with monocular SLAM (Mountney et al., 2006; Totz et al., 2011a) or to create a denser surface (Totz et al., 2011a; Röhl et al., 2012). Earlier works explored the combination of monocular visual cues such as shading with stereoscopic cues (Lo et al., 2008) to overcome the limitations of stereo in homogeneous regions by exploiting the relative information recovered from shading. Recently, by calibrating the pose of the light source relative to the cameras, specular reflections were used to to resolve some of the ambiguity in scale and relative position of shading techniques (Scarzanella, 2012). To our knowledge, the first and only approach to fusion of active and passive methods in the context of laparoscopy has recently been introduced by Groch et al. (Groch et al., 2012) who chose a probabilistic graph cuts based approach to fusing dense noisy point sets ob-

tained from a ToF endoscope with a sparse but accurate point cloid reconstructed with rigid SfM. In general, more work needs to be investigated to determine the best approach to sensor fusion. In terms of clinical workflow optimization, combination of different monocular techniques could be especially interesting. One way would be to reconstruct the albedo for each pixel of the template in DSfM. This would be used with the shading model to predict the input image's pixel colors. The 3D deformation would then be recovered by minimizing the per-pixel difference with the actual input image. Overall, more general fusion frameworks should be investigated especially with increasing attempts at utilizing different visual cues in MIS, such as defocus (Chadebecq et al., 2012) or specularities (Scarzanella, 2012).

Real-time performance:. Although all of the reconstruction methods proposed can be implemented in real time on the GPU, performance remains an important issue in practice. This is particularly essential when considering a clinical application that requires 3D information in order to guide the surgeon or to augment the capabilities of the instruments. For image-guidance real-time performance should be sufficiently fast to maintain video frame rates without lag or impedance to the normal visualization of the surgical site. The hardware loop to perform overlay of such information is practically feasible but requires customized equipment. Usually, video hardware created for broadcasting has been adapted for intra-operative use. Despite acceptable performance, lag remains an issue, and this is compounded by other computational tasks such as registration or biomechanical modelling, which are currently not real-time in general. In the case of robotic control loops, where measurements are used for example to synchronize with the physiological motion of tissues, much higher run times are required to avoid aliasing the motion signals (Ginhoux et al., 2005). Currently, these cannot be achieved for dense parts of the operating field of view in HD video due to the data throughput and computational demands, but predictive models could potentially be a solution.

Clinical workflow integration. Often, clinical studies tend to favor CAS over the conventional procedures in terms of accuracy or radiation exposure (Yaniv and Cleary, 2006), however, the majority of systems have not yet entered clinical practice. The lack of acceptance for clinical use is partly due to a suboptimal integration of the proposed systems into the clinical workflow. A CAS system should not add time or complication to a procedure, but be unobtrusive and simple to operate and should be linked to an ergonomic user interface. In this context, 3D reconstruction methods face various challenges. For active technologies where additional equipment must be introduced into the operating theatre or into the anatomy, significant challenges are miniaturization, integration and regularization for clinical use. Passive techniques bypass some of these problems but must prove to be robust when textural information is not always available for certain tissue types (cf. Fig. 10). Seamless integration of the presented techniques into the clinical workflow further requires minimum setup and calibration times and real-time performance during surgery. In this context, *online* calibration of optical devices without calibration objects remains an unsolved technical challenge, though some preliminary work has been reported (Stoyanov et al., 2005a). A potentially more informed approach would be to use the known shape of objects in the surgical field of view, such as instruments, for calibration. Model-based approaches for instrument detection are required for this to work in practice and some preliminary studies have shown

promising results, though calibration is assumed to be known with these methods (Speidel et al., 2008; Pezzementi et al., 2009).

Non-rigid surface registration. While interventions on sufficiently rigid structures may require only an initial registration at the beginning of the operation, soft-tissue interventions rely on fast, non-rigid registration methods in order to account for the continuously changing morphology of the organs. Unlike approaches based on manual interaction, markers, or calibration of an image modality to the endoscope (cf. sec. 7.2), shape-based registration is potentially well-suited for *non-rigid* registration. Despite many influential publications in surface matching in general (cf. e.g. (Bronstein et al., 2011; Funkhouser and Shilane, 2006; Gelfand et al., 2005; Lipman and Funkhouser, 2009; Wang et al., 2010; Windheuser et al., 2011; Zeng et al., 2010; Zhang et al., 2008), the only fully-automatic non-rigid approaches to intra-operative registration of range data in abdominal procedures have been applied in open surgery (dos Santos et al., 2012) and do not provide real-time performance. To avoid the computational demands of repeating the registration process over time, an alternative registration approach involves continuously updating an initially performed registration via tissue tracking using the endoscopic image information acquired during surgery. This is typically achieved with iterative strategies such as optical flow (Horn and Schunck, 1981; Lucas and Kanade, 1981), which are based on the knowledge of the location of a feature in the previous frame to constrain a search for the corresponding feature in the next frame, assuming a small degree of motion and intensity coherence. Iterative strategies have been combined with predictive models of feature localization based on prior knowledge of anatomical periodicity, machine learning approaches and predictive filtering (Ginhoux et al., 2005; Ortmaier et al., 2005; Bachta et al., 2009; Bogatyrenko et al., 2011; Richa et al., 2010; Giannarou et al., 2012; Mahadevan and Vasconcelos, 2009; Puerto Souza et al.) and have been extensively used in laparoscopic images with varying degrees of success (Sauvee et al., 2007; Elhawary and Popovic, 2011; Ortmaier et al., 2005; Yip et al., 2012).

Biomechanical modelling. While rigid registration algorithms allow computing the pose of *internal* organ structures based on surface information, this sparse sensor information is often insufficient for compensating soft-tissue deformation inside the organ. In the context of the *sparse data exploration problem* (Miga et al., 2011), accurate non-rigid registration can be solved by incorporating a-priori knowledge about the mechanical properties of the tissue via biomechanical modeling. Using elasticity theory, the approach can be formulated as a boundary value problem with displacement boundary conditions generated from intra-operative sensor data. In general, the finite element method (FEM) is used to solve the resulting set of partial differential equations. In several neurosurgical applications, this approach has been successfully applied to compensate the *brain shift* with intra-operative images (cf. e.g. (Chen et al., 2011b; Wittek et al., 2007; Clatz et al., 2005; Skrinjar et al., 2001)).

In contrast to neurosurgery, there are only a few studies on abdominal or laparoscopic interventions that adapt this concept to date (Simpson et al., 2012; Peterlk et al., 2012; Suwelack et al., 2011a; Miga et al., 2011; Pratt et al., 2010; Dumpuri et al., 2010; Cash et al., 2007, 2005). Using biomechanical models for non-rigid registration is challenging as finite element (FE) models are computationally intensive, but have to be solved in real time for CAS while still

being robust and accurate. The application of fast, GPU-based FE solvers in combination with a reduced model complexity is therefore crucial regarding real-time capability. Various FE algorithms exist which can be used for hyper-, visco-, and poroelastic models in the field of real-time soft tissue simulation (cf. e.g. (Marchesseau et al., 2010; Miller et al., 2007)). Both methods have drawbacks regarding robustness and numerical complexity, especially in the context of an intra-operative application. Since previous studies have shown that in this context the material law and its parameterization has very little impact on the registration accuracy as long as a geometrically non-linear model is used (Wittek et al., 2009; Suwelack et al., 2011b), more efficient models, e.g. the corotated FE (Mezger et al., 2009; Suwelack et al., 2011a), can be used, also taking vascular structures inside the organ into account (Peterlk et al., 2012). Another aspect that has to be considered are morphological changes due to cuts which have to be propagated in real time on the FE mesh. A promising and efficient method for real-time cut simulation is e.g. the extended finite element method (X-FEM). Several approaches based on X-FEM can be found in the literature (e.g. (Vigneron et al., 2011; Jerabkova and Kuhlen, 2009)).

Validation. Careful validation, both in controlled environments as well as in clinical scenarios, is crucial for establishing a new system in the clinic (Jannin et al., 2002, 2006). So far, most of the validation studies relating to 3D surface reconstruction have been performed on numerically simulated data (Hu et al., 2007; Wu and Qu, 2007; Mountney and Yang, 2010), on phantom models with known ground truth geometry and motion characteristics (Hu et al., 2007; Wu and Qu, 2007; Richa et al., 2008b; Noonan et al., 2009) or with ground truth data obtained by scanning techniques (Burschka et al., 2005; Hu et al., 2007; Wu and Qu, 2007; Richa et al., 2008b; Noonan et al., 2009; Stoyanov et al., 2010). In this context, ensuring that there is no deformation between data acquisition using scanning and observing the tissue with a laparoscope can be a practical challenge. Experiments on dynamic objects are even more complicated. For this purpose, a combination of mechanical devices and signal generators can be employed to generate repeatable dynamic motions (Richa et al., 2008a; Visentini-Scarzanella et al., 2009; Mountney and Yang, 2010). To register between the experimental and ground truth coordinate systems, high contrast fiducial markers are typically attached to the object under observation. The most challenging scenarios for obtaining reference data are cadaver, *in vivo* and wet lab experiments, which are mostly presented to qualitatively demonstrate practical feasibility (Richa et al., 2008a,b; Noonan et al., 2009; Stoyanov et al., 2010). Quantitative *in vivo* validation could be performed by calibrating the endoscope with an intra-operative CT as shown in (Feuerstein et al., 2008) and acquiring the images under breath-hold.

Regardless of the validation object/subject used, standardization is an important issue to ensure reproducibility and informative value of a study. In order to allow reporting validation experiments in a standaradized manner, Jannin et al. proposed a model for defining and reporting reference-based validation protocols in medical image processing. (Jannin et al., 2006). The model and an associated checklist facilitate the standardization of validation terminology and methodology and thus the comparison of validation studies. Apart from standardized validation protocols, there is also the need for standardized validation objects. To advance future research in 3D reconstruction, standardized data repositories need to be established and shared for comparison of the different techniques. Some efforts relating to stereoscopic reconstruction have been

made in this direction, but these are just preliminary and need further attention (Mountney et al., 2010; Stoyanov et al., 2010). As already mentioned above, some of the authors of this paper have performed a comprehensive evaluation study to assess and compare the accuracy of different reconstruction techniques on the same data. Different hardware and algorithms were applied to acquire *in vitro* data from different organs with various shape and texture.

A further challenge, in the clinical setting is that evaluation metrics and strategies need to be devised for evaluating the difference achieved by using a CAS system as opposed to normal surgery. Indeed studies justifying and explaining the precise and procedure-specific requirements for 3D surface reconstruction are yet to appear and these are needed to highlight what the technology should be aiming for. In the long run, validation of CAS systems in randomized clinical trials are needed to fully prove the benefits of new techniques.

Infrastructure. Today, CAS systems are generally provided as stand-alone solutions and thus cannot be smoothly integrated into the clinical workflow. In recent years, open-source software platforms for medical image processing, such as the *3D Slicer* (www.slicer.org) or the *Medical Imaging Interaction Toolkit (MITK)*[5], have found widespread use in the scientific community. The underlying development process (Schroeder et al., 2004) assures flexible and portable high-quality software. Furthermore, the modular design allows straightforward re-use of existing software components so that users may focus on new developments relevant for a dedicated application. Some frameworks even support running the software as plugin within clinical workstation software (Engelmann et al., 1998) and thus enable its smooth integration into the clinical workflow. Finally, a familiar graphical user interface (GUI) decreases training time and increases acceptance on the part of the physicians who use the system. As pointed out by Cleary and Peters (Cleary and Peters, 2010), common software modules will continue to be needed so that researchers do not have to reinvent the wheel and so that newly developed techniques can be widely disseminated. To address this issue, several prominent research groups and companies have formed a joint initiative to provide the medical imaging community with the next generation open-source toolkit, referred to as *Common Toolkit (CTK)*[6]. However, a common software platform by itself is not sufficient to guarantee the smooth integration of CAS systems into the clinical workflow. In fact, widespread acceptance will only be achieved if the new technologies are effectively integrated into hospital information systems.

Human factors. Today, novel technical solutions often lack widespread acceptance on the part of the physicians who tend to be reluctant to change their habits. Hence, human factors issues relating to the use of new technical equipment need to be addressed. The systems should require a minimum of training, setup time and user interaction. Information overload and the presentation of unfamiliar, fused data sets to the physicians will increase the need for research into human-computer interfaces specific to CAS (Cleary and Peters, 2010). Multidisciplinary partnerships between scientific and clinical personnel are therefore essential.

---

[5] www.mitk.org (Wolf et al., 2005)

[6] www.commontk.org

In conclusion, CAS systems are still a long way from becoming reliable and useful tools that enhance the surgeon's capabilities for laparoscopic procedures. However, the clinical need for enhanced navigation to anatomical targets and higher precision when controlling the surgical instruments to improve the quality of medical procedures, as well as the rapid developments in medical imaging, medical image computing, robotics and computing technologies will continue to move the field forward. As accuracy requirements expand further, with targets for therapy becoming smaller due to improved image resolution and new forms of treatment, and surgery continuing to move toward minimally invasive interventions, the demand for image-guided systems can be expected to further increase in the future (Cleary and Peters, 2010). Once the benefit of the new technologies has been proven in long-term patient studies, and the new systems have been integrated effectively into the clinical workflow, CAS systems will find widespread acceptance in clinical routine.

Ackerman, J.D., Keller, K., Fuchs, H., 2002. Surface reconstruction of abdominal organs using laparoscopic structured light for augmented reality, in: Three-Dimensional Image Capture and Applications V, pp. 39–46.

Albitar, C., Graebling, P., Doignon, C., 2007. Robust structured light coding for 3D reconstruction, in: International Conference on Computer Vision (ICCV), pp. 1–6.

Ali, M., Loggins, J., Fuller, W., Miller, B., Hasser, C., Yellowlees, P., Vidovszky, T., Rasmussen, J., Pierce, J., 2008. 3-D telestration: a teaching tool for robotic surgery. J Laparoendoc Adv A 18, 107–12.

Audette, M.A., Ferrie, F.P., Peters, T.M., 2000. An algorithmic overview of surface registration techniques for medical imaging. Med Image Anal 4, 201–217.

Bachta, W., Renaud, P., Cuvillon, L., Laroche, E., Forgione, A., Gangloff, J., 2009. Motion prediction for computer-assisted beating heart surgery. IEEE T Bio-med Eng 56, 2551–2563.

Bailey, T., Durrant-Whyte, H., 2006. Simultaneous localization and mapping (SLAM): Part II. IEEE Robot Autom Mag 13, 108–117.

Barnard, S.T., Fischler, M.A., 1982. Computational stereo. ACM Computer Surveillance 14, 553–572.

Bartoli, A., Gérard, Y., Chadebecq, F., Collins, T., 2012. On template-based reconstruction from a single view: Analytical solutions and proofs of well-posedness for developable, isometric and conformal surfaces, in: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR). (in press).

Baumhauer, M., Feuerstein, M., Meinzer, H.P., Rassweiler, J., 2008. Navigation in endoscopic soft tissue surgery: Perspectives and limitations. J Endourol 22, 751–766.

Beder, C., Bartczak, B., Koch, R., 2007. A comparison of PMD-cameras and stereo-vision for the task of surface reconstruction using patchlets, in: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1–8.

Benincasa, A.B., Clements, L.W., Herrell, S.D., Galloway, R.L., 2008. Feasibility study for image-guided kidney surgery: assessment of required intraoperative surface for accurate physical to image space registrations. Med Phys 35, 4251–4261.

Bernhardt, S., Abi-Nahid, J., Abugharbieh, R., 2012. Robust dense endoscopic stereo reconstruction for minimally invasive surgery, in: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI): Workshop on Medical Computer Vision (MCV), pp. 198–207.

Besl, P.J., McKay, N.D., 1992. A method for registration of 3-D shapes. IEEE T Pattern Anal 14, 239–256.

Blackall, J.M., Rueckert, D., Calvin R. Maurer, J., Penney, G.P., Hill, D.L.G., Hawkes, D.J., 2000. An image registration approach to automated calibration for freehand 3D ultrasound, in: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 462–471.

Blake, A., Zimmerman, A., Knowles, G., 1986. Surface descriptions from stereo and shading. Image Vision Comput 3, 183–191.

Bleyer, M., Rother, C., Kohli, P., 2010. Surface stereo with soft segmentation, in: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1570–1577.

Bleyer, M., Rother, C., Kohli, P., Scharstein, D., Sinha, S., 2011. Object stereo - joint stereo matching and object segmentation, in: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3081–3088.

Bogatyrenko, E., Pompey, P., Hanebeck, U.D., 2011. Efficient physics-based tracking of heart surface motion for beating heart surgery robotic systems. Int J Comput Assist Radiol Surg 6, 387–99.

Böhme, M., Haker, M., Martinetz, T., Barth, E., 2010. Shading constraint improves accuracy of Time-of-Flight measurements. Comput Vis Image Und 114, 1329 – 1335. Special issue on Time-of-Flight Camera Based Computer Vision.

Bouguet, J.Y., 2012. http://www.vision.caltech.edu/bouguetj/calib_doc/. Accessed online 19-July-2012.

Bouma, H., van der Mark, W., Eendebak, P.T., Landsmeer, S.H., van Eekeren, A.W.M., ter Haar, F.B., Wieringa, F.P., van Basten, J.P., 2012. Streaming video-based 3d reconstruction method compatible with existing monoscopic and stereoscopic endoscopy systems, pp. 837112–837112–10.

Bregler, C., Hertzmann, A., Biermann, H., 2000. Recovering non-rigid 3D shape from image streams, in: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), pp. 690 –696.

Brink, W., Robinson, A., Rodrigues, M., 2008. Indexing uncoded stripe patterns in structured light systems by maximum spanning trees, in: British Machine Vision Conference (BMVC).

Bronstein, A.M., Bronstein, M.M., Guibas, L.J., Ovsjanikov, M., 2011. Shape google: Geometric words and expressions for invariant shape retrieval. ACM TOG 30, 1:1–1:20.

Brown, M., Burschka, D., Hager, G., 2003. Advances in computational stereo. IEEE T Pattern Anal 25, 993 – 1008.

Burschka, D., Li, M., A, M.I., Taylor, R.H., Hager, G.D., 2005. Scale-invariant registration of monocular endoscopic images to CT-Scans for sinus surgery. Med Image Anal 9, 413 – 426.

Cash, D., Miga, M., Glasgow, S., Dawant, B., Clements, L., Cao, Z., Galloway, R., Chapman, W., 2007. Concepts and preliminary data toward the realization of image-guided liver surgery. J Gastrointest Surg 11(7), 844–859.

Cash, D., Miga, M., Sinha, T., Galloway, R., Chapman, W., 2005. Compensating for intraoperative soft-tissue deformations using incomplete surface data and finite elements. IEEE Trans Med Imaging 24(11), 1479–1491.

Castaneda, V., Mateus, D., Navab, N., 2011. SLAM combining ToF and high-resolution cameras, in: IEEE Workshop on Applications of Computer Vision (WACV), pp. 672 –678.

Chadebecq, F., Tilmant, C., Bartoli, A., 2012. Measuring the size of neoplasia in colonoscopy using depth-from-defocus, in: Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS), pp. –.

Chan, D., Buisman, H., Theobalt, C., Thrun, S., 2008. A noise-aware filter for real-time depth upsampling, in: European Conference on Computer Vision (ECCV): Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications, pp. 1–12.

Chen, B.R., Bouchard, M.B., McCaslin, A.F., Burgess, S.A., Hillman, E.M., 2011a. High-speed vascular dynamics of the hemodynamic response. Neuroimage 54, 1021–1030.

Chen, I., Coffey, A., Ding, S., Dumpuri, P., Dawant, B., Thompson, R., Miga, M., 2011b. Intraoperative brain shift compensation: accounting for dural septa. IEEE Trans Biomed Eng. 58, 499–508.

Chen, S., Li, Y., 2008. Vision processing for realtime 3D data acquisition based on coded structured light. IEEE T Image Process 17, 167–176.

Chen, Y., Medioni, G., 1992. Object modeling by registration of multiple range images. Comput Vis Image Und 10, 145–155.

Clancy, N.T., Clark, J., Noonan, D.P., Yang, G.Z., Elson, D.S., 2012. Light sources for single-access surgery. Surg Innov 19, 134 – 44.

Clancy, N.T., Stoyanov, D., Groch, A., Maier-Hein, L., Yang, G.Z., Elson, D.S., 2011a. Spectrally-encoded fibre-based structured lighting probe for intraoperative 3D imaging. Biomedical Optics Express 2, 3119–3128.

Clancy, N.T., Stoyanov, D., Sauvage, V., James, D., Yang, G.Z., Elson, D.S., 2010. A triple endoscope system for alignment of multispectral images of moving tissue, in: Biomedical Optics.

Clancy, N.T., Stoyanov, D., Yang, G.Z., Elson, D.S., 2011b. An endoscopic structured lighting probe using spectral encoding, in: SPIE Novel Biophotonic Techniques and Applications.

Clatz, O., Delingette, H., Talos, I., Golby, A., Kikinis, R., Jolesz, F., Ayache, N., S.K Warfield, F., 2005. Robust nonrigid registration to capture brain shift from intraoperative mri. IEEE Trans Med Imaging 24(11), 14171427.

Cleary, K., Peters, T.M., 2010. Image-guided interventions: Technology review and clinical applications. Annu Rev Biomed Eng 12, 119–142.

Clements, L.W., Chapman, W.C., Dawant, B.M., Galloway, R.L., Miga, M.I., 2008. Robust surface registration using salient anatomical features for image-guided liver surgery: algorithm and validation. Med Phys 35, 2528–2540.

Collins, T., Bartoli, A., 2010. Locally affine and planar deformable surface reconstruction from video, in: International Workshop on Vision, Modeling and Visualization.

Collins, T., Bartoli, A., 2012. Live monocular 3D laparoscopy using shading and specularity information, in: International Conference on Information Processing in Computer-Assisted Interventions (IPCAI), pp. 11–21.

Criminisi, A., Blake, A., Rother, C., Shotton, J., Torr, P.H.S., 2007. Efficient dense stereo with occlusions for new view-synthesis by four-state dynamic programming. Int J Comput Vision 71, 89–110.

Cryer, J.E., Tsai, P.S., Shah, M., 1995. Integration of shape from shading and stereo. Pattern Recogn 28, 1033–1043.

Deguchi, D., Mori, K., Suenaga, Y., ichi Hasegawa, J., ichiro Toriwaki, J., Takabatake, H., Natori, H., 2003. New image similarity measure for bronchoscope tracking based on image registration., in: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 399–406.

Deligianni, F., Chung, A.J., Yang, G.Z., 2006. Non-rigid 2D-3D registration with catheter tip EM tracking for patient specific bronchoscope simulation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 281–288.

Devernay, F., Mourgues, F., Coste-Manière, È., 2001. Towards endoscopic augmented reality for robotically assisted minimally invasive cardiac surgery, in: International Workshop on Medical Imaging and Augmented Reality (MIAR), pp. 16–20.

Dorrington, A.A., Godbaz, J.P., Cree, M.J., Payne, A.D., Streeter, L.V., 2011. Separating true range measurements from multi-path and scattering interference in commercial range cameras, in: SPIE Three-Dimensional Imaging, Interaction and Measurement, pp. 786404–786404–10.

Dumpuri, P., Clements, L.W., Dawant, B.M., Miga, M.I., 2010. Model-updated image-guided liver surgery: preliminary results using surface characterization. Prog Biophys Mol Biol 103, 197–207.

Durrant-Whyte, H., Bailey, T., 2006. Simultaneous localisation and mapping (SLAM): part I the essential algorithms. IEEE Robot Autom Mag 13, 99–107.

Elhawary, H., Popovic, A., 2011. Robust feature tracking on the beating heart for a robotic-guided endoscope. The International Journal of Medical Robotics and Computer Assisted Surgery 7, 459–468.

Engelmann, U., Schröter, A., Baur, U., Schwab, M., Werner, O., Makabe, M.H., Meinzer, H.P., 1998. Openness in (tele-) radiology workstations: The CHILI PlugIn concept, in: International Conference on Computer Assisted Radiology and Surgery (CARS), pp. 437–442.

Falk, V., Mourgues, F., Adhami, L., Jacobs, S., Thiele, H., Nitzsche, S., Mohr, F.W., Coste-Manire, E., 2005. Cardio navigation: Planning, simulation, and augmented reality in robotic assisted endoscopic bypass grafting. Ann Thorac Surg 79, 2040–2047.

Faugeras, O., 1993. Three-Dimensional Computer Vision. MIT Press.

Fayad, H., Pan, T., Clement, J.F., Visvikis, D., 2011. Technical note: Correlation of respiratory motion between external patient surface and internal anatomical landmarks. Med Phys 38, 3157–3164.

Fayad, H., Pan, T., Roux, C., Le Rest, C., Pradier, O., Clement, J., Visvikis, D., 2009. A patient specific respiratory model based on 4D CT data and a Time of Flight camera (ToF), in: Nuclear Science Symposium Conference Record, pp. 2594–2598.

Feuerstein, M., Mussack, T., Heining, S.M., Navab, N., 2008. Intraoperative laparoscope augmentation for port placement and resection planning in minimally invasive liver resection. IEEE Trans Med Imaging 27, 355–369.

Feuerstein, M., Reichl, T., Vogel, J., Schneider, A., Feussner, H., Navab, N., 2007. Magneto-optic tracking of a flexible laparoscopic ultrasound transducer for laparoscope augmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 458–466.

Fichtinger, G., Deguet, A., Massamune, K., Balogh, E., Fischer, G., Mathieu, H., Taylor, R., Zinreich, J., Fayad, L., 2005. Image Overlay Guidance for Needle Insertion in CT Scanner. IEEE T Bio-med Eng 52, 1415–1424.

Figl, M., Rueckert, D., Hawkes, D., Casula, R., Hu, M., Pedro, O., Zhang, D.P., Penney, G., Bello, F., Edwards, P., 2008. Augmented reality image guidance for minimally invasive coronary artery bypass, in: SPIE Medical Imaging: Visualization, Image-Guided Procedures, and Modeling, p. 69180P.

Fischer, J., Arbeiter, G., Verl, A., 2011. Combination of time-of-flight depth and stereo using semiglobal optimization, in: IEEE International Conference on Robotics and Automation (ICRA), pp. 3548 –3553.

Foix, S., Alenya, G., Torras, C., 2011. Lock-in time-of-flight (ToF) cameras: A survey. IEEE Sensors 11, 1917–1926.

Fuchs, S., 2010. Multipath interference compensation in time-of-flight camera images, in: International Conference on Pattern Recognition (ICPR), IEEE Computer Society, Washington, DC, USA. pp. 3583–3586.

Funkhouser, T., Shilane, P., 2006. Partial matching of 3D shapes with priority-driven search, in: Eurographics Symposium on Geometry Processing, pp. 131–142.

Galvez-Lopez, D., Tardos, J., 2011. Real-time loop detection with bags of binary words, in: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE. pp. 51–58.

Garg, R., Roussos, A., Agapito, L., 2011. Robust trajectory space TV-L1 optic flow for non-rigid sequences, in: International Conference on Energy Minimization Methods in Computer Vision and Pattern Recognition, pp. 300 – 314.

Gelfand, N., Mitra, N.J., Guibas, L.J., Pottmann, H., 2005. Robust global registration, in: Eurographics Symposium on Geometry Processing, pp. 197–206.

Giannarou, S., Visentini-Scarzanella, M., Yang, G.Z., 2012. Probabilistic tracking of affine-invariant anisotropic regions. IEEE T Pattern Anal 35, 130–143.

Ginhoux, R., Gangloff, J., de Mathelin, M., Soler, L., Sanchez, M.M.A., Marescaux, J., 2005. Active filtering of physiological motion in robotized surgery using predictive control. IEEE T Robot 21, 67–79.

Glocker, B., Sotiras, A., Komodakis, N., Paragios, N., 2011. Deformable medical image registration: Setting the state of the art with discrete methods. Annu Rev Biomed Eng 13, 219–244.

Gorthi, S., Rastogi, P., 2010. Fringe projection techniques: Whither we are? Opt Laser Eng 2, 133–140.

Grasa, O.G., Civera, J., Montiel, J.M.M., 2011. EKF monocular SLAM with relocalization for laparoscopic sequences, in: IEEE International Conference on Robotics and Automation (ICRA), pp. 4816–4821.

Groch, A., Haase, S., Wagner, M., Kilgus, T., Kenngott, H., Schlemmer, H.P., Hornegger, J., Meinzer, H.P., Maier-Hein, L., 2012. A probabilistic approach to fusion of Time-of-Flight and multiple view based 3D surface reconstruction for laparoscopic interventions, in: International Conference on Computer Assisted Radiology and Surgery (CARS). (in press).

Groch, A., Seitel, A., Hempel, S., Speidel, S., Engelbrecht, R., Penne, J., Höller, K., Röhl, S., Yung, K., Bodenstedt, S., Pflaum, F., dos Santos, T., Mersmann, S., Meinzer, H.P., Hornegger, J., Maier-Hein, L., 2011. 3D surface reconstruction for laparoscopic computer-assisted interventions: Comparison of state-of-the-art methods, in: SPIE Medical Imaging: Visualization, Image-Guided Procedures, and Modeling, p. 796415.

Gröger, M., Sepp, W., Hirzinger, G., 2005. Structure driven substitution of specular reflections for realtime heart surface tracking, in: International Conference on Image Processing (ICIP), pp. 1066–1069.

Gudmundsson, S.A., Aanaes, H., Larsen, R., 2008. Fusion of stereo vision and time of flight imaging for improved 3D estimation. Int J Intell Syst Tech Appl 5, 425–433.

Hager, G., Vagvolgyi, B., Yuh, D., 2007. Stereoscopic video overlay with deformable registration. Medicine Meets Virtual Reality (MMVR) .

Haneishi, H., Ogura, T., Miyake, Y., 1994. Profilometry of a gastrointestinal surface by an endoscope with laser beam projection. Opt Lett 19, 601–603.

Hartley, R.I., Zisserman, A., 2003. Multiple View Geometry in Computer Vision. Cambridge University Press. Second Edition.

Hasegawa, K., Noda, K., Sato, Y., 2002. Electronic endoscope system for shape measurement, in: Kasturi, R., Laurendeau, D., Suen, C. (Eds.), International Conference on Pattern Recognition (ICPR), pp. 792–795.

Hayashibe, M., Suzuki, N., Nakamura, Y., 2006. Laser-scan endoscope system for intraoperative geometry acquisition and surgical robot safety management. Med Image Anal 10, 509–519.

Hayashibe, N., Nakamura, Y., 2001. Laser-pointing endoscope system for intra-operative 3D geometric registration, in: IEEE International Conference on Robotics and Automation (ICRA), pp. 1543–1548.

Henry, P., Krainin, M., Herbst, E., Ren, X., Fox, D., 2010. RGB-D mapping: Using depth cameras for dense 3D modeling of indoor environments, in: Proceedings of the International Symposium on Experimental Robotics (ISER).

Herbort, S., Grumpe, A., Wohler, C., 2011. Reconstruction of non-lambertian surfaces by fusion of shape from shading and active range scanning, in: International Conference on Image Processing (ICIP), pp. 17 –20.

Horn, B.K., Schunck, B.G., 1981. Determining optical flow. Artif Intell 17, 185–203.

Horn, B.K.P., 1970. Shape from Shading: A Method for Obtaining the Shape of a Smooth Opaque Object from One View. Ph.D. thesis. MIT.

Hostica, B., Seitz, P., Simoni, A., 2006. Encyclopedia of Sensors. American Scientific Pub. volume 7. chapter Optical Time-of-Flight Sensors for Solid-State 3D-vision. pp. 259–289.

Hu, M., Penney, G., Edwards, P., Figl, M., Hawkes, D., 2007. 3D reconstruction of internal organ surfaces for minimal invasive surgery., in: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 68–77.

Hu, M., Penney, G.P., Figl, M., Edwards, P.J., Bello, F., Casula, R., Rueckert, D., Hawkes, D.J., 2012. Reconstruction of a 3d surface from video that is robust to missing data and outliers: application to minimally invasive surgery using stereo and mono endoscopes. Med Image Anal 16, 597–611.

Huhle, B., Jenke, P., Straßer, W., 2008. On-the-fly scene acquisition with a handy multisensor-system. Int J Intell Syst Tech Appl 5, 255–263.

Huhle, B., Schairer, T., Jenke, P., Straer, W., 2010. Fusion of range and color images for denoising and resolution enhancement with a non-local filter. Comput Vis Image Und 114, 1336 – 1345.

Iddan, G.J., Yahav, G., 2001. 3D imaging in the studio, in: SPIE Three-Dimensional Image Capture and Applications, pp. 48–56.

Ieiri, S., Uemura, M., Konishi, K., Souzaki, R., Nagao, Y., Tsutsumi, N., Akahoshi, T., Ohuchida, K., Ohdaira, T., Tomikawa, M., Tanoue, K., Hashizume, M., Taguchi, T., 2011. Augmented reality navigation system for laparoscopic splenectomy in children based on preoperative CT image using optical tracking device. International Workshop on Medical Imaging and Augmented Reality (MIAR) 28, 341–346.

Iftimia, N., Brugge, W.R., Hammer, D.X., 2011. Advances in Optical Imaging for Clinical Medicine. Wiley.

Jannin, P., Fitzpatrick, J.M., Hawkes, D.J., Pennec, X., Shahidi, R., Vannier, M.W., 2002. Validation of medical image processing in image-guided therapy. IEEE Trans Med Imaging 21, 1445–1449.

Jannin, P., Grova, C., Maurer, C., 2006. Model for defining and reporting reference-based validation protocols in medical image processing. Int J Comput Assist Radiol Surg 1, 63–73.

Jerabkova, L., Kuhlen, T., 2009. Stable cutting of deformable objects in virtual environments using xfem. IEEE Comput. Graph. Appl. 29, 61–71.

Jin, H., Cremers, D., Wang, D., Prados, E., Yezzi, A., Soatto, S., 2008. 3-d reconstruction of shaded objects from multiple images under unknown illumination. Int J Comput Vision 76, 245–256.

Kahlmann, T., Remondino, F., Guillaume, S., 2007. Range imaging technology: new developments and applications for people identification and tracking, in: Proc. of Videometrics IX-SPIE-IS&T Electronic Imaging, p. 64910C.

Kahlmann, T., Remondino, F., Ingensand, H., 2006. Calibration for increased accuracy of the range imaging camera SwissRanger, in: Proc. of IEVM In International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences, Commission V Symposium 'Image Engineering and Vision Metrology, pp. 136 – 141.

van Kaick, O., Zhang, H., Hamarneh, G., Cohen-Or, D., 2011. A survey on shape correspondence. Comput Graph Forum 30, 1681–1707.

Kawasaki, H., Furukawa, R., Sagawa, R., Yasushi, Y., 2008. Dynamic scene shape reconstruction using a single structured light pattern, in: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1 – 8.

Keller, K., Ackerman, J., 2000. Real-time structured light depth extraction, in: SPIE Three-Dimensional Image Capture and Applications III, pp. 11–18.

Klein, G., Murray, D., 2007. Parallel tracking and mapping for small AR workspaces, in: IEEE and ACM International Symposium on Mixed and Augmented Reality (ISMAR), pp. 225 – 234.

Knaus, D., Friets, E., Bieszczad, J., Chen, R., Miga, M., Galloway, R., Kynor, D., 2006. System for laparoscopic tissue tracking, in: IEEE International Symposium On Biomedical Imaging (ISBI): Macro to Nano, pp. 498–501.

Koch, R., Schiller, I., Bartczak, B., Kellner, F., Köser, K., 2009. Mixin3d: 3d mixed reality with ToF-camera, in: Proc. Dynamic 3D Imaging, pp. 126–141.

Kolmogorov, V., Criminisi, A., Blake, A., Cross, G., Rother, C., 2008. Probabilistic fusion of stereo with color and contrast for bi-layer segmentation. Int J Comput Vision 76, 107.

Konishi, K., Nakamoto, M., Kakeji, Y., Tanoue, K., Kawanaka, H., Yamaguchi, S., Ieiri, S., Sato, Y., Maehara, Y., Tamura, S., Hashizume, M., 2007. A real-time navigation system for laparoscopic surgery based on three-dimensional ultrasound using magneto-optic hybrid tracking configuration. Int J Comput Assist Radiol Surg 2, 483–507.

Kowalczuk, J., Meyer, A., Carlson, J., Psota, E., Buettner, S., Prez, L., Farritor, S., Oleynikov, D., 2012. Real-time three-dimensional soft tissue reconstruction for laparoscopic surgery. Surgical Endoscopy , 1–5.

Kyto, M., Nuutinen, M., Oittinen, P., 2011. Method for measuring stereo camera depth accuracy based on stereoscopic vision, in: SPIE Medical Imaging: Three-Dimensional Imaging, Interaction, and Measurement, p. 78640I.

Lamata, P., Morvan, T., Reimers, M., Samset, E., Declerck, J., 2009. Addressing shading-based laparoscopic registration, in: World Congress on Medical Physics and Biomedical Engineering, pp. 189–192.

Lange, R., 2000. 3D Time-of-Flight Distance Measurement with Custom Solid-State Image Sensors in CMOS/CCD-Technology. Ph.D. thesis. University of Siegen.

Lau, W.W., Ramey, N.A., Corso, J.J., Thakor, N.V., Hager, G.D., 2004. Stereo-based endoscopic tracking of cardiac surface deformation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 494–501.

Lerotic, M., Chung, A., Clark, J., Valibeik, S., Yang, G.Z., 2008. Dynamic view expansion for enhanced navigation in natural orifice transluminal endoscopic surgery, in: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 467–475.

Leven, J., Burschka, D., Kumar, R., Zhang, G., Blumenkranz, S., Dai, X.D., Awad, M., Hager, G.D., Marohn, M., Choti, M., Hasser, C., Taylor, R.H., 2005. Davinci canvas: A telerobotic surgical system with integrated, robot-assisted, laparoscopic ultrasound capability, in: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 811 – 818.

Lindner, M., Kolb, A., 2006. Lateral and depth calibration of PMD-distance sensors. Advances in Visual Computing 4292, 524–533.

Lindner, M., Kolb, A., 2007. Calibration of the intensity-related distance error of the PMD ToF-camera, in: SPIE: Intelligent Robots and Computer Vision XXV, pp. 6764–35.

Lindner, M., Schiller, I., Kolb, A., Koch, R., 2010. Time-of-Flight sensor calibration for accurate range sensing. Comput Vis Image Und 114, 1318–1328.

Lipman, Y., Funkhouser, T., 2009. Möbius voting for surface correspondence. ACM TOG 28, 72:1–72:12.

Lo, B.P.L., Scarzanella, M.V., Stoyanov, D., Yang, G.Z., 2008. Belief propagation for depth cue fusion in minimally invasive surgery, in: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 104–112.

Lucas, B.D., Kanade, T., 1981. An iterative image registration technique with an application to stereo vision, in: International Joint Conference on Artificial Intelligence, pp. 674–679.

Mahadevan, V., Vasconcelos, N., 2009. Saliency-based discriminant tracking, in: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1007–1013.

Maier-Hein, L., Franz, A., dos Santos, T., Schmidt, M., Fangerau, M., Meinzer, H.P., Fitzpatrick, J.M., 2012. Convergent iterative closest-point algorithm to accomodate anisotropic and inhomogenous localization error. IEEE T Pattern Anal 34, 1520–1532. (in press).

Maier-Hein, L., Schmidt, M., Franz, A., dos Santos, T., Seitel, A., Jähne, B., Fitzpatrick, J., Meinzer, H., 2010. Accounting for anisotropic noise in fine registration of time-of-flight range data with high-resolution surface data, in: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 251–258.

Malti, A., Bartoli, A., Collins, T., 2011. Template-based conformal shape-from-motion from registered laparoscopic images, in: Conference on Medical Image Understanding and Analysis (MIUA).

Marchesseau, S., Heimann, T., Chatelin, S., Willinger, R., Delingette, H., 2010. Fast porous visco-hyperelastic soft tissue model for surgery simulation. Prog Biophys Mol Biol 103(2-3), 185–96.

Marescaux, J., Rubino, F., Arenas, M., Mutter, D., Soler, L., 2004. Augmented-reality-assisted laparoscopic adrenalectomy. J Amer Med Assoc 292, 2214–b–2215.

Markelj, P., Tomanzevic, D., Likar, B., Pernus, F., 2010. A review of 3D/2D registration methods for image-guided interventions. Med Image Anal 16, 642 –661.

Marr, D., 1983. Vision: A Computational Investigation into the Human Representation and Processing of Visual Information. Henry Holt & Company.

Marr, D., Poggio, T., 1979. A computational theory of human stereo vision. Proceedings of the Royal Society of London. Series B. Biological Sciences 204, 301–328. http://rspb.royalsocietypublishing.org/content/204/1156/301.full.pdf+html.

Marvik, R., Langø, T., Tangen, G., Andersen, J., Kaspersen, J., Sjølie, B.Y.E., Fougner, R., Fjøsne, H., Hernes, T.N., 2004. Laparoscopic navigation pointer for three-dimensional image-guided surgery. Surg Endosc 18, 1242–1248.

May, S., Fuchs, S., Droeschel, D., Holz, D., Nüchter, A., 2009. Robust 3d-mapping with time-of-flight cameras, in: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 1673 – 1678.

Megali, G., Ferrari, V., Freschi, C., Morabito, B., Cavallo, F., Turini, G., Troia, E., Cappelli, C., Pietrabissa, A., O.Tonet, Cuschieri, A., Dario, P., Mosca, F., 2008. EndoCAS navigator platform: a common platform for computer and robotic assistance in minimally invasive surgery. Int J Med Robot Comp 4, 242–251.

Mersmann, S., Guerrero, D., Schlemmer, H.P., Meinzer, H.P., Maier-Hein, L., 2012. Effect of active air conditioning in medical intervention rooms on the temperature dependency of Time-of-Flight distance measurements, in: Bildverarbeitung für die Medizin (BVM), Springer. pp. 398–403.

Mezger, J., Thomaszewski, B., Pabst, S., Straßer, W., 2009. Interactive physically-based shape editing. Computer Aided Geometric Design 26, 680 – 694.

Miga, M.I., Dumpuri, P., Simpson, A.L., Weis, J.A., Jarnagin, W.R., 2011. The sparse data extrapolation problem: Strategies for soft-tissue correction for image-guided liver surgery, in: SPIE Medical Imaging:Visualization, Image-Guided Procedures, and Modeling, p. 79640C.

Miller, K., Joldes, G., Lance, D., Wittek, A., 2007. Total lagrangian explicit dynamics finite element algorithm for computing soft tissue deformation. Communications in Numerical Methods in Engineering 23, 121–134.

Mirota, D., Wang, H., Taylor, R.H., Ishii, M., Hager, G.D., 2009. Toward video-based navigation for endoscopic endonasal skull base surgery, in: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 91–99.

Mirota, D.J., Ishii, M., Hager, G.D., 2011. Vision-based navigation in image-guided interventions. Annu Rev Biomed Eng 13, 297–319.

Moreno-Noguer, F., Porta, J., Fua, P., 2010. Exploring ambiguities for monocular non-rigid shape estimation, in: European Conference on Computer Vision (ECCV), pp. 370–383.

Mountney, P., Giannarou, S., Elson, D.S., Yang, G.Z., 2009. Optical biopsy mapping for minimally invasive cancer screening, in: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 483–490.

Mountney, P., Stoyanov, D., Davison, A., Yang, G.Z., 2006. Simultaneous stereoscope localization and soft-tissue mapping for minimal invasive surgery., in: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 347–354.

Mountney, P., Stoyanov, D., Yang, G.Z., 2010. Three-dimensional tissue deformation recovery and tracking. IEEE Signal Proc Mag 27, 14–24.

Mountney, P., Yang, G.Z., 2009. Dynamic view expansion for minimally invasive surgery using simultaneous localization and mapping, in: Proc. IEEE Annual International Conference of the IEEE Engineering in Medicine and Biology Society, pp. 1184–1187.

Mountney, P., Yang, G.Z., 2010. Motion compensated SLAM for image guided surgery, in: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 496–504.

Mourgues, F., Vieville, T., Falk, V., Coste-Manière, E., 2003. Interactive guidance by image overlay in robot assisted coronary artery bypass, in: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 173 – 181.

Mutter, D., Soler, L., Marescaux, J., 2010. Recent advances in liver imaging. Expert Rev Gastroenterol Hepatol 4, 613–21.

Nalpantidis, L., Gasteratos, A., 2010. Biologically and psychophysically inspired adaptive support weights algorithm for stereo correspondence. Robotics and Autonomous Systems 58, 457 – 464.

Newcombe, R., Lovegrove, S., Davison, A., 2011. DTAM: dense tracking and mapping in real-time, in: International Conference on Computer Vision (ICCV), pp. 2320 – 2327.

Nicolau, S., Soler, L., Mutter, D., Marescaux, J., 2011. Augmented reality in laparoscopic surgical oncology. Surg Oncol 20, 189–201.

Nicolau, S.A., Brenot, J., Goffin, L., Graebling, P., Soler, L., Marescaux, J., 2008. A structured light system to guide percutaneous punctures in interventional radiology, in: SPIE Optical and Digital Image Processing, p. 700016.

Noonan, D.P., Mountney, P., Elson, D.S., Darzi, A., Yang, G.Z., 2009. A stereoscopic fibroscope for camera motion and 3D depth recovery during minimally invasive surgery, in: IEEE International Conference on Robotics and Automation (ICRA), Piscataway, NJ, USA. pp. 3274–3279.

Nozaki, T., Iida, Y., Morii, A., Fujiuchi, Y., Fuse, H., 2012. Laparoscopic radical nephrectomy under near real-time three- dimensional surgical navigation with C-Arm cone beam computed tomography. Surg Innov , (in press).

Oggier, T., Büttgen, B., Lustenberger, F., Becker, G., Rüegg, B., Hodac, A., 2005. Swissranger SR3000 and first experiences based on miniaturized 3D-ToF cameras, in: Proc. of the First Range Imaging Research Day at ETH Zurich.

Okatani, T., Deguchi, K., 1997. Shape reconstruction from an endoscope image by shape from shading technique for a point light source at the projection center. Comput Vis Image Und 66, 119 – 131.

Ortmaier, T., Groger, M., Boehm, D.H., Falk, V., Hirzinger, G., 2005. Motion estimation in beating heart surgery. IEEE T Bio-med Eng 52, 1729–1740.

Pavlidis, G., Koutsoudis, A., Arnaoutoglou, F., Tsioukas, V., Chamzas, C., 2007. Methods for 3D digitization of cultural heritage. Journal of Cultural Heritage 8, 93–98.

Penne, J., Höller, K., Stürmer, M., Schrauder, T., Schneider, A., Engelbrecht, R., Feußner, H., Schmauss, B., Hornegger, J., 2009. Time-of-flight 3D endoscopy, in: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 467–474.

Perriollat, M., Hartley, R., Bartoli, A., 2011. Monocular template-based reconstruction of inextensible surfaces. Int J Comput Vision 95, 124–137.

Peterlk, I., Duriez, C., Cotin, S., 2012. Modeling and real-time simulation of a vascularized liver tissue, in: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI).

Pezzementi, Z., Voros, S., Hager, G.D., 2009. Articulated object tracking by rendering consistent appearance parts, in: IEEE International Conference on Robotics and Automation (ICRA), pp. 3940 –3947.

Pilet, J., Lepetit, V., Fua, P., 2008. Fast non-rigid surface detection, registration and realistic augmentation. Int J Comput Vision 76, 109–122.

Pizarro, D., Bartoli, A., 2012. Feature-based deformable surface detection with self-occlusion reasoning. Int J Comput Vision 97, 54–70.

Placht, S., Stancanello, J., Schaller, C., Balda, M., Angelopoulou, E., 2012. Fast time-of-flight camera based surface registration for radiotherapy patient positioning. Med Phys 39, 4–17.

Pluim, J.P.W., Maintz, J.B.A., Viergever, M.A., 2003. Mutual-information-based registration of medical images: a survey. IEEE Trans Med Imaging 22, 986–1004.

Pratt, P., Mayer, E., Vale, J., Cohen, D., Edwards, E., Darzi, A., Yang, G.Z., 2012. An effective visualisation and registration system for image-guided robotic partial nephrectomy. Journal of Robotic Surgery 6, 23–31.

Pratt, P., Stoyanov, D., Visentini-Scarzanella, M., Yang, G., 2010. Dynamic guidance for robotic surgery using image-constrained biomechanical models, in: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI).

Puerto Souza, G.A., Adibi, M., Cadeddu, J.A., Mariottini, G.L., . Adaptive multi-affine (ama) feature-matching algorithm and its application to minimally-invasive surgery images, in: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 2371–2376.

Rai, L., Higgins, W.E., 2008. Method for radiometric calibration of an endoscopes camera and light source, in: SPIE Medical Imaging: Visualization, Image-Guided Procedures, and Modeling, p. 691813.

Rapp, H., 2007. Experimental and Theoretical Investigation of Correlating ToF-Camera Systems. Master's thesis. University of Heidelberg.

Rauth, T.P., Bao, P.Q., Galloway, R.L., Bieszczad, J., Friets, E.M., Knaus, D.A., Kynor, D.B., Herline, A.J., 2007. Laparoscopic surface scanning and subsurface targeting: implications for image-guided laparoscopic liver surgery. Surgery 142, 207–214.

Richa, R., Bo, x, L., A.P., Poignet, P., 2010. Beating heart motion prediction for robust visual tracking, in: IEEE International Conference on Robotics and Automation (ICRA), pp. 4579–4584.

Richa, R., Bo, A., Poignet, P., 2011. Towards robust 3D visual tracking for motion compensation in beating heart surgery. Med Image Anal 15, 302 – 315.

Richa, R., Poignet, P., Liu, C., 2008a. Deformable motion tracking of the heart surface, in: IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 3997 – 4003.

Richa, R., Poignet, P., Liu, C., 2008b. Efficient 3D tracking for motion compensation in beating heart surgery, in: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 684–691.

Ringbeck, T., 2009. A performance review of 3D ToF vision systems in comparision to stereo vision systems. Technical Report. PMD Technologies GmbH.

Robinson, A., Alboul, L., Rodrigues, M., 2004. Methods for indexing stripes in uncoded structured light scanning systems. Journal of WSCG 3, 371–378.

Röhl, S., Bodenstedt, S., Suwelack, S., Dillmann, R., Speidel, S., Kenngott, H., Müller-Stich, B.P., 2012. Dense GPU-enhanced surface reconstruction from stereo endoscopic images for intraoperative registration. Med Phys 39, 1632–1645.

Russell, C., Fayad, J., Agapito, L., 2011. Energy based multiple model fitting for non-rigid structure from motion, in: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3009 – 3016.

Salvi, J., Pages, J., Batlle, J., 2004. Pattern codification strategies in structured light systems. Pattern Recogn 37, 827–849.

Salzmann, M., Fua, P., 2011. Linear local models for monocular reconstruction of deformable surfaces. IEEE T Pattern Anal 33, 931 – 944.

dos Santos, T.R., Goch, C.J., Franz, A.M., Meinzer, H.P., Heimann, T., Maier-Hein, L., 2012. Minimally deformed correspondences between surfaces for intra-operative registration, in: SPIE Medical Imaging: Image Processing, p. 83141C.

Sauvee, M., Noce, A., Poignet, P., Triboulet, J., Dombre, E., 2007. Three-dimensional heart motion estimation using endoscopic monocular vision system: From artificial landmarks to texture analysis. Biomed Signal Proces 2, 199–207.

Scarzanella, M., 2012. 3D Reconstruction from Stereo and Photometric Cues in Minimally Invasive Surgery. Ph.D. thesis. Imperial College, London.

Schaller, C., Adelt, A., Penne, J., Hornegger, J., 2009. Time-of-flight sensor for patient positioning, in: SPIE Medical Imaging: Visualization, Image-Guided Procedures, and Modeling, p. 726110.

Schaller, C., Penne, J., Hornegger, J., 2008. Time-of-flight sensor for respiratory motion gating. Med Phys 35, 3090–3093.

Scharstein, D., Szeliski, R., 2002. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. Int J Comput Vision 47, 7–42.

Schmalz, C., Forster, F., Schick, A., Angelopoulou, E., 2012. An endoscopic 3D scanner based on structured light. Med Image Anal 16, 1063–1072.

Schroeder, W.J., Ibanez, L., Martin, K., 2004. Software process: The key to developing robust, reusable and maintainable open-source software, in: IEEE International Symposium On Biomedical Imaging (ISBI), pp. 15–18.

Seitel, A., 2012. Markerless Navigation For Percutaneus Needle Insertions. Ph.D. thesis. German Cancer Research Center (DKFZ) Heidelberg.

Sepp, W., Fuchs, S., 2012. `http://www.dlr.de/rm/desktopdefault.aspx/tabid-4853/6084_read-9201/`. Accessed online 25-June-2012.

Shekhar, R., Dandekar, O., Bhat, V., Philip, M., Lei, P., Godinez, C., E.Sutton, George, I., Kavic, S., Mezrich, R., Park, A., 2010. Live augmented reality: a new visualization method for laparoscopic surgery using continuous volumetric computed tomography. Surg Endosc 24, 1976–1985.

Simpfendorfer, T., Baumhauer, M., Mueller, M., Gutt, C.N., Meinzer, H.P., Rassweiler, J.J., Guven, S., Teber, D., 2011. Augmented reality visualization during laparoscopic radical prostatectomy. J Endourol 25, 1841–1845.

Simpson, A., Dumpuri, P., Jarnagin, W., Miga, M., 2012. Model-assisted image-guided liver surgery using sparse intraoperative data, in: Payan, Y. (Ed.), Soft Tissue Biomechanical Modeling for Computer Assisted Surgery. Springer Berlin Heidelberg. volume 11 of *Studies in Mechanobiology, Tissue Engineering and Biomaterials*, pp. 7–40.

Skrinjar, O., Studholme, C., Nabavi, A., Duncan, J., 2001. Steps toward a stereo-camera-guided biomechanical model for brain shift compensation, in: International Conference on Information Processing in Medical Imaging (IPMI).

Soper, T., Porter, M., Seibel, E., 2012. Surface mosaics of the bladder reconstructed from endoscopic video for automated surveillance. IEEE T Bio-med Eng 59, 1670 –1680.

Soutschek, S., Penne, J., Hornegger, J., 2008. 3D gesture-based scene navigation in medical imaging applications using time-of-flight cameras, in: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR): Workshop on ToF-Camera based Computer Vision, pp. 1 – 6.

Speidel, S., Sudra, G., Senemaud, J., Drentschew, M., Müller-Stich, B.P., Gutt, C., Dillmann, R., 2008. Recognition of risk situations based on endoscopic instrument tracking and knowledge based situation modeling, in: : Visualization, Image-guided Procedures, and Modeling.

Stoyanov, D., 2012a. Stereoscopic scene flow for robotic assisted surgery, in: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI).

Stoyanov, D., 2012b. Surgical vision. Ann Biomed Eng 40, 332–34.

Stoyanov, D., Darzi, A., Yang, G.Z., 2004. Dense 3D depth recovery for soft tissue deformation during robotically assisted laparoscopic surgery, in: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 41–48.

Stoyanov, D., Darzi, A., Yang, G.Z., 2005a. Laparoscope self-calibration for robotic assisted minimally invasive surgery, in: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 114–121.

Stoyanov, D., Mylonas, G., Deligianni, F., Darzi, A., Yang, G., 2005b. Soft-tissue motion tracking and structure estimation for robotic assisted MIS procedures, in: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI). volume 3750, pp. 139–146.

Stoyanov, D., Rayshubskiy, A., Hillman, E., 2012. Robust registration of multispectra images of the cortical surface in neurosurgery, in: IEEE International Symposium On Biomedical Imaging (ISBI), pp. 1643–1646.

Stoyanov, D., Scarzanella, M.V., Pratt, P., Yang, G.Z., 2010. Real-time stereo reconstruction in robotically assisted minimally invasive surgery, in: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 275–282.

Stoyanov, D., Yang, G.Z., 2005. Removing specular reflection components for robotic assisted laparoscopic surgery, in: International Conference on Image Processing (ICIP), pp. 632–635.

Stoyanov, D., Yang, G.Z., 2007. Stabilization of image motion for robotic assisted beating heart surgery, in: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 417–424.

Streckel, B., Bartczak, B., Koch, R., Kolb, A., 2007. Supporting structure from motion with a 3D-range-camera, in: Proceedings of the 15th Scandinavian conference on Image analysis, pp. 233–242.

Su, L.M., Vagvolgyi, B.P., Agarwal, R., Reiley, C.E., Taylor, R.H., Hager, G.D., 2009. Augmented reality during robot-assisted laparoscopic partial nephrectomy: Toward real-time 3D-CT to stereoscopic video registration. Urology 73, 896 – 900.

Suwelack, S., Roehl, S., Dillmann, R., Wekerle, A., Kenngott, H., Mueller-Stich, B., Alt, C., Speidel, S., 2011a. Quadratic corotated finite elements for real-time soft tissue registration, in: MICCAI workshop: Computational Biomechanics for Medicine, pp. 39 – 50.

Suwelack, S., Talbot, H., Röhl, S., Dillmann, R., Speidel, S., 2011b. A biomechanical liver model for intraoperative soft tissue registration, in: SPIE Medical Imaging:Visualization, Image-Guided Procedures, and Modeling.

Suzuki, N., Hattori, A., Hashizume, M., 2008. Benefits of augmented reality function for laparoscopic and endoscopic surgical robot systems, in: MICCAI Workshop: AMI-ARCS, pp. 53 – 60.

Swirski, Y., Schechner, Y.Y., Nir, T., 2011. Variational stereo in dynamic illumination, in: International Conference on Computer Vision (ICCV), pp. 1124–1131.

Szpala, S., Wierzbicki, M., Guiraudon, G., Peters, T.M., 2005. Real-time fusion of endoscopic views with dynamic 3-d cardiac images: a phantom study. IEEE Trans Med Imaging 24, 1207–1215.

Taffinder, N., Smith, S.G.T., Huber, J., Russell, R.C.G., Darzi, A., 1999. The effect of a second-generation 3D endoscope on the laparoscopic precision of novices and experienced surgeons. Surg Endosc 13, 1087–1092.

Tappen, M., Freeman, W., 2003. Comparison of graph cuts with belief propagation for stereo, using identical mrf parameters, in: International Conference on Computer Vision (ICCV), pp. 900 –906 vol.2.

Tardif, J.P., Roy, S., Meunier, J., 2003. Projector-based augmented reality in surgery without calibration, in: Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBS), pp. 548–551.

Taylor, J., Jepson, A.D., Kutulakos, K., 2010. Non-rigid structure from locally-rigid motion, in: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2761 – 2768.

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

Thrun, S., Liu, Y., Koller, D., Ng, A., Ghahramani, Z., Durrant-Whyte, H., 2004. Simultaneous localization and mapping with sparse extended information filters. The International Journal of Robotics Research 23, 693–716.

Totz, J., Fujii, K., Mountney, P., Yang, G., 2011a. Enhanced visualisation for minimally invasive surgery. Int J Comput Assist Radiol Surg .

Totz, J., Fujii, K., Mountney, P., Yang, G.Z., 2012. Enhanced visualisation for minimally invasive surgery. Int J Comput Assist Radiol Surg 7, 423–432.

Totz, J., Mountney, P., Stoyanov, D., Yang, G., 2011b. Dense surface reconstruction for enhanced navigation in MIS, in: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 89–96.

Ukimura, O., Gill, S., 2008. Imaging-assisted endoscopic surgery: Cleveland clinic experience. J Endourol 22(4), 803 – 810.

Ullman, S., 1979. The Interpretation of Visual Motion. MIT Press.

Varol, A., Salzmann, M., Tola, E., Fua, P., 2009. Template-free monocular reconstruction of deformable surfaces, in: International Conference on Computer Vision (ICCV), pp. 1811 – 1818.

Vercauteren, T., Perchant, A., Pennec, X., Ayache, N., 2005. Mosaicing of confocal microscopic *in vivo* soft tissue video sequences, in: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 753–760.

Vigneron, L., Warfield, S., Robe, P., Verly, J., 2011. 3d xfem-based modeling of retraction for preoperative image update. Comp Aid Surg 16, 121–134.

Visentini-Scarzanella, M., Mylonas, G.P., Stoyanov, D., Yang, G.Z., 2009. i-BRUSH: A gaze-contingent virtual paintbrush for dense 3D reconstruction in robotic assisted surgery, in: International Conference on Medical Image Computing and Computer-Assisted Intervention (MIC-CAI), pp. 353–360.

Wang, C., Bronstein, M.M., Bronstein, A.M., Paragios, N., 2010. Discrete minimum distortion correspondence problems for non-rigid shape matching, in: Conference on Scale Space and Variational Methods in Computer Vision (SSVM), pp. 580 – 591.

Warren, A., Mountney, P., Noonan, D., Yang, G.Z., 2012. Horizon stabilized - dynamic view expansion for robotic assisted surgery (hs-dve). Int J Comput Assist Radiol Surg 7, 281–288.

Wengert, C., Cattin, P.C., Duff, J.M., und Gabor Szekely, C.B., 2006. Markerless endoscopic registration and referencing, in: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 816–823.

Windheuser, T., Schlickewei, U., Schmidt, F.R., Cremers, D., 2011. Geometrically consistent elastic matching of 3D shapes: A linear programming solution, in: International Conference on Computer Vision (ICCV), pp. 2134 – 2141.

Wittek, A., Hawkins, T., Miller, K., 2009. On the unimportance of constitutive models in computing brain deformation for image-guided surgery. Biomechanics and Modeling in Mechanobiology 8, 77–84.

Wittek, A., Miller, K., Kikinis, R., Warfield., S., 2007. Patient-specific model of brain deformation: Application to medical image registration. Journal of biomechanics 40(4), 919–929.

Wöhler, C., D'Angelo, P., 2009. Stereo image analysis of non-lambertian surfaces. Int J Comput Vision 81, 172–190.

Wolf, I., Vetter, M., Wegner, I., Böttger, T., Nolden, M., Schöbinger, M., Hastenteufel, M., Kunert, T., Meinzer, H.P., 2005. The Medical Imaging Interaction Toolkit. Med Image Anal 9, 594–604.

Wu, C., Narasimhan, S.G., Jaramaz, B., 2010. A multi-image shape-from-shading framework for near-lighting perspective endoscopes. Int J Comput Vision 86, 211–228.

Wu, C., Wilburn, B., Matsushita, Y., Theobalt, C., 2011. High-quality shape from multi-view stereo and shading under general illumination, in: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), pp. 969 –976.

Wu, C.H., Sun, Y.N., Chang, C.C., 2007. Three-dimensional modeling from endoscopic video using geometric constraints via feature positioning. IEEE T Bio-med Eng 54, 1199–1211.

Wu, T.T., Qu, J.Y., 2007. Optical imaging for medical diagnosis based on active stereo vision and motion tracking. Opt Express 15, 10421–10426.

Xu, Z., Schwarte, R., Heinol, H., Buxbaum, B., Ringbeck, T., 1998. Smart pixel – photonic mixer device (PMD), in: Proc. Int. Conf. on Mechatron. & Machine Vision, pp. 259–264.

Yahav, G., Iddan, G.J., Mandelbaum, D., 2007. 3D imaging camera for gaming application, in: Digest of Technical Papers of Int. Conf. on Consumer Electronics, pp. 1 – 2.

Yang, Q., Engels, C., Akbarzadeh, A., 2008. Near real-time stereo for weakly-textured scenes, in: British Machine Vision Conference (BMVC), pp. 72.1–72.10.

Yang, Q., Yang, R., Davis, J., Nister, D., 2007. Spatial-depth super resolution for range images, in: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1 –8.

Yaniv, Z., Cleary, K., 2006. Image-Guided Procedures: A Review. Technical Report. Georgetown University, Imaging Science and Information Systems Center, Computer Aided Interventions and Medical Robotics.

Yeung, S., Tsui, H., Yim, A., 1999. Global shape from shading for an endoscope image, in: International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI), pp. 318–327.

Yip, M., Lowe, D., Salcudean, S., Rohling, R., Nguan, C., 2012. Tissue tracking and registration for image-guided surgery. IEEE Trans Med Imaging 31, 2169–2182.

Yoon, K.J., Kweon, I.S., 2006. Adaptive support-weight approach for correspondence search. IEEE T Pattern Anal 28, 650–656.

Zeng, Y., Wang, C., Wang, Y., Gu, X., Samaras, D., Paragios, N., 2010. Dense non-rigid surface registration using high-order graph matching, in: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), pp. 382–389.

Zhang, H., Sheffer, A., Cohen-Or, D., Zhou, Q., van Kaick, O., Tagliasacchi, A., 2008. Deformation-driven shape correspondence, in: Eurographics Symposium on Geometry Processing, pp. 1431–1439.

Zhang, R., Tsai, P.S., Cryer, J., Shah, M., 1999. Shape from shading: A survey. IEEE T Pattern Anal 21, 609–706.

Zhang, Z., 2000. A flexible new technique for camera calibration. IEEE T Pattern Anal 22, 1330 – 1334.

Zhou, W., Kambhamettu, C., 2006. Binocular stereo dense matching in the presence of specular reflections, in: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2363–2370.

Zhu, J., Wang, L., Yang, R., Davis, J.E., Pan, Z., 2011. Reliability fusion of time-of-flight depth and stereo geometry for high quality depth maps. IEEE T Pattern Anal 33, 1400–1414.