Surgical Vision

Danail Stoyanov[1,2]
[1]Center for Medical Image Computing, Univeristy College London, WC1 2BT
[2]Hamlyn Centre for Robotic Surgery, Imperial College London, SW7 2AZ
E-mail: danail.stoyanov@ucl.ac.uk

Abstract

The emergence of Minimal Access Surgery (MAS) as a paradigm in modern healthcare treatment has created new challenges and opportunities for automated image understanding and computer vision. In MAS, images recovered from inside the body using specialised devices are used to visualise and operate on the surgical site but they can also be used to computationally infer *in vivo* 3D tissue surface shape, soft-tissue morphology and surgical instrument motion. This information is important for facilitating *in vivo* biophotonic imaging modalities where the interaction between light and tissue is used to infer the structural and functional properties of the tissue. This paper provides a review of the literature for computer vision and image understanding techniques applied to MAS images. The focus of the article is to elucidate a perspective on how computer vision techniques can be used to support and enhance the capabilities of biophotonic imaging modalities during surgery. Note that while MAS encompasses a variety of surgical specialisations this review does not involve procedures performed in the interventional suite. The review has been carried out based on searches in the PubMed and IEEE databases based on the article's keywords.

Keywords – computer vision, quantitative endoscopy, minimal access surgery, *in vivo* biophotonics, optical coherence tomography, confocal endomicroscopy, multispectral imaging, tissue deformation recovery

1. Introduction

Replacing traditional open surgery procedures with less invasive interventions is one of the most important paradigms of modern interventional healthcare delivery [1]. Minimal Access Surgery (MAS) involves the use of elongated instruments to operate on the internal anatomy under direct visualisation of the surgical site using cameras and fibre optics. By limiting access trauma to several small incisions for the camera and the specialised instruments, the patient benefits from shorter hospitalisation periods and improved cosmetic outcomes. Furthermore, surgical procedures can potentially be performed for diagnostic purposes and on patient groups that cannot undergo open surgery due to the risk of co-morbidity [2, 3]. With the potential advantages of MAS in mind, surgical methods are evolving into less invasive procedures such as Natural Orifice Transluminal Endoscopic Surgery (NOTES) [4].

For both diagnostic surgery and surgical disease treatment it is important to locate critical anatomical structures during the procedure in order to preserve nerves and blood vessels, as well as, to target malignant tissues and identify abnormalities. Localisation is challenging in MAS because the surgeon’s field of view is limited to the anatomical surface directly in front of the endoscope [5]. Intraoperative biophotonic imaging modalities which exploit the characteristics of light interacting with tissues can facilitate localisation by providing subsurface structural information at microscale resolutions [6]. In addition, biophotonic techniques can potentially provide functional information and together with the subsurface structural tissue properties this can lead to *in vivo* optical tissue biopsy [7]. However, the deployment of biophotonic techniques in clinical practice faces challenges in the design and miniaturisation of optical and mechanical components as well as in the imaging of dynamic, deformable tissues and the analysis and integration of imaging information within the surgical field of view.

Computer vision techniques for deriving information about the 3D shape of tissue surfaces and their morphology from the endoscopic images in MAS have been investigated for applications in Image-Guided Surgery (IGS) [8, 9]. In IGS, preoperative patient specific medical imaging data is registered to the surgical field of view in real-time [10]. Medical imaging provides information about anatomical structures at the diagnostic and surgical planning stages and this is knowledge is critical for delineating excision margins and avoiding critical regions during surgery. The MAS setup is well posed for augmented visualisation of the surgical site but aligning multimodal patient data is highly challenging in soft-tissue surgery as the anatomy undergoes significant changes between the data acquisition phase and the surgical procedure [11]. In this context, vision methods have been proposed as a means of computing real-time metric measurements from the surgical site for registering preoperative anatomical models and enabling aligned visualisation of patient specific data [8]. Additionally, vision-based metrics about the shape of the surgical site and the position of the surgical instruments makes it feasible to enhance the control of roboticized surgical tools.

By avoiding direct access to the internal anatomy with MAS, the manipulation of tissues using distally controlled surgical instruments becomes challenging. Robotics is one of the key technologies for overcoming this difficulty by providing dexterous and tremor stabilizing tools, better ergonomics and 3D visualisation, and thus facilitates MAS for complex surgical procedures [12, 13]. This is best exemplified by the da Vinci Surgical System by Intuitive Surgical, CA which has been used for MAS in a variety of surgical specialisations [14]. Although the da Vinci has demonstrated the feasibility of operating without large access incisions, direct patient benefits are still limited outside urology and robotic MAS techniques have yet to reach widespread clinical use. Improving the instrumental control capabilities of the surgeon is one of the key challenges requiring new mechatronic designs and also intelligent control modes of operation that adapt to the surgical site. For enhanced control accurate localisation of anatomical targets is required with characteristic, functional and structural information about tissues in real-time during surgery [15].

There is a convergence of requirements for IGS, robotic assisted surgery and *in vivo* biophotonic imaging modalities where all require intraoperative metrics from the surgical site that can potentially be provided by surgical computer vision. The focus of this article is to review the existing work in vision for MAS with perspective on techniques for extracting real-time information from the surgical site that can be used for assisting *in vivo* biophotonics imaging. Vision can be potentially be used to overcome the physical limitations of imaging tissues in highly dynamic and deformable procedures, to provide localisation of microscopic image information and to help the interpretation and classification of imaging pathologies.

2. Background

To extract metric information from the laparoscopic camera one can model the geometry of image projection using a pinhole camera model [16]. This geometric model relates the projection of points in the surgical scene onto the image plane using a set of intrinsic and extrinsic camera parameters as shown in Figure 1. The intrinsic camera parameters describe the optical characteristic of the imaging device such as the focal length and orientation between the image sensor and the centre of projection of the lens. Meanwhile the extrinsic parameters describe the orientation of the device in a reference coordinate system. Typically additional parameters are used to model the effects of lens aberration that distort the image projection. The wide field of view lens system in endoscopes is particularly susceptible to lens distortions that can be effectively modelled using radial and tangential distortion coefficients [17].

Determining the parameters of the camera model is a well understood problem in computer vision and photogrammetry [18, 19] and there are a number of toolkits available to automatically estimate the model coefficients given images of a predefined calibration object

[20-22]. In practice, the model parameters can change due to motion of the scope and optimisation of the field of view through focusing and zooming but real-time estimation of parameters is a highly sensitive problem in the presence of deformation. Online estimation of focal length variations during image acquisition has been studied in the computer vision community [16, 23] but this work has only been applied to MAS images with limited success [24].

In addition to describing the geometry of image formation, the photometric properties of an image can also be exploited to infer higher level information about shape. For such techniques it is necessary to model the image sensor's response to light and the endoscopic light source's distribution function and intensity. Only one method is reported in the literature applying photometric calibration to endoscopes [25] where the authors assume a Lambertian reflectance model and use a Macbeth colour chart to estimate the endoscope's response functions. Additionally, the respective light source and camera positions can also be estimated [26] though limited experimental validation of this technique has been reported. Detailed photometric analysis of MAS images also requires relaxation of the Lambertian model and more complex representation of the Bidirectional Reflectance Distribution Function (BRDF) but this is a challenging task for MAS images.

3. Tissue Surface Shape Estimation

Information about the 3D shape and structure of the surfaces in the surgical site can be recovered from laparoscopic images using different techniques based around the principles of our understanding of the human visual system.

3.1 Stationary Monocular Scope

Monocular images contain information about surface shape in the absence of parallax. This can be observed in images of tissue with poor texture characteristics where the change in perceived illumination over the homogeneous surface is an indicator of the surface's shape.

The aim of Shape-from-Shading (SFS) methods is to computationally infer shape information in a relative coordinate system. The problem is, however, under constrained and made possible only by simplifying the BRDF, assuming Lambertian reflectance and imposing constraints based on the imaging setup in endoscopy.

A common assumption used in SFS techniques for endoscopic images is that the light source and camera centre are coincident because at the scope's tip the physical distance between the two positions is very small. This was originally proposed with the additional constraint was that the projection model was orthographic rather than perspective in [27] and then popularised in [28, 29]. More recent works have enabled the use of the full perspective projection model [30, 31] which is more appropriate for endoscopes. Methods have also focused on relaxing the assumption of coincidence between light and camera [32]. This is important because the relative distance between the endoscope tip and the observed tissues is often small and thus amplifies the difference in irradiance with respect to the position of the light source.

One of the fundamental limitations of SFS is that the inferred shape is not in a metric coordinate frame and thus can only be used for inferring relative information. In addition, methods are typically designed to deal with single surfaces with uniform albedo and this is usually not the case over the image of the surgical site. A potential solution for these unknowns is to use preoperative imaging information in an iterative process using SfS for surface reconstruction in conjunction with a registration process aligning a preoperative model [33].

3.3 Moving Scope Techniques

It is common for endoscopes to move during surgery and, in particular, during diagnostic procedures in natural orifice regions such as bronchoscopy, colonoscopy and rhinoscopy. The camera motion creates parallax between observations of the tissue surfaces and these can be exploited to triangulate the 3D position of points on the tissue surface using techniques known as Structure-from-Motion (SfM). Early work in this area was reported for tracking a bronchoscope by using feature tracking and solving for the scope’s motion using constraints on the epipolar geometry [34]. Similar work based on optical flow has also been reported [35]. The main goal of such methods is to determine the motion of the camera during a procedure but this inherently involves inferring information about the structure of the scene at the same time. Currently, the most advanced systems for SfM in MAS make use of preoperative models of the anatomy to impose additional constraints [36]. Similar work had previously been reported using SfS and attempting to register the observed surface orientation to preoperative data in order to estimate the position of a bronchoscope within the preoperative model [37].

The concept of Simultaneous Localisation and Mapping (SLAM) has also been applied to MAS images in rhinoscopy [38]. The advantage of SLAM is that the uncertainty of 3D reconstruction can be modelled and reduced with multiple observations of a particular site. This approach was adapted for stereoscopic devices in general MAS [39] and recently has also been adapted to directly incorporate a rhythmic motion model for cardiac and respiratory motions [40].

Both SfM and SLAM techniques assume a rigid scene and this is not a valid constraint when dealing with MAS images. In the case of tissue deformation, the simultaneous motion of the tissue and the camera can potentially be filtered using robust statistics [41]. Early work on incorporation deformation in SfM for cardiothoracic surgery was reported based on a genetic algorithm [42, 43] and more recently, Non-Rigid SfM (NR-SfM) techniques have been applied to endoscopic images [44, 45]. The current stability and practicality of these methods is not fully understood but they are a promising step forward for monocular systems and may provide the theory required to build robust reconstruction over full procedures. The main challenges is this area are to build suitable motion priors that can be used to compensate for deforming tissues and to validate whether the baseline created by movement of the scope provides sufficient structure triangulation accuracy in clinical practice.

3.2 Stereoscopic Methods

Robotic surgery platforms like the da Vinci have popularised the use of stereoscopic laparoscopes for 3D visualisation of the surgical site [14]. As well as providing a platform for 3D visualisation, the stereo-laparoscope also can serve as a more reliable metric measurement device using computer vision. An early system for stereo reconstruction used an Field Programmable Gate Array (FPGA) implementation of a normalised cross-correlation algorithm with consistency checking [46]. More recently, a dynamic programming approach was developed and demonstrated to operate in real-time for producing a surface to register with preoperative data [47]. More global optimisation and consideration of temporal continuities has been reported in [48].

In a different approach a method based on a sparse set of salient features has been used to first recover a sparse 3D reconstruction of the surgical site and then propagate this information to achieve a quasi-dense reconstruction [49]. The method can also be used to estimate temporal motion by propagating sparse feature matches temporally and thus recover full 3D scene flow [50] as shown in Figure 2. In this area, there is a growing need to establish a standardised methodology for validation similar to the existing databases available in the computer vision community [51].

Stereo systems are currently the most accessible vision based methods for inferring shape in a single time frame. Stereo laparoscopic devices are used in clinical practice with the da Vinci surgical system and stereoscopic image pairs are not subject to problems due to deformation on a per frame basis. However, the main challenge in computational stereo techniques is establishing accurate point correspondence across the stereo images in the presence of specular highlights, homogeneous surfaces and large disparity discontinuities in the presence of instruments which also introduce occlusions and additional complications are introduced when the surgical site contains dynamic effects such as bleeding and smoke. There are also computational challenges to process full High Definition (HD) video data in real-time though these can, in part, be addressed by Graphical Processing Units (GPU) and parallelisation.

3.3 Active Techniques

Structured light reconstruction systems (SLR) are active vision methods that address the limitations of passive vision techniques by introducing patterns of projected light into the surgical scene. By encoding the light with particular patterns systems can recover 3D depth using a monocular camera [52, 53]. The light pattern can vary in shape and colour in order to provide robust identification in the image and if the projector is calibrated identification of the pattern results in direct 3D computation [54-57]. SLR can also be used with stereo configurations where the projected light adds texture onto the observed surfaces. In this case, the pattern coding does not need to be identified and traditional computational stereo techniques can be used to match the richly textured stereoscopic images.

Another form of active techniques relies on the use of specialized cameras that can measure the phase shift of light at the emitter and receiver and thus deduce the Time-of-Flight (ToF). With this information and the constant speed of light it is possible to infer the depth of the surgical site. Though this technology has been adapted to endoscopic devices the current systems produce limited resolution depth maps that can be difficult to calibrate and error prone when observing tissues with variable light absorption properties [58]. Never the less, this is a promising area of development as ToF may offer better optical penetration in certain conditions such as smoke, which are problematic for other techniques. Methods for registering ToF data for IGS system have recently been reported [59, 60].

4. Motion and Morphology Tracking

4.1 Feature Based Approaches

Tracking sparse landmarks on the tissue surface has been illustrated to work effectively in 2D for short-term laparoscopic video sequences in robotic assisted cardiac surgery [61, 62]. Using a stereoscopic system motion tracking can also be performed in 3D using Lucas-Kanade (LK) [63] iterations to convergence towards the position of tracked templates [64]. The choice of landmark to track be performed manually or using automated salient feature detectors [64].

The problem of locating and tracking salient regions in MAS becomes complicated when the camera moves, tissues deform and there is instrument tissue interaction. This is similar to the matching of widely disparate observation of a scene in outdoor environments and many saliency detectors and descriptors are available for this task [65]. However as notable difference for MAS image is the additional complexity resulting from tissue deformation and combinations of techniques have been used on endoscopic images [36, 66] and a method for learning the most appropriate descriptor for a feature has been reported [67].

4.2 Surface Tracking

Incorporating models of the surface geometry is one potential method for imposing priors onto tracking algorithms and at the same time providing an area of motion estimation rather than a set of discrete points. For estimating the cardiac surface motion in robotic assisted cardiac surgery two basic techniques were proposed[68, 69] and using similar principles a series of more advanced methods have been developed [66, 70].

Instead of using the full warped template to measure the correlation for optimising the surface parameters a feature based approach was used in [71] and a more complex method combining multiple techniques within a framework was developed in [66]. The use of explicit assumptions about the smoothness of the observed soft-tissue surfaces enables the reconstruction of homogenous tissue regions but does not handle discontinuities arising at instrument-tissue boundaries. More advanced surface models can consider biomechanical tissue parameters and use this information to better incorporate preoperative patient specific models, however work in this area this is still preliminary [72].

4.3 Rhythmic Motion Modelling and Prediction

Modelling rhythmic physiological motions arising from respiration and the cardiac cycle can be used to build predictive system to help tissue motion tracking and also to augment the control strategies of surgical robots [62, 73]. Deformation resulting from cardiac and respiratory cycles can be modelled as quasi-periodic or periodic signals and several approaches have been suggested for modelling motion components either individually or in combination [40, 62, 73-75]. The challenge in this area is to provide robust response methods for dealing with unexpected motions for example arising from arrhythmia.

5. Instrument Detection and Tracking

Detecting and tracking surgical instruments in the laparoscopic view is an important task for various applications. In robotic surgery systems, visual servoing requires vision based instrument tracking and a system for simplifying camera control was shown with the development of the AESOP robotic laparoscope holder [76]. The proposed visual instrument tracking method was based on the probabilistic classification of image pixels based on their colour using a Bayesian filter. This was followed by clustering, cluster analysis and temporal filtering. The basic principles of this system illustrate the main lines of thought behind methods developed in subsequent years. By using colour markers mounted onto the instrument tips the reliability and robustness of the detection process can be simplified [77, 78]. Alternative detection strategies can be based on image gradient information where image structures such as lines are assumed to belong to the outline of the laparoscopic instruments [79]. More recent systems for vision based instrument tracking for robotic camera holders have also been reported [80, 81]. For retinal microsurgery instruments can be tracked using the images from the stereomicroscope and a recent technique has been developed using mutual information [82].

In addition to performing 2D tracking in the image space, 3D tracking of the instruments in a metric coordinate system is also possible. With a stereoscopic laparoscope this can be performed using the stereo properties of the images [83]. Monocular approaches deal with the problem either by exploiting perspective cues given some knowledge of the instrument's shape [79, 84] or by introducing markers on the instrument tips that allow the 3D position to be calculated with absolute orientation methods. The design of proximal end marker marker attachments for tracking instruments in image-guided surgery is analysed in [85].

6. View Enhancement Techniques

6.1 Image Stitching

Typically endoscopes have a limited field of view and this inhibits observation of the surgical area and the localisation and navigation capabilities of the surgeon. A vision based solution to the problem is the use of image stitching techniques for mosaicing a set of spatially coherent images together [86, 87]. For laparoscopic images mosaicing methods based on SLAM [88] and on optical flow [89] have been reported for expanding the surgical field of view. For the examination of the bladder methods for creating panoramas of cystoscopy images have been proposed [90]. For a fibre bundle endoscopic system a mosaicing system was developed [91] and for images of the human retina a different mosaic algorithm has been developed based on registration [92].

6.2 Image Enhancement

Processing endoscopic images to improve contrast and visualisation of fine structures has been a longstanding area of interest in endoscopy with early work relying on video processing [93]. However, with the emergence of Narrow Band Imaging (NBI) for emphasising histological features such as capillaries video processing for such image enhancement is less necessary [94].

Resolution in endoscopic systems can be limited due to the size of the scope. Although modern systems have moved to HD imaging for coherent fibre based devices resolution can still be limited and image analysis methods for super resolution have been developed [95]. The challenge with such techniques is that they rely on multiple observations of the scene and this can be difficult and problematic in the presence of view dependent highlights and deformation.

Preprocessing specular highlights to improve the performance of tracking methods has been proposed several authors [96, 97]. Image information obscured by the highlights can be recovered for rhythmically moving tissue, however, this is not practically valuable as specular highlights are a visual cue and hence should not be removed for observation by the surgeon and furthermore, detection by adaptive thresholding is practically a suitable way of negating their potentially error prone influence [97, 98]. Perhaps a more practical application has been suggested for surgical tool detection and inpainting to remove large highlights on the instruments [99].

6.3 Orientation Correction

Navigation in endoscopy and particularly colonoscopy can be challenging and vision system to help guide the surgeon based on identifying the central pathway for the scope have been proposed [100]. With NOTES procedures specialised instruments with embedded Inertial Measurement Sensors (IMU) have been developed to rectify the endoscopy image to a desired alignment [101]. In the case of retroflexing an articulated instrument orientation correction can potentially help the surgeon to align their visual motor axis [102].

7. Emerging Biophotonic Imaging Modalities for MAS

The goal of *in vivo* imaging modalities is to provide full functional, characteristic and geometric information about the surgical site for real-time diagnosis and accurate identification and delineation of surgical targets.

7.1 Confocal Endomicroscopy

Probe-based Confocal Laser Endomicroscopy (pCLE) is a clinically promising imaging modality that can potentially provide *in vivo* histopathology by providing cellular scale information about the structure of tissues in real-time during surgery [103]. While imaging tissues at this scale *in vivo* has tremendous potential, acquiring image at the microscopic scale means that providing sufficient and coherent image coverage over a tissue region is physically difficult. For this reason mosaicking is a particularly important for pCLE [104]. Various mosaicking techniques for pCLE images have been reported [105]. A technique for dealing with cumulative registration errors and accounting for small tissue deformations has recently been reported [106, 107].

The overall aim of pCLE is to provide *in vivo* real-time diagnostic capabilities by imaging tissue structure and allowing the endoscopist to link certain structural characteristic to known pathologies. This can be posed as a recognition and classification problem in vision for automatically guiding diagnosis [108]. However, evidence linking pathologies to certain visible pCLE patterns is still limited and even manual examination of the images by endoscopists is not certain. Therefore posing the problem as an image retrieval system, where given certain observed images during the intervention a set of matching images with previously annotated pathologies is return, has been proposed [109]. Other systems have aimed only at quantifying the observations in pCLE images by detecting vessels or quantifying crypt sizes [110].

In addition to providing mosaicking and image classification, it is important to place the microscopic pCLE observations within the endoscopic field of view that the surgeon is working in. Systems for retargeting the optical biopsy site have been proposed based on SfM and SLAM principles [111, 112] but limited validation has been performed for such techniques.

7.2 Optical Coherence Tomography

Optical Coherence Tomography (OCT) is a quantitative imaging modality capable of penetrating tissues below the exposed surface and obtaining cross sectional imaging of tissue microstructures *in vivo* [113]. An early system for forward facing OCT was implemented in a laparoscope and preliminary images from it were shown on *in vitro* tissues [114]. Since then applications for forward facing probes have mostly been directed towards lumen organs forming parts of the grastrointestinal tract [115]. A recent review of the different designs of imaging probes is available [116].

Particular interest has gone into building diagnostic probes for imaging within the esophagus for providing a means of diagnosing Barrett's esophagus, and imaging in the colon for carcinoma staging [115]. Two probe designs are currently most practical for clinical use: linear scanning catheters [117] and circumferential scanning catheters [118]. These produce different types of cross sectional images but both require sweeping of the probe beam usually involving mechanical actuation to provide a larger imaging region. Instead of mechanical actuation systems using optical trackers mounted on the proximal end of the probe have also been demonstrated as a means of obtaining manual hand-held 2D scans and volumetric 3D scans [119]. Vision based techniques can potentially provide an alternative approach to obtaining an OCT scan by fusing multiple cross sectional samples even in the presence of tissue deformation. By reducing the need for controlled scanning this can potentially enable other applications of OCT imaging such as sentinel node imaging in breast cancer surgery [120].

The field of surgery where OCT imaging is used in routine practice is ocular surgery where OCT can be used to estimate the thickness of the retinal nerve for the treatment of glaucoma. In fact this was the core application area that inspired the development of OCT [121]. Various alignment strategies have been developed for coregistering retinal OCT images [122]. Roboticized OCT probes have been developed for maintaining a stable and fixed distance from the imaging target [123]. With this system alignment between the surgical field of view observed through a stereoscopic microscope and the OCT observations has also been reported [124]. Stereoscopic microscopes can be modelled with the same geometric and photometric tools as endoscopic devices but practically calibration and overlay alignment are more challenging due to the small size of the surgical site and lens system configuration. Single fibre based OCT has also been integrated in a robotic surgical pick instrument for providing a subsurface scan view for better targeting [125].

7.3 Multispectral Imaging

Multispectral imaging is an optical modality based on the acquisition of multiple images of a sample at different wavelengths to compile a complete spectrum of the tissue. The spectrum can be used to monitor the distribution of chromophores in the tissue and provide tissue characterization capabilities [126]. However, the acquisition of a stack of multispectral images may take several hundred milliseconds and during endoscopic investigation or surgery the camera and tissue are not steady but change position, orientation and shape during the acquisition. This causes misalignment of the multispectral image stack and requires correction but the registration of multispectral images can be challenging due to the paucity of textural information, motion blur and low brightness for some spectra.

Gated system for multispectral images have been proposed where acquisition is synchronised with the cardiac cycle via ECG [127]. Another approach is to stabilise the tissue and ensure it is rigid and the camera is static [128]. A different methodology is to use white light images for tracking camera and tissue motion and thus to align the multispectral image stack as shown in Figure 5 [129]. The potential of this imaging modality and fusing it with other modalities is illustrated by a recent rotational scanning probe developed for optical and multispectral imaging [130]. An alternative use for multispectral imaging has also been reported in retinal microsurgery where variations in the illuminating light source were used to reduce phototoxicity exposure during the procedure [131].

9. Summary and Conclusions

The future of MAS procedures is in the convergence of intelligent robotic instrumentation, image-guidance and intraoperative imaging. Biophotonic imaging modalities are promising optical techniques for providing multi-scale functional and structural information about soft-tissues *in vivo*. They can potentially be integrated within current MAS instrumentation but the challenges of imaging dynamic and deformable targets require real-time metric information about the surgical site to be available for biophotonic image alignment, field of view expansion and image interpretation.

Vision is one of the common elements linking surgical MAS technologies because currently images are the main feedback from the MAS site used by the surgeon to operate. The importance and role of vision techniques in image-guidance [8], visualisation [132] and surgical robotics [9, 15] has previously been reported. This article has reviewed the state-of-the-art in computer vision techniques applied to images from MAS with a perspective on how vision can assist biophotonic imaging modalities such as pCLE, OCT and multispectral imaging. Surgical vision methods can be implemented directly on images delivered by surgical cameras such as endoscopes and surgical microscopes, and do not face vast integration resources and challenges for testing in the clinical setting. The main difficulty for surgical vision techniques is that applying computer vision algorithms in dynamic and deformable scenes with complex reflectance properties is not robust and algorithm failure is common [9]. This problem is compounded by practical challenges such as scope repositioning and re-entry, smoke, steam and blood smears occluding the visual field of view, and large instrument occlusions. The main requirements for robust and clinically usable vision techniques can be summarised as:

- System level design where the capabilities and of the vision system are considered in the implementation of the imaging device;
- Efficiency and real-time performance for creating systems that can be applied *in situ* while dealing with large data streams for computational analysis and visualisation with GPU technologies being are a promising solution for this task;
- Validation and requirements analysis to identify the capabilities and performance of vision techniques and to expose realistic application targets. This is currently an area that requires urgent attention from the community to consolidate existing works and create benchmark datasets. Some early work in this area has begun but making more datasets available online and discussing the problem of obtaining in vivo ground truth data would stimulate the community [9];
- Incorporation of prior information from patient specific models and imaging. While this is partially the goal of IGS the integration of such priors can be the solution to building robust system that are able to function throughout a full surgical procedure. Work in this area has been demonstrated in rhinoscopy [36] and bronchoscopy [37].

Many of these challenges highlight that surgical vision for MAS in general is still in the early stages of development. The integration of surgical vision with biophotonic imaging modalities is a potential area that can reach clinical translation before more complex and safety critical concepts such as intelligent robotic instrument control. Vision can provide biophotonic modalities with the capability to image dynamic tissues and fuse multiple image samples at the microscale to form a wider view of the imaging target. The main challenges for this integration to happen *in vivo* are:

- Miniaturisation and integration of optical and mechanical articulation components for providing the physical capability of biophotonic imaging through an endoscope;
- Modelling of the geometry of the imaging system components and their interaction in space and in time;
- Multimodal data processing considering the different image signal characteristics to build robust and repeatable imaging capabilities.

By addressing these areas and with the further progress of more complex computer-integrated systems in the operating theatre the field of surgical vision field has a promising future and could lead to tangible clinical impact. Emerging intraoperative imaging techniques such as ultrasound, optical modalities such as photoacoustic imaging and fluorescence, interventional MRI and hybrid surgery theatres with fluoroscopy capabilities are all contributing to a very dynamic field of research that will have significant impact on the future of interventional healthcare delivery. While it is impossible to provide a comprehensive review of such a broad and multidisciplinary topic that encompasses different fields of study this article has hopefully identified some ideas and technical content that will stimulate further developments in the field.

10. Acknowledgements

The insightful and constructive comments of the anonymous reviewers were very helpful for improving this article. The work was carried out with the financial support of a Royal Academy of Engineering/EPSRC Research Fellowship.

References

[1] Darzi, A. and S. Mackay. Recent advances in minimal access surgery. *British Medical Journal*. 7328:31-34, 2002.

[2] Talamini, M. A., S. Chapman, S. Horgan and W. S. Melvin. A prospective analysis of 211 robotic-assisted surgical procedures. *Surgical Endoscopy*. 10:1521-1524, 2003.

[3] Nasseri, S. S., J. L. Kasperbauer, S. E. Strome, T. V. McCaffrey, J. L. Atkinson and F. B. Meyer. Endoscopic transnasal pituitary surgery: report on 180 cases. *American Journal of Rhinology*. 4:281-7, 2001.

[4] Rattner, D. and A. Kalloo. ASGE/SAGES Working group on natural orifice translumenal endoscopic surgery. *Surg Endosc*. 329-333, 2006.

[5] Baumhauer, M., M. Feuerstein, H. P. Meinzer and J. Rassweiler. Navigation in endoscopic soft tissue surgery: perspectives and limitations. *Journal of Endourology*. 4:751-761, 2008.

[6] Iftimia, N., W. R. Brugge and D. X. Hammer. Advances in optical imaging for clinical medicine. Hoboken, N.J.: Wiley, 2011.

[7] Tearney, G. J., M. E. Brezinski, B. E. Bouma, S. A. Boppart, C. Pitris, J. F. Southern and J. G. Fujimoto. In Vivo Endoscopic Optical Biopsy with Optical Coherence Tomography. *Science*. 5321:2037-2039, 1997.

[8] Mirota, D. J., M. Ishii and G. D. Hager. Vision-Based Navigation in Image-Guided Interventions. *Annual Review of Biomedical Engineering*. 1:297-319, 2011.

[9] Mountney, P., D. Stoyanov and Y. Guang-Zhong. Three-Dimensional Tissue Deformation Recovery and Tracking. *Signal Processing Magazine, IEEE*. 4:14-24, 2010.

[10] Peters, T. M. Image-guidance for surgical procedures. *Physics in Medicine and Biology*. R505-R540, 2006.

[11] Zitová, B. and J. Flusser. Image registration methods: a survey. *Image and Vision Computing*. 11:977-1000, 2003.

[12] Taylor, R. H. and D. Stoianovici. Medical Robotics in Computer-Integrated Surgery. *IEEE Transactions on Robotics and Automation*. 5:765-781, 2003.

[13] Camarillo, D. B., T. M. Krummel and J. K. Salisbury Jr. Robotic technology in surgery: Past, present, and future. *The American Journal of Surgery*. 4, Supplement 1:2-15, 2004.

[14] Guthart, G. S. and J. K. Salisbury. The intuitive telesurgery system: Overview and application. *IEEE International Conference on Robotics and Automation*. 618-621, 2000.

[15] Hager, G., A. Okamura, P. Kazanzides, L. Whitcomb, G. Fichtinger and R. Taylor. Surgical and interventional robotics: part III [Tutorial]. *Robotics & Automation Magazine, IEEE*. 4:84-93, 2008.

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

[17] Heikkila, J. and O. Silven. A Four-step Camera Calibration Procedure with Implicit Image Correction. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 1106-1112, 1997.

[18] Zhang, Z. A Flexible New Technique for Camera Calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 11:1330-1334, 2000.

[19] Tsai, R. Y. An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 364-374, 1986.

[20] Strobl, K. H., W. Sepp, S. Fuchs, C. Paredes and K. Arbter. DLR CalLab CalDe: The DLR Camera Calibration Toolbox. http://www.robotic.dlr.de/callab/. 2011.

[21] Bouget, J.-Y. Camera Calibration Toolkit for Matlab. http://www.vision.caltech.edu/bouguetj/calib_doc/. 2011.

[22] Stoyanov, D. Camera Calibration Tools. http://www.cs.ucl.ac.uk/staff/Dan.Stoyanov/calib/. 2011.

[23] Pollefeys, M., L. V. Gool and M. Proesmans. Stratified Self-Calibration from Image Sequences with Variable Focal Lengths. *European Conference on Computer Vision*. 31-42, 1996.

[24] Stoyanov, D., A. Darzi and G.-Z. Yang. Laparoscope Self-calibration for Robotic Assisted Minimally Invasive Surgery. *International Conference on Medical Image Computing and Computer Assisted Intervention*. 114-121, 2005.

[25] Wu, C., B. Jaramaz and S. G. Narasimhan. A full geometric and photometric calibration method for oblique-viewing endoscopes. *Comput Aided Surg*. 1-3:19-31, 2010.

[26] Stoyanov, D., D. Elson and G.-Z. Yang. Illumination position estimation for 3D soft-tissue reconstruction in robotic minimally invasive surgery. *Proceedings of the 2009 IEEE/RSJ international conference on Intelligent robots and systems*. 2628-2633, 2009.

[27] Rashid, H. U. and P. Burger. Differential algorithm for the determination of shape from shading using a point light source. *Image and Vision Computing*. 2:119-127, 1992.

[28] Deguchi, K. and T.Okatani. Shape Reconstruction from an Endoscope Image by Shape- from-Shading Technique for a Point Light Source at the Projection Center. *IEEE Workshop on Mathematical Methods in Biomedical Image Analysis*. 290-298, 1996.

[29] Okatani, T. and K. Deguchi. Shape Reconstruction from an Endoscope Image by Shape from Shading Technique for a Point Light Source at the Projection Centre. *Computer Vision and Image Understanding*. 2:119-131, 1997.

[30] Tankus, A., N. Sochen and Y. Yeshurun. Perspective Shape-from-Shading via Fast Marching. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2004.

[31] Prados, E. and O. Faugeras. Shape From Shading: A Well Posed Problem? *IEEE Conference on Computer Vision and Pattern Recognition*. 870-877, 2005.

[32] Wu, C., S. Narasimhan and B. Jaramaz. A Multi-Image Shape-from-Shading Framework for Near-Lighting Perspective Endoscopes. *International Journal of Computer Vision*. 2:211-228, 2010.

[33] Lamata, P., T. Morvan, M. Reimers, E. Samset and J. Declerck. Addressing Shading-Based Laparoscopic Registration

World Congress on Medical Physics and Biomedical Engineering, September 7 - 12, 2009, Munich, Germany. 189-192, 2009.

[34] Mori, K., D. Deguchi, J. Sugiyama, Y. Suenagaa, J. Toriwakia, C. R. M. Jr., H. Takabatake and H. Natori. Tracking of a bronchoscope using epipolar geometry analysis and intensity-based image registration of real and virtual endoscopic images. *Medical Image Analysis*. 3:321-336, 2002.

[35] Helferty, J. P. and W. E. Higgins. Combined endoscopic video tracking and virtual 3D CT registration for surgical guidance. *Image Processing. 2002. Proceedings. 2002 International Conference on*. II-961-II-964 vol.2, 2002.

[36] Mirota, D., H. Wang, R. Taylor, M. Ishii and G. Hager. Toward Video-Based Navigation for Endoscopic Endonasal Skull Base Surgery. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2009*. 91-99, 2009.

[37] Deligianni, F., A. Chung and G.-z. Yang. Patient-specific bronchoscope simulation with pq-space-based 2D/3D registration. *Computer Aided Surgery*. 5:215-226, 2004.

[38] Burschka, D., M. Li, R. Taylor and G. D. Hager. Scale-Invariant Registration of Monocular Endoscopic Images to CT-Scans for Sinus Surgery. *International Conference on Medical Image Computing and Computer Assisted Intervention*. 413–421, 2004.

[39] Mountney, P., D. Stoyanov, A. Davison and G.-Z. Yang. Simultaneous Stereoscope Localization and Soft-Tissue Mapping for Minimal Invasive Surgery. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2006*. 347-354, 2006.

[40] Mountney, P. and G.-Z. Yang. Motion Compensated SLAM for Image Guided Surgery. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2010*. 496-504, 2010.

[41] Grasa, Ó. G., J. Civera and J. M. M. Montiel. EKF Monocular SLAM with Relocalization for Laparoscopic Sequences. *IEEE International Conference on Robotics and Automation*. 2011.

[42] Hu, M., G. Penney, P. Edwards, M. Figl and D. Hawkes. 3D reconstruction of internal organ surfaces for minimal invasive surgery. *Med Image Comput Comput Assist Interv Int Conf Med Image Comput Comput Assist Interv*. Pt 1:68-77, 2007.

[43] Hu, M., G. Penney, D. Rueckert, P. Edwards, F. Bello, R. Casula, M. Figl and D. Hawkes. Non-rigid Reconstruction of the Beating Heart Surface for Minimally Invasive Cardiac Surgery. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2009*. 34-42, 2009.
[44] Malti, A., A. Bartoli and T. Collins. Template-Based Conformal Shape-from-Motion from Registered Laparoscopic Images. *Medical Image Understanding and Analysis Conference*. 2011.
[45] Collins, T., B. Compte and A. Bartoli. Deformable Shape-from-Motion in Laparoscopy using a Rigid Sliding Window. *Medical Image Understanding and Analysis Conference*. 2011.
[46] Mourgues, F., F. Devernay, G. Malandain and È. Coste-Manière. 3D reconstruction of the operating field for image overlay in 3D-endoscopic surgery. *International Symposium on Augmented Reality*. 2001.
[47] Hager, G., B. Vagvolgyi and D. Yuh. Stereoscopic Video Overlay with Deformable Registration. *Medicine Meets Virtual Reality*. 2007.
[48] Röhl, S., S. Bodenstedt, S. Suwelack, H. Kenngott, B. Mueller-Stich, R. Dillmanna and S. Speidel. Real-time surface reconstruction from stereo endoscopic images for intraoperative registration. *SPIE Medical Imaging*. 2011.
[49] Stoyanov, D., M. Scarzanella, P. Pratt and G.-Z. Yang. Real-Time Stereo Reconstruction in Robotically Assisted Minimally Invasive Surgery. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2010*. 275-282, 2010.
[50] Cech, J., J. Sanchez-Riera and R. Horaud. Scene flow estimation by growing correspondence seeds. *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. 3129-3136, 2011.
[51] Scharstein, D. and R. Szeliski. A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. *International Journal of Computer Vision*. 1/2/3:7-42, 2002.
[52] K, K. and A. JD. Real-time structured light depth extraction. *Proc SPIE*. 11, 2000.
[53] Fuchs, H., M. A. Livingston, R. Raskar, D. Colucci, K. Keller, A. State, J. R. Crawford, P. Rademacher, S. H. Drake and A. A. Meyer. Augmented reality visualization for laparoscopic surgery. *Medical Image Computing and Computer-Assisted Intervention*. 934-943, 1998.
[54] Zhang, G., J. He and X. Li. 3D vision inspection for internal surface based on circle structured light. *Sensors and Actuators A: Physical*. 1:68-75, 2005.
[55] Albitar, I. C., P. Graebling and C. Doignon. Robust Structured Light Coding for 3D Reconstruction. *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*. 1-6, 2007.
[56] Gioux, S., A. Mazhar, D. J. Cuccia, A. J. Durkin, B. J. Tromberg and J. V. Frangioni. Three-dimensional surface profile intensity correction for spatially modulated imaging. *Journal of Biomedical Optics*. 034-045, 2009.
[57] Clancy, N. T., D. Stoyanov, G.-Z. Yang and D. S. Elson. An endoscopic structured lighting probe using spectral encoding. *Proc. SPIE*. 2011.
[58] Penne, J., K. Holler, M. Sturmer, T. Schrauder, A. Schneider, R. Engelbrecht, H. Feusner, B. Schmauss and J. Hornegger. Time-of-Flight 3-D Endoscopy. *Proceedings of the 12th International Conference on Medical Image Computing and Computer-Assisted Intervention: Part I*. 467-474, 2009.
[59] dos Santos, T., A. Seitel, H.-P. Meinzer and L. Maier-Hein. Correspondences Search for Surface-Based Intra-Operative Registration. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2010*. 660-667, 2010.
[60] Maier-Hein, L., M. Schmidt, A. Franz, T. dos Santos, A. Seitel, B. Jähne, J. Fitzpatrick and H. Meinzer. Accounting for Anisotropic Noise in Fine Registration of Time-of-Flight Range Data with High-Resolution Surface Data. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2010*. 251-258, 2010.
[61] Ortmaier, T., M. Groger and G. Hirzinger. Robust motion estimation in robotic surgery on the beating heart. *Computer Assisted Radiology and Surgery*. 2002.
[62] Ortmaier, T., M. Gröger, D. H. Boehm, V. Falk and G. Hirzinger. Motion Estimation in Beating Heart Surgery. *IEEE Transactions on Biomedical Engineering*. 10:1729-1740, 2005.
[63] Lucas, B. and T. Kanade. An Iterative Image Registration Technique with an Application to Stereo Vision. *International Joint Conference on Artificial Intelligence*. 674-679, 1981.

[64] Stoyanov, D., G. P. Mylonas, F. Deligianni, A. Darzi and G.-Z. Yang. Soft-tissue Motion Tracking and Structure Estimation for Robotic Assisted MIS Procedures. *International Conference on Medical Image Computing and Computer Assisted Intervention*. 139-146, 2005.

[65] Mikolajczyk, K. and C. Schmid. Scale & Affine Invariant Interest Point Detectors. *International Journal of Computer Vision*. 1:63-86, 2004.

[66] Richa, R., A. Bó and P. Poignet. Robust 3D Visual Tracking for Robotic-Assisted Cardiac Interventions. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2010*. 267-274, 2010.

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

[68] Stoyanov, D., A. Darzi and G.-Z. Yang. Dense 3D Depth Recovery for Soft Tissue Deformation During Robotically Assisted Laparoscopic Surgery. *International Conference on Medical Image Computing and Computer Assisted Intervention*. 41-48, 2004.

[69] Lau, W. W., N. A. Ramey, J. Corso, N. V. Thakor and G. D. Hager. Stereo-Based Endoscopic Tracking of Cardiac Surface Deformation. *International Conference on Medical Image Computing and Computer Assisted Intervention*. 494-501, 2004.

[70] Richa, R., P. Poignet and C. Liu. Efficient 3D Tracking for Motion Compensation in Beating Heart Surgery. *International Conference on Medical Image Computing and Computer Assisted Intervention*. 2008.

[71] Stoyanov, D. and Y. Guang-Zhong. Soft tissue deformation tracking for robotic assisted minimally invasive surgery. *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*. 254-257, 2009.

[72] Pratt, P., D. Stoyanov, M. Visentini-Scarzanella and G.-Z. Yang. Dynamic guidance for robotic surgery using image- constrained biomechanical models. *Proceedings of the 13th international conference on Medical image computing and computer-assisted intervention: Part I*. 77-85, 2010.

[73] Ginhoux, R., J. Gangloff, M. d. Mathelin, L. Soler, M. M. A. Sanchez and J. Marescaux. Active Filtering of Physiological Motion in Robotized Surgery Using Predictive Control. *IEEE Transactions on Robotics*. 1:67-79, 2005.

[74] Richa, R., A. P. L. Bó and P. Poignet. Towards robust 3D visual tracking for motion compensation in beating heart surgery. *Medical Image Analysis*. 3:302-315, 2011.

[75] Bachta, W., P. Renaud, L. Cuvillon, E. Laroche, A. Forgione and J. Gangloff. Motion Prediction for Computer-Assisted Beating Heart Surgery. *Biomedical Engineering, IEEE Transactions on*. 11:2551-2563, 2009.

[76] Uecker, D. R., Y. F. Wang, C. Lee and Y. Wang. Laboratory Investigation:Automated Instrument Tracking in Robotically Assisted Laparoscopic Surgery. *Comput Aided Surg*. 6:308-325, 1995.

[77] Guo-Qing, W., K. Arbter and G. Hirzinger. Real-time visual servoing for laparoscopic surgery. Controlling robot motion with color image segmentation. *Engineering in Medicine and Biology Magazine, IEEE*. 1:40-45, 1997.

[78] Tonet, O., R. U. Thoranaghatte, G. Megali and P. Dario. Tracking endoscopic instruments without a localizer: A shape-analysis-based approach. *Comput Aided Surg*. 1:35-42, 2007.

[79] Cano, A., F. Gayá, P. Lamata, P. Sánchez-González and E. Gómez. Laparoscopic Tool Tracking Method for Augmented Reality Surgical Applications Biomedical Simulation. 191-196, 2008.

[80] Krupa, A., J. Gangloff, C. Doignon, M. F. d. Mathelin, M. Guillaume Morel, J. Leroy, L. Soler and J. Marescaux. Autonomous 3-D Positioning of Surgical Instruments in Robotized Laparoscopic Surgery Using Visual Servoing. *IEEE Transactions on Robotics and Automation*. 5:842-853, 2003.

[81] Voros, S., G. P. Haber, J. F. Menudet, J. A. Long and P. Cinquin. ViKY Robotic Scope Holder: Initial Clinical Experience and Preliminary Results Using Instrument Tracking. *Mechatronics, IEEE/ASME Transactions on*. 6:879-886, 2010.

[82] Richa, R., M. Balicki, E. Meisner, R. Sznitman, R. Taylor and G. Hager. Visual Tracking of Surgical Tools for Proximity Detection in Retinal Surgery. *Information Processing in Computer-Assisted Interventions*. 55-66, 2011.
[83] Devernay, F., F. Mourgues and E. Coste-Maniere. Towards endoscopic augmented reality for robotically assisted minimally invasive cardiac surgery. *Medical Imaging and Augmented Reality*. 2001.
[84] Doignon, C., F. Nageotte and M. de Mathelin. The Role of Insertion Points in the Detection and Positioning of Instruments in Laparoscopy for Robotic Tasks. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2006*. 527-534, 2006.
[85] West, J. B. and C. R. Maurer, Jr. Designing optically tracked instruments for image-guided surgery. *Medical Imaging, IEEE Transactions on*. 5:533-545, 2004.
[86] Seshamani, S., W. Lau and G. Hager. Real-Time Endoscopic Mosaicking. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2006*. 355-363, 2006.
[87] Seshamani, S. Direct global adjustment methods for endoscopic mosaicking. *Proc SPIE*. 1:72611D, 2009.
[88] Mountney, P. and Y. Guang-Zhong. Dynamic view expansion for minimally invasive surgery using simultaneous localization and mapping. *Engineering in Medicine and Biology Society, 2009. EMBC 2009. Annual International Conference of the IEEE*. 1184-1187, 2009.
[89] Lerotic, M., A. J. Chung, J. Clark, S. Valibeik and G.-Z. Yang. Dynamic View Expansion for Enhanced Navigation in Natural Orifice Transluminal Endoscopic Surgery. *Proceedings of the 11th International Conference on Medical Image Computing and Computer-Assisted Intervention, Part II*. 467-475, 2008.
[90] Hernández-Mier, Y., W. C. P. M. Blondel, C. Daul, D. Wolf and F. Guillemin. Fast construction of panoramic images for cystoscopic exploration. *Computerized Medical Imaging and Graphics*. 7:579-592, 2010.
[91] Atasoy, S., D. Noonan, S. Benhimane, N. Navab and G.-Z. Yang. A Global Approach for Automatic Fibroscopic Video Mosaicing in Minimally Invasive Diagnosis. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2008*. 850-857, 2008.
[92] Can, A., C. V. Stewart, B. Roysam and H. L. Tanenbaum. A feature-based technique for joint, linear estimation of high-order image-to-mosaic transformations: application to mosaicing the curved human retina. *Computer Vision and Pattern Recognition, 2000. Proceedings. IEEE Conference on*. 585-591 vol.2, 2000.
[93] Nishioka, N. S. and M.-A. Mycek. Initial experience with a real-time video processor for enhancing endoscopic image contrast. *Gastrointestinal Endoscopy*. 1:62-66, 1998.
[94] Hamamoto, Y., T. Endo, K. Nosho, Y. Arimura, M. Sato and K. Imai. Usefulness of narrow-band imaging endoscopy for diagnosis of Barrett's esophagus. *Journal of Gastroenterology*. 1:14-20, 2004.
[95] Lerotic, M. and G. Yang. Super resolution in robotic-assisted minimally invasive surgery. *Comput Aided Surg*. 6:347-56, 2007.
[96] Gröger, M., T. Ortmaier, W. Sepp and G. Hirzinger. Reconstruction of Image Structure in Presence of Specular Reflections. *DAGM-Symposium on Pattern Recognition*. 53-60, 2001.
[97] Stoyanov, D. and Y. Guang Zhong. Removing specular reflection components for robotic assisted laparoscopic surgery. *Image Processing, 2005. ICIP 2005. IEEE International Conference on*. III-632-5, 2005.
[98] Arnold, M., A. Ghosh, S. Ameling and G. Lacey. Automatic segmentation and inpainting of specular highlights for endoscopic imaging. *J Image Video Process*. 1-12, 2010.
[99] Saint-Pierre, C.-A., J. Boisvert, G. Grimard and F. Cheriet. Detection and correction of specular reflections for automatic surgical tool segmentation in thoracoscopic images. *Machine Vision and Applications*. 1:171-180, 2011.
[100] Wang, P., S. M. Krishnan, Y. Huang and N. Srinivasan. An adaptive segmentation technique for clinical endoscopic image processing. *Engineering in Medicine and Biology, 2002. 24th Annual*

Conference and the Annual Fall Meeting of the Biomedical Engineering Society EMBS/BMES Conference, 2002. Proceedings of the Second Joint. 1084-1085 vol.2, 2002.

[101] Höller, K., J. Penne, A. Schneider, J. Jahn, J. Guttiérrez Boronat, T. Wittenberg, H. Feußner and J. Hornegger. Endoscopic Orientation Correction. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2009*. 459-466, 2009.

[102] Warren, A., P. Mountney, D. Noonan and G.-Z. Yang. Horizon Stabilized—Dynamic View Expansion for Robotic Assisted Surgery (HS-DVE). *International Journal of Computer Assisted Radiology and Surgery*. 1-8, 2011.

[103] Le Goualher, G., A. Perchant, M. Genet, C. Cavé, B. Viellerobe, F. Berier, B. Abrat and N. Ayache. Towards Optical Biopsies with an Integrated Fibered Confocal Fluorescence Microscope. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2004*. 761-768, 2004.

[104] Becker, V., T. Vercauteren, C. H. von Weyhern, C. Prinz, R. M. Schmid and A. Meining. High-resolution miniprobe-based confocal microscopy in combination with video mosaicing (with video). *Gastrointestinal Endoscopy*. 5:1001-1007, 2007.

[105] Vercauteren, T., A. Perchant, X. Pennec and N. Ayache. Mosaicing of Confocal Microscopic In Vivo Soft Tissue Video Sequences. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2005*. 753-760, 2005.

[106] Loewke, K. E., D. B. Camarillo, W. Piyawattanametha, M. J. Mandella, C. H. Contag, S. Thrun and J. K. Salisbury. In Vivo Micro-Image Mosaicing. *Biomedical Engineering, IEEE Transactions on*. 1:159-171, 2011.

[107] Loewke, K., D. Camarillo, K. Salisbury and S. Thrun. Deformable Image Mosaicing for Optical Biopsy. *Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on*. 1-8, 2007.

[108] Srivastava, S., J. J. Rodriguez, A. R. Rouse, M. A. Brewer and A. F. Gmitro. Analysis of confocal microendoscope images for automatic detection of ovarian cancer. *Image Processing, 2005. ICIP 2005. IEEE International Conference on*. I-1113-16, 2005.

[109] André, B., T. Vercauteren, A. M. Buchner, M. B. Wallace and N. Ayache. A smart atlas for endomicroscopy using automated video retrieval. *Medical Image Analysis*. 4:460-476, 2011.

[110] Namati, E., J. Thiesse, J. de Ryk and G. McLennan. In vivo assessment of alveolar morphology using a flexible catheter-based confocal microscope. *Computer Vision, IET*. 4:228-235, 2008.

[111] Allain, B., M. Hu, L. Lovat, R. Cook, T. Vercauteren, S. Ourselin and D. Hawkes. A System for Biopsy Site Re-targeting with Uncertainty in Gastroenterology and Oropharyngeal Examinations. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2010*. 514-521, 2010.

[112] Mountney, P., S. Giannarou, D. Elson and G.-Z. Yang. Optical Biopsy Mapping for Minimally Invasive Cancer Screening. *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2009*. 483-490, 2009.

[113] Huang, D. and et al. Optical coherence tomography. *Science*. 5035:1178, 1991.

[114] Boppart, S. A., B. E. Bouma, C. Pitris, G. J. Tearney, J. G. Fujimoto and M. E. Brezinski. Forward-imaging instruments for optical coherence tomography. *Opt Lett*. 21:1618-1620, 1997.

[115] Suter, M. J., B. E. Bouma and G. J. Tearney. High-Resolution Optical Coherence Tomography Imaging in Gastroenterology. *Advances in Optical Imaging for Clinical Medicine*. 187-204, 2011.

[116] Atif, M., H. Ullah, M. Y. Hamza and M. Ikram. Catheters for optical coherence tomography. *Laser Physics Letters*. 9:629-646, 2011.

[117] Bouma, B. E. and G. J. Tearney. Power-efficient nonreciprocal interferometer and linear-scanning fiber-optic catheter for optical coherence tomography. *Opt Lett*. 8:531-533, 1999.

[118] Tearney, G. J., S. A. Boppart, B. E. Bouma, M. E. Brezinski, N. J. Weissman, J. F. Southern and J. G. Fujimoto. Scanning single-mode fiber optic catheter-endoscope for optical coherence tomography. *Opt Lett*. 7:543-545, 1996.

[119] Ren, J., J. Wu, E. J. McDowell and C. Yang. Manual-scanning optical coherence tomography probe based on position tracking. *Opt Lett*. 21:3400-3402, 2009.

[120] Nguyen, F. T., A. M. Zysk, E. J. Chaney, S. G. Adie, J. G. Kotynek, U. J. Oliphant, F. J. Bellafiore, K. M. Rowland, P. A. Johnson and S. A. Boppart. Optical Coherence Tomography: The Intraoperative

Assessment of Lymph Nodes in Breast Cancer. *Engineering in Medicine and Biology Magazine, IEEE*. 2:63-70, 2010.

[121] Hammer, D. X. Advances in Retinal Imaging. *Advances in Optical Imaging for Clinical Medicine*. 85-161, 2011.

[122] Haogang, Z., D. P. Crabb, P. G. Schlottmann, G. Wollstein and D. F. Garway-Heath. Aligning Scan Acquisition Circles in Optical Coherence Tomography Images of The Retinal Nerve Fibre Layer. *Medical Imaging, IEEE Transactions on*. 6:1228-1238, 2011.

[123] Mitchell, B., J. Koo, M. Iordachita, P. Kazanzides, A. Kapoor, J. Handa, G. Hager and R. Taylor. Development and Application of a New Steady-Hand Manipulator for Retinal Surgery. *Robotics and Automation, 2007 IEEE International Conference on*. 623-629, 2007.

[124] Fleming, I. N., S. Voros, B. Vagvolgyi, Z. Pezzementi, J. Handa, R. Taylor and G. D. Hager. Intraoperative Visualization of Anatomical Targets in Retinal Surgery. *Proceedings of the 2008 IEEE Workshop on Applications of Computer Vision*. 1-6, 2008.

[125] Balicki, M., J.-H. Han, I. Iordachita, P. Gehlbach, J. Handa, R. Taylor and J. Kang. Single Fiber Optical Coherence Tomography Microsurgical Instruments for Computer and Robot-Assisted Retinal Surgery. *Proceedings of the 12th International Conference on Medical Image Computing and Computer-Assisted Intervention: Part I*. 108-115, 2009.

[126] Sorg, B. S., B. J. Moeller, Y. Cao and M. W. Dewhirst. Hyperspectral imaging of hemoglobin saturation in tumor microvasculature and tumor hypoxia development. *J Biomed Opt*. 2005.

[127] Nighswander-Rempel, S. P., R. A. Shaw, V. V. Kupriyanov, J. Rendell, B. Xiang and H. H. Mantsch. Mapping tissue oxygenation in the beating heart with near-infrared spectroscopic imaging. *Vibrational Spectroscopy*. 1:85-94, 2003.

[128] Nighswander-Rempel, S. P., R. Anthony Shaw, J. R. Mansfield, M. Hewko, V. V. Kupriyanov and H. H. Mantsch. Regional Variations in Myocardial Tissue Oxygenation Mapped by Near-Infrared Spectroscopic Imaging. *Journal of Molecular and Cellular Cardiology*. 9:1195-1203, 2002.

[129] Neil, T. C., S. Danail, S. Vincent, J. David, Y. Guang-Zhong and S. E. Daniel. A Triple Endoscope System for Alignment of Multispectral Images of Moving Tissue. BTuD27, 2010.

[130] Qiu, L., D. K. Pleskow, R. Chuttani, E. Vitkin, J. Leyden, N. Ozden, S. Itani, L. Guo, A. Sacks, J. D. Goldsmith, M. D. Modell, E. B. Hanlon, I. Itzkan and L. T. Perelman. Multispectral scanning during endoscopy guides biopsy of dysplasia in Barrett's esophagus. *Nat Med*. 5:603-606, 2010.

[131] Sznitman, R., S. Billings, D. Rother, D. Mirota, Y. Yang, J. Handa, P. Gehlbach, J. U. Kang, G. D. Hager and R. Taylor. Active multispectral illumination and image fusion for retinal microsurgery. *Proceedings of the First international conference on Information processing in computer-assisted interventions*. 12-22, 2010.

[132] Stoyanov, D., M. Lerotic, G. Mylonas, A. J. Chun and G.-Z. Yang. Intra-operative Visualizations: Perceptual Fidelity and Human Factors. *IEEE/OSA Journal of Display Technologies*. 4:491-501, 2008.

[133] Wu, J., M. Conry, C. Gu, F. Wang, Z. Yaqoob and C. Yang. Paired-angle-rotation scanning optical coherence tomography forward-imaging probe. *Opt Lett*. 9:1265-1267, 2006.

[134] Clancy, N. T., D. Stoyanov, V. Sauvage, D. James, G.-Z. Yang and D. S. Elson. A Triple Endoscope System for Alignment of Multispectral Images of Moving Tissue. BTuD27, 2010.

Figure Legends

Figure 1. (a) Image of a stereo laparoscope from the da Vinci Surgical System (Intuitive Surgical, CA, USA); (b) geometry of image formation using a pinhole camera model.

Figure 2. (a) Examples of in vivo robotic assisted MIS images taken with the da Vinci surgical system; (b) the corresponding disparity maps for each image computed with the method proposed in this study, where light colours indicate closer to the camera; (c) motion heat maps where warmed colours indicate larger motion [49].

Figure 3. Mosaics constructed from pCLE images acquired with Cellvizio, Mauna Kea Technologies showing (a) biliary strictures; (b) intestinal metaplasia; (c) colon hyperplasia. Images provided by Mauna Kea Technologies courtesy of Prof. Alexander Meining, Klinikum Rechts der Isar, Munich, Germany and Dr. E. Coron, CHU Nantes, France.

Figure 4. (a) Schematic illustration of a forward imaging PARS-OCT probe used in ophthalmic surgery to obtain cross-sectional images of the retina [133]; (b) circular scan images acquired with the PARS-OCT probe from an enucleated porcine eye with removed cornea, lens, and vitreous; (c) linear scan images from the same experiment as (b) showing the posterior segment and the retinal detachment respectively. Images courtesy of Prof. Changhui Yang, Department of Electrical Engineering and Bioengineering, California Institute of Technology.

Figure 5. (top row) A row of images showing the white light images captured by the left camera of the trinocular setup; (middle row) row of corresponding multispectral images showing information captured at 530nm - 620nm with 10nm intervals captured with the system proposed in [134]; (bottom row) the temporal track of a feature overlaid onto of the 620nm multispectral image showing the misalignment motion without correction and the

corresponding multispectral patches after correction also shown as a rendition in the 3D surgical scene with the calibrated trinocular setup and the tracked 3D point on the surface of the tissue.

Figures

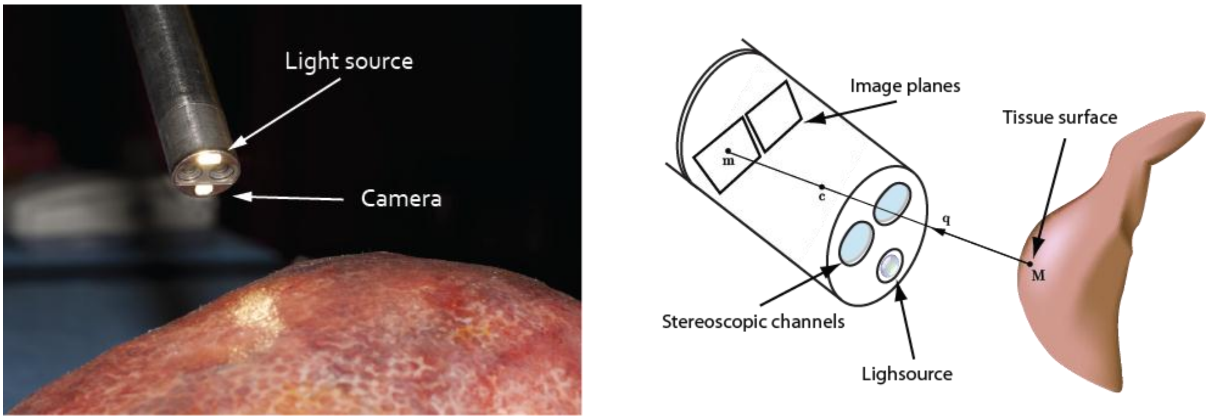

Figure 1. Danail Stoyanov

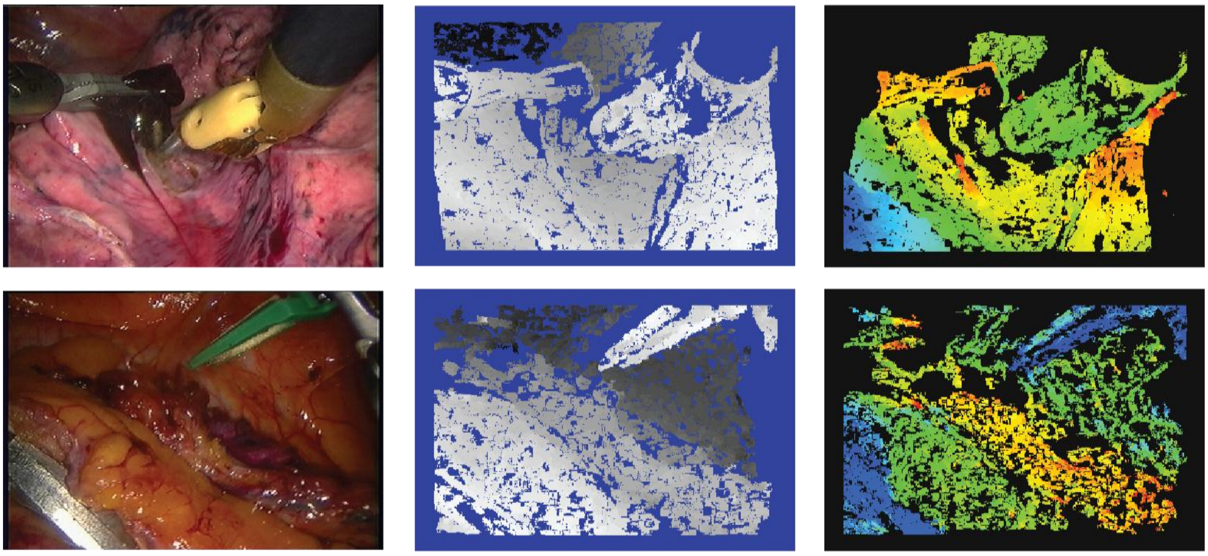

Figure 2. Danail Stoyanov

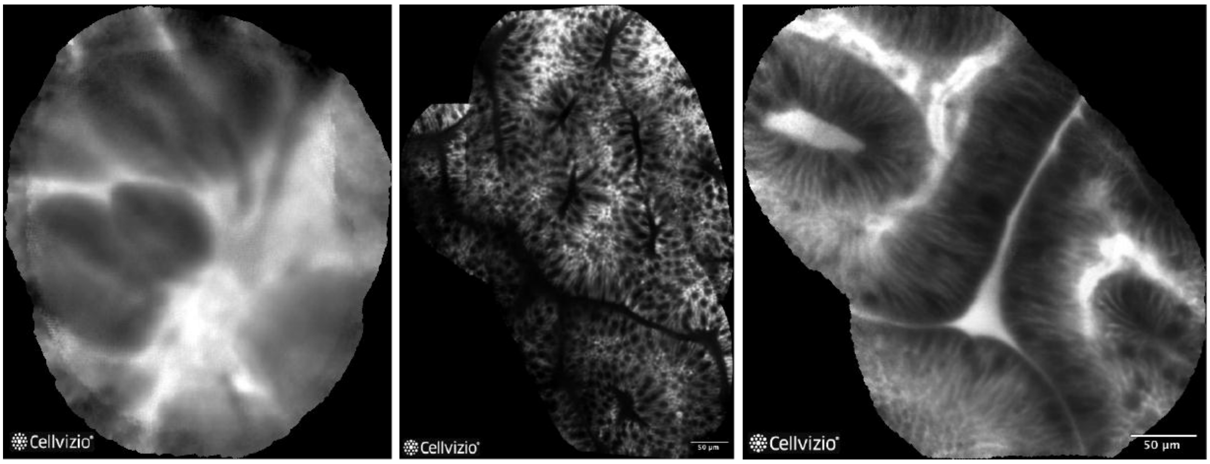

Figure 3. Mauna Kea Technologies

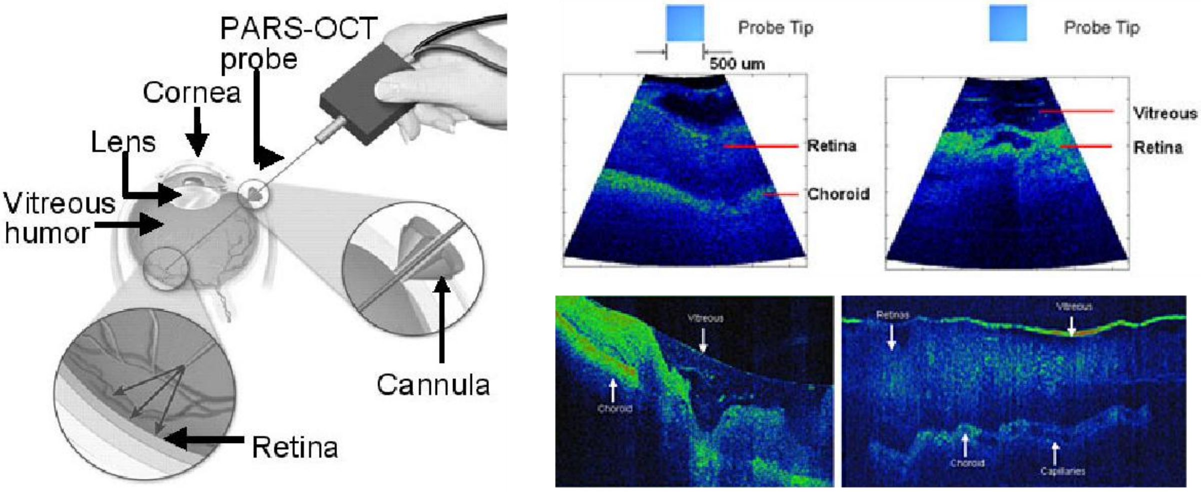

Figure 4. Prof. Changhui Yang

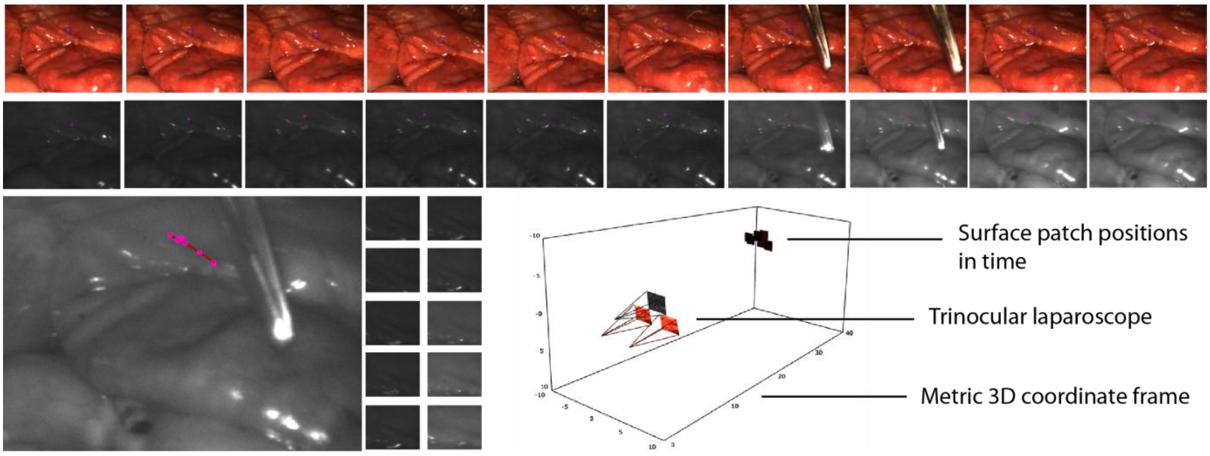

Figure 5. Danail Stoyanov