Computer Methods and Programs in Biomedicine 158 (2018) 135–146

Contents lists available at ScienceDirect

Computer Methods and Programs in Biomedicine

journal homepage: www.elsevier.com/locate/cmpb

SLAM-based dense surface reconstruction in monocular Minimally Invasive Surgery and its application to Augmented Reality

Long Chen[a], Wen Tang[a,*], Nigel W. John[b], Tao Ruan Wan[c], Jian Jun Zhang[a]

[a] Bournemouth University, UK
[b] University of Chester, UK
[c] University of Bradford, UK

ARTICLE INFO

Article history:
Received 14 February 2017
Revised 3 January 2018
Accepted 2 February 2018

Keywords:
SLAM
Surface reconstruction
Augmented Reality
Minimally Invasive Surgery

ABSTRACT

Background and objective: While Minimally Invasive Surgery (MIS) offers considerable benefits to patients, it also imposes big challenges on a surgeon's performance due to well-known issues and restrictions associated with the field of view (FOV), hand-eye misalignment and disorientation, as well as the lack of stereoscopic depth perception in monocular endoscopy. Augmented Reality (AR) technology can help to overcome these limitations by augmenting the real scene with annotations, labels, tumour measurements or even a 3D reconstruction of anatomy structures at the target surgical locations. However, previous research attempts of using AR technology in monocular MIS surgical scenes have been mainly focused on the information overlay without addressing correct spatial calibrations, which could lead to incorrect localization of annotations and labels, and inaccurate depth cues and tumour measurements. In this paper, we present a novel intra-operative dense surface reconstruction framework that is capable of providing geometry information from only monocular MIS videos for geometry-aware AR applications such as site measurements and depth cues. We address a number of compelling issues in augmenting a scene for a monocular MIS environment, such as drifting and inaccurate planar mapping.

Methods: A state-of-the-art Simultaneous Localization And Mapping (SLAM) algorithm used in robotics has been extended to deal with monocular MIS surgical scenes for reliable endoscopic camera tracking and salient point mapping. A robust global 3D surface reconstruction framework has been developed for building a dense surface using only unorganized sparse point clouds extracted from the SLAM. The 3D surface reconstruction framework employs the Moving Least Squares (MLS) smoothing algorithm and the Poisson surface reconstruction framework for real time processing of the point clouds data set. Finally, the 3D geometric information of the surgical scene allows better understanding and accurate placement AR augmentations based on a robust 3D calibration.

Results: We demonstrate the clinical relevance of our proposed system through two examples: (a) measurement of the surface; (b) depth cues in monocular endoscopy. The performance and accuracy evaluations of the proposed framework consist of two steps. First, we have created a computer-generated endoscopy simulation video to quantify the accuracy of the camera tracking by comparing the results of the video camera tracking with the recorded ground-truth camera trajectories. The accuracy of the surface reconstruction is assessed by evaluating the Root Mean Square Distance (RMSD) of surface vertices of the reconstructed mesh with that of the ground truth 3D models. An error of 1.24 mm for the camera trajectories has been obtained and the RMSD for surface reconstruction is 2.54 mm, which compare favourably with previous approaches. Second, *in vivo* laparoscopic videos are used to examine the quality of accurate AR based annotation and measurement, and the creation of depth cues. These results show the potential promise of our geometry-aware AR technology to be used in MIS surgical scenes.

Conclusions: The results show that the new framework is robust and accurate in dealing with challenging situations such as the rapid endoscopy camera movements in monocular MIS scenes. Both camera

* Corresponding author.
E-mail addresses: chenl@bournemouth.ac.uk (L. Chen), wtang@bournemouth.ac.uk (W. Tang), nigel.john@chester.ac.uk (N.W. John), t.wan@bradford.ac.uk (T.R. Wan), jzhang@bournemouth.ac.uk (J.J. Zhang).

https://doi.org/10.1016/j.cmpb.2018.02.006

tracking and surface reconstruction based on a sparse point cloud are effective and operated in real-time. This demonstrates the potential of our algorithm for accurate AR localization and depth augmentation with geometric cues and correct surface measurements in MIS with monocular endoscopes.

1. Introduction

In Minimally Invasive Surgery (MIS), medical procedures are technically demanding, and the difficulty is exacerbated by well-known issues and restrictions associated with MIS, such as the limited field of view (FOV), lack of hand-eye alignment and orientation, and the lack of stereoscopic depth perception in monocular endoscopy. Augmented Reality (AR) technology can help overcome these limitations by overlaying additional information onto the real scene such as annotations at target surgical locations [18], labels [45], measurements of tumour sites [4] or even overlay a 3D reconstruction of anatomy [14,15].

Despite recent advances in powerful miniaturized AR hardware devices and improvements on vision based software algorithms, many issues in medical AR remain unsolved. In particular, the dramatic changes in tissue surface illumination and tissue deformation as well as the rapid movements of the endoscope during insertion and extrusion, all give rise to a set of unique challenges that call for innovative approaches. As with any other technological assisted medical procedure, the accuracy of AR in MIS is paramount.

The miniaturized devices in MIS mean that the Field of View (FOV) captured by a monocular endoscopic camera is usually very small, for example, only 30% to 40% of the whole liver surface is visible in one frame at one time [38]. Traditional AR approaches (i.e. marker-less AR) for MIS are mainly based on feature tracking methods that require those selected feature points to be within the field of view [15]. Given the restricted FOV, the algorithmic limitations of traditional methods can severely affect the precision of AR for procedure guidance. Our proposed geometry-aware AR framework addresses the issue by providing global 3D geometric information of the entire surgical scene so that the information overlay does not depend on the frame by frame local feature extractions, hence, greatly improving the reliability of AR augmentations.

Studies have shown that a typical human uses 14 visual cues to perceive depth, and 11 of the 14 cues do not require binocular vision [11]. For example, depth information can be inferred in monocular vision through occlusions, motion parallax, shadows and texture gradient, and relative size and familiar size etc. The cognitive process of monocular vision enables surgeons to perform laparoscopic under a 2D environment [30]. However, monocular depth cues can only roughly estimate the general depth between objects, the accurate distance between objects cannot be perceived [41]. Although examples of stereoscopic endoscopes do exist, they are not commonly accessible in medical practice [47,50]. We address the aforementioned challenges by providing accurate geometric measurements and artificially generating depth cues through AR technology, which are important improvements in monocular endoscope environment for surgeons to carry out complex procedures. In our AR framework, the distance between objects can be deciphered by relative sizes of AR labels and annotations.

A stereo endoscope can provide stereoscopic vision and such systems are currently available and often integrated into robotic systems (e.g. the da Vinci system from Intuitive Surgical, Inc.). 3D depth information can then be recovered using the disparity map from rectified stereo images during a laparoscopic surgery [8,42,43], so that a 3D reconstruction using a dense point cloud of the laparoscopic scene can be achieved by a propagation method [44] and/or a cost-volume algorithm [6]. Stereo vision based reconstructions, however, can only recover the structure of a local frame without a global overview of the scene, and are very sensitive to noise and luminance changes. Surgeons have to wear 3D glasses or use a binocular viewer on the robotic surgical system. In addition, stereo endoscopic surgery is still too expensive and yet to be widely used in practice. Hence, providing depth cues in monocular endoscope operations will have a significant impact on the accuracy of surgical procedures.

In this paper, we present a novel method and a computational framework to achieve accurate geometry-aware AR through: (i) extracting 3D depth information from camera motions and 3D surface reconstructions; and (ii) using AR technology to fuse rich 3D structural information with a monocular endoscope video stream, such that accurate spatial information in the scene can be derived from the scene geometry, and artificial depth cues can be provided based on the collaboration of the 3D spatial scene with the real-time video streams (i.e. real-virtual overlay and simultaneous mapping). To this end, we explore the potential of the state-of-the-art SLAM framework by modifying and fine-tuning the algorithm for endoscopic camera tracking and mapping, so that the balance between point cloud density and computational performance can be achieved. A 3D surface reconstruction method based on the Moving Least Squares (MLS) smoothing and the Poisson surface reconstruction algorithms are proposed to recover a smooth surface from the unstructured sparse map points extracted from the MIS scene. Simulated laparoscopic sequences generated in a 3D modelling package have been used to evaluate the performance of the proposed framework in terms of robustness of the camera tracking and the accuracy of the surface mesh reconstruction. Camera trajectories are compared with the ground truth camera trajectories, and the 3D surface reconstructions are measured against the 3D models of the simulated laparoscopic scene. The experimental results yield root mean square errors (RMSE) of 1.24 mm for camera trajectories and 2.54 mm for the surface reconstruction.

The obtained global geometric information can be seamlessly integrated into our proposed AR framework, which is capable of achieving AR augmentations at the correct depth and detailed accurate surface measurements. Our method provides new possibilities for novel geometrically informed AR augmentations in monocular endoscopic MIS, including accurate annotations, labels, tumour measurement and artificial depth cues at correct depth locations that are demonstrated with two example applications: i.e. generations of artificial depth cues and the surface measurements of target sites in MIS.

2. Previous work

Recent advances in computer hardware and software technologies have facilitated the use of computer vision techniques for MIS scene guidance and information augmentation. For example, AR guidance systems have been used to visualize preoperative CT images [18,45], for tumour AR visualization in laparoscopic surgery [4] and anatomy structures AR mapping in liver MIS surgery [14,15]. There are, however, some particular challenges faced with AR in MIS. The luminance changes dramatically and an endoscope can move rapidly during insertion and extrusion. Traditional tracking methods for AR in MIS usually involve fea-