COMPUTER METHODS AND PROGRAMS IN BIOMEDICINE 113 (2014) 862–868

journal homepage: www.intl.elsevierhealth.com/journals/cmpb

Stereoscopic visualization of laparoscope image using depth information from 3D model

Atul Kumar[a,b,*], *Yen-Yu Wang*[a,b], *Ching-Jen Wu*[a,b], *Kai-Che Liu*[a,b], *Hurng-Sheng Wu*[a,b]

[a] *Medical Imaging Research Laboratory, IRCAD, Taiwan*
[b] *Department of General Surgery, Chang Bing Show Chwan Memorial Hospital, Taiwan*

ARTICLE INFO

Article history:
Received 11 June 2013
Received in revised form
20 November 2013
Accepted 18 December 2013

Keywords:
Endoscope
Stereo-endoscope
2D–3D conversion
Laparoscopy
Feature based camera tracking

ABSTRACT

Laparoscopic surgery is indispensable from the current surgical procedures. It uses an endoscope system of camera and light source, and surgical instruments which pass through the small incisions on the abdomen of the patients undergoing laparoscopic surgery. Conventional laparoscope (endoscope) systems produce 2D colored video images which do not provide surgeons an actual depth perception of the scene. In this work, the problem was formulated as synthesizing a stereo image of the monocular (conventional) laparoscope image by incorporating into them the depth information from a 3D CT model. Various algorithms of the computer vision including the algorithms for the feature detection, matching and tracking in the video frames, and for the reconstruction of 3D shape from shading in the 2D laparoscope image were combined for making the system. The current method was applied to the laparoscope video at the rate of up to 5 frames per second to visualize its stereo video. A correlation was investigated between the depth maps calculated with our method with those from the shape from shading algorithm. The correlation coefficients between the depth maps were within the range of 0.70–0.95 ($P < 0.05$). A *t*-test was used for the statistical analysis.

1. Introduction

Laparoscopic surgery is indispensable from the current surgical procedures. It uses an endoscope system of camera and light source, and surgical instruments which pass through the small incisions on the abdomen of the patients undergoing laparoscopic surgery. The small size of the incisions remarkably reduces bleeding due to the incision, post-operative complications, and recovery time of the patient.

Conventional laparoscope (endoscope) systems produce 2D colored video images which do not provide surgeons an actual depth perception of the scene. Although surgeons can use indirect evidences of depth such as motion parallax and monocular cues for having the sense of depth in the 2D video image [1], stereo video images can return the actual depth perception to the surgeons. Various studies have shown the advantages of the stereo video over the conventional 2D video in improving the performance of the surgeons [2–5].

The application of computer vision algorithms have been widely used for the additional information such as texture and shape of the organs, from 2D endoscope images [6–11] useful for management of patients. In theory, there are various such algorithms for calculating the depth information and making a stereo image using a single camera or a two camera system. Although the single camera system has been shown to be

* *Corresponding author at*: IRCAD-Taiwan, Chang Bing Show Chwan Memorial Hospital, No. 6-1, Lugong Road, Lukang Township, Changhua County, 505, Taiwan. Tel.: +886 4 7812988; fax: +886 4 7073222.
E-mail address: sharmaatul11@gmail.com (A. Kumar).
0169-2607/$ – see front matter
http://dx.doi.org/10.1016/j.cmpb.2013.12.013

useful for synthesizing a stereo image, the state-of-the-art for stereo endoscope uses two camera systems. Such system uses the principle similar to that of the natural vision of human eyes. Normal human vision uses two eyes to create left and right images of a scene and these two images are used to synthesize a single stereo image by our brain which gives us the perception of depth in the image. Using a similar concept, the two camera stereo endoscope system synthesizes a stereo image from the images obtained with the cameras enclosed in a single endoscope tube [12]. Such stereo endoscope system has limitations which include fixed distance between two cameras in the endoscope which does not allow it to simulate the convergence of the human eyes [12], the diameter of the endoscope tube is not suitable for wide angle camera system [13] and it needs an installation of a complete new system of the stereo endoscope. These disadvantages hinder the current state-of-the-art to become popular among surgeons who are already using single camera system endoscopes.

In this work, the problem was formulated as synthesizing a stereo image of the monocular (conventional) laparoscope image by incorporating into them the depth information from a 3D CT model. Various algorithms of the computer vision were combined for making a system where a 3D shape from the laparoscope image was reconstructed and registered to the 3D CT model of the abdominal organs to identify correspondence between the image pixels and the points of the 3D CT model. Thereafter, the laparoscope camera position was tracked in the 'real' world (reference frame of the laparoscope) and the 'virtual' world (reference frame of the 3D CT model). The depth information of the scene in the virtual world with respect to the camera location was combined to the laparoscope 2D image, to synthesize the left and the right images of the stereo image. A pre-operative CT scan of the abdomen is indicated in many conditions [14], therefore, the presented method would be very useful in such cases, and surgeons would be helped by the stereoscopic view of the organs during laparoscopic surgery.

The remaining of the paper is arranged as follows: Section 2 gives a brief review of the related works. Section 3 describes the complete method for stereo image generation. Section 4 shows the experiments and results with the current method. Section 5 concludes with a discussion and conclusion.

2. Related works

Various approaches attempted to bring the depth perception to the 2D images including shape from shading (SfS), shape from motion (SfM), selection of images from the video sequences as left and right image of the stereo image and homography-based transforms. Each of these methods has their own advantages and limitations.

A review of the algorithms to generate a 3D shape from the 2D image using the information of shading in the image is presented in [15]. Some of these works also included the calculations of the direction of light before creating the 3D shape from the image. The major limitation of such methods is the computation time, which forbids it to be useful for the real-time rendering of the stereo image. Some of these methods use linearization of the SfS function which have concerns of the depth accuracy [16].

In SfM the shape of an object is reconstructed from the matched feature points on its video image sequence. The method described in [17,18] for the 3D reconstruction of a scene needs every frame of the video to contain all the same feature points which may not be possible during endoscopy. In [19] the algorithm reconstructs shape using the estimated camera motion and considers only the translational movement of the camera, which is again not the case in an endoscope movement. Besides, SfM algorithms give a sparse reconstruction of 3D points which makes it difficult to create a smooth 3D shape.

The algorithm in [20,21] uses an optimization of the motion parallax to select two images as stereo-pair from the video image sequence. Such approaches assume the camera motion to be translational only. Similarly, in [17] the algorithm uses two consecutive rectified images from the video sequence to make a stereo-pair for calculating the depth map of the scene. This approach needs an extra tracking device attached to the endoscope camera which makes the system sensitive to errors due to the tracking device. A technique described in [22] to synthesize a stereo image uses the structure from motion and a virtual camera is assigned to each position of the real camera which creates a virtual stereo-pair image using the homographic transformation. Such technique cannot produce true stereo images and requires the continuous movement of the camera.

A machine learning algorithm in [23] trains the computer with the pixel coordinates and the RGB contents of the pixels in an image with the known depth which is further used in the subsequent frames of the video to calculate the depth map. Such technique has not been validated for endoscope images.

The approaches in [24–26] use the technique for visualizing the combined view of the 2D endoscope image and the 3D model from CT or MRI images. Tomazevic et al. [24] implemented the concept of mutual information for registering 2D endoscope image to the 3D model from CT, MRI and multiple view X-ray images. Burcshka et al. [26] used SfM to reconstruct a 3D structure from video image sequence and then register the structure to the CT model using a principle component analysis (PCA)-based 3D–3D registration method. Our approach is similar to that of Helferty et al. [25] which registers the endoscope image with the CT 3D model rendered image and track the endoscope camera position in the reference frame of the CT model. The method described in our work used the 3D CT model of the abdominal organs for the real depth information and combined it with the single camera laparoscope image to synthesize a stereo image. Our work differs from Helferty's work in the following aspects:

1. For the identification of the correspondence between the video image pixels and the point coordinates of the 3D CT model, we reconstruct a 3D shape of structures in one of the video frame using shape from shading (SfS) algorithm and then register it to the 3D CT model using iterative closest point (ICP) algorithm. In Helferty's work the video to CT registration is done using a normalized mutual information algorithm.
2. In Helferty's work the endoscope camera is tracked using the matched feature points of the consecutive video