Computers in Biology and Medicine 67 (2015) 161–171

Contents lists available at ScienceDirect

Computers in Biology and Medicine

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

Endoscopic instrument tracking for surgical simulation training in a controlled environment via a camera and a planar mirror

Erdal Dayak [a], Ulus Çevik [b,*]

[a] Department of Informatics, University of Gaziantep, Gaziantep, Turkey
[b] Department of Electrical-Electronics Engineering, Cukurova University, Adana, Turkey

ARTICLE INFO

Article history:
Received 30 June 2015
Accepted 17 October 2015

Keywords:
MIS simulator
3D pose estimation
Computer vision
Planar mirror
Epipoint geometry

ABSTRACT

Minimally Invasive Surgery (MIS) has many advantages over traditional procedures and thus training with MIS tools via computer simulations has received much attention. These tools are generally grouped into two major categories: Physical training-boxes, and Computer vision/Virtual Reality (VR) tools. In this study, a computer vision based simulator is proposed which uses a training box that is composed of a single camera and a planar mirror. Occlusions are appropriately handled by the use of the epipoint geometry. The average 3D positional error was 0.96 mm ($\pm$ 0.44 mm) at 1280×960 resolution, and 1.18 mm ($\pm$ 0.52 mm) at 320×240. So, the error is minimally affected as the resolution decreases. The proposed method has some advantages over relevant literature methods, such as an improved accuracy (approximately 60%) even at low resolutions with a low processing time (approximately 30%). Therefore, the proposed method appears as a promising and low cost (approximately 50%) alternative for computer vision based MIS training tools.

1. Introduction

For a few decades, Minimally Invasive Surgery (MIS) has had a great popularity due to its advantages over traditional surgical techniques, such that it reduces hospital stays and causes in less pain after the operation [1,2]. However, such closed operations have also their difficulties when compared to open surgeries. Among such complexities, the leading one is that the surgeons do not have a chance to observe the operation by bare eyes instead they have to accomplish the operation via a camera. This type of operation requires surgeons to be specially trained to adapt their hand-eye coordination skills for the camera assisted closed surgery. Therefore, tools to train surgeons for MIS have been emerged as a necessity.

There are many commercial MIS simulation and training tools in the market. These tools are generally grouped into two major categories: Physical training-boxes, and Computer vision/Virtual Reality (VR) tools [3]. Sometimes, Augmented Reality (AR) techniques are also considered as the third category [4]. Physical training facilities are mainly utilized to improve surgeons' hand skills and to quantitatively measure their operational performance indices, such as accuracy and speed. In the literature, there are many studies on measuring the effectiveness of the trainees [5–7]. However, physical box trainers have some limitations such that they require maintenance since the materials in the box are irreversibly damaged while the trainee performs operations, such as cutting and clipping [8]. As an alternative to physical box trainers, VR simulators offer a cheaper option that does not require such maintenance costs. However, actual commercial prices of these products may differ depending on the marketing strategies. Literature studies, such as [5], proved that VR simulators may shorten the learning stage of laparoscopic operations, specifically for porcine cholecystectomy. Despite being cheap and effective, VR simulators have also drawbacks. They are unable to provide an accurately realistic operation, and a lack of interaction with the trainee (e.g. haptic feedback). VR simulators generally adopt two approaches: sensor-based, and computer vision-based. In the former one, the estimation of a position, movement, rotation and the state of surgery tools are done via some sensors, like optic and magnetic ones. In the latter one, all estimated parameters are obtained via computer vision/image processing techniques.

In computer vision based VR MIS simulations, the most important task is to accomplish 3D pose estimation of laparoscopic/endoscopic instruments (LEIs) accurately in a reasonable amount of time. This task usually includes some sub-tasks such as selecting the LEI in the taken still image (e.g. a single frame taken from a streaming camera), calculating its 2D position, and converting it to a 3D position. 3D pose estimation can be done either via a single camera (monocular image processing) or multiple cameras such as stereoscopic image processing with two cameras.

* Corresponding author. Tel.: +90 5326256979.
E-mail address: ucevik@cu.edu.tr (U. Çevik).

http://dx.doi.org/10.1016/j.compbiomed.2015.10.012
0010-4825/© 2015 Elsevier Ltd. All rights reserved.

In the single camera approach, perspective projection basics and geometric features of the objects are used to perform 3D pose estimation. However, in the stereoscopic method, two camera images are merged in order to estimate the depth of the object. In our study, we also use a single camera, but, we obtain a stereo image using a mirror image. Our method is, thus, a hybrid approach of above mentioned methods. It is much cheaper than the methods given in previous studies since it does not require multiple cameras, and special sensor based LEIs.

In the literature, there are several studies that have the same goals of 3D pose estimation from streaming camera images. For instance, Allen et al. [9] proposed a method that uses a single camera. In their method, LEI detection is done via a thresholding process, and the pose estimation is based on the vanishing point on the shaft edges. This method fails to track the neck of the LEI when it is too close to the camera. In another similar study, Loukas et al. [4] also suggested a method that uses a single camera. In the study, they primarily attempt to detect the neck part of the LEI, and the pose estimation is based on the fundamental perspective projection geometry. The LEI is detected via a marker of different color which is put on the neck of the LEI. Main problem of the method is that it fails in tracking when the neck part of the LEI vanishes due to some reasons such as LEI occlusions, penetration of the tip into a tissue, and the presence of similar colors, with the marker, in the scene. Moreover, since the perspective projection is the only way to perform 3D pose estimation, the performance diminishes when low resolution images are used. The study of Shin et al. [10] utilizes a single camera too, however, apart from previous studies, they put a marker made up of at least 3 different colors. Using the Haralick algorithm, they perform 3D pose estimation in addition to the calculations of roll and pitch angles of the LEI. However, the method was not able to handle occlusion situations [10]. Moreover, the error increases as the LEI gets farther from the center of the view of the camera. In order to reduce the error, based on the mentioned problems, they use high-resolution images (e.g. 1920×1080 at 60 fps) which slow down the detection process. In the study of Pérez et al. [11], two orthogonal cameras are used to a overcome the mentioned problems while 3D pose estimation. They report that the LEI occlusion problem was resolved in their method. However, the error of the method reaches to the maximum level when one of the LEI gets too close to one of the cameras. Moreover, it should be noted that the method is utilized for motion tracking purposes only.

So, according to the relevant literature, there are some major problems in the studies that utilize a single camera view for MIS simulation purposes. Such as, because there is only one single view of the LEIs, a great error is observed while estimating the 3D pose of the LEIs. Thus, in order to reduce the errors, such studies are generally compelled to work with only high resolution streaming media which slows down the image processing task. As a result, single camera studies usually require computer systems with a high power processing unit (CPU), or a graphical processing unit (GPU) in order to complete the task in real time which increases the expected cost. In the proposed method, a single camera view is supported with a planar mirror view. Thus, two views from different angles are obtained. This approach comes with several advantages. Primarily, this kind of stereo view is analyzed using epipoint and epiline geometric features which help recover lost information (e.g. depth) as a result of the 2D projection of the 3D LEI. This way, occlusion problems (e.g. overlapping LEI) can be resolved. Furthermore, in contrast with relevant studies that utilize single camera views, the proposed method can also work with low resolution images as well, with lower errors. When the proposed method compared with the studies that use two cameras, it has also some advantages over them. Two camera based studies take two images as an input and process them together that requires much more processing than our method, since our method processes only one image at a time. Therefore, this study is superior in terms of processing burden and speed. Moreover, it costs less compared to the two camera solutions.

2. Materials and methods

2.1. The flowchart of the study

The flowchart of the work is given in Fig. 1. As can be seen, the flowchart is divided in groups, and each group is given a letter symbol. They show the name of the section in which the processes within this group are explained in detail. Group A is explained in Section 3, group B in Section 4, and group C in Section 4.1.

In this work, the video that consists of the real and the mirror images of LEIs is processed frame by frame in the order given in the flowchart. Here, each frame is an $X \times Y$ matrix, and each element in this matrix is a 3D vector containing the RGB values of each pixel within the video frame. Where, X is the width, and Y is the height of the frame in terms of pixels. In processing, using the real and mirror images of the LEIs, in a frame, 2D coordinates of the starting, and ending points of the neck region marker of each LEI are obtained. Then, these coordinates are converted to 3D.

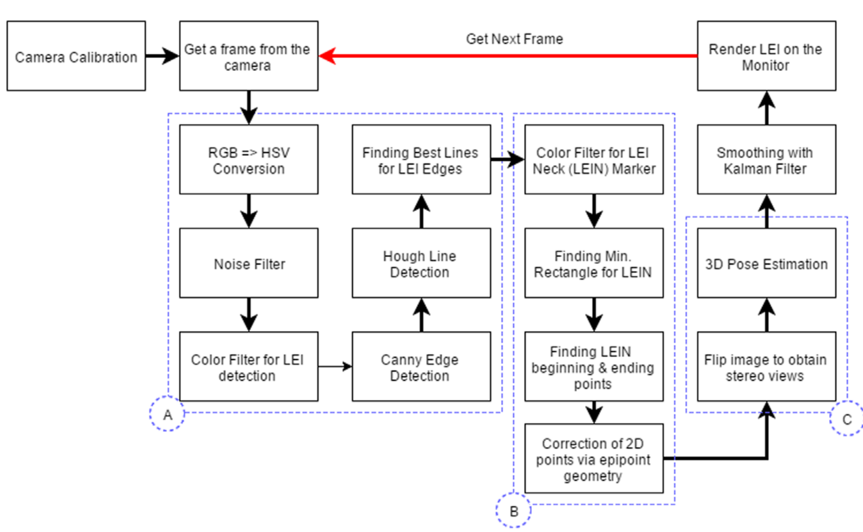

Fig. 1. Flowchart of the study.