2014 CMBEC37 Conference
Vancouver, BC
May 21 – 23, 2014

TOWARD FEATURE-BASED ENDOSCOPIC REGION ALIGNMENT WITH AUGMENTED REALITY

Jiawei Zhou and Shahram Payandeh
School of Engineering Science, Simon Fraser University

INTRODUCTION

The application of computer technology in Minimally Invasive Surgery (MIS) has provided numerous advantages for both patient and surgeon. A particular focus, namely designing an interactive surgeon computer interface (SCI) is crucial to image-guided surgery which can avoid inefficient communication between surgeons and their assistants. Moreover, a suitable SCI is capable of providing visual enhancement to facilitate surgeons with patient data management and surgical strategy planning by being combined with augmented reality (AR) technique.

In general, both open surgery and MIS type surgeries are performed in various operating theaters. In open surgery, surgeons are free to use their hands to contact medical data during surgery. However, this is not the case in MIS since operation is accomplished through usage of long-stem surgical tools. In [1], a non-contact mouse system is designed for open surgery to replace standard computer mouse functions with hand gestures. Surgeons can put their hands inside a pre-defined workspace to interact with computer using various hand gestures. To assist the image-guided vascular surgery, [2] develops a system making use of Microsoft Kinect sensor to realize the touchless interaction between surgeons and computer. [3] proposes a hybrid user interface for MIS which is able to provide new optimal solutions for surgical tasks by combining orthogonal slice views into in-situ visualization. They then later develop an interactive virtual mirror providing desired views of the 3D object from any viewpoint[4]. A neural network framework is presented in [5] to identify and classify gestures of surgical tools and their motions in MIS.

Compared to traditional interfaces such as keyboard and mouse, the above mentioned SCI methods have shown advantages in addressing aspect related to image-guided surgery. However, most of the above approaches require additional hardware such as depth sensors and multiple cameras. In [6], we develop a non-robotics interactive system which can be used to enhance the practice of MIS without adding additional hardware to the surgical theater. The proposed SCI allows surgeons to browse patients' pre-operative medical images by touching virtual menus which are overlaid on the real endoscopic scene (Figure 1). To provide intra-operative guidance, our system can also help surgeon manually register patient's overlaid pre-operative images on the surgical scene. However, during the motion of endoscope, the registered images are prone to be lost and they need to be manually realigned again by surgeon. Therefore, a robust region tracking algorithm is necessary to tackle the endoscope motion problem during MIS [7]–[9].

In this paper, we explore the feature-based endoscopic region alignment based on our proposed non-robotics system. First we present an overview of our interactive SCI and then we provide some preliminary study for Augmented Reality (AR) in in-vitro environments. The basic idea of in-vivo implementation is stated including some related aspects of the theory of Scale Invariant Feature Transform (SIFT) and its application related to the surgical scene.

OVERVIEW OF INTERACTIVE SCI

In [6] and [10], a non-robotics system is developed on Visual Studio 6.0 platform working for monocular endoscope. The system consists of the surgical tool tracking algorithm and the interactive SCI. OpenCV is applied for video frame processing and tracking calculation. For interactive SCI, OpenGL is used to overlay

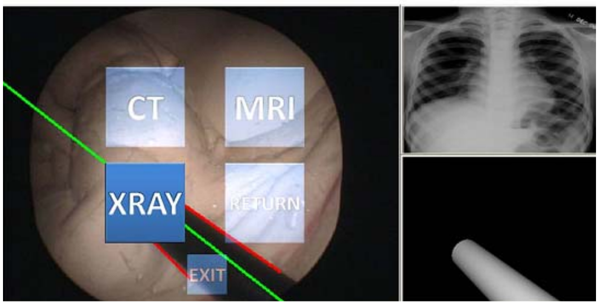

Figure 1: The interactive SCI presented in our previous work. It shows the case when 'XRAY' option is selected, the Xray image is shown in the right.

virtual graphics on the real endoscopic scene and to render 3D reconstructed surgical tool model during tracking. Once the surgical tool moves in the camera view, its features such as tool's tip and two edges are detected by the image processing module. The location of the tool is then computed from these features and is sent to the tracking module.

One example for the interactive events is to allow the surgeons to browse patient's pre-operative images without moving their eyes from endoscopic scene on monitor (Figure 1, 2a). Besides activation for image browsing, the surgeon can also manipulate a virtual object using the surgical tool to match a pre-defined target. Figure 2b illustrates the manually registration procedure. The white organ mesh is the pre-defined target. The orange virtual liver is 'sticking' on the surgical tool. The task for the user is to use the surgical tool as a "joystick" in order to displace the virtual liver to match with the target.

PRELIMINARY IN-VITRO AR STUDY

Before we study the application of AR to the actual surgical tasks, a preliminary in-vitro study is conducted. Here, a simple marker-based AR system is developed on a mobile platform (namely, Blackberry Playbook platform). The reason for selecting a mobile device is that its camera movements can easily be configured to replicate the expected movements of the endoscope used in real MIS situations. Generally, a basic marker-based AR system contains the following components: (1) image acquisition; (2) marker detection; (3) virtual-real object alignment; (4) augmented

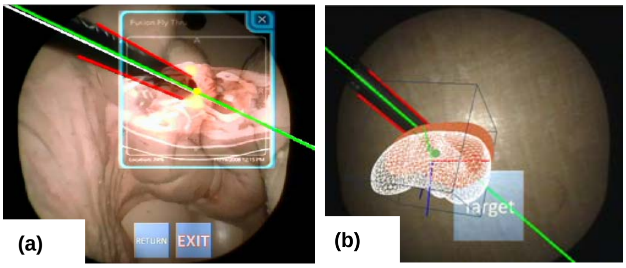

Figure 2: Snapshots of interactive events through our SCI. (a) Pre-operative image browsing example. When the tool moves, the image is always shown on the tool's tip position. (b) Virtual object-target registration example. The user can manually register the orange liver to the white target by using the tool.

image display. The marker is regarded as the region of interest (ROI) in camera view which can be used to navigate the virtual object to its position in real world image.

The marker detection method in our study is composed of image pre-processing, image segmentation and corner detection. In pre-processing step, a Gaussian smooth filter is applied on the whole image to eliminate the noise from input. The binarization of the blurred image is to highlight the marker's structure and the border between marker and background. Edges of the marker are identified using Canny detector and its corners are extracted by Harris corner detection algorithm[11]. Geometrical approaches are applied to extract other features of the marker including its outer border and inner structure. Figure 3a1 and 3a2 show the detected results in different camera view. The corners are indicated as crosses where blue is for outer corners, yellow is for inner corners and green is for the marker's center. The border of the marker is labeled in thick green line. Specifically, blue line and red line represent the X-axis and Y-axis of marker frame respectively. To study how to integrate 3D virtual object with real environment, we attempt to overlay a virtual cube on the marker. The center of the cube is defined to coincide with that of the marker and they share the same orientation. Figure 3b1 and 3b2 display the virtual cube under different camera views.

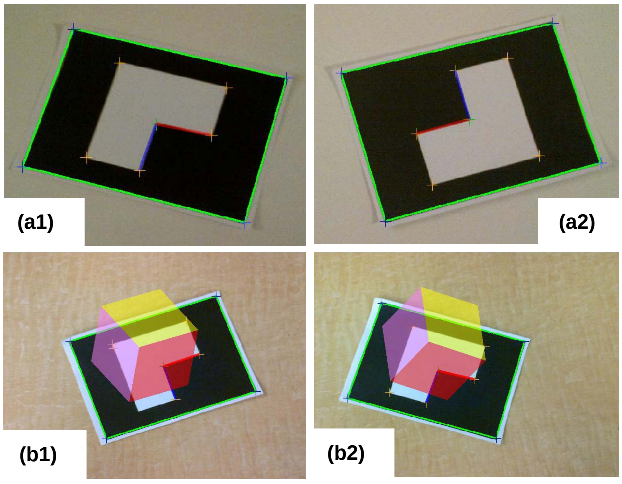

Figure 3: Example of marker-based AR system results. (a) Marker's border and orientation are detected under different camera views. (b) Overlay a virtual cube on real world scene according to the detected marker.

IN-VIVO IMPLEMENTATION BASED ON SIFT

To explore the idea of AR application to real MIS scene, the problem we have to face is the connection between virtual object and actual surgical scenes. In our preliminary in-vitro study, the binary marker is used as the landmark for locating the virtual object. However, this method is not possible in MIS because it's impossible to put such a marker inside patient's abdominal cavity. An alternative method is to consider some specific regions in human's body as the 'marker'. As a result, we can locate the virtual object such as patients' image data or the virtual tissue/organ model on this 'marker'. When the endoscope view changes, the 'marker' is able to be detected between the image frames so that the overlaid virtual object can be tracked between video streams. One challenge in region alignment between frames is how to match the same region across two images where the scale and orientation of the region are changed. To solve region alignment problem, we apply one of the popular image matching algorithm SIFT[12] in this paper. The keypoints, which are the features of region invariant to image scale, translation and rotation, are extracted and given distinctive descriptors in different frames. The keypoint descriptor contains not only the position but also the orientation of the keypoint and can be used to identify each keypoint from another. By matching the keypoint descriptors between frames, it is possible to track the same region with these keypoints.

The SIFT algorithm is composed of building Gaussian scale-space, detecting and locating keypoints, assigning reference orientation to keypoints and producing keypoint descriptor. To obtain scale invariance, a multi-level scale space is created by down-sampling. For each octave (i.e. each level), the image is smoothed by a stack of Gaussian kernels with different variances so that there are a set of blurred images in each octave. To detect the local extrema in scale-space, the difference-of-Gaussian(DoG) images are produced by subtracting adjacent Gaussian images in each octave. Maxima and minima of DoG images are detected by comparing the current pixel with its 26 surrounding neighbors. These neighbors are from the current scale and adjacent scales as well. To make the keypoints rotationally invariant, gradient computation of a keypoint neighborhood is implemented. For a given image I, the magnitude of pixel (x,y) is

$$M_{x,y} = \sqrt{\left(I_{x+1,y} - I_{x-1,y}\right)^2 + \left(I_{x,y+1} - I_{x,y-1}\right)^2} \qquad (1)$$

where $I_{x,y}$ is the intensity value of pixel (x,y). The direction at pixel (x,y) is computed as

$$\theta_{x,y} = \tan^{-1}((I_{x,y+1} - I_{x,y-1}) / (I_{x+1,y} - I_{x-1,y})) \qquad (2)$$

For each sample point around a specific keypoint, its orientation is weighted by its magnitude. An orientation histogram is used to classify these computed orientations into different directions. The highest peak in the histogram is selected as the dominant direction for the specific keypoint. To uniquely identify one keypoint to another, a keypoint descriptor is assigned to each keypoint. In [12], a 128-element feature vector is designed for each keypoint. This keypoint descriptor comes from a 16x16 sample array around each keypoint. It contains the gradient magnitude and orientation distribution in its keypoint neighborhood. To find the best matching keypoint between frames, we measure the

Euclidean distance for two 128-element descriptors. The keypoint with minimal distance is considered as the best matching one.

To test SIFT algorithm [13] in in-vivo application, we select 10 352x240 sample video streams containing real surgical scenes. Due to the high computation cost, we extract 20 frames from each video stream for the offline experiment. For each image pair, it takes about 5~6s to run the matching. Figure 4 is the example of region matching in various frames. The green circle with the green line segment represents the keypoint descriptor. The size of the circle indicates the magnitude of the keypoint while the direction of the circle is the keypoint's orientation. The matching keypoints are connected by red line.

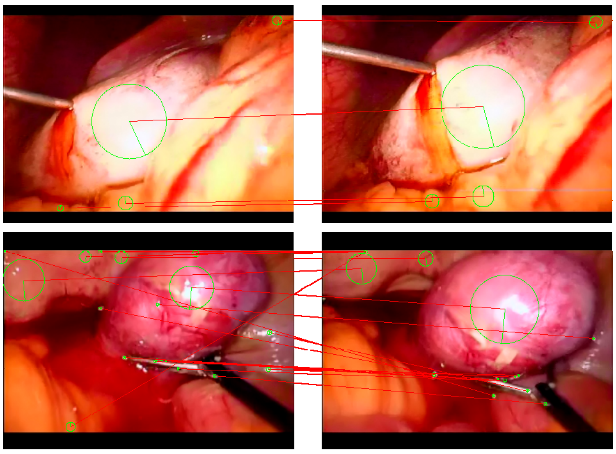

Figure 4: SIFT application in real surgical scene. Two examples are given in two rows respectively. SIFT can mostly match the moving regions in two different frames.

DISCUSSION AND CONCLUSION

In this paper, we present a brief overview of our interactive SCI which can realize some simple manipulation of virtual object using the surgical tool. A basic marker-based AR system is presented in our preliminary study to overlay a virtual cube on the detected marker under different camera views. To explore the AR technique into real surgical scene, a feature-based matching method SIFT is applied to align regions under various endoscope viewpoint. From the experimental results, SIFT method can basically match regions in real surgical scenes. However, this method is time-consuming which is not suitable for the real-time implementation. In the future work, an affine invariant extension of SIFT(ASIFT) will be tested in in-vivo environment. The ASIFT is able to enhance its performance under substantial viewpoint change and reduce the complexity of the original SIFT. Other algorithms such as Simultaneous Localization and Mapping(SLAM) and Random Sample Consensus(RANSAC) can also be tested on the real surgical scene and combined with AR display.

REFERENCES

[1] T. Fong, S. Grange, and C. Baur, "A Non-Contact Mouse for Surgeon-Computer Interaction," *Technol. Heal. Care*, vol. 12, no. 21, pp. 245–257, 2004.

[2] K. O'Hara et al., "Touchless interaction in surgery," Commun. ACM, vol. 57, no. 1, pp. 70–77, Jan. 2014.

[3] J. Traub et al., "Hybrid navigation interface for orthopedic and trauma surgery.," Med. Image Comput. Comput. Assist. Interv., vol. 9, no. Pt 1, pp. 373–80, Jan. 2006.

[4] B. C. Navab N, Feuerstein M, "Laparoscopic virtual mirror new interaction paradigm for monitor based augmented reality," in *Virtual Reality Conference, VR'07. IEEE*, 2007, pp. 43–50.

[5] S. Payandeh, J. Hsu, P. Doris,"Toward the design of a novel surgeon–computer interface using image processing of surgical tools in minimally invasive surgery," Int. J. Med. Eng. Inform., vol. 4, pp. 1–24, 2012.

[6] X. Sun and S. Payandeh, "Toward development of 3D surgical mouse paradigm," *2011 IEEE/RSJ Int. Conf. Intell. Robot. Syst.*, pp. 2096–2101, Sep. 2011.

[7] P. Mountney, D. Stoyanov, and G. Yang, "Three-dimensional tissue deformation recovery and tracking," *Signal Process. Mag. IEEE*, vol. 27, no. 4, pp. 14–24. 2010

[8] S. Giannarou and G. Yang, "Content-based surgical workflow representation using probabilistic motion modeling," *Med. Imaging Augment. Real. Springer Berlin Heidelb.*, pp. 314–323, 2010.

[9] S. Giannarou, M. Visentini-Scarzanella, and G.-Z. Yang, "Probabilistic tracking of affine-invariant anisotropic regions.," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 130–43, Jan. 2013.

[10] X. Sun, "Image guided interaction in minimally invasive surgery," Master The, 2012, Simon Fraser University.

[11] C. Harris and M. Stephens, "A combined corner and edge detector.," *Alvey Vis. Conf.*, 1988.

[12] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, Nov. 2004.

[13] "SIFT demo online." [Online]. Available: http://www.ipol.im/.