

# Visual Tracking of Laparoscopic Instruments

Jiawei Zhou and Shahram Payandeh

The Experimental Robotics and Imaging Laboratory, Simon Fraser University, Canada

Email: jza102@sfu.ca; shahram@cs.sfu.ca

**Abstract**—In order to enable the standard set-up used in the training and the practice of minimally invasive surgery procedures to be more immersive, it is necessary to design and develop an interactive surgeon computer interface. In this paradigm, visual tracking of the surgical instrument takes the center stage in design of such an interface. In this paper, we present a novel 3D visual tracking system using a single endoscopic camera. 2D feature recognition and stochastic approach are applied to assist the tracking system and the 3D reconstructed model of instrument is also provided. Two tracking algorithms based on standard Kalman filter and extended Kalman filter are proposed and compared in the article. The performance of the proposed tool tracking algorithms are evaluated and discussed using both emulated and video footage of the actual surgical environment. The results show our methods can track the objective well in “Good” quality scenes and those scenes with strong contrast between tool and background.

**Index Terms**—minimally invasive surgery, Image tracking, Kalman filter.

## I. INTRODUCTION

Owing to the technological progress in robotics, computer vision, optics and biomaterials, the appearance of a modern surgical technique, minimally invasive surgery (MIS), has brought new methods and more possibilities to medical therapy. In comparison with conventional open surgery, MIS is performed through tiny incisions using endoscope and special instruments and has more benefits like less transfusions, shorter hospitalization time as well as more rapid recovery. However, surgeons also have to face new challenges brought by MIS such as limited field of view, hand-eye coordination and loss of depth information.

To overcome some of these challenges, augmented reality technology which merges the virtual objects into real world is developed to assist medical therapy. In computer-aided surgery, surgeon computer interface (SCI) is the crucial part which can provide 3D visual enhancement to facilitate surgeons with their image management and surgical planning. A real-time hand gesture system called Gestix for medical image manipulation is developed collaboratively by Ben-Gurion University of the Negev and Washington Hospital Center [1]. UNC Chapel Hill offers an interactive, 3D stereo needle insertion guidance system to help surgeons with

intra-operative needle insertion during tumor resection [2]. Another example is the da Vinci Surgical System made by Intuitive Surgical, Mountain View, CA, USA. Stereo camera is utilized in its 3D HD Vision system which provides a virtual extension of the surgeon’s hands and eyes into the patient’s body. Although stereo photography shows its advantage in 3D image capture and object reconstruction, it’s not quite popular in clinical application because of its high cost.

In any design of SCI, visual tracking is critical to interactive operation since the manipulation of superimposed virtual objects in the real scene is based on the tracked information. In general, the tracking approaches can be categorized into bottom-up and top-down procedures. In bottom-up procedure, the target’s position is located by extracting feature of the target, such as edges, corners and colors, from image sequences. This kind of tracking methods include color-based tracking [3], contour-based tracking and motion history image (MHI) method [4]. The drawback of these methods is its inability to deal with tracking in a clustered and complex environment. Another group of tracking is top-down process. They not only make use of targets’ salient features but also involve prior information about the object. A probabilistic Condensation algorithm is presented by Remi W. *et al* in [5] to track the surgical instrument during a laparoscopic surgery. Christophe D. *et al* address a real-time segmentation method to detect surgical instruments inside the abdominal cavity based on joint hue saturation color feature [6], [7]. In [8], Sandrine V. *et al* propose a method to detect the instrument using information on the 3D position of the insertion point of an instrument into the abdominal cavity. This method is based on the shape of the instruments rather than their color. Cano *et al* design a projective model to achieve 3D tracking of laparoscopic tools using 2D information of the image [9].

In our previous work [10], four possible methods including Camera-based tracking, pose estimation, MHI and Particle filter are studied for tracking instruments with a single endoscopic camera in an in-vitro setting. In this paper, we present a novel real-time 3D visual tracking system. The system utilizes 2D feature recognition as well as stochastic approach to track a single surgical instrument under a single endoscopic camera and reconstruct the 3D pose of the instrument using projective geometry [9]. Two tracking algorithms based on standard and extended Kalman are proposed. Given the complexity of the real surgical scenes and the adequate scene magnification

which is offered through the endoscopic lenses, our work is focused on scenes where the tool has adequate surgical motion space.

The remaining of the paper is organized as follows. Section II shows some preliminary studies including 2D feature recognition and 3D pose of the tool reconstruction before we implement tracking algorithm. Section III introduces the two surgical tool tracking methods based on Kalman filter. Experimental studies of tracking results and global monitoring part are presented in Section IV and V under both emulated and real surgical environment. Conclusion and future work is proposed in the end.

## II. PRELIMINARY STUDIES

Before we present the proposed visual tool tracking method, some preliminary studies are introduced first in this section. This part of work includes how we reconstruct 3D pose of a single surgical tool making use of its 2D features extracted from the projective plane.

### A. 2D Feature Recognition

In order to know the 3D location of the surgical tool with respect to endoscopic camera, we need to find where the tool is in 2D projective plane (image plane). Considering the geometrical structure of general surgical tool, we assume the tool is uniform cylindrical shape with known physical diameter. In 2D projective plane, the tool can be described as two edges, a midline and a tip point. Once we detected two edges of the tool in image plane, the midline of the tool then can be calculated from these edges and the tip of the tool can also be found along the midline.

To calculate tool's midline, some preprocessing steps are needed in order to extract edges of the tool. Generally, the background in real endoscopic surgery contains images of the organs and tissues. Considering this unique feature of the surgical site, we split the original images into three single images where each of them contains information only from one color channel (red, green or blue). Then we just use the red channel image so that the details in background can be blurred. From our experimental study, a binary filter is added on red channel image to locate the tool's position since, in general, the color of the surgical tools are different from the background images (e.g. dark black or gray color material). Canny filter is then applied on the binary image and is followed by Hough transform to detect the edges of the tool. After edge extraction, the location of the midline of the tool is easy to compute using standard description of a line located in x-y plane:  $y=mx+b$ . To locate the tool's tip, we conduct a 1D edge detection along the segmented mid-line of the tool. Since the tool is uniformly black while the background is brighter, the tip is localized at the sudden change of the intensity. To avoid the reflection spot and side effect, some noise reduction filters are implemented to reduce such effect. To search for the tip location, both x and y directions in image plane are considered. Generally, we start searching along x direction (left to right). However, when the tool is almost vertical in the scene, the search is then along y direction. To make the

algorithm works better, we split the two situations depending on the slope of tool's midline.

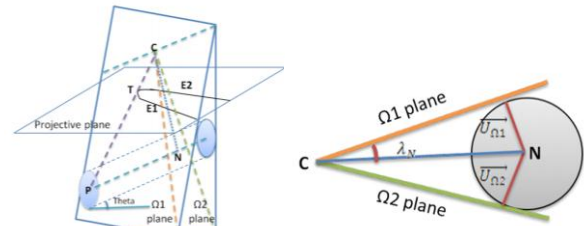


Figure 1. 3D projective model (left) and its transverse section (right).

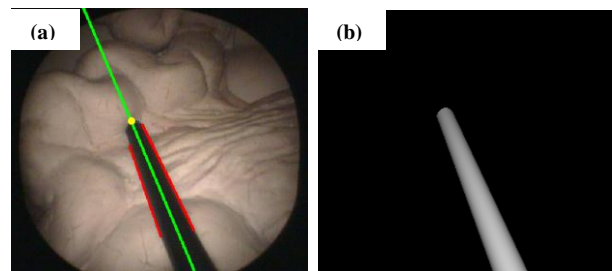


Figure 2. (a) 2D feature recognition and (b) 3D reconstructed tool rendered in OpenGL.

### B. Spatial Description of the Surgical Tool-tip

Given the physical diameter of the tool and its segmented edges, we can obtain depth information and relative position between the tool and the endoscopic camera based on 3D projective geometry. The projective model is shown in Fig. 1. In this model, C is the camera center and T is the tip of the tool in 2D image with its 3D coordinate P. E1 and E2 are two edges of the surgical tool projected on the image plane. N is the perpendicular foot from camera to tool.  $\theta$  is the angle between the physical tool with respect to the image plane. Two tangential planes  $\Omega_1$  and  $\Omega_2$  including one of the tool's edges and the camera vector to the edge respectively are defined. First we need to calculate the normal vector to each tangential plane through cross product between the edge and the camera vector to the edge. From the normal vectors we obtained, the angle  $\lambda_N$  can be calculated. By utilizing the known vectors and physical diameter of tool, we are able to extract both the 3D location of the tip point and the orientation of the tool. For detailed derivation, please refer to [11]. Fig. 2 displays 2D detected features and 3D reconstructed tool.

### C. Surgical Tool Tracking Problem

In real surgical operation, the surgeon may want to know the 3D location of the surgical tool to help him measure the distance between tissue and tumor. Following previous methodology, a simple approach to track tool's motion is to extract edges and calculate tip frame by frame. However, this approach is not efficient and is sensitive to the effect of random noise in the image (e.g. variation in illuminant). According to the color theory [12], the color space is changing under different illuminant. In the actual scene from the surgical procedure, the color in the image may

have different RGB proportion which can lead to artifact in canny edge detection. Other possible noise is from reflection. Metal used in the design of some of the surgical tools and light reflection from the tissues and organs can create some “light reflection area” which also results in unwanted edges. Fig. 3a is an example of miss-detection under noise. When encountering with a number of selected edge segments, it becomes very challenging to identify the correct ones belonging to tool.

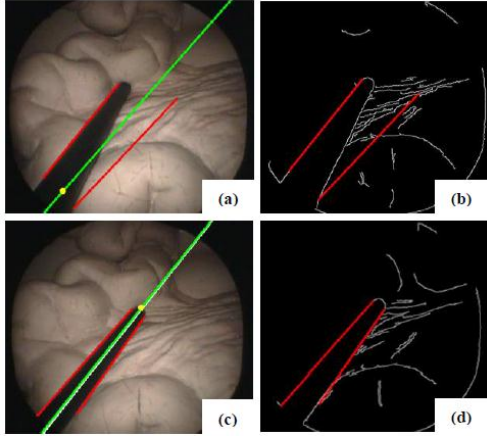


Figure 3. Comparison of 2D feature recognition with and without Kalman filter. (a) is the detection result without Kalman filter. The red lines in (b) indicate the edges extracted as tool's edges. (c) is the detection result using Kalman filter and (d) is its labeled canny edge image.

To deal with the random effect of noise in tracking problem, a stochastic framework needs to be developed to estimate the location of the tool based on its edge description. In general, we can assume that the tool's motion is within the limits of the required surgical task. Moreover, the noise is assumed to be independent, white and with normal probability distribution. The noise covariance is also assumed to be constant. In the following section, a Kalman filter, owing to its simplicity and reduced computational complexities, is used to track the location of midline of the tool. Fig. 3c shows an example of the tracking result using proposed in the following Kalman filter.

### III. KALMAN FILTER TRACKING METHODS

In this section, we present two feature-based methods to track single surgical instrument during MIS. These methods utilize different mathematical framework namely standard Kalman filter and extended Kalman filter respectively.

#### A. Kalman Filter Method

In general, Kalman filter can be defined by a set of relationships providing an efficient recursively computational method to estimate the state of process [13]. As we mentioned in previous section, the midline of the tool is chosen as the desired state which needs to be estimated. This line can be represented as a linear equation  $y = mx + b$  where  $m$  is the slope and  $b$  is the intercept in the image plane. The state vector at the current time  $k$  is

defined as  $X_k = [m, b, m', b']^T$  where  $(m', b')$  denotes the first derivative of the slope-intercept parameters. The tracking system is composed of time update (prediction) and measurement update (correction). For each frame, the location of tool's midline is predicted from previous frames and the system model we obtained is then modified according to the selected measurements.

In update phase, both the priori estimated state vector  $\hat{X}_k^-$  and priori estimated error covariance  $P_k^-$  are provided from the previous time  $k-1$ . The prediction equation is defined as

$$\hat{X}_k^- = A\hat{X}_{k-1} + W_{k-1} \Rightarrow \begin{bmatrix} m_k \\ b_k \\ m'_k \\ b'_k \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} m_{k-1} \\ b_{k-1} \\ m'_{k-1} \\ b'_{k-1} \end{bmatrix} + W_{k-1} \quad (1)$$

$$P_k^- = AP_{k-1}A^T + Q$$

$$p(W_{k-1}) \sim N(0, Q)$$

where  $W_{k-1}$  is process noise with the noise covariance  $Q$  and  $A$  is transition matrix.

Before we correct the priori value  $\hat{X}_k^-$  and  $P_k^-$ , we need to actually measure the process through image processing of the surgical scene to obtain  $Z_k$ , which is known as the measurement, to generate an improved posterior state estimate. Corresponding to the state vector we defined, here, the real location of tool's midline is used as the measurement  $Z_k = [m, b]^T$  which can be expressed as

$$Z_k = HX_k + V_k \Rightarrow \begin{bmatrix} m_{meas k} \\ b_{meas k} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} m_k \\ b_k \\ m'_k \\ b'_k \end{bmatrix} + V_k \quad (2)$$

$$p(V_k) \sim N(0, R)$$

where  $V_k$  is measurement noise with the noise covariance  $R$  and  $H$  is measurement matrix. During the 2D feature recognition, the image after Canny filter contains both useful and noisy information. To filter out the edges that unlikely belonging to the tool, the predicted midline is utilized. If the difference in the slopes between the possible edge and the predicated midline is smaller than a threshold and its edge point is close enough to the midline, we then assume this edge belongs to the tool. Once the edges are filtered, we classify the remaining lines to be either left or right edge and convert them into the slope-intercept form. After finding out the measurement  $Z_k$ , the state model is corrected according to correction equations

$$K_k = P_k^- H^T (HP_k^- H^T + R)^{-1}$$

$$\hat{X}_k = \hat{X}_k^- + K_k (Z_k - H\hat{X}_k^-)$$

$$P_k = (I - K_k H) P_k^- \quad (3)$$

where  $K$  refers to gain matrix. In the end of each recursion, the corrected posterior value  $\hat{X}_k$  and  $P_k$  turn to be the input of next time  $k+1$ . Fig. 4 shows the Kalman tracking result with some preprocessing steps. In Fig. 4d, the white

line is the corrected midline of the surgical tool and the green line is the predicted one.

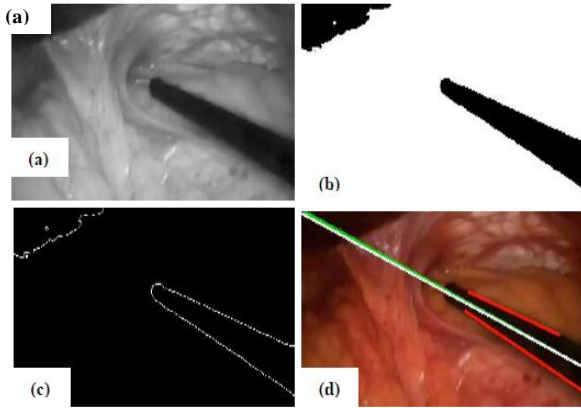


Figure 4. Different steps in the tracking algorithm. (a) Red channel image, (b) binary image, (c) canny image and (d) tracking result.

The initialization of Kalman filter requires information from the two edges of the first frame. The initial midline is calculated based on these two edges detected. Once the tracking is lost, the system then re-initializes the Kalman filter from the frame after the lost one.

#### B. Extended Kalman Filter Method

Though the standard Kalman filter tracking algorithm performs well in most of the cases, it still has some limitations. For example, when tool moves from horizontal direction to vertical direction, both slope  $m$  and intercept  $b$  of tool's midline increase rapidly and nonlinearly since the both of them tends to be infinite in vertical orientation. This position change is easy to introduce large error in tracking system. One possible solution is to describe the lines in their polar parameters  $\rho$  and  $\theta$  with respect to the image frame which is presented in Hough transform [14]. Each line passing through the given point  $A(x, y)$  can be expressed as

$$\rho = x \cos(\theta) + y \sin(\theta) \quad (4)$$

where  $\rho$  is the perpendicular distance of the line to the origin and  $\theta$  is the angle between the line and the reference axis. Since the representation of line is nonlinear, an extended Kalman filter is required in the new tracking algorithm. This kind of nonlinear Kalman filter model is widely used in various fields such as unmanned aerial vehicles navigation [15] and lane marking detection [16, 17]. In [18], Steven *et al.* has presented a combined tracking method using Kalman filter and Hough transform to detect the edges of moving objects. In this paper, similar method is used to track two edges of moving tool.

Similar to standard Kalman filter, extended Kalman filter also consists of prediction and correction parts. Here, the state vector is defined as  $X_k = [x, y, u, v, w, \theta_1, \rho_1, \theta_2, \rho_2]^T$  where  $(x, y)$  is the rotation center of tool's motion in the image plane. We divide the tool's movement into translation and rotation where  $(u, v)$  are defined as the translational velocity along  $x$  and  $y$  axis respectively and  $w$  is the rotational velocity.

Moreover,  $(\rho_i, \theta_i)$  ( $i=1,2$ ) are referred to as the line parameters of two edges of the tool. To simplify the computation, we assume both translation and rotation have constant speed. For current time  $k$ , the rotation center has been translated from  $(x_{k-1}, y_{k-1})$  at speed  $(u_{k-1}, v_{k-1})$ . Then the state vector at time  $k$  has the form

$$\hat{X}_k^- = [x_k, y_k, u_k, v_k, w_k, \theta_{1k}, \rho_{1k}, \theta_{2k}, \rho_{2k}]^T = f(\hat{X}_{k-1}) \quad (5)$$

in which  $u_k, v_k$  and  $w_k$  are constant and

$$\begin{aligned} x_k &= x_{k-1} + u_{k-1}, y_k = y_{k-1} + v_{k-1}, \\ \theta_{i(k)} &= \theta_{i(k-1)} + w_{k-1}, \\ \rho_{ik} &= r_i(k-1) = \rho_{i(k-1)} - x_{k-1} \cos(\theta_{i(k-1)}) - y_{k-1} \sin(\theta_{i(k-1)}) \\ &\quad + x_k \cos(\theta_{ik}) + y_k \sin(\theta_{ik}), \\ i &= 1, 2 \end{aligned} \quad (6)$$

The measurement we choose here is the location of two edges

$$Z_k = [\theta_{1k \text{ real}}, \rho_{1k \text{ real}}, \theta_{2k \text{ real}}, \rho_{2k \text{ real}}]^T \quad (7)$$

The transition matrix  $A$  and measurement matrix  $H$  are Jacobian matrices defined as

$$\begin{aligned} A_{[i,j]} &= \frac{\partial f_{[i]}}{\partial X_{[j]}}(\hat{X}_k^-), H_{[m,j]} = \frac{\partial h_{[m]}}{\partial X_{[j]}}(\hat{X}_k^-) \\ m &= 1 \dots 4, i, j = 1 \dots 9 \end{aligned} \quad (8)$$

The Hough space is defined by polar parameter  $(\theta, \rho)$ . Given a point  $P(x, y)$  in the Cartesian image space, we can draw a corresponding curve (i.e. sinusoids) in Hough space which indicates all the possible lines in the image space passing through the point  $P(x, y)$ . For those points which are collinear in the image space, their corresponding curves in the Hough space will intersect at a common point  $(\theta, \rho)$ . These intersections are called peaks. Generally, the Hough space is stored using a matrix whose element values indicates the number of curves passing through this point  $(\theta, \rho)$ . Peaks are expressed as local maxima in this matrix which correspond to straight lines in image space.

Since the tracking is conducted inside a human body where the rigid surgical tool is surrounded by soft organs and tissue, we assume that the surgical tool has most salient line structure which means the two maximum peaks represent two edges of the tool. For frame  $k$ , the estimated peak position from frame  $k-1$  is used to restrain the area of searching the real location of edges. Part of the Hough Transform matrix can be found in Appendix. The peak searching area is restrained inside the red rectangle according to the estimated state vector from frame  $k-1$  and the circles indicate the two maximum peaks we found inside this area. The measure  $Z_k$  derives from these two peaks and is used to correct the predicted peak position for frame  $k$ .

#### IV. EXPERIMENTAL RESULTS

The standard Kalman filter algorithm is implemented on Visual Studio 6.0 platform. OpenCV is applied for image



processing and tracking calculation including edge detection and Kalman filter update. For the interactive interface, we use OpenGL to overlay virtual menu on the real endoscopic environment. Moreover, OpenGL is also utilized to render 3D reconstructed tool model during tracking. This system can work both on the real-time in-vitro captured image sequences and recorded actual surgical procedures. For the real-time case, images are captured by the Karl Storz Supercam 9050B CCD color camera. Illumination is provided by the Karl Storz 615 xenon light source through a fibre optic cable. The light intensity we set during tracking is 100. Matrox Imaging Library (MIL) is responsible for image acquisition, camera calibration and image format transition. The extended Kalman filter method is conducted in Matlab and it returns two edges as tracking results without midline indication as well as 3D reconstruction.

First of all, to validate the accuracy of visual tracking results, we compare the 3D reconstructed tool's tip location with the ground truth. As shown in Fig. 5, several white markers with different heights are placed on a grid and red points on the top of these markers are defined as the ground truth. The 3D coordinates of these red points are measured from the grid and its height. When we use the tool to point at these points under endoscopic camera, the system returns the estimated 3D tip coordinates. In our experiment, 6 groups of locations from 6 different depths are tested. For each group, we fixed the depth and selected 10 random points with variant x-y coordinates. Table I shows the error between estimated tip position and ground truth.

TABLE I. GROUND TRUTH EXPERIMENTAL RESULT

	Mean error(mm)	Std. error(mm)
<b>X</b>	-0.9725	1.4008
<b>Y</b>	-0.0931	1.8087
<b>Z</b>	-0.7088	3.0164

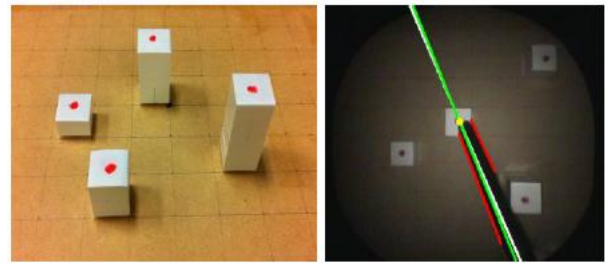


Figure 5. 3D reconstructed tip location validation using ground truth.

After the ground truth experiment, we use emulated endoscopic environment to test the tracking performance of two different Kalman methods. The ideal images are grabbed on a clean table where no objects in the background. Fig. 7 demonstrates the tracking results using these methods under ideal environment. The tracking success rates (tracked frames/200 frames) for standard and extended Kalman filter are 90% and 80% respectively. To simulate surgical environment, the tool is put in a plastic abdominal model. As can be seen in Fig. 7 (c) and (d), both methods also perform well in this case. The tracking success rates are 86% and 70% which has a slight decline.

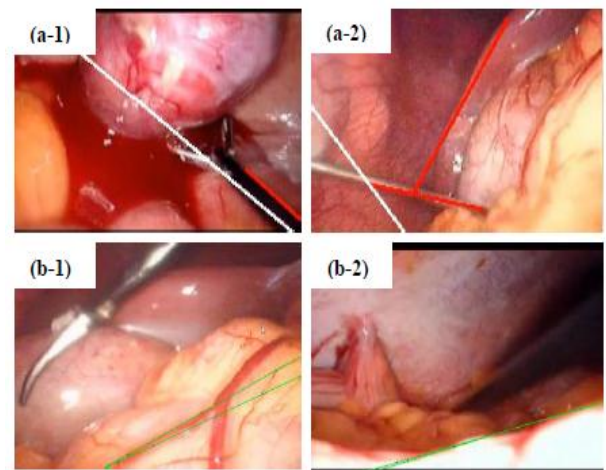


Figure 6. Cases when both methods fail in tracking. (a) is for standard Kalman filter method when testing real video No.6 and No.7 and (b) is for extended Kalman filter method when testing real video No.3 and No.9.

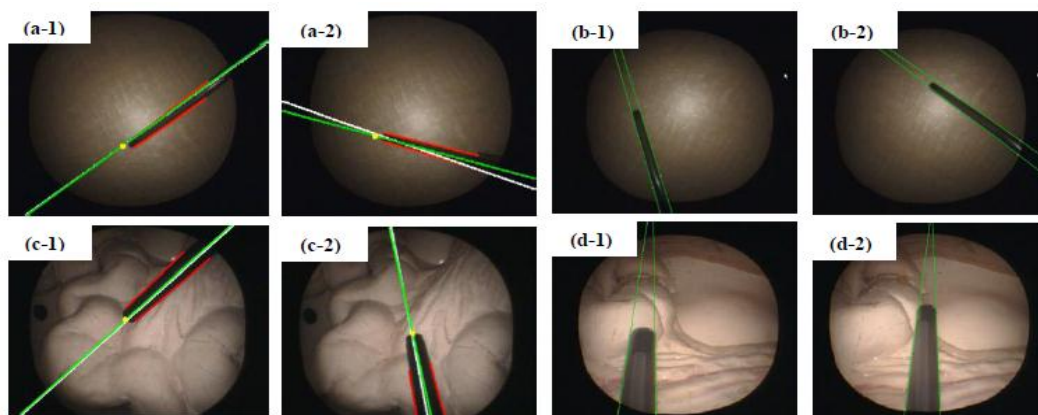


Figure 7. Tracking results using two Kalman filter methods under ideal and emulated environment. (a) and (b) list some frames under ideal environment. (c) and (d) are under emulated environment. (a) and (c) are for standard Kalman filter. (b) and (d) are for extended Kalman filter

To test the performance of two tracking methods we presented in real surgical environment, we have chosen 10 recorded real endoscopic and laparoscopic video streams. Each of them lasts for about 10s containing 200~300 frames. The standard method can almost track the tool with 60% tracking success rate when the frame has strong contrast between background and objective (Fig. 8). However, it's sensitive to the shadow surrounding the tool (Fig. 9c). Moreover, the color and material of the tool also influence the tracking result of standard method (Fig. 9d). In comparison, extended method is robust to these factors but it's easy to lose tracking when the environment containing similar line structures. For those more complex videos which are usually contaminated with blood, smoke and tissues, both two methods failed. Furthermore, when the tool is moving at a rather fast speed both methods failed

(Fig. 6b-2). Sample video under ideal environment is uploaded on the website <http://www.sfu.ca/~jza102>. More clear videos will update in the future.

## V. GLOBAL MONITORING FOR QUALITY OF SURGICAL TOOL TRACKING

Table I lists the tracking success rate when we track a single tool in ideal environment, emulated environment and real surgical environment respectively. The tracking task turns to be challenging when we cope with the real surgical cases. For some real videos like No.7~No.10, both methods failed due to the complex surgical scene such as low contrast between background and metallic tool, fast moving tool, smoke noise and so on.

TABLE II. TRACKING PERFORMANCE OF TWO KALMAN FILTER METHODS UNDER EMULATED AND REAL SURGICAL ENVIRONMENT

	Tracking Success rate (frames/200 frames)					
	Ideal environment	Emulated environment	Real video No.1 Little shadow Dark tool	Real video No.2 Strong contrast Dark tool	Real video No.3 Vessel Dark tool	Real video No.4 Partially occlusion Metallic tool
KF	90%	86%	30%	72%	62%	Fail
EKF	80%	70%	55%	43%	Fail	60%
	Real video No.5 Strong contrast Dark tool	Real video No.6 Blood Dark tool	Real video No.7 Low contrast Metallic tool	Real video No.8 Smoke Metallic tool	Real video No.9 Fast moving tool	Real video No.10 Complex background
KF	65%	20%	Fail	Fail	Fail	10%
EKF	37%	18%	Fail	Fail	Fail	Fail

Note: KF ---- Kalman filter method, EKF ---- extended Kalman filter method

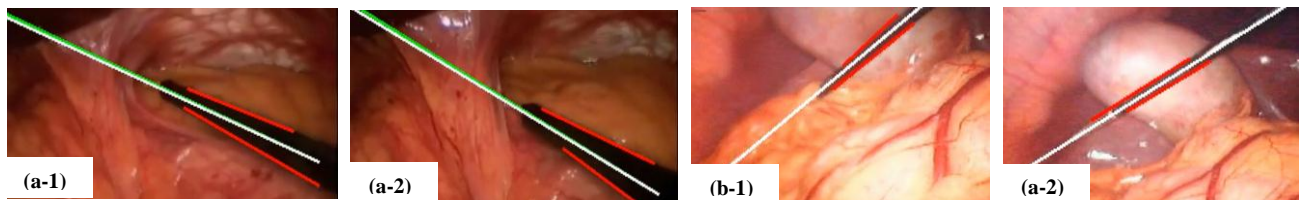


Figure 8. Tracking results of standard Kalman filter method when frames with strong contrast between objects and background. (a) is the tracking results in different frames from real video No.2. (b) is the tracking results in different frames from real video No.3.

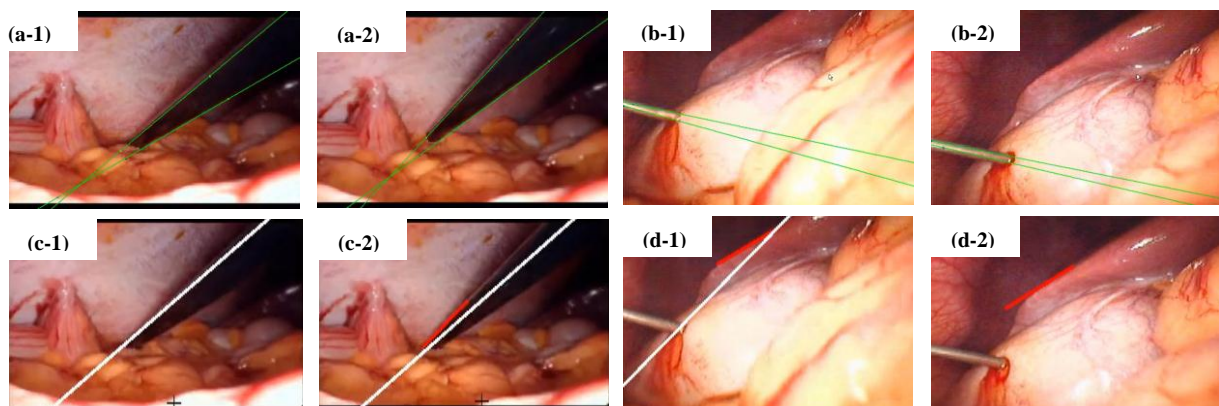


Figure 9. Cases when extended Kalman filters performs better. (a) shows tracking results of different frames from real video No.1 using extended Kalman filter method. (b) shows tracking results of different frames from real video No.4 using EKF. (c) and (d) are the tracking results of the same videos using standard Kalman filter method.

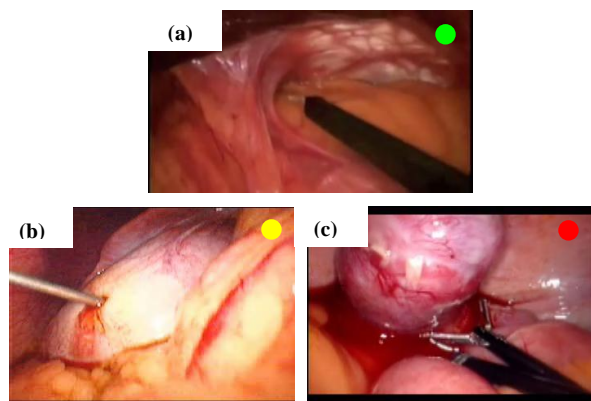


Figure 10. Surgical scenes are classified into three levels where green icon indicates “Good” image quality (real video No.2), yellow icon is for “OK” case (real video No.4) and red icon is a warning for “Bad” image quality (real video No.10).

To measure the natural image quality, there are three categories of methods: full-reference, reduced-reference and no-reference [19]. The first two methods require a reference together with the processed image or video which limits their application in image quality evaluation. No-reference methods are not relevant to the reference image which is of our interest. The existing no-reference approaches focus on both global features like image histogram and local features like the spread of edges. In our case, the measurement is based on line segments. In consideration of the application of our tracking algorithm which aims to create an interactive surgeon-computer interface to provide more information to surgeon during MIS, we classify the image quality into three levels (Fig. 10). The top level is defined as “Good” quality. In this kind of scenes, the objective we want to track has strong color contrast with image background and there’s no similar line structure in the scene. One example is the ideal environment and emulated environment. The lower level is defined as “OK” quality where the scene contains not so much confounding factors which we think our methods can also work at an acceptable success rate. Those scenes which are full of blood, tissue and multiple instruments are classified as the “Bad” quality. As an initial study and to assess the image quality in our application, we consider both the number and length of line segments in their canny images. If the image contains large number of line segments such as 50~100 segments and the line segments’ length are very short, for example, smaller than 30 pixels, the quality is classified into “Bad”. Fig. 4(c) is the example for “Good” quality where we can see clear edges are detected in that canny image.

The three image quality levels are indicated by an icon on the corner of the scene. Particularly, green is for “Good”, yellow is for “OK” and red is for “Bad”. During the initialization and tracking, the system first recognizes the quality of scene. If the green icon is on, it means the surgeon can activate tool tracking module and overlay the virtual menu on real surgical scene. In other scenes and during the tracking phase, when the icon turns yellow,

surgeon can also conduct tracking and virtual menu interaction. However, the performance of tracking is not guaranteed in this case. The red icon is a warning signal for the surgeon which tells him that this scene is not suitable for tool tracking and the virtual menu function is deactivated.

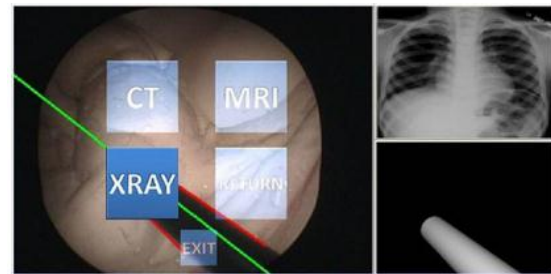


Figure 11. Surgeon computer interface of our visual tracking system.

Another part of global monitoring is the initialization of tracking set-up according to different types of MIS and different surgical tools. For example, surgeon can tell the system whether the MIS is colonoscopy or laparoscopy and whether the tool he uses is dark one or metallic one. After inputting this kind of information, the system can choose different tracking method to use. Here, extended Kalman filter method is ready for tracking metallic tool and standard Kalman filter method is for dark tool case.

## VI. CONCLUSION AND FUTURE WORK

This paper introduces a novel visual tracking system to track the surgical instrument under emulated and real surgical environment. Two tracking algorithms based on standard and extended Kalman filter respectively are compared and both of them perform well in “Good” quality scenes like ideal and emulated environment and also those scenes with strong contrast between tool and background. Fig. 11 shows the application of the tracking results. However, they have their own drawbacks when dealing with complicated surgical scenes. For example, standard Kalman filter is sensitive to the material of tool which means it’s easy to track dark tool compared with metallic tool. The extended Kalman filter is quite unstable when the background contains similar line structures. In our future work, particle filter which is less sensitive to image noise will be applied in our tracking system. Though the global monitoring part can choose the better tracking method for different scenes, it’s necessary to develop a more robust algorithm where system can fine tune itself to suit the tracking environment.

As the global measurement of image quality is the basis to conduct visual tracking and surgeon computer interaction, more methods can be explored in our future work to assess the scene in a more reliable and comprehensive way. On the other hand, we can further explore the feature and color space of tracking to present combined tracking algorithm which can cope with more complex surgical scenes like “OK” quality scenes even “Bad” quality scenes. Moreover, multiple instruments tracking is also under study in our laboratory.



## APPENDIX

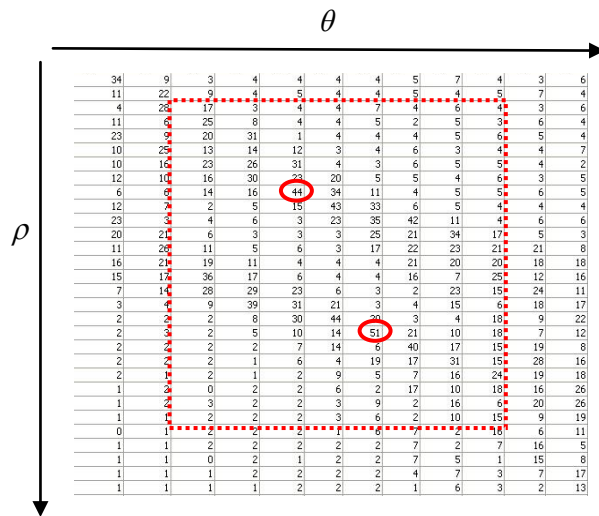
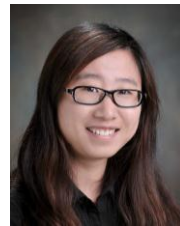


Figure 12. Searching peaks in hough space

## REFERENCES

- [1] J. Wachs, *et al.*, "Gestix: A doctor-computer sterile gesture interface for dynamic environments," *Soft Computing in Industrial Applications*, pp. 30-39, 2007.
- [2] M. Rosenthal, *et al.*, "Augmented reality guidance for needle biopsies: A randomized, controlled trial in phantoms," in *Medical Image Computing and Computer-Assisted Intervention Lecture Notes in Computer Science*, Springer, 2001, pp. 240-248.
- [3] O. Tonet, T. Ramesh, G. Megali, and P. Dario, "Image analysis-based approach for localization of endoscopic tools," in *Proc. Surgetica*, vol. 5, 2005, pp. 221-228.
- [4] G. R. Bradski and J. W. Davis, "Motion segmentation and pose recognition with motion history gradients," *Machine Vision and Applications*, vol. 13, no. 3, pp. 174-184, 2002.
- [5] R. Wolf, J. Duchateau, P. Cinquin, and S. Voros, "3D tracking of laparoscopic instruments using statistical and geometric modeling," in *Proc. Medical Image Computing and Computer-Assisted Intervention*, 2001, pp. 203-210.
- [6] C. Doignon, P. Graebing, and M. de Mathelin, "Real-time segmentation of surgical instruments inside the abdominal cavity using a joint hue saturation color feature," *Real-Time Imaging*, vol.11, no. 5-6, pp. 429-442, 2005.
- [7] C. Doignon, F. Nageotte, and M. de Mathelin, "The role of insertion points in the detection and positioning of instruments in laparoscopy for robotic tasks," in *Proc. Medical Image Computing and Computer-Assisted Intervention*, 2006, pp. 527-534.
- [8] S. Voros, J. A. Long, and P. Cinquin, "Automatic detection of instruments in laparoscopic images: A first step towards high-level command of robotic endoscopic holders," *The International Journal of Robotics Research*, vol. 26, no. 11-12, pp. 1173-1190, 2007.
- [9] A. Cano, *et al.*, "Laparoscopic tool tracking method for augmented reality surgical applications," *Biomedical Simulation*, pp. 191-196, 2008.
- [10] X. Sun and S. Payandeh, "Toward development of 3D surgical mouse paradigm," in *Proc. International Conference on Intelligent Robots and Systems, IEEE*, 2011, pp. 2096-2101.
- [11] X. Sun, "Image guided interaction in minimally invasive surgery," Thesis, in Engineering Science, Simon Fraser University: Burnaby, 2012.
- [12] C. Godau and B. Funt, *The Logvinenko Object Color Atlas in Practice. Color Research & Application*, vol. 37, no. 2, pp. 117-125, 2010.
- [13] G. Welch and G. Bishop, "An introduction to the kalman filter," University of North Carolina at Chapel Hill, Chapel Hill, NC, 27599-3175, TR95-041, vol. 7, no. 1, 1995.
- [14] P.V.C. Hough, "Method and means for recognizing complex patterns," US Patents 3069654, 1962.
- [15] G. G. Rigatos, "Nonlinear kalman filters and particle filters for integrated navigation of unmanned aerial vehicles," *Robotics and Autonomous Systems*, vol. 60, no. 7, pp. 978-995, 2012.
- [16] V. Voisin, M. Avila, B. Emile, S. Begot, and J. C. Bardet, "Road markings detection and tracking using hough transform and kalman filter," in *Advanced Concepts for Intelligent Vision Systems*, Springer, 2005, pp. 76-83.
- [17] A. Borkar, M. Hayes, and M. T. Smith, "Robust lane detection and tracking with RANSAC and Kalman filter," in *Proc. 16<sup>th</sup> IEEE International Conference on Image Processing*, 2009, pp. 3261-3264.
- [18] S. Mills, T. Pridmore, and M. Hills, "Tracking in a hough space with the extended kalman filter," in *Proc. BMVC*, 2003.
- [19] C. Chen, A. C. Kot, and H. Yang, "A two-stage quality measure for mobile phone captured 2D barcode images," *Pattern Recognition*, vol. 46, no. 9, pp. 2588-2598, September 2013.



**Jiawei Zhou** was born in Shanghai, China. She received a B.Eng. Degree in Biomedical Engineering, Shanghai Jiaotong University, China in 2008. After graduation, she continued her study and received her M. Eng Degree in the same university in 2011. Now she's pursuing her second M.A.Sc Degree at School of Engineering Science in Simon Fraser University, Canada.

During her study in Shanghai Jiaotong university, she worked as a Research Assistant in Medical Image Processing Laboratory with Prof. Binjie Qin from 2008 to 2011. Her research field focuses on Medical image registration and segmentation and she published a paper related to non-rigid registration for MRI in Proc. SPIE 8009, ICDIP 2011. And from 2011, she is a member of Experimental Robotics Laboratory and TA in School of Engineering Science, SFU. Her research interests include image tracking and augmented reality application in mobile device.



**Shahram Payandeh** is the Professor from School of Engineering Science, Simon Fraser University, Canada. He received M.A.Sc. Degree and Ph.D. Degree in University of Toronto, Canada, in 1986 and 1990 respectively. He is the group leader of Experimental Robotics Laboratory with plenty of publications. His research interests are on Robotics, Multi-view tracking, Haptic devices and haptic rendering.