An Experimental Study of Visual Tracking in Surgical Applications

Jiawei Zhou and Shahram Payandeh
Experimental Robotics and Imaging Laboratory, School of Engineering Science, Simon Fraser University, Burnaby, BC, Canada

Keywords: Visual Tracking, Minimally Invasive Surgery, Adaptive Gaussian Mixture Model, Particle Filter.

Abstract: Tracking surgical tools in mono-endoscopic surgery can offer a conventional (non-robotics) application of this type of procedure a versatile surgeon-computer interface. For example, tracking the surgical tools can enable the surgeon to interact with the overlaid menu which allows them to have access to medical information of the patient. Another example is the capability that such tracking can offer where the surgeon through surgical tool can manually register per-operative images of the patient approach on the surgical site. This paper presents the results of some of the tracking schemes which we have explored and analysed as a part of our studies. Tracking framework based on both Gaussian and non-Gaussian framework are explored and compared. Although majority of the approaches can offer a robust performance when used in the real surgical scene, the method based on Particle Filter is found to have a better success rate. Based on these experimental results, the paper also offers some discussions and suggestions for future research.

1 INTRODUCTION

The research and development in computer-aided surgical application can dramatically promote the delivery and training of modern medicine. For example, image guided surgical navigation can assist the surgeons in performing minimally invasive surgery (MIS) through tiny incisions. Moreover, a computer-based simulation system with Augmented Reality (AR) is able to offer a safe and realistic learning environment to medical students instead of using a real-life context. The critical part for these applications is the implementation of a user computer interactive system (UCIS) based on surgical requirements.

To avoid unnecessary physical contact with the environment, some research groups have focused on the development of a gesture-based UCIS. In this kind of UCIS, the standard user computer interaction devices, such as mouse and keyboard, are replaced with sensing devices (e.g. Microsoft Kinect sensor) which can capture the motion and the hand gestures of the user in the workspace (Gallo et al. 2011; Ruppert et al. 2012). Another kind of UCIS integrates AR technology into surgical and medical requirements. These UCISs are able to superimpose the medical image data or 3D virtual medical model directly onto the view of the surgeon and to spatially register the image or model of the patient (Navab et al. 2007; Su et al. 2009).

Although a number of robotic surgical systems with 3D stereo cameras have been developed to assist the surgeon during MIS, most hospitals still prefer utilizing a traditional and non-robotic surgical system with a monocular camera due to limited budgets. This paper is focused on the enhancement of such non-robotic monocular MIS set-ups. In the previous work, a real-time interactive system for a non-robotic monocular endoscope MIS was developed to enhance the practice of MIS training without adding extra hardware to the existing setup (Sun 2012). In this system, the surgical instrument is being considered as the input control variable so a robust tracking algorithm is important to localize the instrument in the field of view.

To achieve an accurate position of the surgical tool, a number of challenges must be overcome. These challenges include limited measured information from endoscopic video signals, the complexity of the surgical scene, reflection of light, occlusion of surgical tools and so on. To cope with these problems, a number of research groups attach specific markers on the surgical instrument to help with the tracking (Tonet et al. 2007) and others utilize segmentation techniques and feature detection to assist tracking (Doignon et al. 2005; Cano et al. 2008).

Zhou J. and Payandeh S..
An Experimental Study of Visual Tracking in Surgical Applications.
DOI: 10.5220/0005346506080613
In *Proceedings of the 10th International Conference on Computer Vision Theory and Applications* (VISAPP-2015), pages 608-613
ISBN: 978-989-758-091-8

For both marker-based and marker-less tracking methods, the information from an image is not enough to obtain a reliable result. Instead, the information contained in a sequence of images should be taken into account to improve the accuracy of the tracking. To estimate the position of the surgical tool, we introduced some probability based estimators, such as the Kalman Filter (KF) and the Extended Kalman Filter (EKF) to the visual tracking system in the previous work (Zhou & Payandeh 2014). The application of these estimators is able to return more accurate and reliable tracking results. This paper is a further development of the experimental study based on the previous visual tracking work. An adaptive Gaussian Mixture Model (AGMM) method is implemented to track the surgical instrument under the assumption of Gaussian background components. Moreover, to explore better tracking performance for surgical application, a more general tracking scheme based on Particle Filter (PF) is presented. This framework is further combined with the AGMM as the Hybrid approach to provide 2D feature information for the tool during tracking. These methods are experimentally evaluated in both an in-vitro scene and an in-vivo setting, and also compared with the results from the previous work.

2 METHODS

In this section, we present an overview of three visual tracking approaches for MIS instrument localization. A Gaussian-type tracking method based on AGMM is firstly introduced, followed by a more general PF tracking scheme with a weight-based resample strategy. To explore better tracking performance, a Hybrid approach which combines the PF framework and the AGMM is also implemented.

2.1 AGMM Method

To detect a moving object in image sequences, one type of tracking methods is based on background subtraction. The AGMM method is a successful application in the visual tracking field. The basic idea for the AGMM is to set up a background model which can be used to distinguish the foreground object from background environment. The region of the moving object is highlighted by calculating a reference image and subtracting each new frame from this reference image. The AGMM was originally proposed in (Stau & Romano 1998) and was further developed by introducing a shadow detection scheme (Kaewtrakulpong & Bowden 2002). Our laboratory has successfully applied the AGMM method in a people surveillance system (Dai 2012). In this paper, we use the AGMM to track the moving surgical tool.

In the AGMM, the pixel values in the scene background are modelled using a mixture of adaptive Gaussian components. Given an arbitrary pixel value x_t, the ith Gaussian density function at time t is $\eta(x_t, \mu_{i,t}, \sigma_{i,t}^2)$ with a mean value $\mu_{i,t}$ and a standard deviation $\sigma_{i,t}$. The probability of a particular pixel p_0 having value x_t is defined as

$$P(x_t) = \sum_{i=1}^{k} w_{i,t}\eta(x_t, \mu_{i,t}, \sigma_{i,t}^2) \qquad (1)$$

where $w_{i,t}$ is the weight for the ith Gaussian component and k is the number of components. To cope with slight changes in the background, such as changing illuminants, an adaptive background model is necessary. To allow a foreground object to become part of the background later, the Gaussian distribution having the lowest weight is replaced with a new Gaussian function. This new Gaussian component is given a low normalized weight which will be used in the time $t+1$. Meanwhile, the mean and variance of the other remaining Gaussian components for the time $t+1$ are also updated.

To deal with the shadow noise in the returned foreground object region, a shadow elimination step is added to the output of AGMM. Some morphological operations such as opening and closing are also applied to the shadow detected image. After these post-processing steps, the contour of the moving surgical instrument is accurately extracted frame by frame.

2.2 PF Tracking Framework

The idea of the PF is to generate a group of weighted samples (particles) to approximate the posterior probability density function (PDF). These particles are weighted according to the weighting function. This function is created based on the measurement from the image data. Generally, higher weights are given to more reliable particles. If the number of particles are large enough, we can recover the unknown posterior PDF for the state-space using its approximation after several iterations. Since the PF method does not assume any linearity of the system or Gaussian noise distributions, it is widely used to track objects in general and medical applications (Tehrani Niknejad et al. 2012; Ito et al. 2013). In the previous work, a colour-based PF method was used to track a coloured marker in the emulated surgical scene (Sun 2012). In this paper, a modified PF framework is used to determine the possible region for the moving surgical instrument.

For our tracking problem, the target object is selected to be the rectangle region Rec which indicates the possible area for the surgical tool. The state vector X at the current time t is described as $X_t = [x_C, y_C, x'_C, y'_C, w_R, h_R, w'_R, h'_R]_t^T$ where (x_C, y_C) is the coordinate of the rectangle center. w_R and h_R are the width and height of the rectangle. x'_C and y'_C are its instantaneous velocities along the x and y directions respectively. w'_R and h'_R are the instantaneous change of width and height. To measure the similarity of the target and the particles, we choose the colour distribution in the HSV space as the tracking metric since the region of the instrument tends to have more dark or bright components than other regions. By calculating the 2D Hue-Saturation colour histogram from various areas, we can find those areas containing the same object owing to their similar colour distributions.

The PF consists of prediction and update phases. The particles at current time t are first propagated from particles at time $t-1$ according to a first-order auto-regressive system dynamic model. In the update phase, we compare the 2D colour histogram of each sample area with the reference histogram. The reference one is calculated in the initialization step. The weighting function is designed based on the histogram similarity. Higher weights are given to those areas with similar colour distribution to the reference region. From the predicted particles and their normalized weights, we can estimate the possible region for the moving surgical instrument using their expected value.

To avoid the degeneracy problem in PF, we use an adaptive resampling strategy to eliminate particles with low weights. We first sort the predicted particles in a descending order according to their weight values so the particles with higher weights have the priority to be chosen. To generate the new particle set, we take the first element from the particle queue and make n_{update} copies of it in the new particle set. The number n_{update} is calculated from

$$\begin{aligned} n_{update}[j] &= \min(h \in \mathrm{Z} \mid h \geq n_{update}[j]) \\ \sum_{j=1}^{m} n_{update}[j] &= N \end{aligned} \tag{2}$$

where m is the number of elements we selected from the particle queue. Figure 1 illustrates an example of weighted-based resampling. The original particle set has 5 particles indicated in different colours. The height represents its weight. In (b), the particles are organized in a descending order. After resampling, the new particle set contains the copies from those particles with high weights (green, blue and gray ones). The particles with low weights (red and orange) are removed from the particle set.

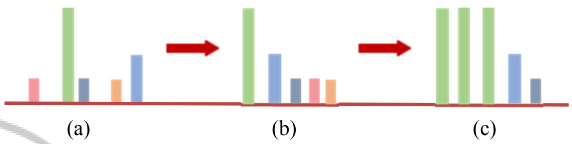

Figure 1: Illustration of the weight-based resample strategy. (a) Original particle set. (b) Ordered particle queue. (c) New particle set.

2.3 A Hybrid Method

The AGMM method is able to detect the moving foreground object in a stable background but it may fail tracking if the background or the viewing condition is changing. The PF approach can be used for more general surgical scenes. However, it requires the initial state vector, usually obtained from the user. To overcome these two drawbacks, a hybrid method integrating the PF and the AGMM is presented. After we get the estimated region from the PF framework, 2D feature detection is applied within the region so that the tip and two edges of the instrument can be returned.

For a given video stream, we assume that the backgrounds in every constant short time period (e.g. 1~2s) are relatively stable so that the AGMM is able to detect the moving object. Before we start the PF part, the AGMM is first applied to determine the initial position of the moving surgical tool and returns a rectangular region containing the tool. We use this rectangular region to compute the reference histogram and to generate particles in the initialization of the PF. After the initialization, the PF is responsible for the surgical tool tracking as stated in Section 2.2. If the tracking is lost, the AGMM is applied again to relocate the rectangular region that encloses the initial position of surgical tool, and the PF is re-initialized. After we obtain the region of interest from the PF, we use the feature detection method to find the tool's tip position. The details for the feature detection can be referred to the previous work (Sun 2012).

3 EXPERIMENTAL RESULTS

To evaluate the performance of all the three tracking methods we have presented (AGMM, PF, and the Hybrid one) for surgical and medical usage, the experiments were conducted under an in-vitro training environment and an in-vivo environment. The in-vitro training cases were captured from the surgical emulated setup in our laboratory (Sun 2012).

For in-vivo surgical experiments, we used several surgical video clips from the online video atlas of Dr. Julio Alejandro Murra Saca Medical Clinic (see reference for the website). All the results are also compared with the KF and EKF methods which are proposed in the previous work (Zhou & Payandeh 2014).

3.1 Tracking in in-vitro Training Environment

In this part of the experiments, we assume the endoscope is stationary and the background is rarely changing (i.e. as a part of the surgical training environment). First of all, we tested the tracking of a moving surgical tool in the ideal scene. For this type of scene, there are no objects in the background. In the next stage, to simulate the surgical environment, we test the tool tracking in a more complex scene. The ideal background is replaced with a training abdominal cavity model. Due to the hardware and software limitations, only the KF approach is tested in real-time and the other methods are tested offline.

The tracking results for all the methods are demonstrated in Figure 2. For the KF approach, it returns the 2D features of the tool including the tool's edges, midline and tip. The EKF method is able to detect two edges of the tool. The AGMM and PF methods indicate the region containing the moving surgical tool. The Hybrid method can detect the 2D features of the tool based on the interest region.

3.2 Tracking in in-vivo Surgical Environment

For the in-vivo experiments, we used 10 endoscopic video clips captured from the real surgical scenes. Each of them lasts for about 10s containing 200~300 frames. In these initial experiments, we focus on the scenes with short duration where the motion of instrument is slow and smooth without rapid change in orientation, scale and position. The example of tracking results are displayed in Figure 3.

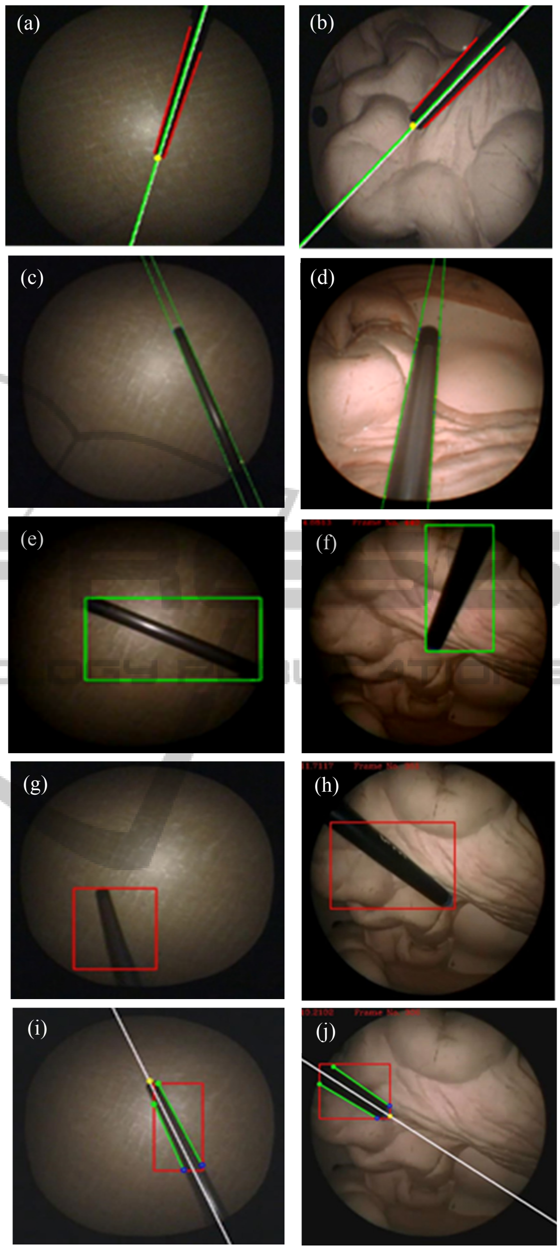

Figure 2: The tracking results of all the five methods in the in-vitro training environments. The 1st column shows the results in ideal scene and the 2nd column displays the results in emulated scene. Each row corresponds to different methods. From top to bottom, they are using the KF, EKF, AGMM, PF and Hybrid approach respectively.

To evaluate the tracking performance of each method, we select the tracking success rate, defined as the number of tracked frames per 200 frames, as the criterion for tracking performance. The tracking success rates for all the five tracking approaches under both the in-vitro training environment and the in-vivo real surgical environment are listed in Table 1. The in-vitro training environment is composed of the ideal scene and emulated scene. The in-vivo environment includes 10 surgical scenes labelled from video No.1 to video No.10. For the videos No.1, No.2, No.3, No.5 and No.6, the surface of the tool is dark and matt. For videos No.4 and No.7, the tool is made of a metallic material. Video No.8 involves multiple tool tasking. In video No.9, the tool is moving at a fast speed and video No.10 has a complex background.

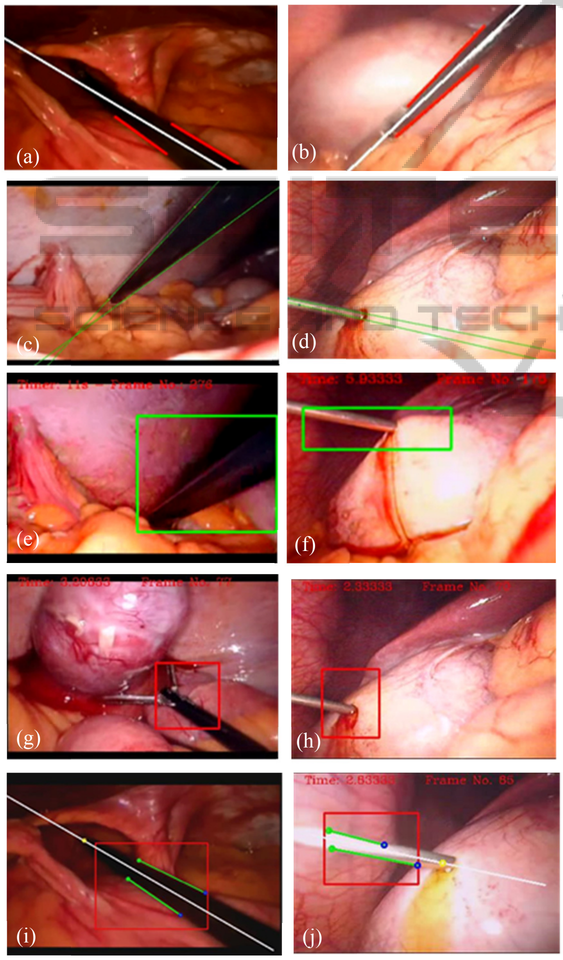

Figure 3: Examples for the tracking results of all the five methods in the real surgical scenes. Each row corresponds to different methods. From top to bottom, they are using the KF, EKF, AGMM, PF and Hybrid approach respectively.

4 DISCUSSION AND CONCLUSION

In this paper, we have proposed experimental studies of various visual tracking techniques for MIS and related training. The tracking performance of all the methods (AGMM, PF and the Hybrid one), is evaluated under different environments. For the in-vitro experiment including the ideal scene and emulated scene, they all have successful performance when tracking single surgical instrument. In the previous work, the KF and EKF methods perform well when the instrument moves slowly and the background is clean but they are sensitive to the noise from the background. Compared to the KF and EKF methods, the methods presented in this paper are more robust with respect to such noises as long as the background is stable or has only a slight change. However, all the methods have limitations when coping with the real surgical scene. The AGMM method is able to keep tracking under a stable background but it cannot deal with the situation when the background changes rapidly. The PF method has the best tracking results in comparison with all the other methods even though it fails in the complex surgical scene and the fast tool motion. However, the manual initialization may lead to errors when the initial region has a similar colour histogram with another area in the background. The Hybrid method works well in the in-vitro environment without manual initialization and is able to come back from the lost tracking situation. Based on the tracked region, it is able to provide the tool's feature information even if the tracked region just covers a small part of the tool. Nevertheless, the Hybrid method cannot be applied to real surgical scenes due to its unreliable initialization. Due to the motion of both the background and the surgical tool, the AGMM easily gives wrong information to the PF framework which may lead to tracking failure.

In spite of the fact that none of the tracking methods in our experimental study can be used as a practical solution for the general surgical tool tracking problem, these methods have shown their possibility in a stable working condition which is the case for the surgical training box. Based on the results from our experimental study, a more robust tracking approach can still be developed using the PF framework. To obtain more reliable tracking results, a new measurement is required to consider both the colour distribution and feature information of the moving tool. A better detection algorithm is also needed to find the location of the tool's tip. One possible solution is to include the tip location into the

Table 1: The tracking performance of the KF, EKF, AGMM, PF and Hybrid method under various experimental environments.

Method	**Tracking Success Rate (frames/200 frames *100%)**				
	KF	**EKF**	**AGMM**	**PF**	**Hybrid**
Ideal Scene	90%	80%	98%	80%	83%
Emulated Scene	86%	70%	96%	90%	92%
Video No.1	30%	48%	42%	45%	20%
Video No.2	48%	33%	53%	50%	15%
Video No.3	52%	Fail	38%	60%	Fail
Video No.4	Fail	50%	55%	67%	40%
Video No.5	45%	31%	48%	65%	50%
Video No.6	20%	18%	23%	56%	17%
Video No.7	Fail	Fail	10%	58%	10%
Video No.8	Fail	Fail	Fail	53%	Fail
Video No.9	Fail	Fail	Fail	Fail	Fail
Video No.10	10%	Fail	12%	15%	Fail

state vector to minimize the error in feature detection. Moreover, to realize the automation of the tracking, the PF framework needs a more reliable initialization strategy.

REFERENCES

Cano, A. M. et al., 2008. Laparoscopic Tool Tracking Method for Augmented Reality Surgical Applications. In *4th International Symposium, ISBMS , Proceedings*. pp. 191–196.

Dai, X., 2012. *Intelligent People Surveillance Framework*.Master Thesis, Simon Fraser University.

Doignon, C., Graebling, P. & de Mathelin, M., 2005. Real-time Segmentation of Surgical Instruments inside the Abdominal Cavity using a Joint Hue Saturation Color Feature. *Real-Time Imaging*, 11(5-6), pp.429–442.

Gallo, L. et al., Controller-free Exploration of Medical Image Data : Experiencing the Kinect. In *2011 24th International Symposium on Computer-Based Medical Systems (CBMS)*. Bristol: IEEE, pp. 1–6.

Ito, T., Anzai, D. & Wang, J., 2013. A Modified Particle Filter Algorithm for Wireless Capsule Endoscope Location Tracking. In *Proceedings of the 8th International Conference on Body Area Networks*. pp. 536–540.

Kaewtrakulpong, P. & Bowden, R., 2002. An Improved Adaptive Background Mixture Model for Real- time Tracking with Shadow Detection. *Video-Based Surveillance Systems. Springer US*, pp.135–144.

Navab N, Feuerstein M, B. C., 2007. Laparoscopic Virtual Mirror New Interaction Paradigm for Monitor Based Augmented Reality. In *Virtual Reality Conference, VR'07. IEEE*. pp. 43–50.

Ruppert, G. C. S. et al., 2012. Touchless Gesture User Interface for Interactive Image Visualization in Urological Surgery. *World Journal of Urology*, 30(5), pp.687–91.

Stau, W. E. L. G. C. & Romano, R., 1998. Using Adaptive Tracking to Classify and Monitor Activities in a Site. *Computer Vision and Pattern Recognition, 1998. Proceedings. IEEE Computer Society Conference on.*, 2.

Su, L.-M. et al., 2009. Augmented Reality during Robot-assisted Laparoscopic Partial Nephrectomy: toward Real-time 3D-CT to Stereoscopic Video Registration. *Urology*, 73(4), pp.896–900.

Sun, X., 2012. *Image guided interaction in minimally invasive surgery*. Master Thesis, Simon Fraser University..

Tehrani Niknejad, H. et al., 2012. On-Road Multivehicle Tracking Using Deformable Object Model and Particle Filter With Improved Likelihood Estimation. *IEEE Transactions on Intelligent Transportation Systems*, 13(2), pp.748–758.

Tonet, O. et al., 2007. Tracking Endoscopic Instruments without a Localizer: A Shape-analysis-based Approach. *Computer Aided Surgery*, 12(1), pp.35–42.

Zhou, J. & Payandeh, S., 2014. Visual Tracking of Laparoscopic Instruments. *Journal of Automation and Control Engineering Vol*, 2(3), pp.234–241.

Julio Alejandro Murra-Saca MD. Notes on Cyber Gastroenterology, [Online]. Available: http:// www.murrasaca.com/Laparoscopic.htm.