Visual Tracking of Laparoscopic Instruments in Standard Training Environments

Brian F. ALLEN [a] Florian KASPER [a] Gabriele NATANELI [a] Erik DUTSON [b]
Petros FALOUTSOS [a]

[a] *Department of Computer Science, University of California, Los Angeles*
[b] *Department of Surgery, University of California, Los Angeles*

Abstract. We propose a method for accurately tracking the spatial motion of standard laparoscopic instruments from video. By exploiting the geometric and photometric invariants common to standard FLS training boxes, the method provides robust and accurate tracking of instruments from video. The proposed method requires no modifications to the standard FLS training box, camera or instruments.

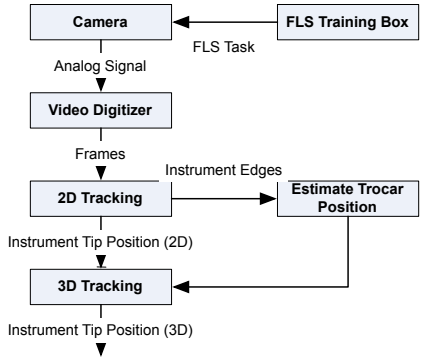
Keywords. Laparoscopic Surgery, Surgery Training, Machine Vision

Introduction

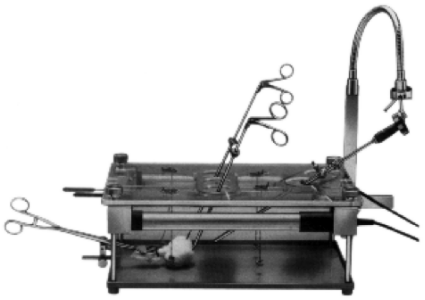
Laparoscopic surgery is the most common and widely available minimally invasive surgical technique employed today. With a fiber-optic camera and specialized instruments, entire procedures can be accomplished through keyhole incisions. In comparison to open surgery, laparoscopic procedures are less invasive, require shorter periods of hospitalization and entail faster recovery time and less pain for the patient. However, such benefits do not come without costs.

In the case of a laparoscopic surgery, perhaps the primary trade-off is the difficulty of the operation and the need for a specialized repertoire of motor skills. To address the difficulty of training and evaluating the skill of surgeons, theSociety of American Gastrointestinal and Endoscopic Surgeons (SAGES) adopted the Fundamentals of Laparoscopic Surgery (FLS) as a standardized toolset for certification and assessment. FLS is a set of experimentally validated training tasks and equipment [6], providing a standardized means to assess the motor skills specific to laparoscopy. Such objective measure of skill is particularly important in light of studies that show that training surgeons have little ability to self-assess [5].

FLS assessment gauges manual skills entirely on two features of task performance: movement efficiency (measured by the time taken to complete the task) and a precision measure specific to the task. Precision measures include transient, observed actions, such as dropping a block in the peg transfer task, as well as after-the-fact measures, such as divergence from the target circle in the cutting task, or security of a suture knot. Improvement in the accuracy of assessment has been demonstrated by considering more information than FLS records. In



(a) Overview of the process showing data flow.



(b) A standard FLS box trainer.

Figure 1.

particular, tracking the full *spatial motion* of the instruments during the course of the task performance provided significant gains by considering metrics such as the path length instrument tips travelled [8].

Unfortunately, the equipment needed to acquire detailed spatial tracking data is expensive and specialized. The researchers have predominately employed either (1) precise magnetic tracking [1], (2) mechanical linkages attached to the instruments [7], or (3) virtual reality (VR) simulators with joysticks replacing laparoscopic instruments [11]. Note that (1) and (2) require physical attachments to instruments, while VR simulators typically rely on joysticks that simulate actual laparoscopic instruments. Notably, and most comparable to our work, Tonet et al. [9] considered tracking actual instruments using computer vision. However, that method requires modifying the instruments by affixing a ring of Lambertian material at a know position. In addition, machine vision techniques for laparoscopy have been proposed to control robotic camera holders [10], and for visual-servoing of laparoscopic robots [4]. In this work, we make use of several methods employed by other authors. Voros et al. proposed the use of a probabilistic Hough transform [10] for tracking instruments to automate control of a laparoscope. Doignon et al. [3] describe a least-squares fit of the instrument positions across a series of images to estimate the trocar position.

The main contribution of this work is the synthesis of a complete system for tracking tools in FLS training boxes, including the accurate detection of the instrument shafts within the image, the estimation of tool-tip position along the shaft, the automatic registration of the trocar's position, and the geometric computation of the camera-space position. This method, summarized in figure 1(a), and is specifically tailored to tracking laparoscopic instruments in standard FLS trainer boxes. Our goal is purposefully less ambitious than attempts to track instruments in general settings, such as *in vivo*. We happily exchange generality for reliability and accuracy in this particularly useful setting.

1. Methods and Materials

The primary equipment of FLS is a "box trainer," pictured in figure 1(b), with several ports and a fixed camera. Our system accepts the video recorded by the camera included in the standard FLS training box.

1.1. Image-Space Position (2D)

Distinct photometric features of the FLS toolset allow us to robustly track the 2D position of instrument tool-tips within each frame of video. Our algorithm has three phases: (1) color space analysis and extraction of the instrument contours, (2) line fitting to estimate the direction of each instrument shaft , (3) linear search to identify the most probable position of the tool-tip along each instrument.

In the standard FLS setup, both the pegs and instrument shafts have a distinct black color. A simple thresholding operation provides a binary probability map of both the pegs and the shafts (shown in figure 2(b)), which we then filter with the application of the erosion and dilation morphological operators. By carefully picking the number of iterations for these operators, we isolate the contours of the two instruments in one step, as shown in figure 2(c). The number of iterations is determined automatically, as described in Section 1.1.2. Automated tuning greatly improves the robustness of this step.

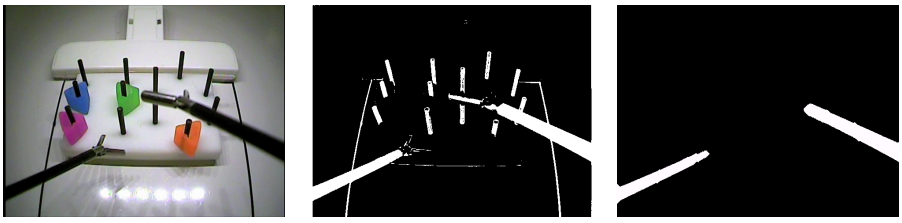
By applying the Hough transform on the isolated instrument maps, we extract the lateral contours of each shaft (shown in figure 2(d)). Considering that the instruments are always posed diagonally in the frame, we use the inclination of the lateral contours to group them as belonging to the left and right instruments. We fit a line by least-squares that corresponds to the major axis of each instrument, to each group. The forward direction (from the image borders to the center) of each axis defines a line along which we are going to search for the instrument tool-tips. Figure 2(e) shows the best-fit lines for the example frame.

1.1.1. Searching for the Instrument Tool-Tips with a Confidence Estimate

The demarcation point between the instrument and the tool-tip is clearly defined by the abrupt transition between the black color of the shaft and the lighter color of the tool-tip metal body. For added robustness, we search for this point along the direction of each instrument in two functional domains: (1) color space, and (2) gradient space. If we call α the angle between the instrument axis and the the Y-axis, the directional gradient of the image along this angle is given by convolving the image with a rotated forward differencing kernel:

$$\begin{bmatrix} cos(\alpha) - sin(\alpha) & cos(\alpha) & cos(\alpha) + sin(\alpha) \\ -sin(\alpha) & 0 & sin(\alpha) \\ -cos(\alpha) - sin(\alpha) & -cos(\alpha) & -cos(\alpha) + sin(\alpha) \end{bmatrix}$$

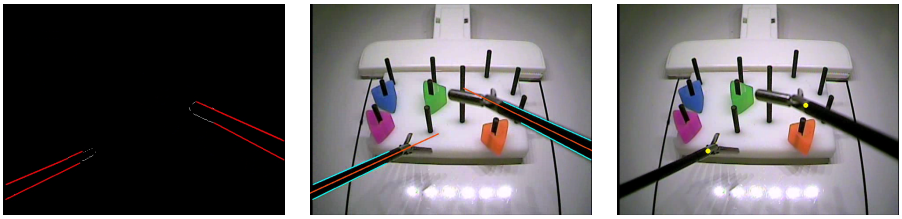
The point T_G found in the gradient domain is consistently more accurate than T_C found in the color space. Therefore, we always use T_G for tracking the position of the tool-tip. On the hand, we use T_C to produce an estimate of the confidence we have in T_G. We found experimentally that the accuracy of tracking is greatly affected by a shift in the color space characteristics of the instrument region, due to the tool-tips getting out of focus. Hence, by estimating the discrepancy



(a) Unmodified frame from the FLS camera during a training task.

(b) Binary probability map of black regions.

(c) Binary mask with isolated instruments.



(d) Extracted lateral contours of instruments.

(e) Instrument direction estimated using line-fitting.

(f) Tracked position in 2D.

between T_C and T_G, which are two measurements of the same quantity, we obtain a rather reliable estimate of the confidence of T_G. We express this notion as $P(T_G) = 1 - \frac{\|T_G - T_C\|}{\beta}$ where β is a normalization constant.

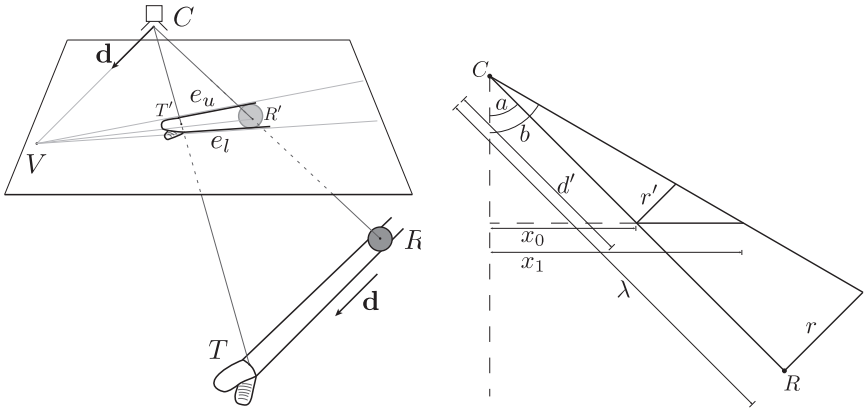
The linear search for T_G assumes that there is a single sharp peak in the gradient. However, this assumption is often violated by the presence of specular highlights along the instrument shaft. Noting that such highlight are typically clustered in the region of the shaft closer to the edge of the screen, we mitigate their effect by starting the search from the middle of the instrument axis as opposed to the beginning.

1.1.2. Automated Tuning

One parameter that greatly affects the robustness of our approach is the number of iterations for the erosion operator: too many steps remove the instrument regions completely, while too few leave additional noise in the initial binary mask computed in the first step of the algorithm. To address this problem, we consider the raw binary probability map of the pegs and instruments, use a heuristic to remove the instrument contours, and determine the minimum number of erosion steps required to remove all the noise. We repeat this approach for a window of frames to find the best value for the given video sequence.

1.2. Camera-Space Position (3D)

The key idea that allows locating the 3D position of the tool-tip from a single frame of video is recognizing that the vanishing point of the edges of the instrument's image provides the 3D direction of the instrument $\mathbf{d}$ [2]. That is, the vector from the camera (i.e., the origin of the camera frame using the pin-hole model) to the vanishing point is equal to the direction of the instrument itself. Figure 2(g) illustrates this property, with R representing the fixed point through which the instrument passes. Likewise, the diagram illustrates that the 3D position of the



(g) The geometry of the image formation of the instrument.

(h) Calculation of the depth λ of the trocar. The shown plane contains the major axis of the projected ellipse of the trocar ($\overline{x_0 x_1}$) and the camera.

Figure 2.

tool-tip is the intersection of two lines: the line that passes through the trocar R in the direction of the instrument $\mathbf{d}$, and the pre-image of the tool-tip (i.e., the line passing through both the camera point and the image of the tool-tip). This approach assumes that the camera-space position of the trocar R is known. Unfortunately, it is not possible to locate R from a single frame.

1.2.1. Edges' Vanishing Point and Direction of the Instrument

Once the images of the framing edges of the instrument (e_u, e_l) are found, the vanishing point is $V = e_u \times e_l$, assuming lines e_u, e_l and point V are in 2D homogeneous coordinates. Thus, all lines in the scene that are parallel to the direction of the instrument $\mathbf{d}$ will have images that pass through V. Now consider the line that passes through the camera point C and is parallel to $\mathbf{d}$: $C + t\mathbf{d}$. The image of this line must also pass through V, as V is the vanishing point for the direction $\mathbf{d}$. Therefore, the line $C\bar{V}$ is equivalent to $C + t\mathbf{d}$. Since the world frame is simply the camera frame, the direction of the instrument is simply $\mathbf{d} = \frac{V-0}{||V||}$.

1.2.2. Position of Tool-Tip

The tool-tip point T is the point on the instrument that corresponds to the distal end of the instrument (see figure 2(g)). The tool-tip is some unknown distance k from R in the direction of the tool, $T = R + s\mathbf{d}$. But note that T is also located on the pre-image of point T', i.e., on the line $C + t(T' - C) = t(T' - 0) = t\mathbf{b}$ with $\mathbf{b} \equiv (T' - 0)$. The procedure for locating T' in the image will be considered in the next section.

Ideally, T is simply at the intersection of lines $L_1(s) = R + s\mathbf{d}$ and $L_2(t) = t\mathbf{b}$, however such precision is unlikely. Instead, consider the points on each line closest to the other line, $L_1(\tilde{s})$ and $L_2(\tilde{t})$. The segment $\overline{L_1(\tilde{s})L_2(\tilde{t})}$ is uniquely perpendicular to both $L_1(s)$ and $L_2(t)$.

$$\tilde{s} = \frac{\langle \mathbf{d}, \mathbf{b} \rangle \langle \mathbf{b}, (R-0) \rangle - \langle \mathbf{b}, \mathbf{b} \rangle \langle \mathbf{d}, (R-0) \rangle}{\langle \mathbf{d}, \mathbf{d} \rangle \langle \mathbf{b}, \mathbf{b} \rangle - (\langle \mathbf{d}, \mathbf{b} \rangle)^2} \tag{1}$$

$$\tilde{t} = \frac{\langle \mathbf{d}, \mathbf{d} \rangle \langle \mathbf{b}, (R-0) \rangle - \langle \mathbf{d}, \mathbf{b} \rangle \langle \mathbf{d}, (R-0) \rangle}{\langle \mathbf{d}, \mathbf{d} \rangle \langle \mathbf{b}, \mathbf{b} \rangle - (\langle \mathbf{d}, \mathbf{b} \rangle)^2} \tag{2}$$

Taking the midpoint of $\overline{L_1(\tilde{s})L_2(\tilde{t})}$ as the estimate of T gives

$$T = \frac{(R + \tilde{s}\mathbf{d}) + (\tilde{t}\mathbf{b})}{2}. \tag{3}$$

1.2.3. Locating the Trocar from a Sequence of Images

So far we have assumed that the position of the trocar R (the center of the region of space through which all instruments pass) is known. To determine R , the framing edge pairs (e_u^i, e_l^i) for each frame i are collected. If there were no errors in the (e_u^i, e_l^i), the image of the trocar would be the locus of points on the image plane between the edges for all frames. Due to inevitable noise, the actual image of the trocar is smaller than the observed locus. To more robustly find the trocar's image, an image point R' is found as the point closest to all of the (e_u^i, e_l^i), that is, for $E = \{e_u^i, e_l^i\}, \forall i$ and $v^i \perp l^i$ for all $l \in E$,

$$R' = \arg\max_{p \in I} \sum_i \left| \text{proj}_v(l_0^i - p) \right|. \tag{4}$$

With the center of the image of the trocar R' determined, the ellipse centered at R' with one axis of $(R' - 0)$ that best matches the set of E is found. Define $x_0 \equiv ||R' - [w/2, h/2]^T||$ and $x_1 \equiv x_0 + m$, where w, h are the width and height of the image, and $2m$ is the length of the major axis of the ellipse. The geometry of the trocar's projection is shown in figure 1(a), in the plane containing the ellipse's major axis and the camera. Defining a, b, r', d' as in figure 1(a), the depth of R, λ is determined by

$$r' = d' \tan(b) \qquad b = \tan^{-1}(x_1) - \tan^{-1}(x_0)$$

$$d' = \frac{x_0}{\sin(a)} \qquad \lambda = d' \frac{r}{r'}.$$

With both λ and the image of the trocar R', the 3D position of the trocar is known.

2. Results

For our experiments, we captured several video sequences of the FLS peg transfer task with the standard camera included in the box trainer and a completely unaltered setup. The illumination is provided by an array of LED lights included in the box. One group of tasks was performed by an expert surgeon featuring controlled smooth motions, while a second group was performed by a novice and

is affected by jittery non-smooth motions. In both cases, we recorded robust 2D tracking of the instrument tips that were validated visually.

Figure 2(f) shows the tracked position (in yellow) of the two instrument tips from the unmodified FLS video, shown in figure 2(a). The accompanying video shows the performance of our tracker for a short clip with thumbnails of the intermediate steps. The measure of confidence of the tracked position allows us to automatically disable tracking of an instrument tip when it is no longer visible in the scene. The tracker is unable to track the position of the instrument tip accurately when the instrument is too close to the camera and thus very blurry. However, in such cases, the measure of confidence is very low, as expected.

3. Conclusion

In this paper we presented a complete system for tracking the 3D position of the instrument tips of a standard FLS box trainer. Our approach is robust, does not require any physical alteration of the toolset, and works with the standard camera included in the kit. In the future, we would like to combine our existing tracking capabilities with a more thorough analysis of the entire scene as a means to produce a more accurate assessment of FLS tasks.

References

[1] B. Allen, V. Nistor, E. Dutson, G. Carman, C. Lewis, and P. Faloutsos. Support vector machines improve the accuracy of evaluation for the performance of laparoscopic training tasks. *Surgical endoscopy*, 24(1):170–178, 2010.

[2] A. Cano, P. Lamata, F. Gayá, and E. Gómez. New Methods for Video-Based Tracking of Laparoscopic Tools. *Biomedical Simulation*, pages 142–149, 2006.

[3] 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. *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2006*, pages 527–534, 2006.

[4] A. Krupa, C. Doignon, J. Gangloff, and M. de Mathelin. Combined image-based and depth visual servoing applied to robotized laparoscopic surgery. In *Proc. of the 2002 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2002.

[5] VA Pandey, JHN Wolfe, SA Black, M. Cairols, CD Liapis, and D. Bergqvist. Self-assessment of technical skill in surgery: the need for expert feedback. *Annals of The Royal College of Surgeons of England*, 90(4):286, 2008.

[6] J. Peters, G.M. Fried, L.L. Swanstrom, N.J. Soper, L.F. Sillin, B. Schirmer, K. Hoffman, et al. Development and validation of a comprehensive program of education and assessment of the basic fundamentals of laparoscopic surgery. *Surgery*, 135(1):21–27, 2004.

[7] J. Rosen, J.D. Brown, L. Chang, M. Barreca, M. Sinanan, and B. Hannaford. The Blue Dragon-a system for measuring the kinematics and the dynamics of minimally invasive surgical tools in-vivo. In *Proceedings- IEEE International Conference on Robotics and Automation*, volume 2, pages 1876–1881. Citeseer, 2002.

[8] C.D. Smith, T.M. Farrell, S.S. McNatt, and R.E. Metreveli. Assessing laparoscopic manipulative skills. *The American Journal of Surgery*, 181(6):547–550, 2001.

[9] O. Tonet, R.U. Thoranaghatte, G. Megali, and P. Dario. Tracking endoscopic instruments without a localizer: A shape-analysis-based approach. *Computer Aided Surgery*, 12(1):35–42, 2007.

[10] 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*, 26(11-12):1173, 2007.

[11] JD Westwood, HM Hoffman, D. Stredney, and SJ Weghorst. Validation of virtual reality to teach and assess psychomotor skills in laparoscopic surgery: Results from randomised controlled studies using the MIST VR laparoscopic simulator. *Medicine Meets Virtual Reality: art, science, technology: healthcare and evolution*, page 124, 1998.