A New Framework for Vision-Enabled and Robotically-Assisted Minimally Invasive Surgery

Yuan-Fang Wang[†] Darrin R. Uecker[‡] Yulun Wang[‡]

† Department of Computer Science ‡ Computer Motion Inc.

University of California 130D Cremona Drive

Santa Barbara, CA 93106 Goleta, CA 93117

Abstract

This paper presents our on-going research at bringing the state-of-the-art in vision and robotics technologies to enhance the emerging minimally invasive surgery, in particular the laparoscopic surgical procedure (Figure 1). A framework that utilizes intelligent visual modeling, recognition, and servoing capabilities for assisting the surgeon in maneuvering the scope (camera) in laparoscopy is proposed.

The proposed framework integrates top-down model guidance, bottom-up image analysis, and surgeon-in-the-loop monitoring for added patient safety. For the top-down directives, high-level models are used to represent the abdominal anatomy and to encode choreographed scope movement sequences based on the surgeon's knowledge. For the bottom-up analysis, vision algorithms are designed for image analysis, modeling, and matching in a flexible, deformable environment (the abdominal cavity). For reconciling the top-down and bottom-up activities, robot servoing mechanisms are realized for executing choreographed scope movements with active vision guidance.

The proposed choreographed scope maneuvering concept facilitates the surgeon's control of his/her visual feedback in a handless manner, reduces the risk to the patient from inappropriate scope movements by an assistant, and allows the operation to be performed faster and with greater ease. In this paper, we describe the new framework and present some preliminary results on laparoscopic image analysis for segmentation and instrument localization, and on instrument tracking.

Keywords: laparoscopy, surgery, robotics, image analysis, tracking

1 Introduction

There has been a revolution in medical surgery in recent years toward "minimally invasive surgery" [1, 2]. In particular, laparoscopy (Figure 1.a), a type of minimally invasive surgery, has been widely used for gall bladder removal, hernia repair, and laparoscopically assisted hysterectomy [1, 2]. In laparoscopy, several small incisions are made on the patient to accommodate surgical instruments such as scalpels, scissors, and staple guns. The surgeon's visual feedback is provided by a video scope inserted through the patient's navel. The scope acquires video images of the bodily cavity which are displayed in real time on a monitor. This setup enables the surgeon to operate instruments through the small incisions, as opposed to a large incision for direct viewing.

Laparoscopic procedures reduce the trauma inflicted on the patient during surgery, significantly shorten the time for the patient to recuperate, and can lower the cost of the treatment. Because of the tremendous benefit gained over the traditional surgical procedures, it is fast gaining popularity.

Though laparoscopic surgery has proven to be beneficial to the patient, this patient-oriented technology nonetheless has increased the difficulty of performing the procedures for the surgeon. One main reason for the increased difficulty is that the surgeon's visual feedback is suboptimal because of poor scope (camera) positioning. The current mode of laparoscopic surgery is that an assistant holds and positions the scope in response to the verbal directions from the surgeon (Figure 1.a). The method of operation is inefficient and frustrating for the surgeon because the commands are often interpreted and executed imprecisely or incorrectly by the assistant. Furthermore, as laparoscopic images are highly magnified, slight hand trembling induces annoying jitter in the video display. Consequently, a waste of man power and a high risk to the patient result.

To improve the current mode of laparoscopic surgery, many mechanical scope positioning systems have been proposed [3, 4, 5, 6, 7]. The general idea is to have a robot holding the scope and responding to the positioning commands issued by the surgeon through a
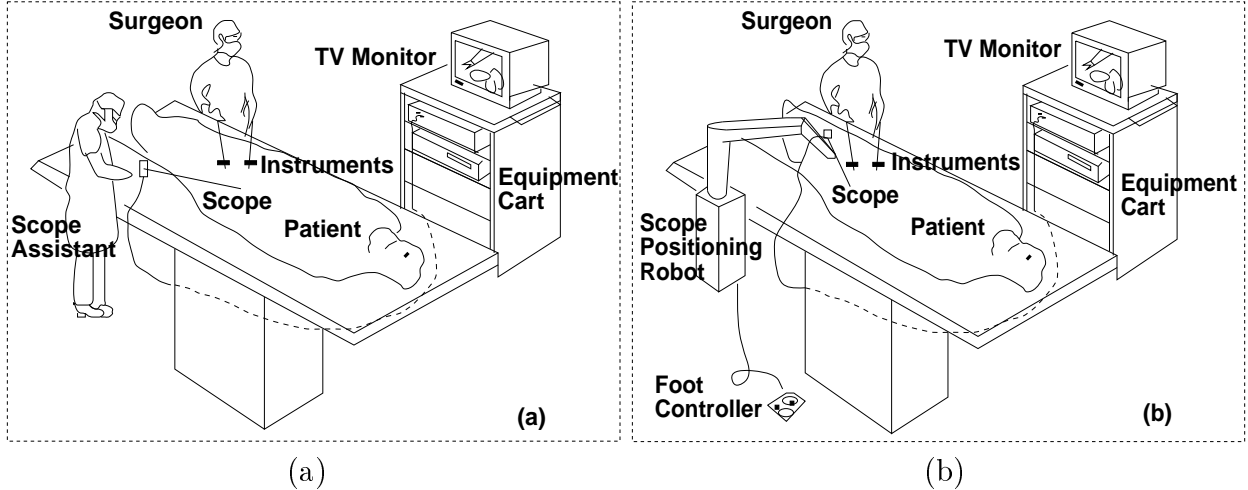
Figure 1: (a) Traditional laparoscopy performed by a surgeon and a scope assistant, and (b) robotically-assisted laparoscopy where a robot replaces the scope assistant.

hand-held controller, a foot pedal, or other interface mechanisms such as a speech interface (for example, see Figure 1.b). This mode of operation improves the visual feedback to the surgeon by giving the surgeon direct control of his/her visual feedback and eliminating the assistant from the loop. The procedure can thus be performed faster and with greater ease.

However, given the surgeon direct visual control has the undesired side effect that the surgeon is constantly being distracted to maneuver the scope. Often times, a seasoned assistant can anticipate the surgeon's viewing need to position the scope without the surgeon's intervention. This is especially true during the procedures (e.g., suturing) where the scope aiming and movements are repetitive and follow a fixed pattern. (E.g., for suturing, zooming in when the surgeon is tying a knot and zooming out when the surgeon is pulling on the suture.) Current mechanical positioners rely completely on the surgeon's interactive commands and lack the intelligence to automate such exercises.

This paper presents a framework to address this "intelligence gap" between a robotic and a human assistant. *The main objective is to develop "choreographed" scope maneuvering capability in laparoscopy with active vision guidance. In particular, a framework utilizing intelligent visual modeling, recognition, and servoing capabilities for assisting the surgeon in maneuvering the scope (camera) in laparoscopy is proposed.* We argue that for procedures in laparoscopy where the surgeon's viewing need is well understood and can be categorized,

the scope movements could best be choreographed in advance and then be "called-back" and executed automatically, with real-time vision guidance and monitoring by the surgeon. And mechanical devices are ideally suited for such operations which follow a fixed pattern, and are repetitive and learnable. *We believe that this approach combines the best of both worlds in providing the surgeon with a directly-controlled and stable visual feedback (through a mechanical positioning device), and on-demand choreographed scope movements (through the emulation of an experienced scope assistant).* With over one million laparoscopic procedures performed each year in the U.S., improvement in the scope positioning with the proposed system will lead to increased patient safety and decreased operating time, with a potential significant cost saving.

In this paper, we describe the new framework and present some preliminary results on laparoscopic image analysis for segmentation and instrument localization, and on instrument tracking. Furthermore, the instrument localization and tracking algorithm we developed [8, 9, 10, 11] has been subject to rigorous test in the operating room environment. It "has been very effective and has not induced errors in technical performance of procedures, and has shortened the time required for specific procedural tasks" [12, 13].

The reminder of this paper is organized as follows: Section 2 provides an overview of the proposed framework. Section 3 illustrates a particular application of the framework in localizing and tracking instruments in laparoscopic images. Section 4 provides some preliminary results, and Section 5 contains concluding remarks.

2 The Framework

Two principles underlie the design of the proposed intelligent scope maneuvering system: *hierarchical task decomposition for modular design and construction, and human-in-the-loop servoing control for added safety.* The system architecture is sketched in Figure 2.a. We envision the system will comprise many functional modules organized roughly in four hierarchical layers with predefined communication and interaction patterns: sensing & modeling, integration & coordination, guidance & control, and supervising & planning. We will de-
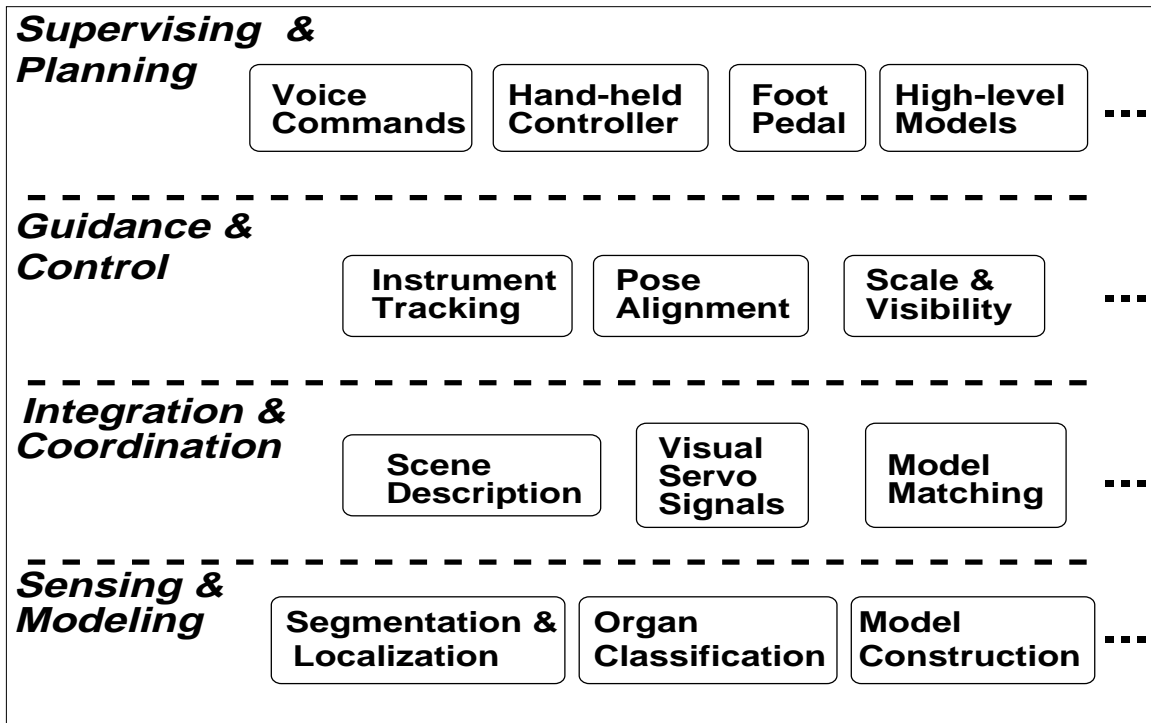
scribe the functionality of each layer in more details below, followed by an example (Figure 2.b) of how such a system can accomplish the scope maneuvering during the initial insertion of a trocar/cannula [1, 2].

Sensing & Modeling   This lowest layer is composed of functional modules for processing the visual information from the scope for recognition and modeling, and for scope (camera) motion control. Major functionalities provided will be:
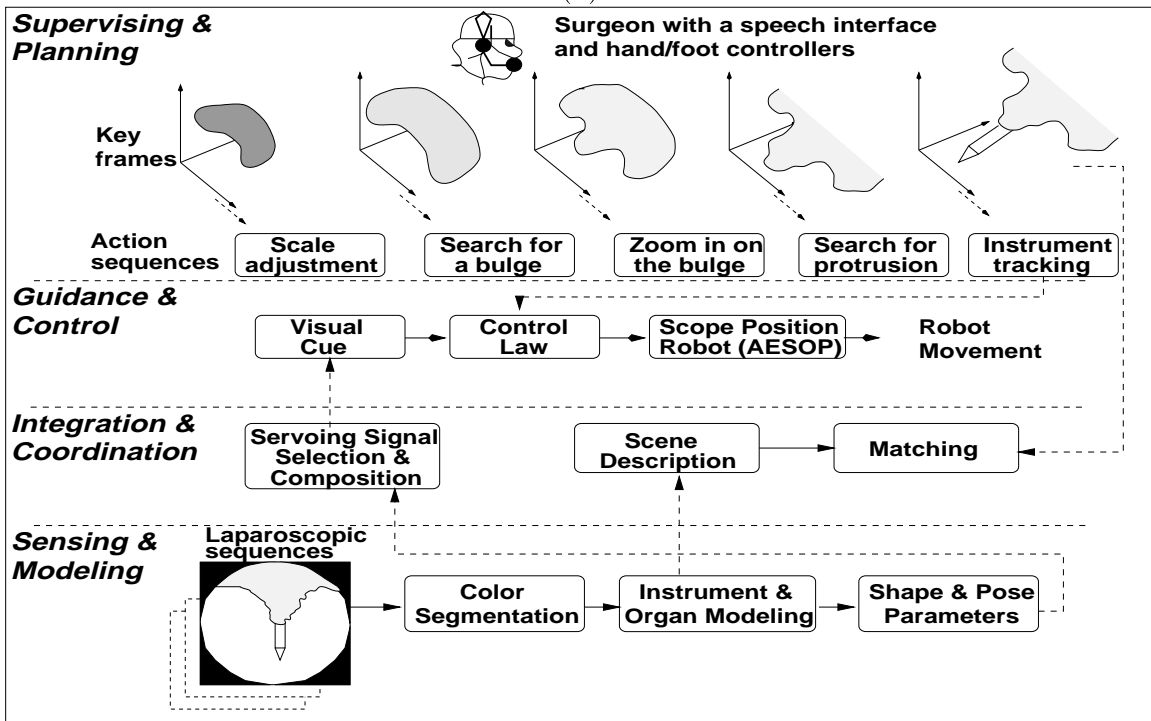
• *Segmentation and localization:* for extracting instruments, organs, and other anatomical landmarks from the laparoscopic images, using color, shape, and texture information.

• *Shape modeling:* for describing the shapes, poses, and dynamics of various instruments, organs, and anatomical landmarks. Domain knowledge will be heavily relied upon here. For example, the shaft of an instrument must be of a cylindrical shape to pass through the cannula opening on the abdominal wall. Hence, an instrument shaft appears as a rectangle or a trapezoid in images. Of particular importance is to portray the shape and deformation of the flexible abdominal anatomy. The global shapes of various organs and anatomical landmarks can be modeled as hierarchical spline patches (e.g., the abdominal wall), generalized cylinders (e.g., intestine and appendix), and superellipsoids (e.g., spleen, liver, gall bladder, etc.), with possible local shape deformation.

• *Scope servoing:* for extracting the shape, size, and pose parameters from individual organ and instrument models for assembling the robot control signals at the higher layers.

Integration & Coordination   This layer is responsible for (1) integrating visual cues over both the spatial and temporal domains into a scene description, (2) organizing visual cues in a suitable form for computing the robot control signals, and (3) for choreographed scope motions, correlating the scene description with the high-level scene models to determine the correct time stamps and action sequences.

Guidance & Control   This layer bridges the top-down directives and the bottom-up image processing activities. It is responsible for interpreting the directives from the supervising

(a)

(b)

Figure 2: (a) The proposed scope maneuvering system architecture, and (b) the architecture as applied to the choreographed sequence of the insertion of a main trocar/cannula.

surgeon and high-level scene models for properly utilizing and reconciling the sensor information from the lower layers to generate suitable scope movement sequences. For example, the surgeon might issue a command to follow a particular instrument (e.g., the one currently being used). This layer then employs the proper control law and utilizes the sensor feedback to keep the instrument centered. Another example is that the high-level models might initiate a choreographed scope movement sequence for the dissecting operation. Then this layer is responsible for directing the lower layers to locate a grabbing instrument and a cutting instrument (e.g., a scalpel), and invoking a proper control law for maintaining the instruments' relative positions to the organ in between.

Supervising & Planning   This topmost layer represents the human-in-the-loop monitoring activities, and choreographed activity planning based on the high-level scene models. A high-level model comprises a visual component with key frames and a knowledge-based component with action sequence annotations (e.g., see Figure 2.b). It is responsible for generating the choreographed sequences, with low-level modules providing the needed "trigger" information in terms of time stamps and scene descriptions. The surgeon can always issue commands, say, through a speech interface, to override the directives from the high-level models. This human-in-the-loop supervisor mode is essential for the safety of the patient.

An Example of Initial Trocar/Cannula Insertion   We will now illustrate how such a system can be used in positioning the scope during the initial insertion of a main trocar/cannula. A trocar has a sharp pointed conical end for penetrating the abdominal parietes (Figure 3). In laparoscopy, the optimal site for insertion is the immediate sub-umbilical region. Typically, three to six such openings are made [1, 2], and they can then be used to accommodate other instruments. It is most important that during the trocar insertion, the surgeon is monitoring the punctuation site closely to avoid accidental damage to the internal organs.

Referring to Figure 2.b, the surgeon aims the scope to view the vicinity of the trocar punctuation point and issues a voice command to the effect of "initiating the choreographed

7

trocar/cannula insertion sequence." The high-level model then takes over the control of aiming the scope. The visual component of the model may comprise key frames of the undisturbed abdominal wall, the strained abdominal wall from the initial trocar penetration, and the abdominal wall with a trocar present. Attached to these snapshots are directives for scale adjustment, zooming the camera onto the bulge on the abdominal wall, and tracking the trocar movement, respectively (Figure 2.b).

Using the pre-planned choreographed sequences, the high-level model directs the lower layers to (1) analyze images of the abdominal wall to construct a model using hierarchical spline surfaces (the sensing & modeling layer[1]), (2) determine a proper scale (distance to the abdominal wall) by zooming the scope to cover an adequate viewing area (both the guidance & control and integration & coordination layers), (3) initiate a search for a bulge on the abdominal wall which signals the initial penetration of a trocar/cannula (both the integration & coordination and sensing & modeling layers), (4) if such a bulge is detected, maneuver the scope to zoom in onto the bulge (the guidance & control layer), (5) start searching for a metal protrusion along the length of the bulge (the sensing & modeling and integration & coordination layers), and (6) extend the view volume to include the trocar penetration when a trocar presence is detected (all three lower layers).

3 Instrument Localization and Visual Servoing

We have applied the above framework for a particular task in laparoscopy, that of instrument localization and tracking. This capability is important in laparoscopy because for patient's safety the surgeon's view must always include the operating instrument. It is cumbersome and distracting for the surgeon to use, say, a foot pedal to control the robot to re-aim the scope all the times. Hence, a semi-automated, on-demand instrument following mechanism is very useful. Furthermore, it enables the surgeon to guide the camera by repositioning an instrument (i.e., using the instrument as a pointer). It should be noted that this ability is deployed on an "on-demand" basis and can be disengaged at any time by the surgeon's

---

[1]Only layers with major actions during the particular subsequence are noted in this example.
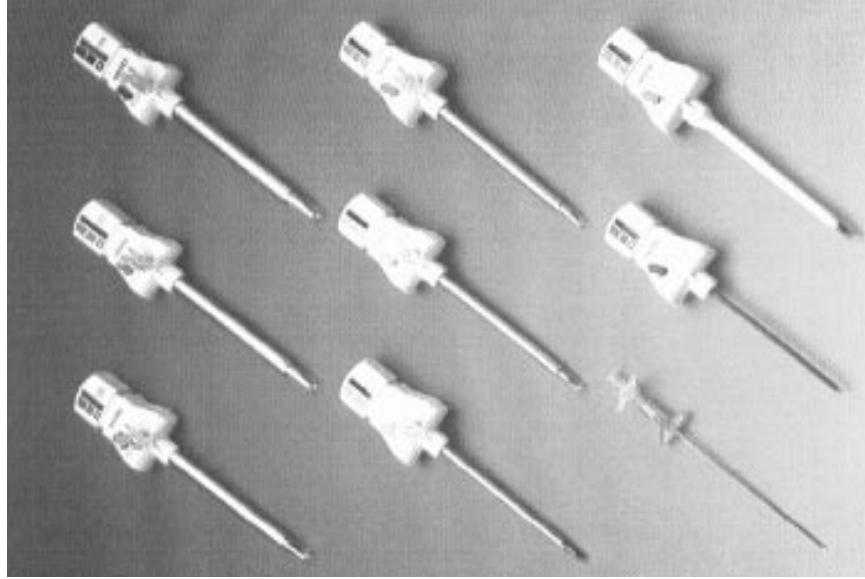
Figure 3: Disposable trocar/cannula of different sizes.

verbal command.

For the camera to track the motion of an instrument, the instrument should first be localized and followed in the laparoscopic sequence. Whenever the image location of the instrument deviates from the reference location (e.g., the center of the image), a robot control signal is generated to steer the scope to compensate for the instrument motion automatically.

For instrument localization and tracking in laparoscopic sequences, we have digitized a large number of image sequences taken during real laparoscopic surgery to serve as our test data. we have obtained a large sample of color signatures from over two hundred thousand instrument and organ pixels in these sequences, and have computed the color statistics of various instruments and organs. These color statistics have allowed us to classify, group, and label instrument pixels.

More specifically, In our case there are two classes: organ (ω_1) and instrument (ω_2). We estimated the *a priori* probabilities $p(\omega_1)$ as 0.7, and $p(\omega_2)$ as 0.3. This corresponds to the typical scenarios where the surgeon operates two instruments simultaneously and each instrument occupies roughly 15% of the image space. Each sample (i.e., a pixel) provides a 3-dimensional measurement vector $\mathbf{x} = (R, G, B)$, where the components represent the red, green, and blue color intensities. Furthermore, we assumed $p(\cdot \mid \omega_i), i = 1, 2$

9

to be a multi-variate normal distribution with mean μ_i and variance Σ_i, and estimated $\mu_i = \sum_{j=1}^{N_i} \mathbf{x}_{ij}/N_i, \Sigma_i = \sum_{j=1}^{N_i} (\mathbf{x}_{ij} - \mu_i)(\mathbf{x}_{ij} - \mu_i)^T/(N_i - 1)$, where $\mathbf{x}_{ij}$ represents the measurement vector of samples in class i, and N_i the sample size of class i, $i = 1, 2$. Then using a simple Bayesian classifier [14], a pixel, with a measurement vector $\mathbf{x}$, is assigned to class i if $log\, p_i(\mathbf{x}) > log\, p_j(\mathbf{x})$, where

$$log\, p_i(\mathbf{x}) = -\frac{1}{2}(\mathbf{x} - \mu_i)^T \Sigma_i^{-1}(\mathbf{x} - \mu_i) - \frac{3}{2}log\, 2\pi - \frac{1}{2}log \mid \Sigma_i \mid + log\, p(\omega_i)\ .$$

Furthermore, we have computed the shape and motion parameters of each instrument region detected in an image sequence to help instrument recognition and motion prediction. We compute the 0th, 1st, and 2nd-order moments for each instrument region. If the ratio of M_{20} to M_{02} exceeds a certain threshold—which indicates an elongated shape—the region is classified as an instrument region, where M_{20} and M_{02} denote the second-order moments computed in the object-centered coordinate system [15]. We then compute a bounding box for an instrument region either as a rectangle (far-field case) or a trapezoid (near-field case). The object motion is tracked over a laparoscopic sequence by following the motion of the bounding box.

The servoing algorithm for instrument tracking is depicted in Figure 4.a. The specific sensing & modeling layer's function is for segmenting, grouping, labeling, and tracking instrument regions in images, as described in the preceding paragraphs. The integration & coordination layer is for isolating the desired instrument and computing its tip position (x, y). The guidance & control layer compares the instrument tip's current position (x, y) against a canonical, reference location (x_d, y_d), (e.g., the center of the image). $(\delta x, \delta y)$ is the error signal which is used to compute the robot control signal $(\delta\theta, \delta\varphi, \delta\rho)$. The gain in this algorithm is used for robustness.

Note that the physical constraint imposed on the scope by the abdomen entry point allows only three degrees-of-freedom, (θ, ϕ, ρ), for manipulating the camera (Figure 4.b): zooming in/out is a change in ρ, panning left/right is a change in θ, and panning up/down is a change in ϕ. The Jacobian matrix which relates $(\delta x, \delta y)$ to $(\delta\theta, \delta\varphi, \delta\rho)$ can be shown to

be [8, 9, 10]

$$\mathbf{J} = \begin{bmatrix} -xy\sin\phi + y\cos\phi & -\frac{\rho}{Z} - (1 + x^2) & \frac{x}{Z} \\ -x\cos\phi - \sin\phi(1 + y^2) - \frac{\rho\sin\phi}{Z} & -xy & \frac{y}{Z} \end{bmatrix} . \qquad (1)$$

Experimental results are shown in the next section.

4   Experimental Result

The development platform is a mockup OR with an AESOP scope positioning robot [7] (Figure 5), several laparoscopic instruments, a video scope, and a flexible mannequin torso to emulate the human abdomen. The tracking action is initiated by a simple voice command ("AESOP track") from the supervising & planning layer. The sensing & modeling layer then performs segmentation, modeling, and tracking of instruments in the laparoscopic images. The integration & coordination layer filters the inputs from the sensing layer to select the instrument for tracking (using temporal correlation). The particular instrument's position and size, as reported by the integration & coordination layer, are used to form the input vector at the guidance & control layer where a suitable control law is employed for maneuvering the scope (Equation 1). When the tracked instrument's position and/or shape deviate from the desired values (e.g., the instrument is too far from the center of the image or becomes too small), an error signal is generated. The guidance & control layer uses the error signal to compute and direct a robot movement that compensates for the deviation automatically.

Two typical image analysis results, using video-taped sequences of real laparoscopic surgery, are shown in Fig. 6. The left column displays stages in instrument tracking in a sequence with two near-field instruments, while the right column with a single instrument in the far field. Fig. 6.a shows the original images. Fig. 6.b depicts the results of color classification, where each pixel was classified as either an organ (white) or an instrument (black) pixel using a pre-trained Bayesian classifier [14]. Median filters were then used to remove spurious noise points. The results are shown in Fig. 6.c.

Each localized instrument region then inherited a unique identifier (represented by different gray levels in Fig. 6.d). Shape parameters were estimated. The instrument model is
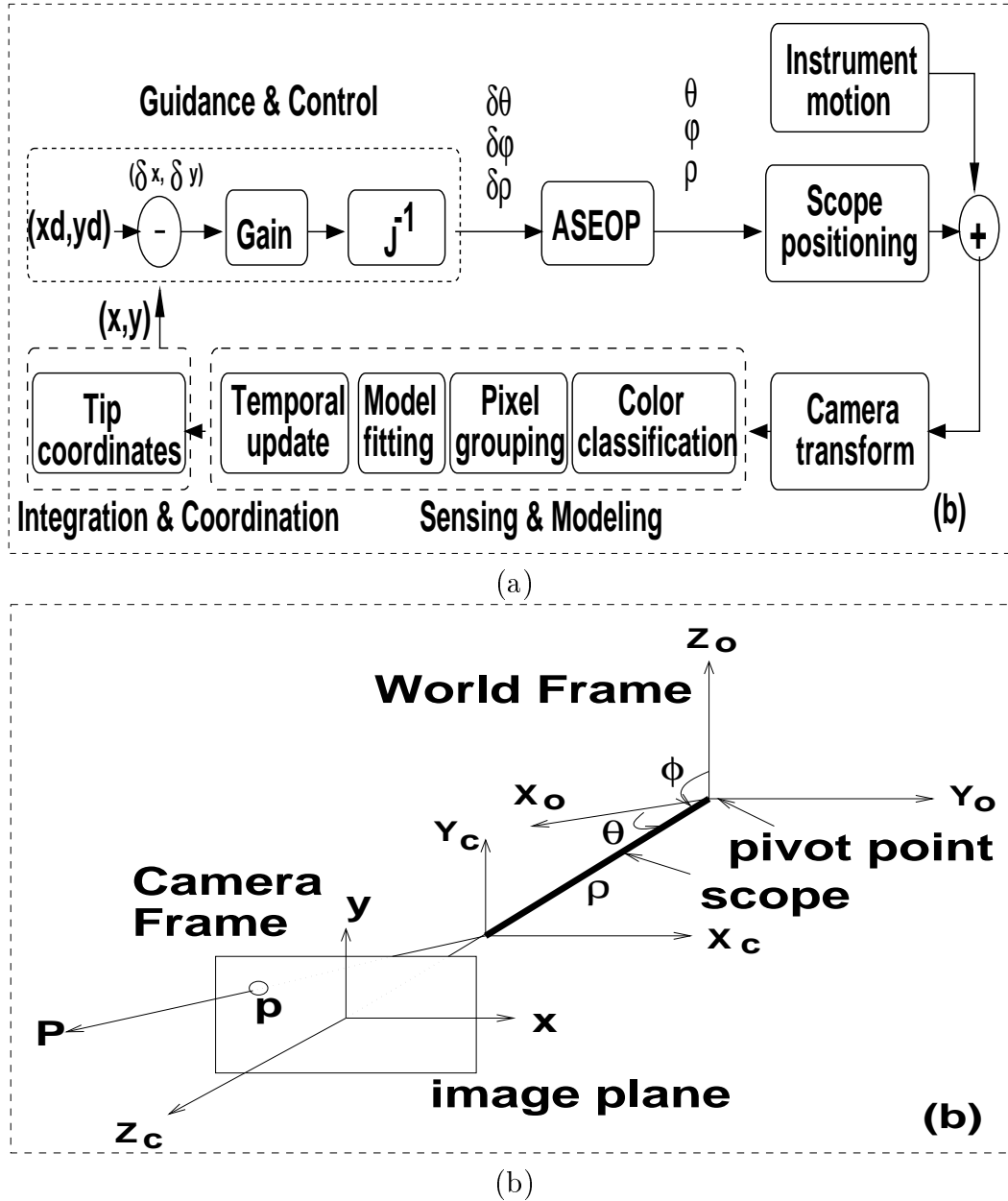
Figure 4: (a) The block diagram for instrument tracking and (b) the mathematical coordinate system used in instrument tracking.

Figure 5: The AESOP experimental platform.

depicted graphically as a bounding box that encloses an instrument region. The left column of Fig. 6.e shows the bounding boxes where a trapezoid was used for instruments close to the camera. The right column of Fig. 6.e shows the bounding box in the far field case where a rectangular shape was assumed. The bounding boxes' locations over time were tracked to propagate the instrument labels. Fig. 7 shows 9 sample images in a typical sequence with the instrument bounding boxes superimposed.

The ability to center a moving instrument is shown in Figure 8. Figure 8.a depicts the path of the instrument during tracking. Figure 8.b shows the deviation of the tip position from the image center (100,100) which gradually converged to zero. These figures clearly show the ability of the algorithm to track an instrument in motion.

5   Concluding Remarks

We believe that the proposed concept of choreographed scope maneuvering with vision and robotic guidance offers numerous advantages, and the potential payoffs can be quite substantial. The revolution toward minimally invasive surgery is gathering momentum and the number of laparoscopic procedures performed will increase unabated for well into the next
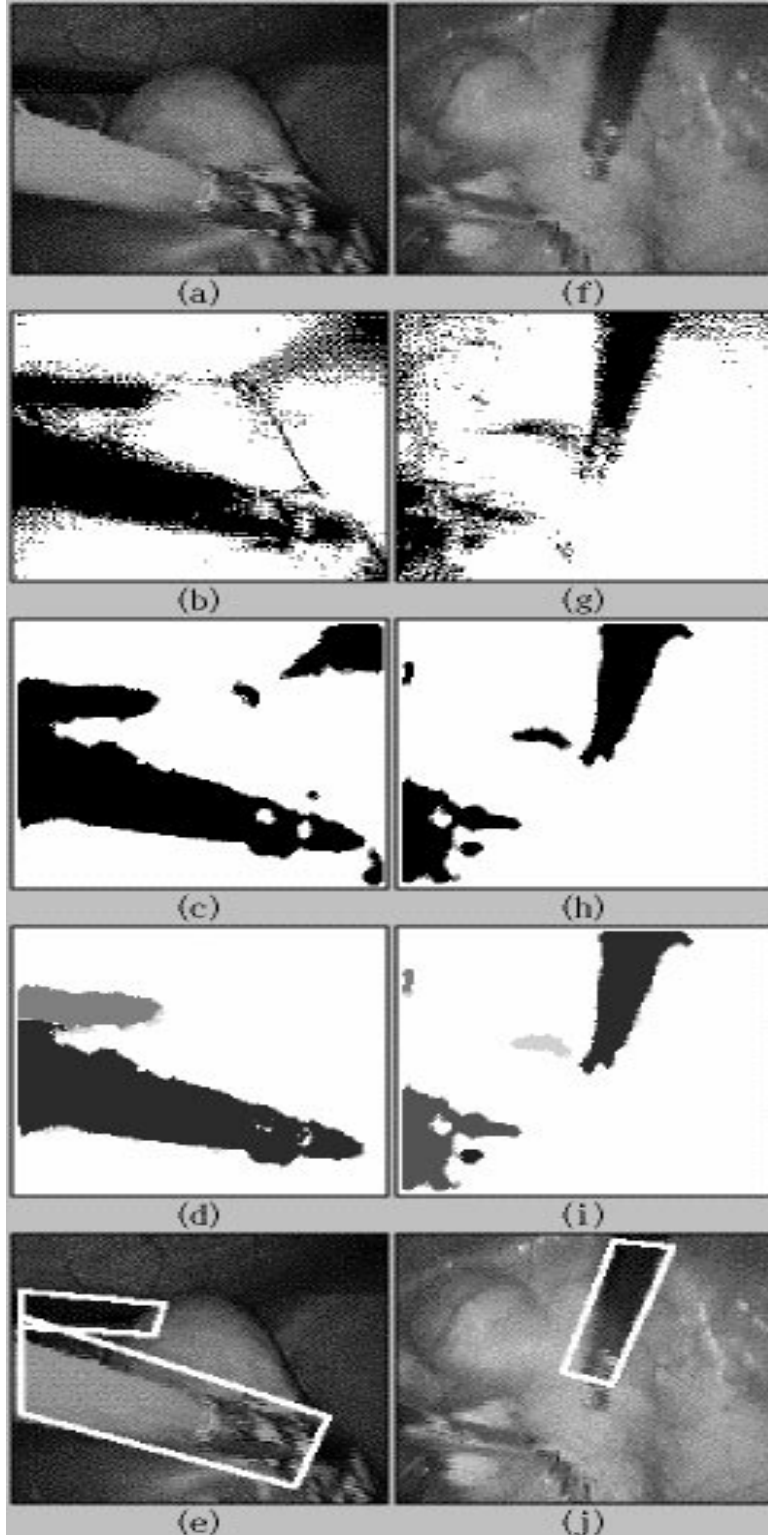
Figure 6: Left column: near field case, right column: far field case. **(a)** Original image. **(b)** Binary image from color classification. Black and white pixels represent surgical instruments and organs, respectively. **(c)** Directional median filters were used to suppress noise. **(d)** Labeled image, different gray levels represent isolated regions. **(e)** Computed bounding boxes.
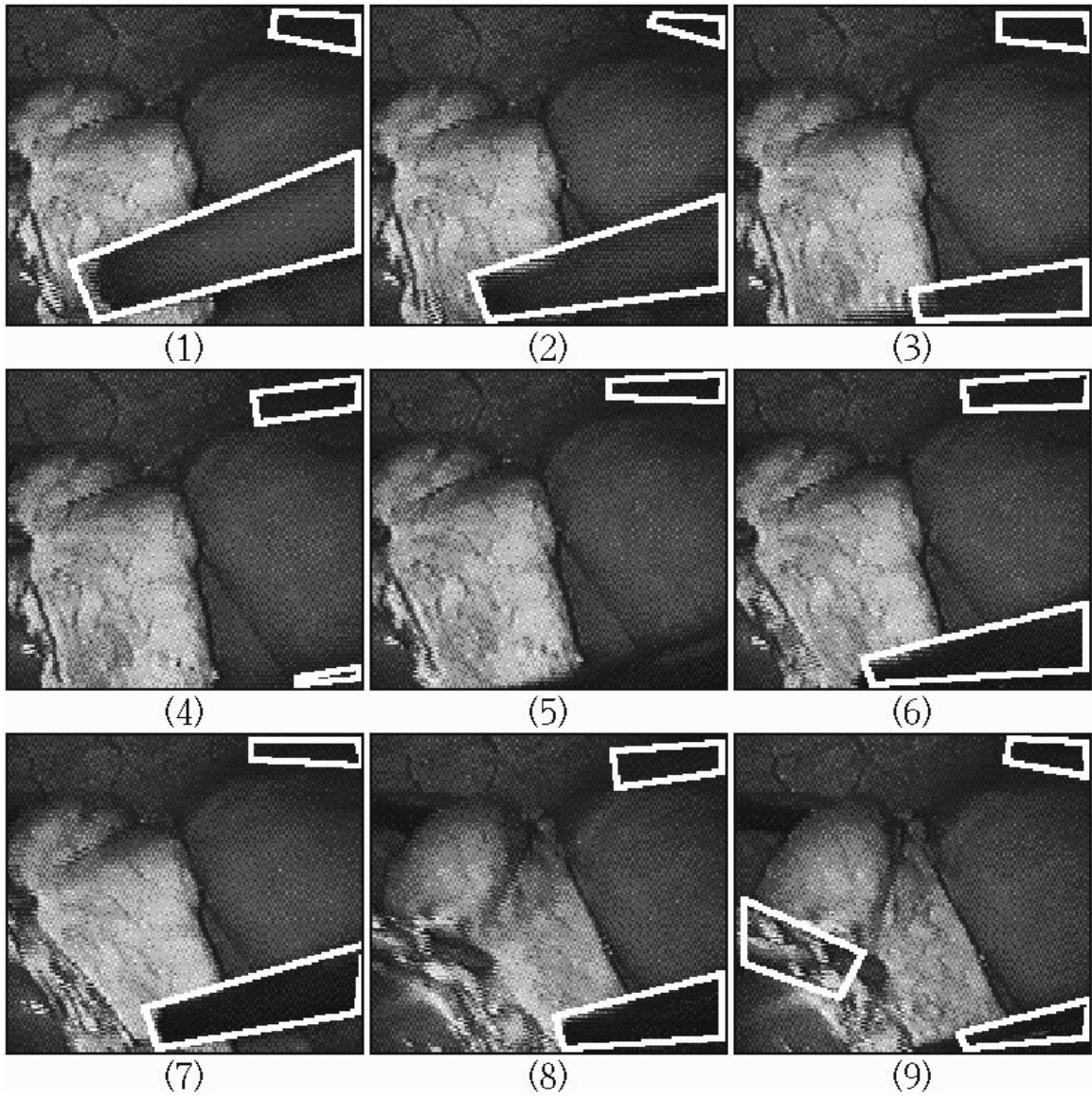
14

Figure 7: A sequence of 9 images showing the movements of instrument bounding boxes.
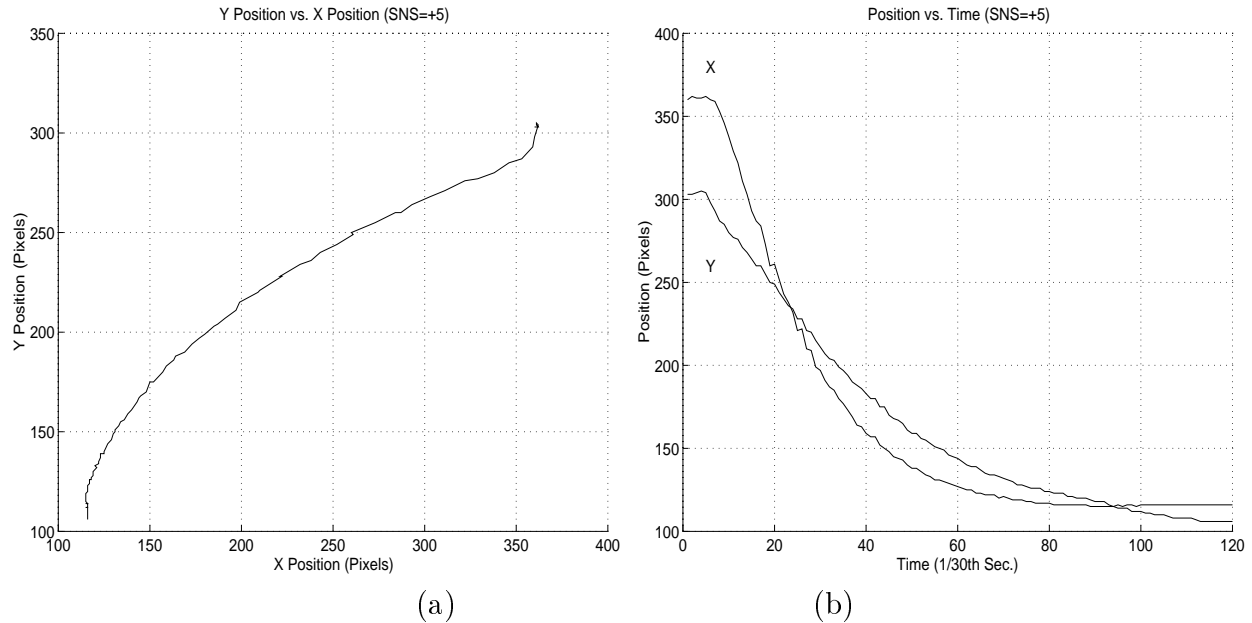
Figure 8: **(a)** Path of the instrument being tracked in the image plane. **(b)** Error in feature location vs. time.

century. Furthermore, we predict the onset and future expansion of the robotically-enhanced surgical technologies will drastically increase the sophistication of laparoscopic surgery, to the point that computing assistance becomes indispensable. Hence, we feel that the research is both timely and highly relevant.

The proposed choreographed scope maneuvering concept facilitates the surgeon's control of the visual feedback in a handless manner, reduces the risk to the patient from inappropriate scope movements by an assistant, and allows the operation to be performed faster and with greater ease.

Cost savings by adopting such a technology can be tremendous. With a typical operating room charge of $25-$30 per minute any improvements that save time also save money. It has been estimated that by employing a simple foot-controlled mechanical scope positioning device, eliminating some or all of the assistant, scrub nurse, and scope assistant in laparoscopic operations, and accounting for time saved in the operating room, savings of (conservatively) $100 per procedure can be achieved (by shedding just a few minutes off an operation). Employing sophisticated on-demand choreographed scope maneuvering to further improve the

visual feedback to the surgeon, even greater savings are possible. *With over one million laparoscopic surgeries performed each year in the U.S., this translates into an annual saving nationwide in hundreds of millions.*

In this paper we describe the framework and provide results of a particular implementation of the framework on instrument localization and tracking. Our future work will be to extend the framework for more intelligent scope maneuvering assitance.

6   BIBLIOGRAPHY

References

[1] J. F. Hulka and H. Reich. *Textbook of Laparoscopy, 2nd Ed.* W. B. Saunders, Philadelphia, PA, 1994.

[2] J. G. Hunter and J. M. Sackier (eds.). *Minimally Invasive Surgery.* McGraw-Hill, New York, 1993.

[3] Armstrong Company Literature.

[4] R. Hurteau, S. DeSantis, E. Begin, and M. Gagner. Laparoscopic Surgery Assisted by a Robotic Cameraman: Concept and Experimental Resutlts. In *Proc. Int. Conf. Robot. and Automat.*, pages 2286–2289, San Diego, CA, Jan. 1994.

[5] J Petelin and W. L. Chernoff. Computer Assisted Instrument Control. In *Proc. Symp. Medicine Meets Virtual Reality II*, San Diego, CA, Jan. 1994.

[6] R. Taylor, J. Funda, B. Eldridge, K. Gruben, D. Larose, and S. Gomory. Image-guided Command and Control of a Surgical Robot. In *Proc. Symp. Medicine Meets Virtual Reality II*, San Diego, CA, Jan. 1994.

[7] Y. Wang. AESOP: Automated Endoscope for Optimal Positioning. Technical Report 2, Computer Motion Inc., 1993.

[8] C. Lee, D. R. Uecker, Y. F. Wang, and Yulun Wang. Image Analysis for Automated Tracking in Robot-Assisted Endoscopic Surgery. In *Proceedings of International Conference on Pattern Recognition*, pages 88–92, Jerusalem, Israel, Oct. 1994.

[9] D. R. Uecker, C. Lee, Y. F. Wang, and Yulun Wang. A Speech-Directed Multi-Modal Man-Machine Interface for Robotically Enhanced Surgery. In *Proc. 1st Int. Symp. Medical Robot. and Comput. Assisted Surgery*, pages 176–183, Pittsburgh, PA, Sep. 1994.

[10] D. R. Uecker, C. Lee, Y. F. Wang, and Yulun Wang. Automated Instrument Tracking in Robotically-Assisted Laparoscopic Surgery. *J. of Image Guided Surgery*, 1(6), Dec. 1996.

[11] Y. F. Wang, D. Uecker, and Yulun Wang. Choreographed Scope Maneuvering in Robotically-Assisted Laparoscopy with Active Vision Guidance. In *IEEE Workshop on Applications of Computer Vision*, pages 187–192, Saratoga, FL, Dec. 1996.

[12] W. P. Geis, H. C. Kim, E. J Brennan Jr., and P. C. McAfee. Robotic Arm Enhancement to Accommodate Improved Efficiency and Decreased Resource Utilitzation in Complex Minimally Invasive Surgical Procedures. In *Proc. Symp. Medicine Meets Virtual Reality IV*, Jan. 1996.

[13] L. R. Kavoussi, R. G. Moore, , J. B. Adams, and A. W. Partin. Urologists at Work: Comparison of Robotic vs. Human Laparoscopic Camera Control. *J. Urology*, 154:2134–2136, Dec. 1995.

[14] R. O. Duda and P. E. Hart. *Pattern Classification and Scene Analysis*. Wiley, New York, 1973.

[15] B. K. P. Horn. *Robot Vision*. The MIT Press, Cambridge, MA, 1986.