Choreographed Scope Maneuvering in Robotically-Assisted Laparoscopy with Active Vision Guidance

Yuan-Fang Wang†
† Department of Computer Science
University of California
Santa Barbara, CA 93106

Darrin R. Uecker‡ Yulun Wang‡
‡ Computer Motion Inc.
130D Cremona Drive
Goleta, CA 93117

Abstract

This paper presents our research at bringing the state-of-the-art in vision and robotics technologies to enhance the emerging laparoscopic surgical procedure (Figure 1). 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.

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.

1 Introduction

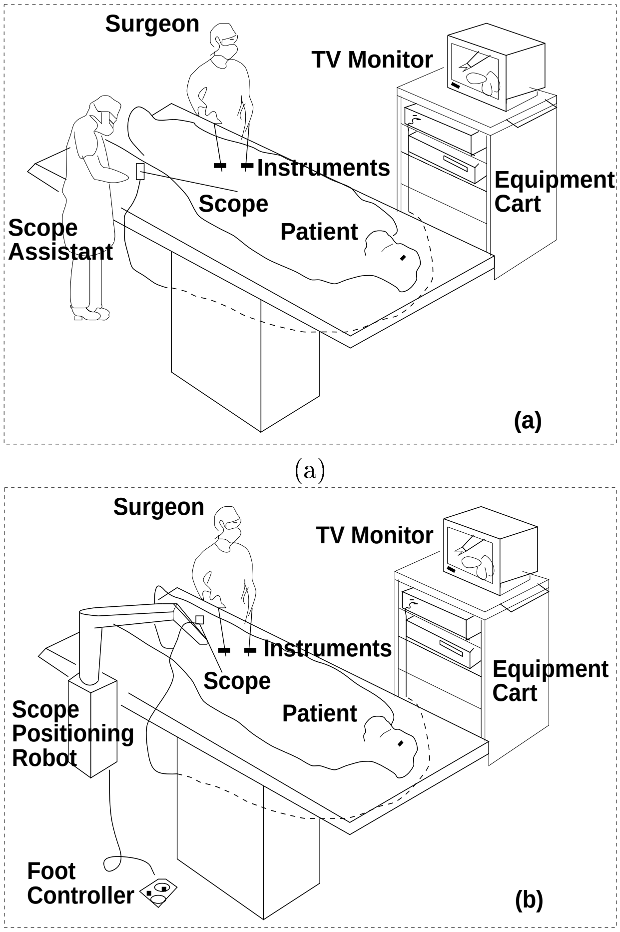
There has been a revolution in medical surgery in recent years toward "minimally invasive surgery" [2, 3]. In particular, laparoscopy (Figure 1.a), a type of minimally invasive surgery, has been widely used for gall bladder removal, hernia repair, and laparoscopically assisted hysterectomy [2, 3]. In 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, this patient-oriented technology 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 are proposed [1, 4, 6, 7, 10]. The general idea is to have a robot holding the scope and responding to the positioning commands issued by the surgeon through a 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.



(a)

(b)

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.

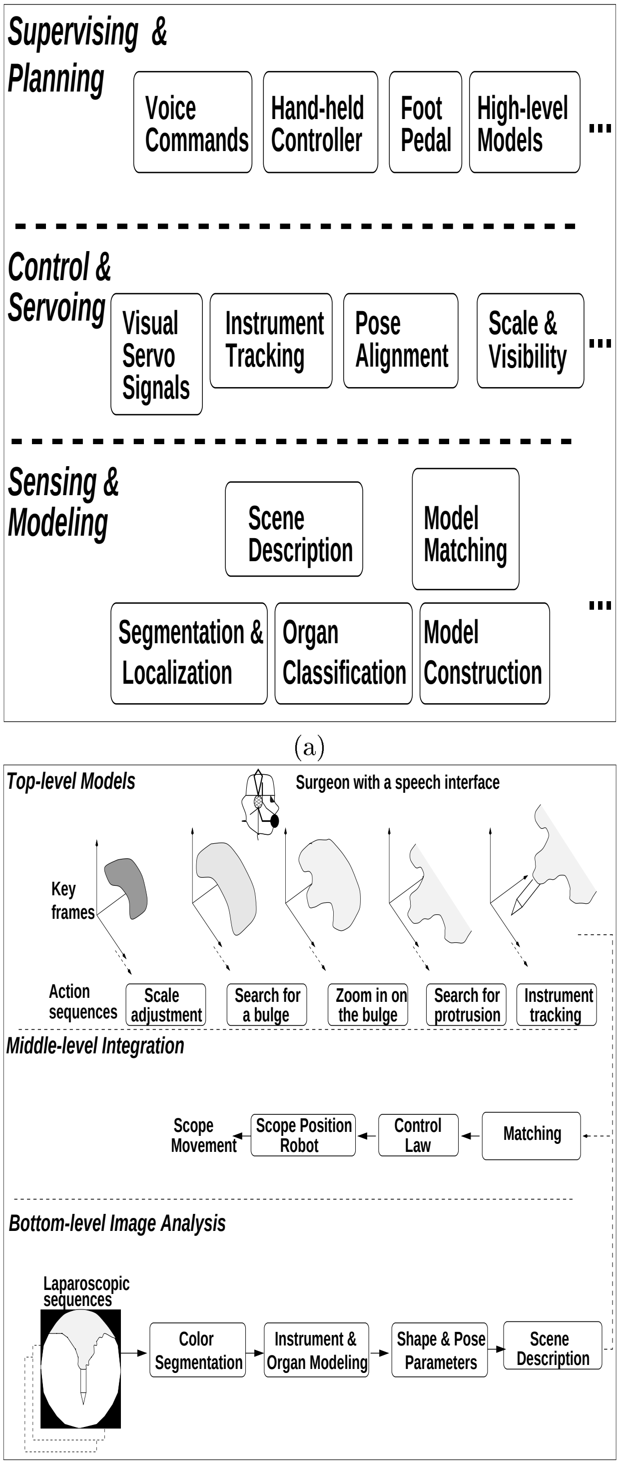
(E.g., for suturing, zooming in when the surgeon is tying a knot and zooming out when the surgeon is pulling on the suture.) Current mechanical positioners rely completely on the surgeon's interactive commands and lack the intelligence to automate such exercises.

The 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 cost saving in hundreds of millions.

2 Approaches

Two principles underlie the design of the proposed scope positioning 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 describe 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 [2, 3].

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 will be modeled as hierarchical spline patches (e.g., the abdominal wall), gen-



(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.

eralized cylinders (e.g., intestine and appendix), and superellipsoids (e.g., spleen, liver, gall bladder, etc.), with possible local shape deformation.

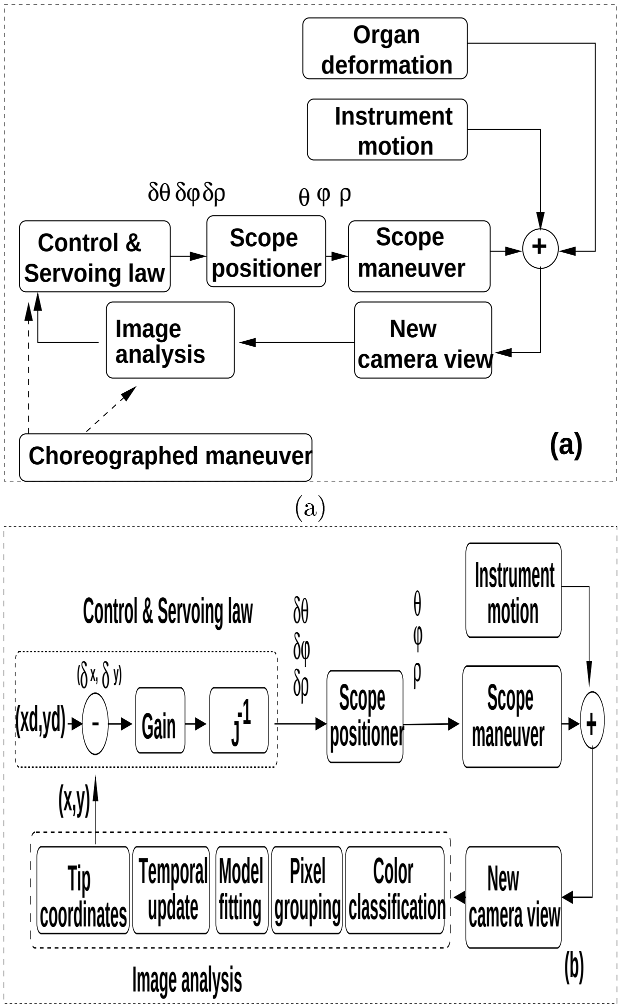
- *Scope servoing:* for extracting shape, size, and pose parameters from individual organ and instrument models for assembling the robot control signals at the higher layers.

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 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.

We implement this visual servoing function as the control loop depicted in Figure 3.a. As depicted in the figure, under the guidance of the supervising & planning layer, the abdominal scene is analyzed to extract successively more abstract and concise visual information for generating the servoing signals.

As a concrete example, the servoing algorithm for instrument tracking is depicted in Figure 3.b. The specific sensing & modeling layer's function is for segmenting, grouping, labeling, and tracking instrument regions in images. 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 $(\delta\theta, \delta\varphi, \delta\rho)$. Note that the physical constraint imposed on the scope by the abdomen entry point allows



(a)

(b)

Figure 3: (a) The block diagram of the visual servoing function, and (b) the block diagram as applied to instrument tracking.

only three degrees-of-freedom, (θ, ϕ, ρ), for manipulating the camera (Figure 4): 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 gain in this algorithm is used for robustness. The Jacobian matrix which relates $(\delta x, \delta y)$ to $(\delta\theta, \delta\varphi, \delta\rho)$ can be shown to be [5, 9, 8]

$$\mathbf{J} = \begin{bmatrix} -xysin\phi + ycos\phi & -\frac{\rho}{Z} - (1+x^2) & \frac{x}{Z} \\ -xcos\phi - sin\phi(1+y^2) - \frac{\rho sin\phi}{Z} & -xy & \frac{y}{Z} \end{bmatrix}, \tag{1}$$

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-

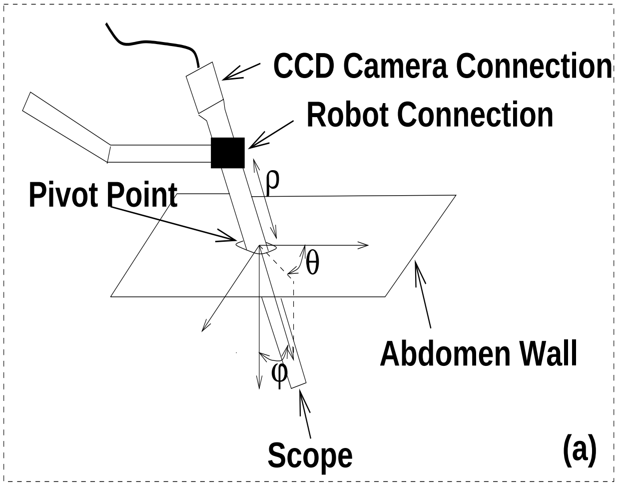


Figure 4: Sliding constraint imposed by the pivot point on the laparoscope.

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 5). In laparoscopy, the optimal site for insertion is the immediate subumbilical region. Typically, three to six such openings are made [2, 3], 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 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, zoom-

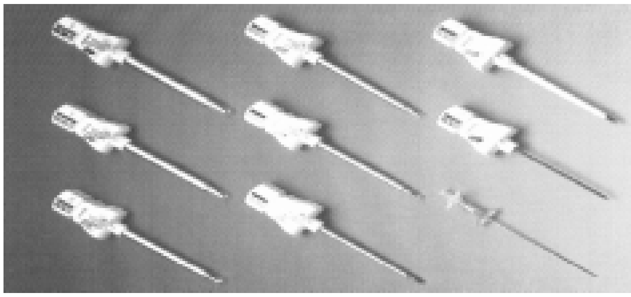


Figure 5: Disposable trocar/cannula of different sizes.

ing 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 Experimental Result

Currently, we are realizing a choreographed scope maneuvering sequence for instrument localization and tracking. This capability is of a fundamental importance in laparoscopy as for the safety reason the surgeon's view should always include the operating instrument. Furthermore, this capability can be used by the surgeon to guide the camera by repositioning an instrument (i.e., using the instrument as a pointer).

The development platform is a mockup OR with an AESOP scope positioning robot [10] (figure 6), 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. This is accomplished by a Jacobian matrix which relates the change of the image appearance (i.e., shape and location) of an instrument to the scope's degree of freedom in motion. 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.

Figure 6: The AESOP experimental platform.

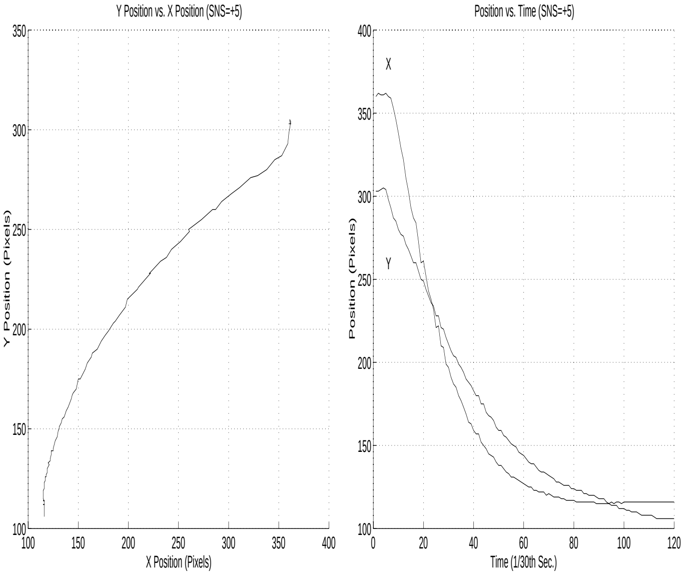
The ability to center a moving instrument is shown in Figure 7. Figure 7.a depicts the path of the instrument during tracking. Figure 7.b shows the deviation of the tip position from the image center (100,100) which gradually converged to zero. These figures clearly show the ability of the algorithm to track an instrument in motion.

4 Concluding Remarks

We believe that the proposed concept of choreographed scope maneuvering with active vision guid-

---

[1] Only layers with major actions during the particular subsequence are noted in this example.



(a) (b)

Figure 7: **(a)** Path of the instrument being tracked in the image plane. **(b)** Error in feature location vs. time.

ance 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 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.*

5 BIBLIOGRAPHY

References

[1] Armstrong Company Literature.

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

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

[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] C. Lee, D. R. Uecker, Y. F. Wang, and Yulun Wang. Image Analysis for Automated Tracking in Robot-Assisted Endoscopic Surgery. In *Proc. Int. Conf. Pattern Recognit.*, pages 88–92, Jerusalem, Israel, Oct. 1994.

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

[7] 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.

[8] D. R. Uecker, C. Lee, Y. F. Wang, and Yulun Wang. Automated Instrument Tracking in Robotically-Assisted Laparoscopic Surgery. to appear in *J. of Image Guided Surgery.*

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