A Hybrid View in a Laparoscopic Surgery Training System

Chuan Feng, Jerzy W. Rozenblit, PhD
Electrical and Computer Engineering Department
The University of Arizona
Tucson, Arizona 85721
Email: {fengc, jr}@ece.arizona.edu

Allan J. Hamilton, MD
Arizona Simulation Technology
and Education Center
The University of Arizona
Tucson, Arizona 85721
Email: allan@surgery.arizona.edu

Abstract

In this paper, a hybrid view application is proposed — a subsystem of a computerized laparoscopic surgery training system. To minimize the potential hazards of laparoscopic surgery, an assistive training system is being developed. A digital camera and magnetic position sensors are used to detect laparoscopic instruments in the system. The hybrid view is a component of this system which overlays the positions of organs and objects with the path history of the instruments. This method could help confirm erroneous movements made by surgeons and provide more useful information than separate sensors. This may minimize the cognitive overload on the surgeons. Initial experimental results are presented to show the feasibility of the proposed method.

1 Introduction

Laparoscopic surgery is performed with an endoscope and several long, thin instruments through small incisions. Due to its minimally invasive nature, laparoscopic surgery offers advantages such as shorter recovery time and reduced pain compared with traditional methods. However, inexperienced surgeons often lack a correct perception of an instrument's position due to a restricted vision field, hand-eye coordination problems, limited work space and the lack of tactile sensation. Those issues make laparoscopic surgery a difficult skill for medical students and residents to master.

There has been some research on the effectiveness of different kinds of training and guidance. Traditional surgical training methods such as using animals and cadavers have limitations because animals do not have the same anatomy as a human being and cadavers can not provide correct physiological responses. Surgery simulation is increasingly perceived as a valuable addition to traditional medical training. According to [3], laparoscopic training translates into approximately 30-35% more efficiency as measured by operative time and decreased complication rates compared to a control group not receiving simulation training.

One representative simulation tool-set is called the Pelvi-trainer [4]. It is a box that simulates the abdomen, with apertures for the insertion of instruments and camera. Trainees use real instruments to practice basic skills by manipulating objects or interacting with artificial tissue and anatomical models, using a video display for visualization. The main limitation of this approach is the absence of objective performance assessment, a feature available in some Virtual Reality (VR) systems, which use computer to simulate the whole training procedure. Hamilton [5] reports that speed was the only end-point measured within the pelvi-trainer while the virtual reality (VR) simulator reports error for each task performance [6]. Limitation of the VR simulations include inadequate realism of the virtual environment, inaccurate haptic feedback and the exorbitantly high cost of these systems.

Our vision for the proposed Virtual Assistant Surgical Training (VAST) system [1] is to bridge the gap between pelvi-trainers and VR systems, combining the advantages of both approaches to design a system that is simple and effective. We propose a knowledge-based sensor system to provide training prior to surgery and assistance during surgery. Our design features the embedding of micro-sensors into the instruments employed for simulation training. The detection and recording of instrument movement permits our system not only to measure a trainee's progress in acquiring psychomotor skills and compare these data to normative databases, but also to evaluate instrument effectiveness in reducing error. From a training perspective, the sensor-based system tracks and returns information on various performance metrics such as position and velocity of instruments, total path length of motion, erratic movements, time taken, number of attempts, dexterity, etc.

Fig. 1 contrasts the Knowledge-based VAST System with the traditional approach. In the VAST system, the sur-

Proceedings of the 14th Annual IEEE International Conference and
Workshops on the Engineering of Computer-Based Systems (ECBS'07)
0-7695-2772-8/07 $20.00 © 2007 IEEE

geon acts upon the patient or simulator through instruments and receives visual and force feedback from the supervisor both in the operating room and training settings. The supervisor represents the sensing interface and the knowledge-based computer system. It consists of a sensor fusion engine at the front-end and a knowledge based inference system at the back-end. In this paper, we discuss the issues related to the sensors.

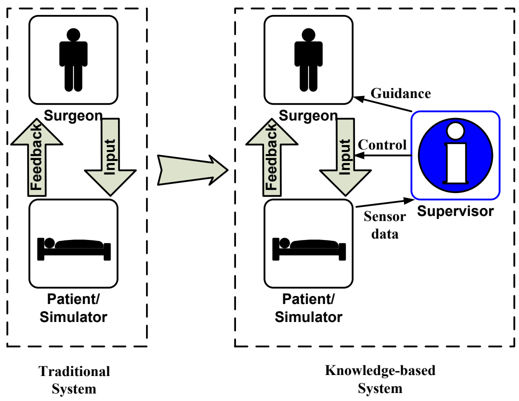

Figure 1. VAST system

A prototype system is being developed which is capable of high fidelity motion tracking of surgical instruments and basic performance assessment analysis. To our knowledge, this is the first approach that uses different kinds of sensors to assist in accurate tracking of instruments' positions and movements. The system fuses data from sensors and provides information to surgeons. To enable the data to be correctly understood, the various sensor outputs must be fused to provide a single representation.

Two kinds of sensors are implemented within the training system, a CCD camera and magnetic kinematics sensors. The Hybrid View, shown in the Fig. 2, which overlays the positions of organs and objects with the path history of the instruments, is one of the components of the VAST system. Immediate visual feedback obtained from the hybrid view helps surgeons check and evaluate the entire training process to minimize potential hazards, identify problem areas and find solutions.

For successful surgery, the laparoscopic camera should be moved from time to time to adjust the view of the surgical site. How to maintain data integrity during camera movement is the key issue of the hybrid view. Some researchers track the instruments' position using image analysis [2]. However, reliability is questionable because image analysis is highly related to the lighting conditions and attitude of the instruments. This can also make the tracking results discontinuous. For example, if the instrument is blocked by tissue or blood, vision tracking is unavailable. On the other hand, magnetic sensors can provide continuously accurate 3D position information on the instrument but lack the mapping relationship between the 3D space and 2D image. In order to solve the issue without more specialized hardware, a multi-level processing method was used.

In the remainder of the paper, we discuss our proposed approach in more detail. In section 2, the hybrid view generation system is presented in detail. In Section 3, several experiments are described must prove the feasibility of the proposed method. Section 4 is the conclusion.

2 System Design

2.1 System Architecture

In our application, shown in Fig. 3, a system with three processing levels generates the hybrid view.

The sources of information which include different kinds of sensors and knowledge database, are indicated on the left side of Fig. 3. A CCD camera and $microBIRD^{®}$ 6-DOF magnetic kinematics sensor are used in our application [7]. The CCD camera is connected to an endoscope which provides live image of the operating site. The kinematics sensing system includes a magnetic field transmitter, two position sensors (1.3mm in diameter) which can be mounted on the tip of the laparoscopic instruments and a PCI interface data processing card. The transmitter remains fixed to pro-

Figure 2. Hybrid View Sketch Map

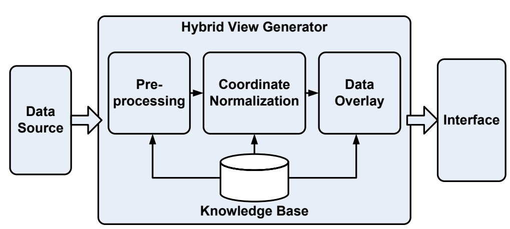

Figure 3. Hybrid View Generating System Framework

Proceedings of the 14th Annual IEEE International Conference and Workshops on the Engineering of Computer-Based Systems (ECBS'07)
0-7695-2772-8/07 $20.00 © 2007 IEEE

vide a Cartesian frame of reference for position tracking.

The right side of the figure shows the system interface. The interface consists of a subsystem interface and a human computer interface (HCI). The subsystem interface exchanges information with the VAST system and the HCI allows a human operator to input commands and get information from the generator. A *hybrid view* will be output through the HCI so that it can be watched by surgeons. The hybrid data will also be sent to the inference module of VAST through the subsystem interface. Therefore, further evaluation and feedback can be made. In a training setting, the trainee is required to repeat the motion within the safety bounds until a motion rule check passes. This enforced learning process should help the trainee master the necessary skills.

Three processing levels are shown in the system architecture. The first level is Source Preprocessing, which synchronizes the information flow from different sensors and distributes data to appropriate processes. The second processing level is Coordinate Normalization, which transforms sensor data into a consistent set of coordinates and estimates the position and kinematics of the instruments. In the third processing level, Data Overlay, data are fused together so the hybrid view can be generated. Detailed design information is provided below.

2.2 Level 1: Source Preprocessing

Source Preprocessing is the first level of the hierarchical processing model, which guarantees the synchronization of different data sources. Because the overall objective is generating the correct instrument track over the operating site, the camera drives this process. When a new frame is captured by the camera, the Source Preprocessing level forces the microBIRD sensor to obtain the corresponding 3D coordinates of the instruments and to distribute the data to the next processing levels.

Image filters are applied for noise elimination. A median filter is used to minimize the noise because it reduces the effect of unrepresentative pixels with almost no degradation to the underlying image.

Color segmentation method is used to finish the image preprocessing. Before color segmentation, a color cluster training process [11] is necessary. First, a color that does not appear within typical laparoscopic images is chosen to mark the instrument. Then an image of the color-marked instrument will be taken under real working light conditions. The mark on the instrument can then be outlined manually to provide a color cluster of the mark within the RGB color space.

During the working period, all pixels within the image are distinguished by color segmentation. Each pixel whose RGB value falls within the pre-training color cluster will be treated as one element of instrument area set I. Therefore, the color image can be converted into a binary data matrix which distinguishes the exact pixels of instruments. Each pixel in the image is mapped to the correspondent position of the data matrix.

2.3 Level 2: Coordinate Normalization

If there are no abnormal signals from the Source Preprocessing level, level 2 — Coordinate Normalization will do further processing. This level transforms sensor data into a consistent set of coordinates and estimates the position and kinematics of the instruments.

Before an operation, a rule module is applied to infer the information sent through the previous processing level. Sample rules are shown here:

- *if* microBIRD sensor data are incorrect
  then refinement to kinematics sensor data is disabled.
- *if* image data are incorrect
  then refinement to camera data is disabled.
- *if* both kinds of sensor data are incorrect
  then cancel the current working cycle.

The purpose of the module is to synchronize the functional modules according to the source data. Only the validated sensor data can be used for further processing.

2.3.1 Coordinate transformation of the kinematics sensors

The intent of the data fusion system is to generate a consistent Hybrid View, so transforming the sensory data into the image coordinates is an necessary procedure. Camera calibration is the first step when trying to map 3D information to a 2D image taken by a camera.

In order to provide real time computing ability, a simple linear mapping algorithm called Direct Linear Transformation (DLT) is implemented, which uses a linear camera model to extract pixel coordinates from the 3D data. The DLT mapping function can be expressed as equation 1:

$$u = \frac{L_1 X_w + L_2 Y_w + L_3 Z_w + L_4}{L_9 X_w + L_{10} Y_w + L_{11} Z_w + 1}$$
$$v = \frac{L_5 X_w + L_6 Y_w + L_7 Z_w + L_8}{L_9 X_w + L_{10} Y_w + L_{11} Z_w + 1} \quad (1)$$

Where $L_1 \ldots L_{11}$ are the parameters of transformation, which are the combination of the camera is intrinsic and extrinsic parameters. (X_w, Y_w, Z_w) is the 3D space coordinate, and the (u, v) is 2D image pixel. This equation can be solved by the least-square algorithm [8].

The drawback of DLT is that the calibration method cannot adapt to camera movements during the operation. To

Proceedings of the 14th Annual IEEE International Conference and
Workshops on the Engineering of Computer-Based Systems (ECBS'07)
0-7695-2772-8/07 $20.00 © 2007 IEEE

solve this issue, we need to modify the camera parameters in real time.

Because intrinsic parameters will not change when the camera moves, our approach was to modify the extrinsic parameters only. A virtual camera algorithm is applied as follows:

Assume the camera is fixed in a specific position, after the calibration. An isomorphic mapping function can be built as follows: 2:

$$\lambda \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} = f(X_c) = f(\begin{pmatrix} x_c \\ y_c \\ z_c \\ 1 \end{pmatrix}) \tag{2}$$

Where u, v are the pixel coordinates in image plane and x_c, y_c, z_c are the 3d point coordinates in world system, which is called S_c. X_c is the expended vector $(x_c, y_c, z_c, 1)^T$.

When the camera moves, the world system S_c will not move, so there exists a transform matrix M, which indicates the position and orientation of the camera movement. The mapping function is:

$$X'_c = M \cdot X_c \tag{3}$$

The original mapping function in 2 can be used. However, the point X_c has been transformed into a new coordinate X'_c within the new world system. So the point X'_c will be mapped to a new pixel $P' = (u', v')$ as below:

$$\lambda \begin{pmatrix} u' \\ v' \\ 1 \end{pmatrix} = f(X'_c) = f(M \cdot X_c) \tag{4}$$

From equation 4, we can confirm that the camera calibration parameters L need not be modified any more. We only need to continuously refresh the transformation matrix M so that it can express the transform relationship. In the next processing level, a detailed solution for refreshing the transformation matrix is described, which relates to image analysis.

2.3.2 Coordinate transformation of the camera data

The image data matrix sent from the Source Preprocessing level is a 2D binary data set that indicates the result of color segmentation. The position of the instrument must be acquired from the image data. The center of mass is used to indicate the instrument's position. Within the binary image, the center of mass can be calculated according to equation 5:

$$P = \frac{1}{M} \sum m_i r_i \tag{5}$$

Where M is the sum of the particle pixels; r_i is the position of the ith pixel, which is the distance from the pixel to the origin of the image; m_i is the value of the ith pixel; P is the center of mass.

2.4 Level 3: Data Overlay

The purpose of the Data Overlay level is to generate the hybrid view from different data sources.

As we previously mentioned, the transformation matrix M is used to indicate the shift and rotation of the camera. M is set to a unit matrix when the system is initialized since there is no camera movement. When the camera is moved according to surgical need, the matrix must be refreshed in real time.

If the 3D coordinates and the corresponding 2D pixels are known, the only unknown of the equations e:transform9 is the M, which consists of 6 moving parameters of the camera: $(x, y, z, \theta_{pitch}, \theta_{yaw}, \theta_{roll})$. These points, with known 3D to 2D mapping relationship are called reference points. Thus, when the camera moves, the first three reference points can be used to generate equations of M. The Newton method is used to solve the nonlinear equations. It is easily implemented in the system without any significant computational power consumption.

According to the feature of DLT, the matrix obtained from the above algorithm can only fit a small area close to the current position of the instruments. Therefore, the matrix needs to be modified continuously even when the camera is fixed. In order to efficiently modify the transformation matrix, the error E is defined as below:

$$E := (\Delta u, \Delta v) = (u - u', v - v') \tag{6}$$

Where (u, v) is the pixel position got from image analysis and (u', v') is the pixel got from 3D to 2D mapping. E is the vector indicating the difference between 2 separate data sources. The modification rule is indicated in formula 7.

$$\begin{array}{cc} if & \sum |E| > T_E \\ then & refresh \\ else & continue \end{array} \tag{7}$$

Where T_E is the threshold of accumulated error, and $sumE$ is the integration of errors, which can eliminate random noise and calculation inaccuracy.

After obtaining the transformation matrix of the 3D to 2D mapping, the hybrid view is generated. At the beginning, information sent from the lower processing level is analyzed by the knowledge based engine. If no data are available, the current fusion cycle will be ignored and nothing will be generated. If only one data source is verified, the tracks will be generated according to the one source. Otherwise, the camera data is used to draw the hybrid view; at the same time, data from both sources are used to refresh the 3D to 2D transformation matrix. The flow chart is shown in Fig. 4.

3 Experiment

To evaluate the proposed method, several experiments have been conducted. We will describe the camera calibration, image analysis and the 3D-2D transformation results separately in this section.

3.1 Camera calibration

Camera calibration is the first step of the hybrid view fusion process. In our application, DLT is used to obtain the camera parameters. To acquire the camera parameter vector L, a calibration testing board is applied. The calibration board contains a rectangular coordinate network of 1×1 inch in size with one inch margins. Thus the 3D coordinates in the world space and the corresponding 2D pixels can be determined easily.

The experiment process is described below:

- Fix the camera and calibration board in the space.
- Capture a picture of the calibration board by the camera.
- Determine $n(n > 6)$ pixels, as shown in Fig 5.
- Determine the related 3D points coordinates.
- Calculate the calibration parameters L.

Table. 1 shows the calibration error. The accuracy is high enough and sufficient for the hybrid view application.

3.2 Image analysis

The image analysis experiment is shown in Fig. 6:

The left picture is the image captured from the camera. Our aim is to determine the position of the color mark area.

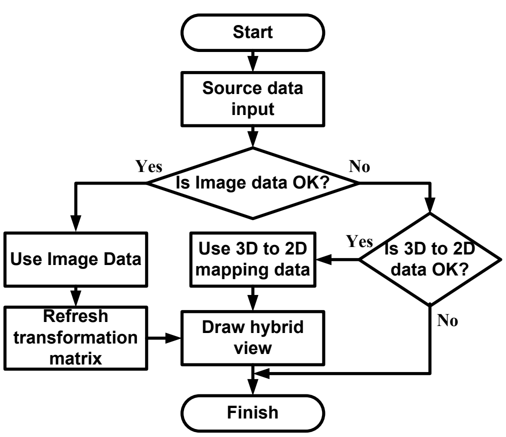

Figure 4. Hybrid view generation flow chart

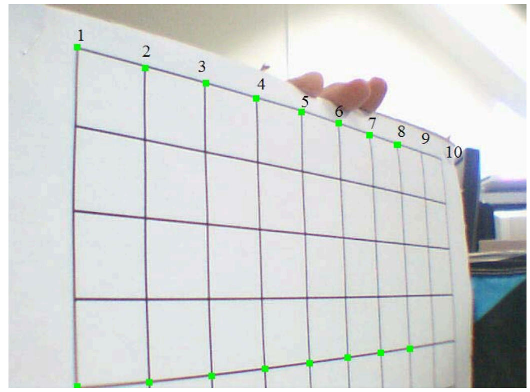

Figure 5. Camera Calibration Board

Table 1. Camera calibration results

3D coordinate (mm)	2D (pixel)	3D-2D map (pixel)	error (pixel)
(7.18,-10.16,7.18)	(495,430)	(496.4,431.2)	1.8411
(5.39,-20.32,5.39)	(447,162)	(446.7,161.7)	0.4036
(3.59,-10.16,3.59)	(419,441)	(417.9,441.2)	1.1608
(0.00,-20.32,0.00)	(306,117)	(305.3,118.5)	1.6611
(-1.80,-10.16,-1.80)	(251,464)	(251.0,462.4)	1.6467
(-3.59,-20.32,-3.59)	(168,77)	(170.1,77.2)	2.0993
(-5.39,-10.16,-5.39)	(84,479)	(84.0,483.5)	4.5198

Figure 6. Image Analysis

Proceedings of the 14th Annual IEEE International Conference and
Workshops on the Engineering of Computer-Based Systems (ECBS'07)
0-7695-2772-8/07 $20.00 © 2007 IEEE

The middle picture shows the binary image after color segmentation (the red color mark has been segmented well). The right picture is the result of mass point calculation. The green cross indicates the mass point of the color mark.

3.3 3D to 2D transformation

3D-2D transformation is the most complex part within the system. It is also the most important part. The calibration board is used to show the feasibility of the algorithm described above. The experiment is set up as follows:

- Capture a picture at position 1.
- Calculate camera parameters L according to the reference points.
- Move camera to position 2 and capture a new picture.
- Calculate a transfer matrix using 3 reference points.
- Calculate pixels from 3D space according to camera parameters and transferring matrix.

The diagram in Fig. 7 shows the experiment process. Two pictures captured by the same camera in two different positions are shown. The geometric relations between the camera and calibration board are shown at the bottom of the picture. The camera takes the first picture at position 1, then moves to position 2. During the experiment, the calibration board is fixed, so the 3D coordinates of the reference points will not change. Three reference points are chosen in the second picture to calculate the transformation matrix. After the calculation, a virtual path in the left picture is mapped to the right picture. Although they are the same points in 3D space, due to the movement of the camera, the hybrid paths are different. This is the reason we use the hybrid view, which correctly describes the 3D to 2D mapping relationships.

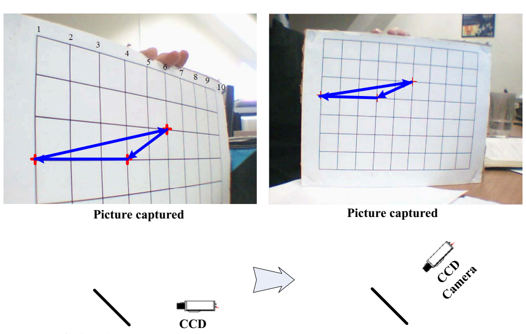

Figure 7. 3d to 2D mapping experiment

After mapping the 3D space coordinates onto the 2D image pixels, some errors can be found, as shown in Fig. 8. The reasons for the errors are twofold: one is the calibration method itself, the other is the transfer matrix. We use a linear transformation method to describe the mapping function which discards the nonlinear distortion of the lens, thus errors are unavoidable. The transformation matrix is acquired by the Newton method, which is a nonlinear optimization method, not an exact solution. According to the error diagram in Fig. 9, the points near the reference points are more accurate than the points more distant. Therefore, we need to refresh the 3D-2D transformation matrix frequently.

In the VAST system, the microBIRD sensor is used to in place of the calibration board for acquiring the 3D space coordinates. All the other operations are the same as the experiments described above.

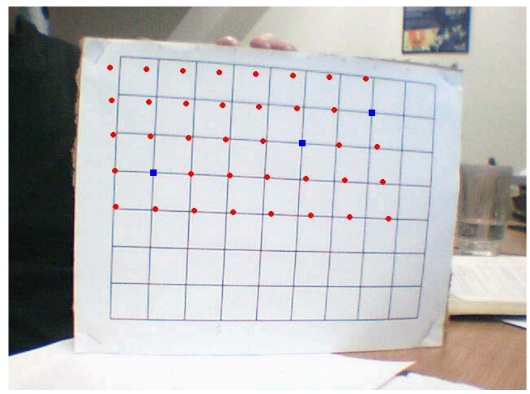

Figure 8. Pixels obtained from transformation

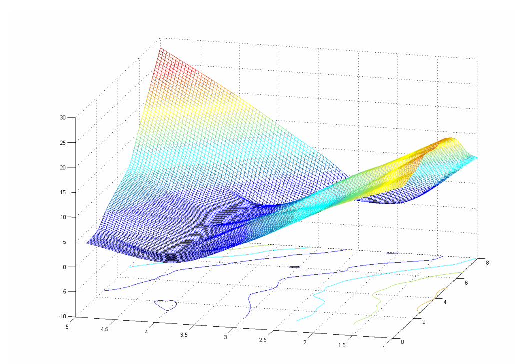

Figure 9. Error Diagram of transformation

4 Conclusion

In this paper, a novel method of the realizing a hybrid view within a virtual, assistive surgical training system has been presented. The hybrid view, which is a subsystem of a computerized laparoscopic surgery training system, helps

Proceedings of the 14th Annual IEEE International Conference and
Workshops on the Engineering of Computer-Based Systems (ECBS'07)
0-7695-2772-8/07 $20.00 © 2007 IEEE

confirm the erroneous movements without more expensive sensors while reducing the complexity of the system. In the proposed method, we use a multi-level processing model to describe the architecture of the system. Intensity information of the image analysis, 3D space Euclidean transformation, and digital camera calibration methods have been presented. Experimental results demonstrate the feasibility of the proposed methods. Further advances need to be made to provide three dimensional vision through a stereo endoscope and haptic information from the tip of the instruments. Currently, we are testing the system with medical students and residents.

References

[1] C. Feng, H. Haniffa and et al., "Surgical Training and Performance Assessment Using Motion Tracking," *EMSS 06*, 2006.

[2] G. Wei, K. Arbter and G. Hirzinger,"Real-time visual servoing for laparoscopic surgery. Controlling robot motion with color image segmentation," *IEEE Engineering in Medicine and Biology Magazine*, Vol. 16, Issue. 1, pp. 40-45, 1997.

[3] J. Korndorffer, Jr, J. Dunne, and et al., "Simulator Training for Laparoscopic Suturing Using Performance Goals Translates to the Operating Room," *Journal of the American College of Surgeons*, Vol. 201, Issue. 1, pp. 23-29, 2005.

[4] S. Shrivastava, R. Sudarshan, and et al., "Surgical Training and Performance Evaluation using Virtual Reality based Simulator," *Virtual Concept 2003, France*, 2003.

[5] E.C. Hamilton, D.J. Scott and et al., "Comparison of video trainer and virtual reality training systems on acquisition of laparoscopic skills Journal Surgical Endoscopy," *Surgical Endoscopy Journal*, Vol. 16, pp. 406 - 411, 2002.

[6] Simbionix Ltd., *http://www.simbionix.com/*, 2006.

[7] Ascension Technology Co., "microBIRD Technical Reference Guide," *http://www.ascension-tech.com/* , 2005.

[8] Y.I. Abdel-Aziz, and H.M. Karara, (1971). "Direct linear transformation from comparator coordinates into object space coordinates in close-range photogrammetry," *Proceedings of the Symposium on Close-Range Photogrammetry*, pp. 1-18, 1971.

[9] J. Leonard, "Sensor Fusion for Surgical Applications," *The 15th AESS/IEEE Dayton Section Symposium*, pp.37-44, 1998.

[10] L. Joskowicz and R.H. Taylor, "Computers in imaging and guided surgery," *IEEE Computational Science and Engineering*, pp.65-72, 2001.

[11] X. Zhang and S. Payandeh, "Application of Visual Tracking for Robot-Assisted Laparoscopic Surgery," *Journal of Robotic Systems*, pp.315-328, 2002

[12] B. Law, M.S. Atkins and etal. "Eye gaze patterns differentiate novice and experts in a virtual laparoscopic surgery training environment," *Proceedings of the 2004 symposium on Eye tracking research and applications*, pp. 41-48, 2004.

[13] F. Tendick and M.C. Cavusoglu, "Human-Machine Interaces for Minimally Invasive Surgery," *19th IEEE EMBS*, pp. 2771-2776, 1997.

[14] J. Llinas and D.L. Hall, " An introduction to multisensor data fusion," *Procedings of the IEEE*, pp.6-23, 1998.