Application of Visual Tracking for Robot-Assisted Laparoscopic Surgery

Xiaoli Zhang and Shahram Payandeh

Experimental Robotics Laboratory
School of Engineering Science
Simon Fraser University
Burnaby, BC, V5A 1S6, Canada
xzhanga@cs.sfu.ca, shahram@cs.sfu.ca

Abstract: With the increasing popularity of laparoscopic surgery, the demand for better modes of laparoscopic surgery also increases. The current laparoscopic surgery mode requires an assistant to hold and manipulate the endoscope through commands from the surgeon. However, during lengthy surgery procedures, accurate and on time adjustment of the camera cannot be guaranteed due to the fatigue and hand trembling of the camera assistant. This paper proposes a practical visual tracking method to achieve automated instrument localization and endoscope maneuvering in robot-assisted laparoscopic surgery. Solutions concerning this approach, such as, endoscope calibration, marker design, distortion correction and endoscope manipulator design are described in detail. Experimental results are presented to show the feasibility of the proposed method.

Keywords: Laparoscopic Surgery, visual tracking, calibration, barrel distortion

1. Introduction:

Recently, laparoscopic surgery gained increasing popularity because of its minimally invasive nature. In laparoscopic surgery, a surgical operation can be performed with an endoscope and several long, thin, rigid instruments (see Figure 1) through several small
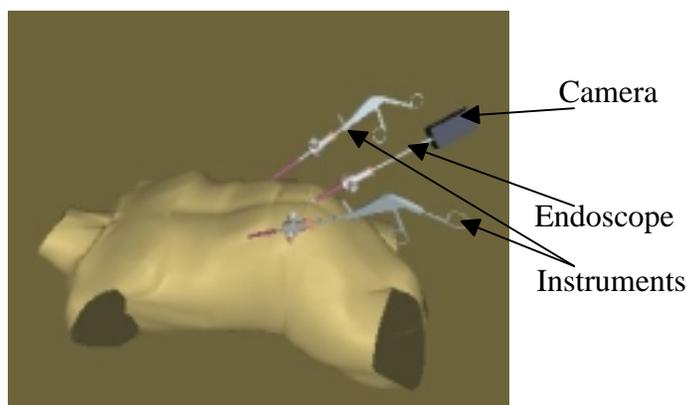
Figure 1. Tools needed for laparoscopic surgery.

incisions. Compared with traditional surgery, laparoscopic surgery provides less pain, faster recovery and small injuries. However, to control the surgeon's field of view, an assistant surgeon is needed to hold and manipulate the camera. During lengthy

procedures, accurate and on time adjustment of the camera cannot be guaranteed due to the fatigue of the camera assistant. Instead, using a robot to control the camera will result in less erroneous camera motion and can off-load routine tasks.

A typical laparoscopic surgery using robot camera assistant is shown in figure 2. The endoscope relays images of internal organs to a camera, and the images are displayed on a video screen placed in front of the surgeon. However, the lack of intelligence in the

Monitor for viewing the
operation site
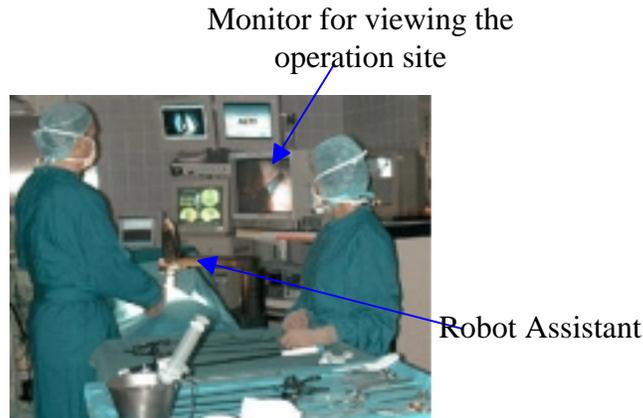
Robot Assistant

Figure 2. Laparoscopic surgery using a robot camera assistant.

design of current robot camera controllers presents a challenging problem. For example, the surgeon in general is often distracted by issuing positioning command to the robot manipulator via hand-held controllers, a foot pedal or other interface such as voice control, e.g., voice controlled robot AESOP[1].

To improve the current robot-assisted laparoscopic surgery system, we propose to use visual tracking techniques to enable automatic robotically control of the camera. In order to track the instrument, we need to know the three-dimensional (3D) positional information. In general, such 3D information can be obtained from stereo-endoscopy (the use of two separate optical channels, each with its own lens optics). However, this approach doubles the hardware, thus it is not widely accepted by clinicians. We propose a method that uses a single endoscope to obtain the 3D positional information of the instrument.

Visual tracking is an active research area addressed by many researchers. Prewer and Kitchen [2] used edge correspondence to track the image features. Azarbayehani and Pentland [3] obtained the 3D locations of the face and hands by tracking skin blobs. Reddi and Loizou [4] used optical flow field to control the camera's motion to keep a target at the center of the camera's field of view. By using prior knowledge about how humans move, Leventon [5] built a Gaussian probability model to reconstruct the 3D motion of a human figure from a monocular image sequence. However, only a few researchers addressed visual tracking in laparoscopic surgery: Wang et al. [6] used color statistics to classify, group and label instrument directly, Wei et al. [7] designed a marker and using color selection and color segmentation to detect and locate the marker.

Due to the diversity of the instruments, it is hard to track the tip of instruments directly as described in [6]. In this paper, we propose to design a marker and attach the marker to the tip of the instrument. When needed, visual tracking techniques can be

applied to track the marker as the instrument moves. Instead of using color information for visual tracking as proposed in [6] and [7], we use monochrome data for processing. Generally, monochrome image processing is quicker, simpler and robust. The 3D positional information of the marker then will be used to realize automatic tracking of the instruments.

There are many difficulties associated with using markers for instrument localization. First, images taken from the endoscope are distorted so it is difficult to obtain the correct measurement of the images. Second, tools may enter the endoscope field of view from different directions and at various angles. Finally, using images from mono-endoscope to reconstruct 3D spatial information is a challenging task. These problems will be addressed and solved in later parts of this paper.
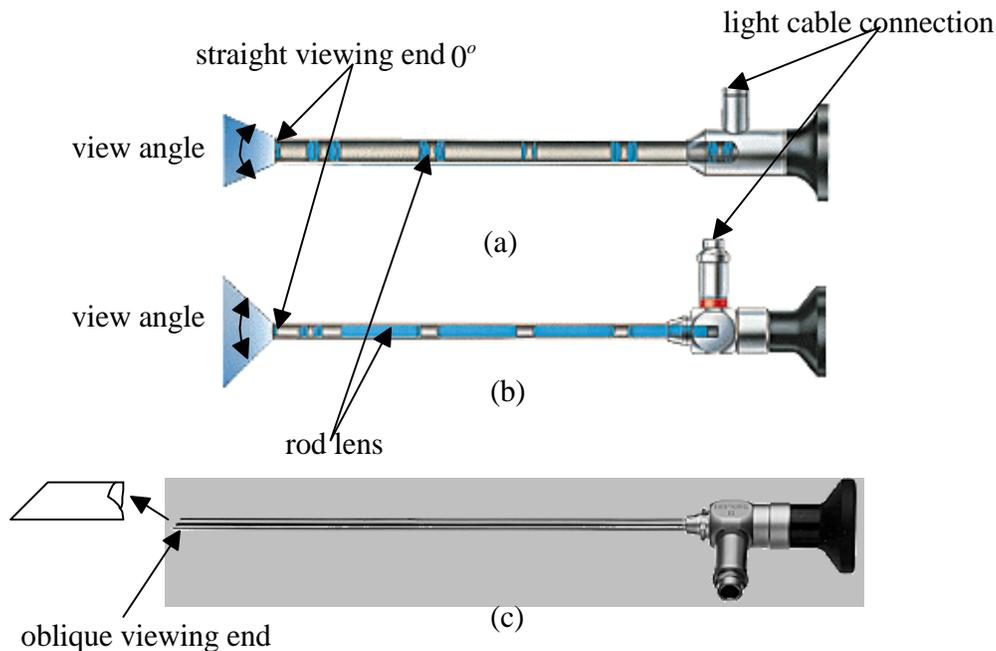
Figure 3. Rod Lens endoscope (a) An endoscope featuring a conventional lens system. Refraction occurs as light passes through small lenses seated at distant intervals. (b) A HOPKINS rod lens endoscope. The rod lenses are seated at close intervals, resulting in better light transmission and reduced diameter. (c) Forward-oblique Endoscope 45°.

The remainder of this paper is organized as follows: Section 2 presents a method for endoscope calibration. In section 3, marker design and tracking algorithm are discussed. In section 4, endoscope image distortion correction method is described in detail. To investigate the tracking results, the design and simulation of a camera manipulator are described in the section 5. Finally the experimental results are shown and future work is discussed.

2. Calibration Method

Defining the pixel-to-world mapping is known as camera calibration. Once the camera has been calibrated, we can transform the image pixel coordinates to their real-

world equivalents. Camera calibration is a critical first step in many applications such as dimensional measurement of mechanical parts, tracking, camera-on-robot configuration and robot vehicle guidance. Camera calibration has been investigated extensively by many researchers (e.g., [8], [9]). However, methods for camera calibration using rigid endoscope have not been addressed before.

In general, endoscope images have particular characteristics. The endoscope viewing angle, typically between $65^o - 75^o$, as shown in Figure 3, is narrow compare to the wide field of view of human eyes. Thus, vision is limited by the narrow endoscope field of view. To enlarge the field of view of endoscopes, wide-angle lens are employed in the endoscope. However wide-angle lenses intrinsically cause barrel distortion (i.e., transformation of straight line into curves). Barrel distortion cause objects in the image to be distorted along radial lines from the image center. This type of distortion is non-linear: image areas farther away from the center appear significantly smaller. This paper proposes a calibration method that assumes the image located within the small center area is distortion free.

Typical calibration parameters can be classified into two classes: extrinsic parameters and intrinsic parameters. This section presents methods for determining intrinsic parameters, namely, the effective focal length, the real image center and the scale factor. The extrinsic parameters, that is, the rigid body transformation from the world coordinate system to the camera coordinate system is discussed later.

Figure 4 illustrates the basic geometry of the camera model with (x, y, z) as the world coordinate system. In robotic system, this coordinate frame can be placed at the robot base. (x_c, y_c, z_c) is the camera coordinate system, $P(x, y, z)$ is the 3D coordinates of a

(a)

(b)   (c)
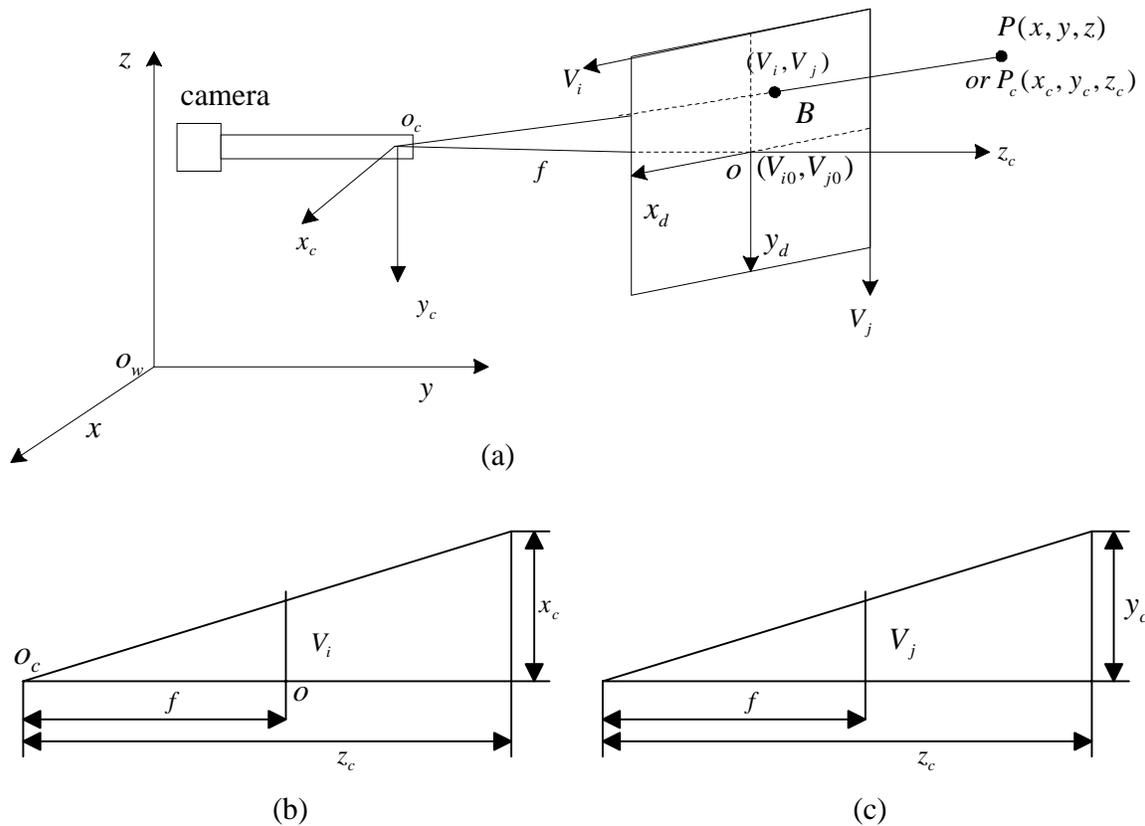
Figure 4. Camera geometry with perspective projection.

object P in the world coordinate system, $P(x_c, y_c, z_c)$ is the coordinate of point P in the camera coordinate system with its z_c axis the same as the optical axis. (x_d, y_d) (not shown in Figure 4) is the image coordinates of $P(x_c, y_c, z_c)$ projected into the image plane if a perfect pinhole model assumption is used [10]. However, since the unit for the coordinate used in the computer, (V_i, V_j), is the number of pixels in the frame memory, additional scale factors need to be specified that relate the camera coordinate system to the computer image coordinate system (using pixel units).

The focal length f is the distance between the image plane center, o, and the optical center, o_c. Here we define the image center as the point where the camera optical axis passes through the image plane. In theory, the image center should be the center of the image plane. However, in practice, especially in the endoscope image, the real image center is distorted due to the imperfect grind of endoscope lens.

In the following, first we propose a simple and fast method to obtain the real image center. Then, a method for calculating the calibration parameters such as the effective focal length and the scale factor. Figure 5 shows the calibration flow chart.

(x_c, y_c, z_c) 3D camera coordinate system

Step1. Find the real image center

Step 2. parameters to be calibrated : focal length

Step 3. Parmeters to be calibrated: scale fectors

(V_i, V_j) image buffer coordiante
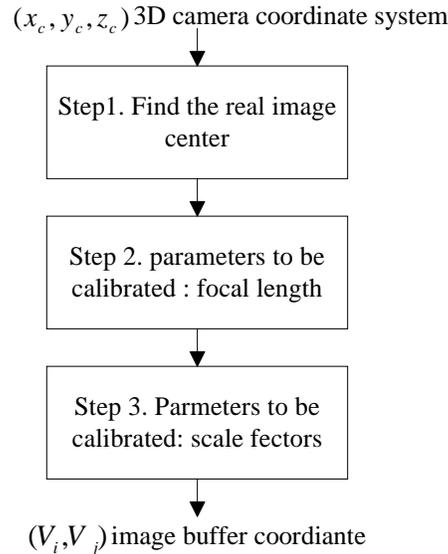
Figure 5. Three steps in endoscope calibration.

The first problem that needs to be addressed is the method for calculating the real image center. Since abdominal cavity is dark, an external light source is used, see Figure 3. A fiber optic cable transmits light to the side of the endoscope. The light source is evenly distributed through the rod lens. If we point the endoscope perpendicular to a white background, the light source forms a white circle in the image as shown in Figure 6. (a). Inspired by this feature, we propose a simple method to obtain the real image center by pointing the endoscope to a white background and recording the image. The center of the white circle in the image is defined as the real image center as seen in Figure 6. (b). Other work, such as Asari et al. [11], considered the distortion in endoscope image and proposed to use straight-line patterns. Then chose the intersection of lines that remain straight in the endoscope image as the real image center. However, their methods need

complex calculations. The feasibility of the proposed method for determining the real

(a)   (b)
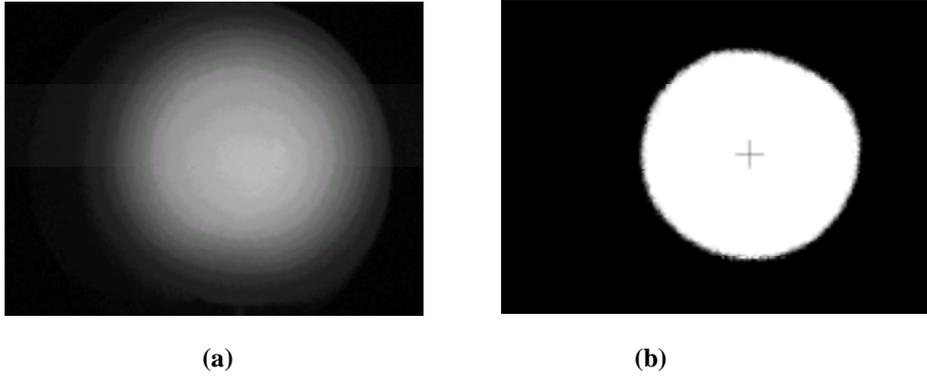
**Figure 6. Endoscopic image (a) original gray scale image,
(b) binary image (the image center has been marked).**

image center is evaluated later in the experimental results section.

The second parameter to be calibrated is the effective focal length. As shown in Figure 4, the transformation from the 3D camera coordinate (x_c, y_c, z_c) to the image coordinate (V_i, V_j) based on the perspective projection with pinhole camera geometry yields the following equations:

$$\frac{\overline{oo_c}}{z_c} = \frac{f}{z_c} = \frac{(V_i - V_{i0})S_x}{x_c} = \frac{(V_j - V_{j0})S_y}{y_c},$$

$$x_c = \frac{z_c(V_i - V_{i0})S_x}{f}, \; y_c = \frac{z_c(V_j - V_{j0})S_y}{f},$$

(1)

where $\overline{oo_c}$ is the distance between the center of the image buffer to the center of the camera coordinate frame, (V_{i0}, V_{j0}) is the image center in the pixel unit, S_x, S_y are the scale factors that map (x_d, y_d) in the camera coordinate frame to (V_i, V_j) in the computer image coordinate frame. When V_i, V_j, X_c and Z_c are known parameters the effective focal length can be calculated from equation (1). Also, when f, V_i, V_j and x_c, y_c are known parameters, we can calculate the depth information, Z_c, of an object in
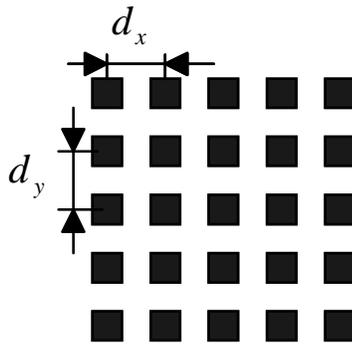
Figure 7. The pattern image for calibration.

the field of view. Scale factors S_x, S_y can be obtained using the following formula

$$x_d = (V_i - V_{i0})S_x, \; y_d = (V_j - V_{j0})S_y.$$ (2)

In general, manufacturers of CCD cameras supply the information of center-to-center distance between adjacent sensor elements in the Y direction (i.e., scale factor S_y) as a fixed value: $S_y = N_{fY} / d_y$, where N_{fY} is the number of pixels in the Y direction; d_y is the dimension of CCD in the Y direction. The scale factor S_x is an uncertain value due to various reasons, such as slight hardware timing mismatch between image acquisition hardware and camera scanning hardware, or the imprecision of the timing of TV scanning itself. Here we propose a simple and fast method to obtain the relationship between S_x and S_y.

By capturing an image of known dimension (e.g., the square grids, d_x and d_y are the actual center-to-center distance in millimeters of adjacent squares, see Figure 7), we can compute their corresponding, D_x and D_y, in pixel units in the computer image coordinate. The following relationship exists: $D_x = S_x d_x, D_y = S_y d_y$.

Solving for S_x, we get

$$S_x = S_y \times D_x / D_y.$$ (3)

In actual implementation, we can measure several squares' center-to-center distances and take the mean value to reduce the possible error in image processing.

3. Lens Distortion and Correction Methods

No lens can produce perfect images. Common imperfections are aberrations that degrade the quality or sharpness of the image and lens distortions that deteriorate the geometric quality (or positional accuracy) of the image. Endoscope image has a fundamental lens distortion due to the wide-angle design of the endoscope's objective lens. Wide-angle lenses are used in endoscope because they provide larger viewing fields. However, lens distortion will result in erroneous measurement for image positions in the resulting images. To eliminate the distortion effect, corrections should be applied to the measurement of the resulting endoscope images.

Lens distortions are classified as either radial or tangential. Radial distortion, as its
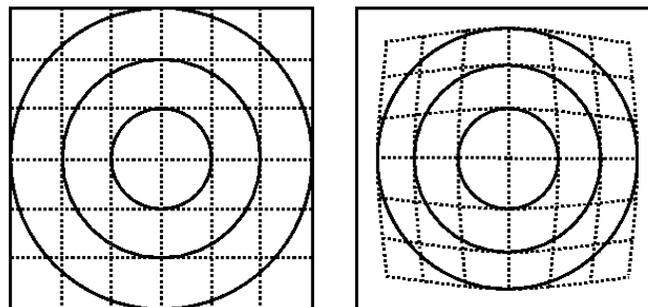
Figure 8. Barrel type distortion (The left image is an undistorted grid pattern, and the right image is the same pattern viewed after radial distortion).

name implies, causes image position to be distorted along radial line from the optical axis. Radial distortion includes barrel distortion, pincushion distortion or the combination of these two types. Tangential distortion is due to imperfect centering of the lens components and other manufacturing defects in a compound lens, and is also called decentering. That is, the optical center and the lenses are not strictly collinear. The pixel shift is away from the optical center and the new position lies at a new angle location as measured from the optical center. Tangential lens distortions are generally very small and are seldom corrected for.

Barrel type radial distortion is common in laparoscopic images, see Figure 8. Areas further from the center of the field of view appear smaller than they really are. In this paper, we only consider the barrel type distortion. We propose a distortion correction method that approximates the distortion curve by a polynomial and uses the least squares method to find the coefficients for this polynomial. With this experimentally determined polynomial, we can correct the endoscope image accordingly.

3.1 Radial Lens Distortion Model

The radial lens distortion model consists of two image planes. The distorted image plane is represented by (V_i', V_j'), while the distortion correction image plane is represented by (V_i, V_j). Their corresponding centers are represented by (V_{i0}', V_{j0}') and (V_{i0}, V_{j0}). The center of the distortion-correction image can be chosen arbitrarily. Unlike the method proposed by [12] that selects these two centers at different locations, in this paper, we select both centers the same as the real image center. The real image center is defined as the point where straight lines remain straight as they pass through this point in the distorted image plane.

Using a polar coordinate system with its origin at the real image center (V_{i0}, V_{j0}), a point P' in the distorted image plane can be represented as

$$r' = \sqrt{(V_i' - V_{i0}')^2 + (V_j' - V_{j0}')^2},$$

$$\theta = \arctan(\frac{V_j' - V_{j0}'}{V_i' - V_{i0}'}),$$

(4)

where magnitude, r', is the distance of P' to the center (V_{i0}', V_{j0}'). The corresponding point, P', in the corrected image plane is P. Point P can be represent in polar coordinate as (r, θ). Because we assume that the distortion is pure radial, the polar angle is unchanged in the distorted and corrected image planes. Hence: $\theta = \theta'$.

The objective of the distortion model is to find the relationship between P and P'. As shown in Figure 9, the relationship between r and r' is

$$r' = r - \Delta r,$$ (5)

where Δr can be represented as an odd-ordered polynomial series[13]:

$$\Delta r = k_1 r + k_2 r^3 + k_3 r^5 + k_4 r^7 + ...$$ (6)

where $k_i (i = 1,2,3...)$ are the expansion coefficients. After obtaining the coefficients, the new pixel location in the corrected image plane can be calculated as:

$$V_i' = V_i - \Delta V_i = V_i(1 - \frac{\Delta r}{r}) = V_i(1 - k_1 - k_2 r^2 - k_3 r^4 - ...)$$

$$V_j' = V_j - \Delta V_j = V_j(1 - \frac{\Delta r}{r}) = V_j(1 - k_1 - k_2 r^2 - k_3 r^4 - ...).$$

(7)

To obtain the polynomial series for mapping the distorted image to the distortion correction image, first, we need to obtain the coefficients for the polynomial.

3.2 Coefficients Estimation

The polynomial coefficients define the shape of the curve. They can be calculated by nonlinear regression analysis, such as the least squares method, to obtain the best curve fit to a given data set (obtained from experiments). A pattern image (see Figure 7) was used to obtain the given data set. Since we assume that the distortion is pure radial, the distortion is circularly symmetric. Without the loss of generality, testing squares that lie in the horizontal or vertical of the image can represent the general distortion pattern for the whole image. For a given data set $S = \{C_1, C_2, ..., C_i, ... C_N\}$, let there be N columns of testing squares, for each $C_i(r_i, \Delta r_i)$, r_i represents the radial distance between the center of the square from the real image center, Δr_i is the distortion at a radial distance r_i. Consider the polynomial series of degree $2M + 1$:

$$\Delta r_i = k_0 r_i + k_1 r_i^3 + k_2 r_i^5 + ... + k_j r_i^{2j+1} + ..., \quad j = 0, ... M$$

(8)

The deviation of a point from the above equation is

$$F_i = y_i - \sum_{j=0}^{M} k_j r_i^{2j+1} .$$

(9)
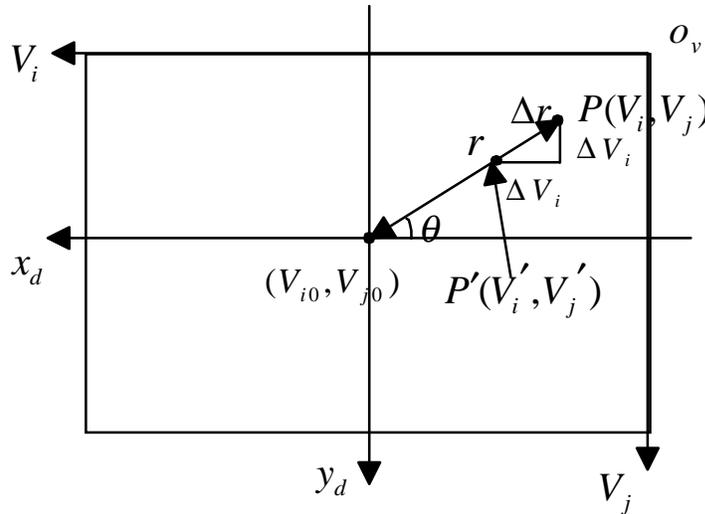
Figure 9. Components of Radial Distortion. The image point (V_i, V_j) is radically displaced by Δr to the new position (V_i', V_i').

The least squares problem is then to find the values of k_j, $j = 0,..,M$, so as to minimize

$$\sum_{i=1}^{N}[y_i - (\sum_{j=0}^{M} k_j r_i^{2j+1})]^2 \qquad (10)$$

Hence, k_j can be calculated by

$$\frac{\partial F_i}{\partial k_j} = 0, \qquad for \quad j = 0,1,..., M. \qquad (11)$$

From equation (9), $M+1$ simultaneous equations are obtained and represented in the matrix form:

$$AK = Y. \qquad (12)$$

where

(a)   **(b)**
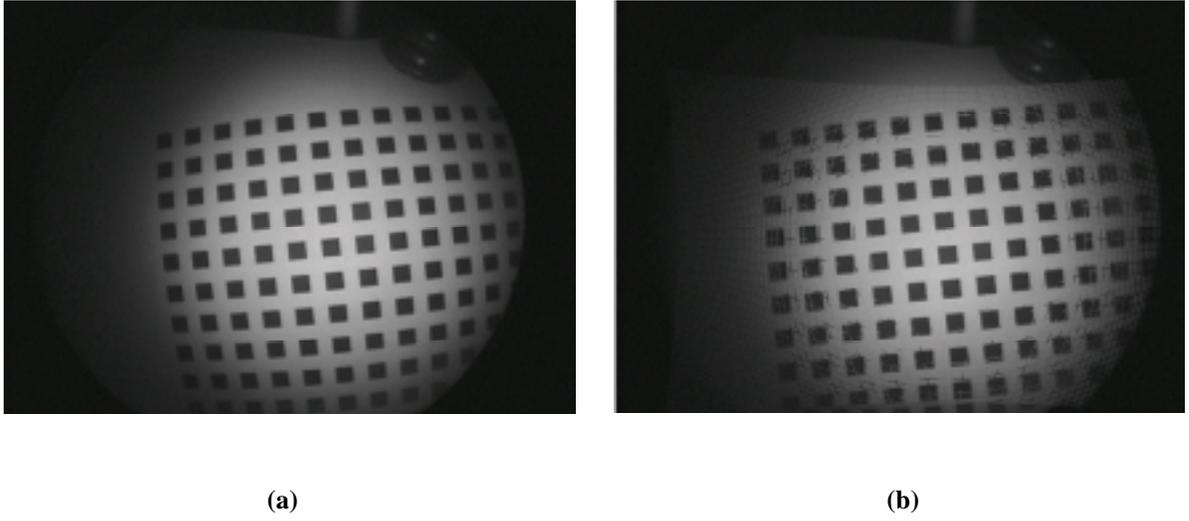
Figure 10. (a) The original image, (b) the image after distortion correction (using forward mapping).

$$A = [r_i^{2j+1}]_{(N+1)\times(M+1)}, \; for \; i = 0,..N, \; j = 0,...,M,$$

$$K = [k_0, k_1,..., k_M]^T,$$

$$and \; Y = [y_0, y_1,..., y_N]^T.$$

Equation (12) can be calculated as

$$K = (A^T A)^{-1} A^T Y. \qquad (13)$$

The matrix K consists of most probable values for unknowns, $k_0, k_1,..., k_M$.

Once the expansion coefficients are computed, all pixels from the distorted image are mapped onto the corrected image. However, the direction of mapping is one problem to be considered. One approach is mapping from the distorted image plane to the corrected image plane using the formula of $r = r' + \Delta r'$ or

$$r = k_1' r' + k_2' r'^3 + k_3' r'^5 + k_4' r'^7 + ... \qquad (14)$$

where r' is the radial distance measured in the distorted image plane. r is the radial distance measured in the corrected image plane. However, this method of mapping will result in blank pixels in the corrected image due to the non-linear expansion of the image. The original image has to be mapped into a new enlarged image. See Figure 10 (b),

notice there are many blank pixels in the image, especially in the peripheral areas. This problem can be avoided by using an inverse-mapping method, that is, mapping from the corrected image plane to the distorted image plane. The coefficient estimation method described in equation (7) used the inverse mapping method. In equation (7), r was measured in the corrected image plane. Thus, for every pixel in the corrected image plane, the corresponding location in the distorted image is obtained using the polynomial (8). The information (e.g., gray level) for that pixel location is assigned to the pixel in the corrected image plane. In the event that the pixel positions calculated using inverse mapping are non-integers, we simply round them to the closest integers. Because the corrected image enlarged the distorted image in barrel type radial distortion, several pixels in the corrected image may possibly find the same pixel in the original image. In this manner, all pixels in the corrected image plane can find their corresponding pixel values, thus generating a complete undistorted image.

4. Marker Design

As mentioned in the introduction section, researchers have proposed several types of tracking methods such as optical flow (the optic flow field is the 2D distribution of apparent velocities that associated with the variation of brightness patterns on the projection), feature points correspondence and model-based tracking. In our application, due to the diverse design of instruments, it is difficult to locate the instrument directly. To solve this problem, a black strip marker is designed. The marker is attached to the tip of instruments, which is then identified for the tracking task.

In general, the tracking task includes recognizing markers and calculating its relative position in the camera coordinate system. In real surgery, the instrument may enter the field of view from different positions and at different angles, using shape analysis and pattern matching would be infeasible, for there is no fixed shape suited for tracking. Instead, the image contrast information is used alone for instrument segmentation to locate the marker's position in the image. Two factors should be considered in the marker design. For one thing, in real-time image tracking, the marker should be simple and easy to locate. For the other, the marker's size should be chosen according to the size of the instruments.

Considering the above constrains, we design the marker as shown in Figure 11. The strips in Figure 11 are the designed markers. They are of the same size, each with the
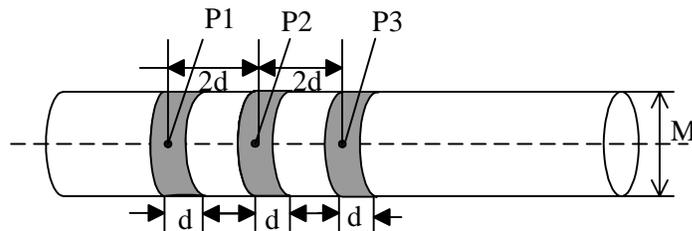
Figure 11. The designed marker.

width of d ($M/3 < d < M/2$). The diameter of the instrument is M. Points $P1$, $P2$ and $P3$ are the centers of each strip projected onto the image plane. These strips' center-to-center distance is $2d$.

Designing the marker in this shape has several advantages. First, the shape of the marker is simple and its contrast with the instruments and background is large; thus, it is easy to identify the marker. Second, the tool's diameter M is used to acquire depth information. Regardless how the instrument rotates along its axis (see the dotted line in Figure 11), we can always find two peripheral points on the marker that can represent the tool's diameter. Finally, even if an organ or other instruments block one or two strips, we can still get the instrument positional information form the remaining strips (at least one strip should be in the endoscope field of view).

When needed, the tracking task is to move camera so as to position an instrument feature, such as markers, at the desired location of the endoscope field of view and also keep the depth of the chosen instrument feature as a given value as the instrument moves.

Considering Figure 11, during the tracking procedure, one of the points among $P1$, $P2$, and $P3$ is located, the located point has not been blocked and is the closest to the tip of the instrument. To determine which point is the closet one to the tip, we can compare their diameter M value in the image. We assume that in laparosocopic surgery, generally the tip of the instrument is the point that is the furthest away from the endosocpe; thus, the smallest M is the one closed to the tip of the endoscope. Based on the change of the chosen point in V_i and V_j directions in the image plane and the calibration parameters described in section 2, we can calculate the displacement in real world unit to which the endoscope should be moved. From the change of diameter M in the image plane, the depth information of the marker can also be obtained using formula (1). This information is utilized as feedback for manipulating the endoscope so as to keep the tip of the instrument at a desired position in the endoscope field of view.

5. Endoscope Holder Design

In laparosocopic surgery, the endoscope is inserted into the abdominal cavity through a trocar. Constrains at the pivot point only allow four degrees of freedom for
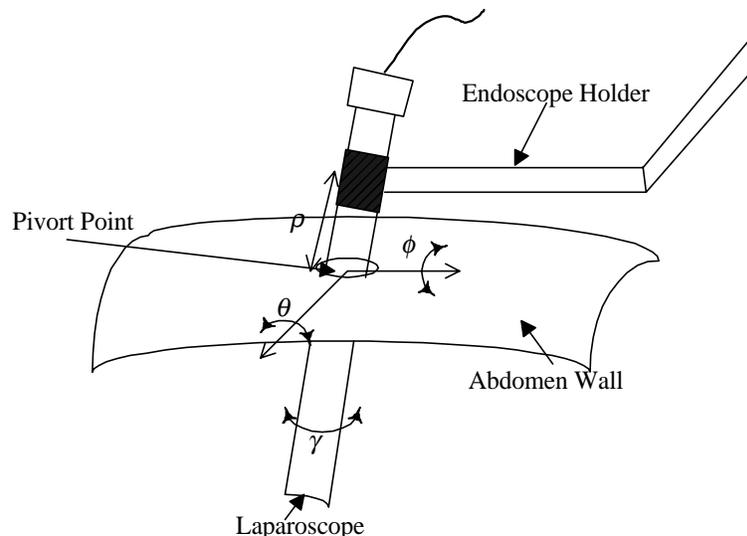
Figure 12. 4 DOF constraints imposed by the pivot point on the endoscope.

manipulating the endoscope, that is, three degree of freedom of angular movement (θ, ϕ, γ) around the incision point and one translational degree of freedom ρ in or out the incision point (see Figure 12). These spherical movement constrains should be carefully considered prior to the design of the endoscope holder.

Considering Figure. 12, we notice that rotation, γ, is only used to adjust the orientation of the image. Generally this angle will not change during the surgery. Based on this observation, the endoscope holder design can be simplified as 3DOF design with two rotational DOF and one translational DOF.

For robot manipulators, there are two types of joints, P: prismatic (prismatic joints exhibits sliding or linear motion) and R: revolute (revolute joints exhibit rotary motion about an axis). The endoscope holder has to have spherical movements. Among different combinations of prismatic and revolute joints, 3DOF RRP produces a spherical coordinate robot, which fits the purpose of this work. Another notable problem is that the designed endoscope holder should grantee the fixed position at the pivot point.

Considering all the constraints stated above, a designed endoscope holder is illustrated in Figure 13. This design satisfies the kinematic constrains. To ensure that the endoscope holder will not change its position at the incision point, we can adjust point, P, to the same height as the incision point. The world coordinate frame, (x, y, z), is set at the base of the endoscope holder. Generally, the camera is mounted at the end of the endoscope outside the abdomen. The endoscope transmits the image as seen at the tip of the endoscope inside the abdomen. Hence, we can set the camera coordinate frame, (x_c, y_c, z_c), at the center of the tip of the endoscope.
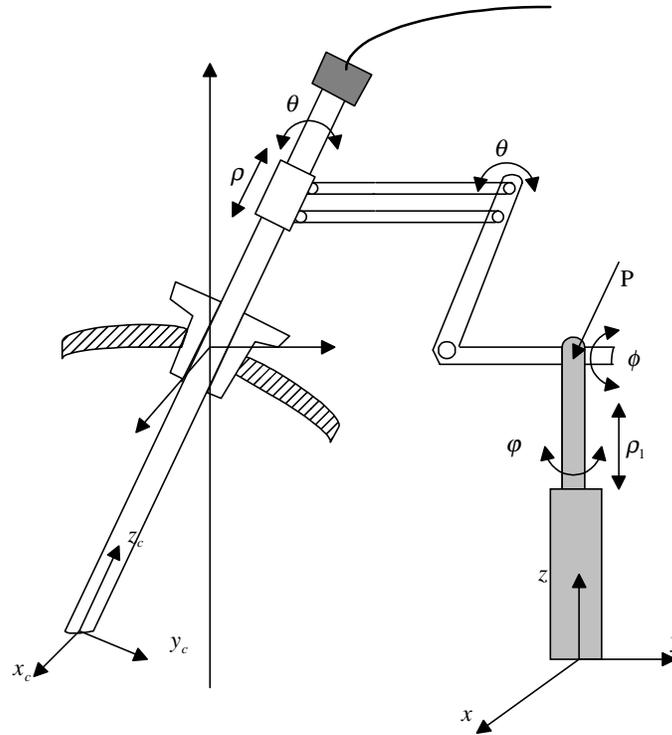
Figure 13. The designed scope holder.

The tracking task is to obtain the three displacement parameters, $\Delta\theta, \Delta\phi$ and ρ, based on the image features. These parameters are used to control the endoscope holder. As shown in Figure 14, here the desired position for the marker is chosen as the image center (V_{i0}, V_{j0}). When tracking, first, we determine the errors, ΔV_i and ΔV_i, between the current marker location and the desired location, and then transform them to real world coordinates using the calibration parameters described in section 2. Second, the following formula exists for small angles: $\Delta\alpha = \Delta d / \rho$. Replace Δd with ΔV_i and ΔV_j, $\Delta\alpha$ with $\Delta\theta$ and $\Delta\phi$, we obtain:

$$\Delta\theta = \Delta V_i / \rho$$
$$\Delta\phi = \Delta V_j / \rho$$

(14)

The translation parameter, ρ, is calculated from the current diameter M in the image plane using formula (1). Last, the endoscope can be adjusted accordingly to keep the instrument at a desired location under the endoscope field of view.
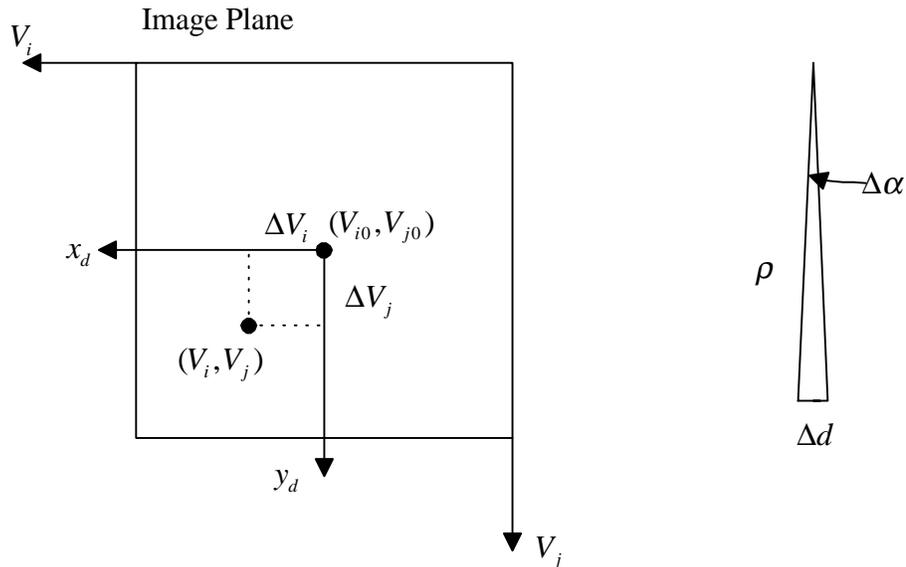
Figure 14. The geometric approach for obtaining three displacement parameters.

6. Experimental Results

To evaluate the proposed method, several types of experiments are conducted. In this section, we describe the camera calibration, endoscope image distortion-correction, endoscope holder simulation and visual tracking results separately.

An electronic endoscopy system, comprising a Karl Storz endoscope[12] is used in capturing the image. A testing pattern-image containing a rectangular array of black squares of 4×4 mm in size, separate by 4mm in the horizontal and vertical direction (see figure 7) is used for calibration and distortion correction purpose.

The horizontal line

The vertical line
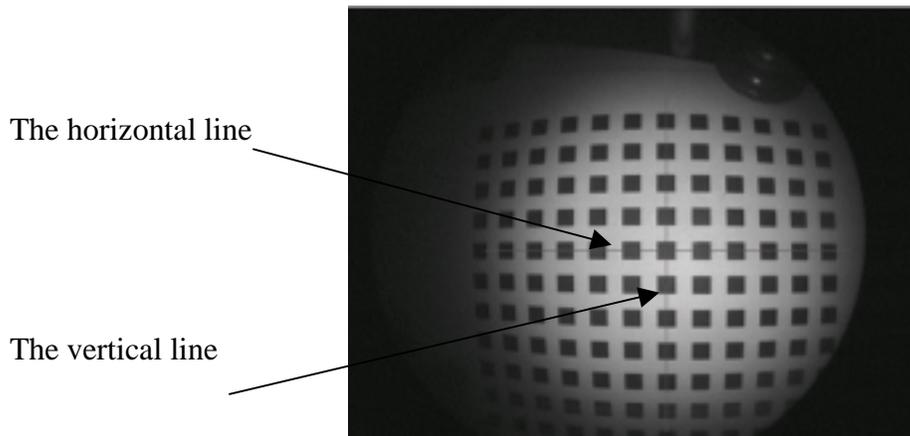
Figure 15. Verify the proposed center determination method.

6.1 Calibration Results

A discussion of the calibration parameters, which are the real image center, the scale factors, and the focal length, and how they are obtained, is found in this section.

1) Locate the real image center.

Before the experiment, a grid pattern is used to evaluate the feasibility of the proposed center determination method. First, the real image center is located by the proposed method. Then the software draws two crossed long lines with their intersections at the real image center using an overlay image. Third, place the pattern image under the ensoscope field of view. The pattern image should be placed by picking one square in the pattern image and aligning its sides with the two marked long lines. At the same time, align the chosen square's center with the real image center (see Figure 15). As we can see in figure 15, lines passing through real image center remain straight in the endoscope image. Numerical analysis also shows that for one row of squares that travel along the horizontal line, their centers have the same V_j coordinates. For one column of squares that travel along the vertical line, their centers have the same V_i coordinates, while other columns or rows of squares do not have this feature. It shows the proposed method is feasible.

(a)   (b)
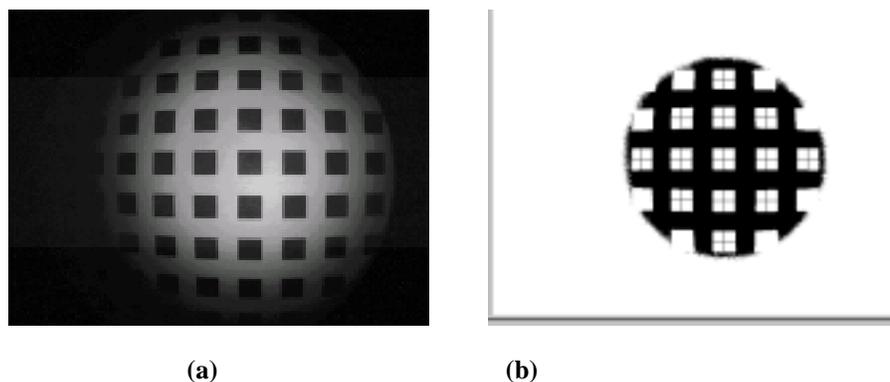
Figure 16. (a) Original gray scale image, (b) Binary image, the blob center have been marked.

Applying this method, the real image center could be obtained as follows. An image is recorded by pointing the endoscope perpendicular to a white background. The captured image is then processed by binarization using a threshold. The center of the white circle in the binarized image is chosen as the real image center (see Section 2, Figure 6).

2) Obtaining image coordinates of the calibration points.

The centers of squares are chosen as the points for calibration. Because this work is only interested in obtaining the intrinsic parameters, we only need to measure the adjacent square's center-to-center distance. Their values in the world coordinate unit are known.

Image coordinates of the calibration points are computed as follows:

a) Acquire a gray scale image (see Figure 16 (a)).

b) Threshold the image and calculate the center of every blob (see Figure 16 (b)). Because the threshold value is not critical and the illumination system is kept stable in our application (It can also be selected automatically by analysis of intensity histograms if necessary).

3) Calculating camera intrinsic parameters.

This section describes steps for calculating camera intrinsic parameters as well as the calculated intrinsic parameters of the endoscope used in this work.

a) Scale factors.

In this step, first we examine the validity of the assumption that the distortion in the center area is negligible. The closer the endoscope is to the object, the more distorted the image appears. After obtaining the real image center, the test grid is placed perpendicularly 50mm (the typical range for endoscope surgery is 40mm-100mm) away from the endoscope, with the center of one square at the real image center. Figure 16 shows the original gray scale endoscope image and the image after being binarized and marked the center of each blob. To reduce the influence of lens distortion, we use the 9 central squares for calculation. As shown in Table 1, from the coordinates of the center of

Table 1 V_i, V_j coordinates of the center of 9 squares in the central area (pixel unit)

(298, 196)	(363, 168)	(430, 169)
(297, 234)	(364, 234)	(431, 234)
(297, 300)	(363, 301)	(431, 300)

these 9 squares, their corresponding V_i or V_j coordinates are only at most 1 pixel difference. Given the first column in the table as an example, the coordinates of the three points along V_i direction are 298, 297, 297, only one pixel's difference from each other. This difference is negligible for our current application. We use mean values of the center-to-center distance between adjacent squares along V_i and V_j directions to calculate the scale factors.

$$\frac{S_x}{S_y} = \frac{66.67}{65.83} \simeq 1$$

b) Focal length.

Considering the above steps, we note that the scale factors in V_i and V_j directions are almost the same. In this task, we do not need to calculate the value of focal length in the world coordinate unit, the focal length can be calculated as:

$$f = \frac{(V_i - V_{i0})Z_c}{X_c} = \frac{66.67 \times 50}{8} = 416.7 \text{(pixel)}$$

3) Calibration Data

The depth information can be obtained by using the calculated focal length and equation (1) as described in Section 2. The image of the markers are record at various distances from the endoscope (40, 60, 70, 80 and 90 mm). The experimental data are shown in Table 2. Errors in this depth measurement experiment comes from different sources, such as the image processing error, the accuracy of physically measured distance, and the actual angle of the laparoscope with the experimental plane (we assume it as 90°).

Table 2. Experimental data via the proposed algorithm.

No.	Actual distance (mm)	Calculated distance (mm)	Errors (mm)
1	40	42.5	-2.5
2	60	60.5	-0.5
3	70	69.4	0.6
4	80	78.1	1.9
5	90	88.2	1.8

6.2 Distortion Correction Results

For the proposed distortion correction procedure, the same calibration chart containing a rectangular array of black squares is used.

The testing chart is placed perpendicularly to the camera's optical axis. Figure 17(a) shows the captured image. We compute the polynomial coefficients using the method proposed in Section 3. The corrected image is shown in Figure 17(b). The image is

(a)   (b)
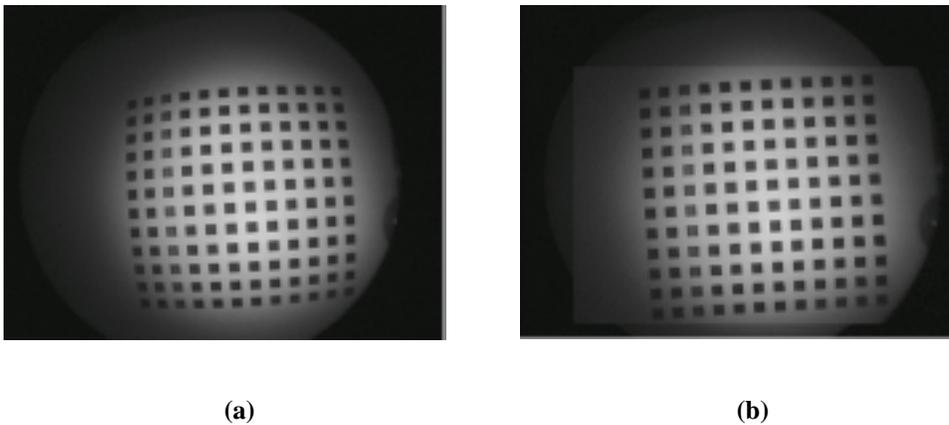
Figure 17. (a) Distorted image of testing image, (b) corrected image corresponding to (a) using inverse mapping.

magnified after correction, and lines are straightened especially in the peripheral part.

The least total error reduces when the order of the polynomial becomes higher. Nonetheless, with the increase of the order of polynomial, the computation time also increases. The variation in least-square total error was negligible beyond the 3rd polynomial order (see Table 3), therefore, we pick the polynomial order to be 3 as a compromise between computation time and error.

Table 3. Variation in least total error with respect to order of expansion polynomial .

Polynomial order n	1	3	5	7
Least square total error	1.9909	0.1445	0.04196	0.04113

The relationship between the radial distance of the testing points (in the experiment, we chose squares lies in the same row) from the real image center before and after distortion correction is shown in Table 4. Note that the calculated value has a decimal part. Since the image buffer accepts integer pixel locations, the floating number is rounded to its closest integer. After the distortion correction, the radial distance is more evenly distributed and closer to the actual radial distance.

Table 4. The radial distance of testing points from the real image center before and after distortion correction.

Idea Radial Distance(pixel)	22	56	84	112	140
Radial Distance (pixel) (before correction)	22	55	82	108	134
Radial Distance (pixel) (after correction)	22.4009 (22)	56.0093 (56)	84.0326 (84)	111.6783 (112)	140.1538 (140)

The above figures and experimental results show that the proposed distortion correction method yields satisfactory results for visual and computer analysis.

6.3 Endoscope Holder Simulation Results

To visualize and validate the endoscope holder design. OpenGL was used to simulate the animated control of the endoscope holder through the data obtained from the image tracking process.

Using Visual C++ split window techniques; the simulated endoscope holder and the real endoscope image captured in real time are shown simultaneously in Figure 18. The left side of the image shows the endoscope image while the right side shows the animated endocope holder (the small block represents the mark attached to the instrument). The endoscope holder moves according to the positional information obtained from the marker that attached in the instrument indicated in the left image. The endoscope holder can be controlled in real time by moving the tool under the endoscope.
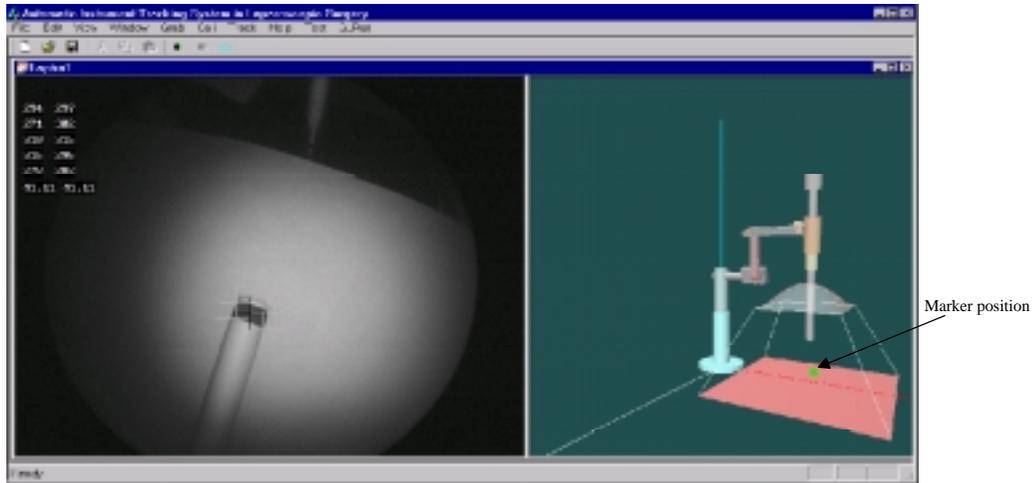
Figure 18. Endoscope holder simulation result.

6.4 Tracking Results

To estimate errors in automated instrument localization and tracking with the proposed method, several measurement methods are applied to evaluate the results. Besides direct measurement, pcBird system[12] (a position and orientation measurement system) is also used to obtain positional data for comparison with the tracking results.

Figure 19 shows the experimental setup. The endoscope is fixed into a stand in the visual tracking experiment. The pcBird comprises a transmitter, a receiver (sensor). The pcBird measures the position and orientation of a sensor by transmitting a pulsed DC magnetic field measured by the sensor. From the measured magnetic field characteristics, the chip computes the sensor's position and orientation with respect to the center of the transmitter. In the experiment, the sensor is mounted in the upper part of a tool, while the
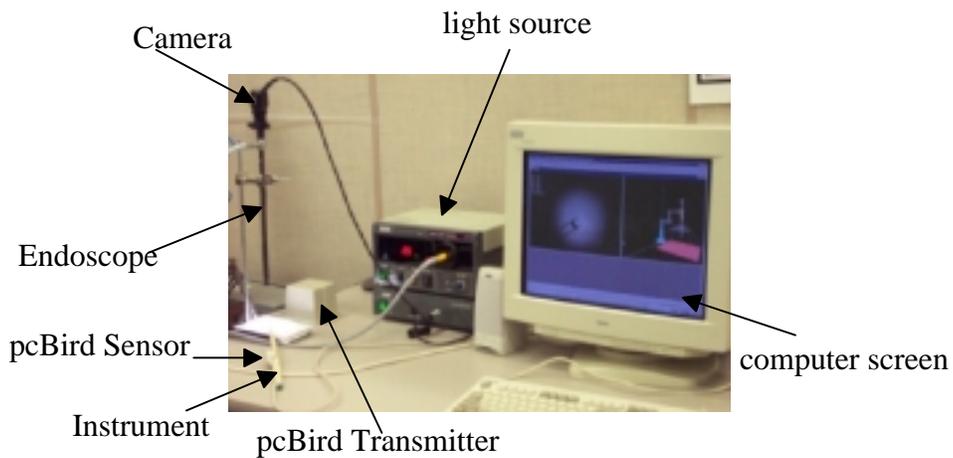
Figure 19. Experimental setup for the tracking experiment
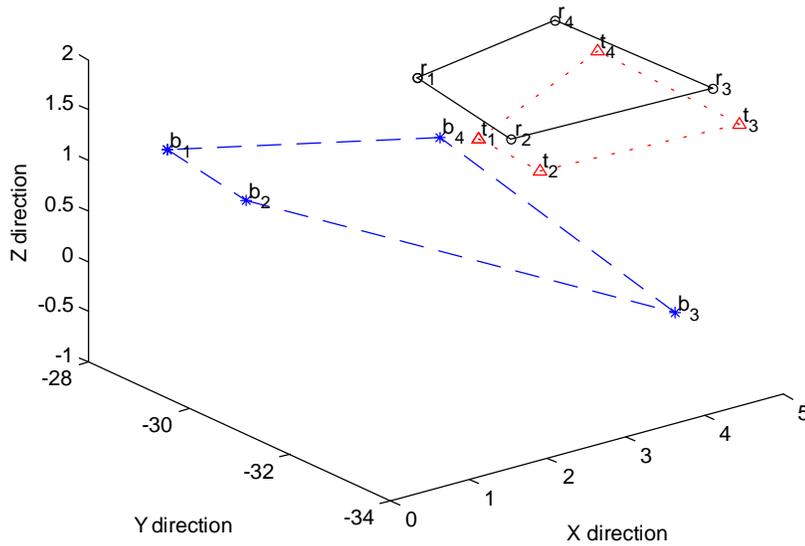
Figure 20. Comparison of tracking results (initial b represent the pcBird tracking results, t represents the proposed tracking results, r represents the ruler measured real positions).

maker is attached to the tip of the tool. Since the relative position of the sensor and the marker is fixed, the marker's position can be derived from the sensor's position. Thus, the marker's 3D position obtained from the pcBird and from the proposed image tracking method can be compared.

We define a world coordinate frame at the center of the transmitter. The camera coordinate frame is defined at the tip of the endoscope. Four points with different location are given for the experiment. We place the tool at the given location and let the center of the marker (select one strip of the marker) overlap with the given points. In this experiment, we capture ten frames in each location. The mean tracking error is defined as the average difference over a whole sequence between the actual measured position and the position obtained from the proposed tracking methods. A similar approach is applied to the pcBird position and orientation measurement system. The results are shown in Figure 20 and Table 5. The error is measured in terms of cm and in the x, y and z directions. From Figure 20, the visual tracking results are better than the pcBird position and orientation measurement system. One reason is related to the restrictions in using the pcBird. The pcBird is supposed to be used in a non-metallic environment. However the operating room environment cannot guarantee this restriction. The errors for both tracking methods are affected by imprecise manual placement of the marker at the known position.

Table 5. Average Error in Instrument Tracking

Sequence No. (10 frames)	Mean tracking error (cm)					
	x direction		y direction		z direction	
	pcBird	Image Tracking	pcBird	Image Tracking	pcBird	Image Tracking
1	- 2.0817	- 0.0286	1.7060	- 1.2817	-0.6661	-0.2999
2	- 1.9400	- 0.1123	2.2452	- 0.7557	-0.7152	-0.1177
3	- 0.3100	- 0.2132	0.2786	- 0.8689	-2.2257	-0.1177
4	- 0.9635	0.0200	0.7644	- 0.8301	-1.1250	-0.1177

7. Concluding Remarks

In this paper, some novel ideas of using visual tracking to realize automated instrument localization and endoscope maneuvering in robot-assisted laparoscopic surgery are presented. Implementation methods are described in detail. Good experimental results demonstrate the feasibility of the proposed methods. Some more experiments can be conducted in future work in the visual tracking part, such as, testing the tracking accuracy when the tools enter the endoscope field of view from different angles.

In our proposed method, we use the intensity information of the image for the marker localization, the specular reflection due to lights projected into the shining objects does not affect the tracking results. This is because in gray scale images, the intensity difference between the shining objects and the marker is even larger than surroundings, which is a desired feature for locating the marker.

The general goal of this work is to further improve the existing laparoscopic surgery system, to build a human – machine interface for the surgeon to automatically manipulate the endoscope. The system can be applied to surgical training simulators and telepresence (telepresence systems are used in medicine to manipulate equipment at a remote site. The surgeon has the sense of actually being at the site performing the procedure).

Our future research goal is to utilize this technique together with other sensors, such as laser sensors to reconstruct the surface of the surgical site, so to overcome the lack of 3D depth cue of the current laparoscopic surgery systems. The research will benefit the development of the laparoscopic surgery system in the long run.

References:

[1] L.Mettler, M. Ibrahim and W. Jonat, "One Year of Experience Working with the Aid of a Robotic Assistant the Voice-controlled optic holder AESOP in Gynaecological Endoscopic Surgery", *Human Reproduction*, Vol. 13, No. 10, 1998, pp. 2748-2750.

[2] David Prewer, Les Kitchen. "A Simple Fast Edge-Based Visual Tracker". *Technical Report* 97/20, pp. 1-14. http://citeseer.nj.nec.com/59426.html

[3] A. Azarbayejani and A. Pentland, "Real-time self-calibrating stereo person tracking using 3-D shape estimation from blob features", Proceedings of International Conference on Pattern Recognition , Aug. 1996, pp. 627—632.

[4] Reddi, S. and G. Loizou, "Analysis of camera behavior during tracking", IEEE Transactions on Pattern Analysis and Machine Intelligence, volume 17 (1995), number 8, pp. 765 – 778.

[5] M. E. Leventon, "Bayesian estimation of 3-d human motion from an image sequence" Technique report . July 1998. http://www.merl.com/reports/TR98-06/

[6] Y. Wang, D. R. Uecker and Y. Wang, "Choreographed Scope Maneuvering in Robotically-Assisted Laparoscopy with Active Vision Guidance", IEEE Workshop on Application of Computer Vision, Saratoga, FL, Dec. 1996, pp. 187 – 192.

[7] G. Wei, K. Arbter and G. Hirzinger, "Real Time Visual Servoing for Laparoscopic Surgery", IEEE in Engineering and Biology, Jan. 1997, pp. 40 – 45.

[8] C. Chen, S. Stitt and Y.F. Zheng, "Robotic Eye-in-Hand Calibration by Calibrating Optical Axis and Target Pattern", Journal of Intelligent and Robotic System, 12, 1995, pp. 155-173.

[9] R. Y. Tsai, "A Versatile Camera Calibration Technique for High- Accuracy 3D Machine Vision Metrology Using Off-the-Sheld TV Cameras and Lenses", IEEE Journal of Robotics and Automation, Vol. RA-3, No. 4. Aug. 1987, pp. 323-344.

[10] James D. Foley, Andries van Dam, Steven K. Feiner, John F. Hughes, Computer Graphics: Principles and Practice, Addison-Wesley 1995.

[11] K.V. Asari, S. Kumar and D. Rashakrishnan, "Technique of Distortion Correction in Endoscopic images Using a Polynomial Expansion", Medical &Biological Engineering & Computing, vol.37, 1999, pp 8-12.

[12] W. E. Smith, N. Vakil and S. A. Maislin, "Correction of Distortion in Endoscope Images", IEEE Transactions on Medical Imaging, Vol. 11. No. 1 1992, pp. 117- 122.

[13] P. Wolf, Elements of Photogrammetry, Mcgraw-Hill Inc.1983.

[14] KARL STORZ Endoscopy Canada Ltd, http://www.karlstorz.com/.

[15] Ascension Technology Corporation, http://www.ascension-tech.com.