Proceedings of the 28th IEEE
EMBS Annual International Conference
New York City, USA, Aug 30-Sept 3, 2006

Laparoscopic Image Analysis for Robotic Arm Guidance

Zacharias E. Gketsis, Dimitrios Tzagkas, Petros V. Hatzilias and Michalis E. Zervakis, *Member, IEEE*

***Abstract*— Some of the most important problems arising during a laparoscopic surgery relate to the limited perception of depth and field of view (fov). In this paper we explore the possibilities of dealing with these problems using appropriate re-modeling of the laparoscopic equipment and image processing algorithms, as to increase the perception ability of the surgeon within the operation space. We demonstrate that by using two camera units and appropriate estimation of the epipolar geometry between the camera views, we can acquire an accurate 3D map of the scene, as well as increased field of view. The benefits for the surgeon include the enhancement of the visible operating space and the increased perception of this space, with the ability of accurately estimating distances of sensitive abdomen regions and organs from the end of the laparoscope tip. Several simulation and real-world experiments are presented as to validate the potential of the proposed scheme**.

I. Introduction

LAPAROSCOPIC surgery is becoming increasingly important in a variety of biomedical areas. Besides reducing the duration and cost of operation, it provides benefits to the experienced surgeon in terms of accuracy and efficiency, as well as to the patient in terms of fast recovery and reduced pain. Significant research has been directed towards improving the surgical equipment and the camera quality [1, 2, 3, 4]. Nevertheless, there still exist open problems that reduce the functionality of equipment attained by the surgeon. Two of these problems relate to the lack of a 3D view with full distance information from vital organs and the restricted field of view attained by the camera. An increase of field of view, as well as the extraction of 3D coordinates from two cameras facing the same scene would enhance the surgeon's perception and his/her operational capability without imposing a new burden for them, as with the use of special 3D glasses.

These issues are explored in this paper, taking under consideration the restrictions and limitations of laparoscopy. In this environment, the abdomen wall operates as a constraint in the surgeon's movements. Furthermore, it provides the point of rotation for the laparoscope. Through the trocars the surgeon moves the surgical tools inside a conical region. Distortion-correction is often performed real-time, so that the surgeon sees on the monitor an undistorted version of the video sequence. However, in order to increase his/her field of view, the surgeon needs to move the laparoscope in the patient's abdomen, with the extra danger of approaching the area of vital organs. In order to alleviate such problems, we investigate the use of a two-camera laparoscope, where the cameras are mounted side-by-side on the near-end of the laparoscope, with only one of their axes in parallel [5]. This arrangement, with the conical region of movement and the views of the two cameras, are presented in Figure 1. With accurate estimation of the epipolar geometry between the two views, we can acquire the 3D map of the scene under consideration and estimate the distance of organs from the end tip of the laparoscope. Furthermore, using the two cameras we can easily generate a panoramic view and an increased field of vision and operation for the surgeon.

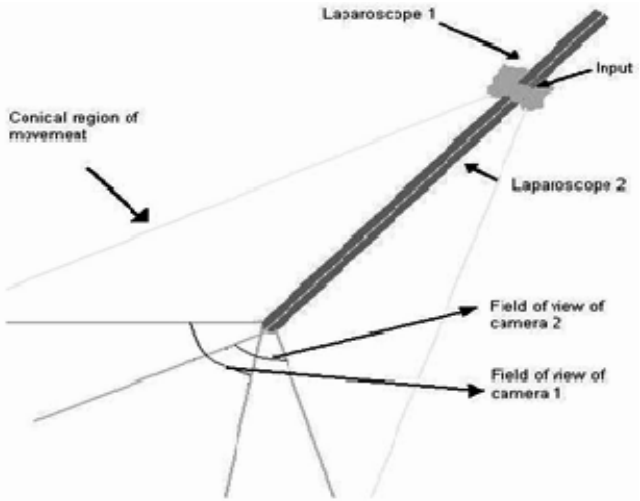

Fig. 1. Laparoscope's conical region of movement and field of view of the two cameras.

The paper is organized as follows. Section II describes the reconstruction processes for the 3D coordinates and the panorama. Section III illustrates our results for both virtual and real scene configurations.

II. 3D Reconstruction

A. Geometric model for our Stereo System

In this section the 3D coordinates of the tracked feature points are reconstructed from a stereo system. Let two cameras with image coordinates placed at the center of the

Manuscript received July 04, 2006. This work was supported in part by the EU IST project BIOPATTERN, Contact No: 508803.

Z.E. Gketsis, D. Tzagkas and M.E. Zervakis are with the Department of Electronics and Computer Engineering, at the Technical University of Crete, Chania, 73100, Greece (corresponding author M.E. Zervakis:
tel: +30-28210-37206; fax: +30-28210-37542;
e-mail: michalis@danai.systems.tuc.gr).

P.V. Hatzilias is with the District Hospital of Chania, Director of the Urology Department, Chania, 73100, Greece.

image sensor and denoted by (X_1, Y_1, Z_1) and (X_2, Y_2, Z_2), respectively, with the world coordinate system denoted by (X, Y, Z). The simplest stereo concatenation has the two cameras in parallel, with only a baseline displacement b between them, as in Figure 2. Let also (V_{1i},V_{1j}) and (V_{2i},V_{2j}), in pixel units, denote the image coordinates at **x** and **y** directions for the left and right cameras, respectively, and (V_{1i0},V_{1j0}) , (V_{2i0},V_{2j0}) denote the left and right camera centers.

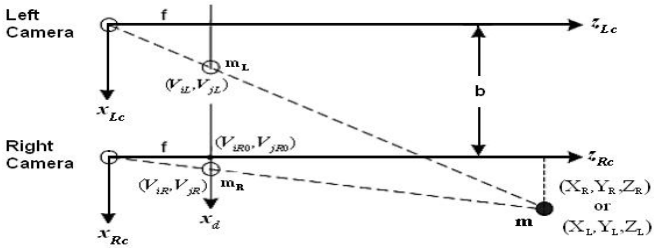

Fig. 2. Geometric model for the stereo system

The reconstruction formulas for X, Y and Z coordinates of the 3D reconstructed point **m** can be extracted using homogeneous coordinates. Thus, for the image plane of the right camera (V_{2i},V_{2j}) we obtain:

$$u=K\cdot P\cdot m \Leftrightarrow$$

$$\begin{bmatrix} w\cdot V_{2i} \\ w\cdot V_{2j} \\ w \end{bmatrix} = \begin{bmatrix} f_x & -f_x\cdot\cot(\theta) & V_{2i0} & 0 \\ 0 & \frac{f_y}{\sin(\theta)} & V_{2j0} & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} \cos(\phi) & 0 & \sin(\phi) & t_x \\ 0 & 1 & 0 & t_y \\ -\sin(\phi) & 0 & \cos(\phi) & t_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \quad (1)$$

The unitary matrix **P** denotes the 4x4 homogeneous transform that encapsulates the transformation from world to camera coordinate systems. The above form includes a shift and a rotation of the camera around its **y** axis, as in the camera concatenation of Figure 1. Furthermore, **K** denotes the (intrinsic) camera *calibration matrix* inducing the skew angle θ, which is most often almost 90°, with deviations introduced by sensor manufacturing (not square pixels).

In the simple case of parallel cameras (Figure 2) the transformation model is simplified. Thus, for the right camera on its own coordinate system (X_2, Y_2, Z_2) we obtain:

$$\begin{bmatrix} w\cdot V_{2i} \\ w\cdot V_{2j} \\ w \end{bmatrix} = \begin{bmatrix} f_x & 0 & V_{2i0} & 0 \\ 0 & f_y & V_{2j0} & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} X_2 \\ Y_2 \\ Z_2 \\ 1 \end{bmatrix} \Rightarrow \begin{matrix} X_2=\frac{(V_{2i}-V_{2i0})\cdot Z_2}{f} \\ Y_2=\frac{(V_{2j0}-V_{2j})\cdot Z_2}{f} \end{matrix} \quad (2)$$

Considering the corresponding coordinate relations for the left camera (X_1, Y_1, Z_1) and the relationships between the two camera coordinate systems:

$$X_2-X_1=b \ , \ Z_2=Z_1=Z$$

we obtain the estimate of the **z** coordinate as:

$$Z=\frac{f\cdot b}{(V_{2i}-V_{1i})}=\frac{f\cdot b}{\text{disparity}} \quad (3)$$

The focal length f (in pixels) is a known parameter (computed from calibration) and the baseline distance b between the two cameras is fixed (in mm). The only unknown is the disparity value, which is the difference (in pixel units) between the **x** coordinates of a point present in the left camera's image plane and the same point present in the right camera's image plane. If we know the disparity for two matching points in the two cameras, then the computation of the **z** distance of that point in the world coordinate system can be readily derived from eqn. (3). Several algorithms have been proposed for this, so called, *correspondence problem*. Some of them operate in the region of isolated points and some others operate in the area of the entire image plane.

Our proposed approach is based on a combination of these two schemes, as presented in the next section. Another point of concern in dealing with eqn (3) is the inverse relation between the point distance and the disparity induced in the two camera planes. Due to limited pixel resolution, the estimation error increases with the actual distance from the camera planes. This is also evidenced in the examples presented in Section III. Nevertheless, for the range of distances to be handled in laparoscopic applications, this error appears tolerable.

B. The Correspondence Problem

Given two images formed in two image planes the problem is posed as follows: for a point $\mathbf{u_1}$ in the first plane, which point $\mathbf{u_2}$ in the second plane corresponds to $\mathbf{u_1}$? The matching becomes more difficult as the difference in position (and orientation) of the stereo views increases.

A typical approach for matching seeks to find feature points in each image and then attempts to match them on the basis of their surrounding regions. For instance, the Harris corner detector [6] could be used to extract feature points with a strong 2D component, for which matching become simple and efficient through correlation. A small *correlation window* (pixel patch) is centered on each feature point of the first image and a *normalized cross-correlation* algorithm is then employed to find the best matches in the second image. This process is also applied in reverse order and matches are accepted if they exhibit a maximum in both comparisons.

The previous algorithmic scheme for solving the correspondence problem is quite simple and fast, but operates on isolated image regions. As such, its efficiency is low and it often results in false matches producing heavy outliers in the disparity estimation process. To reduce the effects of outliers in the estimation process, several regression approaches have been proposed employing parametric modeling of the matching relationship and involving multiple matched windows for the estimation of model parameters.

Let a point (V_{1i}, V_{1j}) in the first image (left camera) with homogeneous coordinates $\mathbf{u_1}$ appearing in the second image at (V_{2i}, V_{2j}) with homogeneous coordinates $\mathbf{u_2}$**.** There is a 2D homography $H_p:\Re^2\rightarrow\Re^2$ that links these two points as:

$$u_2=H_p u_1 \quad (4)$$

This image warping is in essence a transformation that changes the spatial configuration of an image. In our

application we consider the particular type of warp namely the Euclidean warp, also called the Euclidean similarity transform:

$$\begin{bmatrix} V_{2i} \\ V_{2j} \\ 1 \end{bmatrix} = \begin{bmatrix} s\cdot\cos\alpha & s\cdot\sin\alpha & t_x \\ -s\cdot\sin\alpha & s\cdot\cos\alpha & t_y \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} V_{1i} \\ V_{1j} \\ 1 \end{bmatrix} \quad (5)$$

Rearranging $\mathbf{H_p}$ in vector form ($\mathbf{h_p}$) and the vector $\mathbf{u_1}$ in matrix form ($\mathbf{Z_1}$), this equation is also written in terms of the parameter vector as:

$$u_2 = Z_1 h_p \quad (6)$$

The problem of parameter estimation in (5) or (6) has been extensively considered in motion estimation for video segmentation [7]. Least squares and robust regression schemes have been employed for the derivation of the parameter vector from a number of matched points. Furthermore, clustering schemes in the parameter space have been considered for a more global consideration of the spatial effects of the warping transform.

Experimenting with such models in the restricted environment of laparoscopic applications, we found it more efficient for accurate estimation to employ a scheme that also takes under consideration the epipolar geometry of the stereo system. The epipolar geometry can be represented in algebraic terms by the so called 3x3 *Fundamental Matrix* **F** [8]. Knowledge of this matrix enables us to accurately map each feature point on the first image plane, to its corresponding point on the other plane. The epipolar line $\boldsymbol{l_2}$ is defined on the second image from the image point $\mathbf{u_2}$ and the epipole $\mathbf{e_2}$, which is the point of intersection of the line joining the camera centers (the baseline) with the image plane. Hence:

$$l_2 = [e_2]_x u_2 = [e_2]_x H_p u_1 = F u_1 \quad (7)$$

where $\mathbf{F} = [\mathbf{e}_2]_x\mathbf{H}_p$ defines the Fundamental matrix. Notice that **F** being 3x3 matrix has only rank 2 (because $[\mathbf{e}_2]_x$ has rank 2), whereas $\mathbf{H}_p$ has rank 3. Recall that in homogeneous coordinates a line is represented by the vector perpendicular to it. Thus, the dot product between the homogeneous coordinates of a point on the line and the homogeneous coordinates of the line is zero, leading to:

$$u_2^T \cdot l_2 = u_2^T \cdot F \cdot u_1 = 0 \quad (8)$$

Alternatively, we can write $\mathbf{Af = 0}$ by rearranging **F** in vector form (**f**) and the vectors $\mathbf{u_1}$ and $\mathbf{u_2}$ in matrix form (**A**).

Given enough corresponding 2D point matches $\mathbf{u_{1i} \leftrightarrow u_{2i}}$ we can set up a set of equations that allow us to solve for **F**. Note that **F** can only be determined up to a scale factor so eight matching points are sufficient. Examples of robust methods are M-Estimators, Least-Median-Squares (LMedS), RANdom SAmpling Consesus (RANSAC), MLESAC (Maximum LikElihood SAmple Consensus) and MAPSAC (Maximum A Posteriori SAmple Consensus) which can be used both in the presence of outliers and in bad point localization [8].

Random sample consensus (RANSAC) provides a general technique for model fitting in the presence of outliers. We used RANSAC algorithm because it models epipolar geometry closer to reality, than other methods mentioned [8]. The algorithm defines the *minimal* number of points needed to specify the model (e.g. n=8, for the case of the 8-point algorithm) and fits the model to a randomly selected minimal subset. It applies the transformation to the complete set of points and counts inliers. If the number of inliers exceeds a set threshold, the algorithm flags the fit as appropriate and stops.

C. Panoramic Image generation

One application for image warping is merging of several images into a complete mosaic –e.g. to form a panoramic view [9]. Our formulation of Euclidean warp readily allows the creation of panoramic images from the two camera views. In case of small baseline distances, as in our case, the Fundamental matrix enables the efficient estimation of correspondences between the two images. The estimation of parameters in eqn (4) or (8) readily provides the transform for mapping one image on the coordinates of the other, allowing for direct stitching of images and panorama generation.

D. Proposed Implementation

Through our investigations, we highlight two points of novel contribution. The first is towards reducing the error induced in distance computation through eqn (3), due to the limited pixel resolution. For this problem, we employ a perspective transform [3] before any other computation, as to decouple the resolution appearing on the camera planes from the actual distance of the point considered.

The second area of investigation relates to the real time implementation of the distance computation scheme. We employ a retrospective operation similar to the RANSAC approach, where the model is first derived from a minimal number of points and then it is validated by direct application on the point-set of interest. Towards this direction, one could easily notice the strong coupling of the affine transform parameters in eqn (5) with the depth of field under consideration (**z**-distance). For fixed camera concatenation, the parameter set could be predetermined and used in the form of look-up tables as a function of distance. Our proposed approach for real-time implementation relies on a validation scheme of the following form. From a small number of points in the two camera images, determine **z**-distance using simple neighborhood matching. For the "crudely" estimated distance recover the transform parameters and backproject points from the left to the right camera. Finally, determine a distance metric for points on the right camera from their corresponding backprojected counterparts. This metric should be within tolerance limits if the initial distance estimation is accurate, in which case the result is presented to the surgeon with a set confidence. The proposed implementation strategy is simple, fast and quite reliable for small distances, as discussed in the next

section.

III. Simulation And Results

We have created a virtual space using the program Virtual Reality and Simulink Toolbox of MATLAB (*The Mathworks,Inc*). For evaluation purposes we develop a simulation model of AESOP1000 robotic arm [10], with two cameras mounted on the far-end of the arm. Through VRML language we locate the arm's model in various virtual spaces for extracting 3D coordinates of the various objects present in the scene. Through ray tracing, we simulate the views of two cameras observing the virtual space. Several experiments in different environments have been conducted, some of which are presented here. In one experiment, we simulate a space with two kidneys (figure 3). The distance between the cameras and the kidneys' centers of mass is 200mm. The baseline between the two cameras is now 20mm. The estimated X, Y, Z coordinates for several feature points are provided. In particular, the estimation of the z-distance is quite accurate for points at different depths.

For real-space applications, we use a set of Creative PCCAM 550 digital cameras with 1.3Mpixels CCD sensor. The focal lengths for 640x480 image resolution were found to be $f_x \approx f_y \approx 750$ pixels. The object of interest appears in distance Z_{real}=148mm. The cameras' centers translation is 4.5mm (baseline distance).

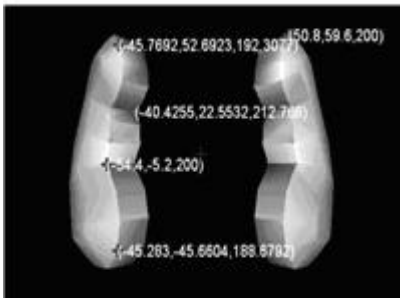

Fig. 3. Results of the 3D reconstruction algorithm in virtual environment

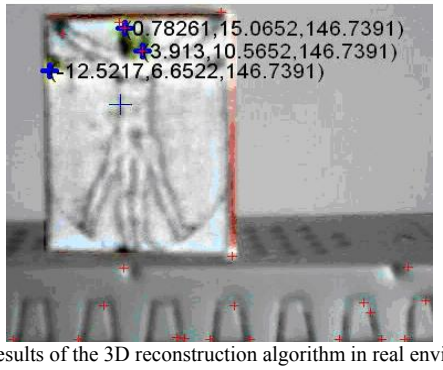

Fig. 4. Results of the 3D reconstruction algorithm in real environment: Z_{real}=148 mm, Z_{est}=146.7391 mm

In all our experiments (in either virtual or real spaces), we observe an increase of the deviation error for the distance of sampling points of interest. A plot of this error in terms of the actual distance is presented in Figure 5, providing measures of ambiguity in distance estimation through our proposed scheme. In Figure 5, we present the absolute error $|\text{mean}(\mathbf{Z}_{est})-Z_{real}|$ as a variant of Z_{real}. For distances smaller than 200mm, the mean error is lower than 4mm, which is acceptable by the surgeon.

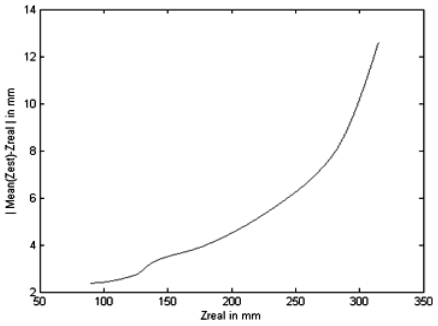

Fig. 5. Absolute error $|\text{mean}(\mathbf{Z}_{est})-Z_{real}|$ in millimeters with respect to Z_{real}

IV. Conclusions

This paper focuses on modeling of a stereo vision system for robotic arm guidance applications, such as Laparoscopic surgery. We investigate the use of a two-camera laparoscope and exploit the epipolar geometry between the camera views as to acquire an accurate 3D map of the scene, as well as increased field of view. For real time implementation we propose a retrospective scheme for parameter estimation. We demonstrate the application of the proposed scheme in both virtual and real environments. The estimation accuracy increases with the baseline distance and decreases with the depth of view. Under the limitations of the laparoscopic set-up, the accuracy achieved is reasonable and acceptable by surgeons.

References

[1] Zhengyou Zhang, "A flexible new technique for camera calibration," Microsoft research, March 1999.

[2] Chao Zhang, James P. Helferty, Geoffrey McLennan, William E. Higgins, "Non-linear distortion correction in endoscopic video images," *ICIP 2000*.

[3] H. Tian, T. Srikanthan, K. Vijayan Asari, S. K. Lam, " Study on the effect of object to camera distance on polynomial expansion coefficients in barrel distortion correction," *Proceedings of the 2002 IEEE Southwest Symposium on Image Analysis and Interpretation,* 2002, pp. 255-259.

[4] James P.Helferty, Chao Zhang, Geoffrey McLennan, W.E.Higgins "Videoendoscopic distortion correction and its application to virtual guidance of endoscopy," *IEEE Trans. Med. Imaging,* Vol. 20, July. 2001, pp. 605-617.

[5] Richard Hartley and Andrew Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, Second edition, pp. 239-261.

[6] Chris Harris and Mike Stephens, "A combined corner and edge detector," *Proceedings of the Fourth Alvey Vision Conference,* Manchester, pp 147-151, 1998.

[7] D. Zhang and G. Lu, "Segmentation of Moving Objects in Image Sequence: A Review," *Circuit, Systems, and Signal Processing*, 20(2), 143–189, 2001.

[8] P.H.S. Torr and D.W. Murray, "The development and comparison of robust methods for estimating the Fundamental matrix," *International Journal of computer vision,* September 1997.

[9] M.B. Stegmann, "Image Warping," Informatics and Mathematical Modelling, Technical University of Denmark, 2001.

[10] P.Chatzilias, Z. Kamarianakis, S. Golemati, M.Christodoulou, "Robotic control in hand-assisted laparoscopic nephrectomy in humans– a pilot study," *Proceedings of the 26th Annual International Conference of the IEEE EMBS,* 2004, pp. 2742-2745.