31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

Stereo Endoscopy as a 3-D Measurement Tool

Matthew Field, Duncan Clarke, Stephen Strup, MD, and W. Brent Seales

Abstract— **Stereoscopic endoscopes restore much of the sense of depth perception lost in minimally invasive surgery techniques. They also open the possibility of anatomical reconstruction from in-vivo video using established computer vision techniques. However, limitations in computational power, as well as imaging constraints due to specularities, color, and movement, make full three-dimensional reconstruction difficult to achieve in real time.**

We demonstrate that a subset of this reconstruction problem, the ability to take accurate anatomical measurements, can be trivially achieved using current stereo-endoscopes and well-known algorithms. After careful calibration of the camera pair, we track the surgical instruments, which present the strongest features in the operating scene. The three-dimensional positions of the tips of the instruments are automatically calculated in real-time, and the computation of the distance between them allows for accurate measurements of any anatomical region within reach of the instrument tips.

I. INTRODUCTION

Minimally invasive surgery (MIS) has greatly improved outcomes for patients, but has significantly complicated the task of surgeons. In particular, the limited field of view, lack of depth perception, and reduced tactile response all make performing MIS much more challenging than traditional open surgery. Technological advances continue to close this gap, particularly in the visualization of the surgical field. High-definition cameras and better optics have greatly improved the quality of endoscopic video. Stereoscopic endoscopes return a sense of depth and open myriad possibilities for applying computer vision to intra-operative video. We demonstrate that a calibrated stereo endoscope can take measurements of anatomy during surgery that rival current techniques involving the introduction of another object (a tape measure) into the scene.

A. Related Work

Most previous work in endoscopic reconstruction has focused on monocular scopes [1], [2]. Typically, this work has relied on an external optical tracking systems to locate the camera relative to the patient in an attempt to register intra-operative imagery with pre-operative scans. Fuchs et al [3] used a structured light reconstruction that combined a standard scope with a secondary instrument that emitted a light stripe. While the results were promising, it required an extra instrument to be inserted into the patient to achieve its reconstruction. Devernay et al [4] calibrated a stereoscope similar to the one in our work as part of an augmented reality system for heart surgery. All of these methods showed promise, but required additional equipment beyond the standard operating suite. Our algorithms run on stereoscopic video alone, and introduce little additional burden on operating room personnel.

Instrument tracking has also been studied, particularly as it applies to robotic surgery [5], [6], [7]. Most use color classification to distinguish the instruments from the anatomy. Our method relies on the use of optical markers on the instrument barrels, but future refinements should eliminate this need.

This work was supported by project STITCH, funded by DoD TATRC grant W81XWH-06-1-0761.

M. Field and D. Clarke are with the Center for Visualization and Virtual Environments at the University of Kentucky. Email: field@vis.uky.edu, dclarke@fremontassociates.com

S. Strup is the William S. Farish Professor of Surgery at the University of Kentucky. Email: sstru2@email.uky.edu

W. B. Seales is the Gill Professor of Computer Science and Director of the Center for Visualization and Virtual Environments at the University of Kentucky. Email: seales@netlab.uky.edu

II. METHOD

Our system makes use of the properties of stereoscopic cameras to take measurements of features using laparoscopic video. The first step in this process is the calibration of the stereo camera pair. The second involves tracking of the surgical instruments, the most easily recognizable objects within the scene. Measurements are then taken by placing the instruments adjacent to the regions of interest and computing the distance (in 3-space) between the instrument tips.

For these tests we captured video from the stereo endoscope of the *da Vinci*® Surgical System, a surgical robot manufactured by Intuitive Surgical. Our current results are based on the system's standard analog NTSC 640x480 video output.

A. Camera Calibration

The first step in 3D reconstruction from stereoscopic video is the calibration of the cameras. Calibration determines the properties of the camera by modeling it as a pinhole camera. The focus and center of projection of such a model describes the way in which light from the 3D scene is projected onto a 2D image plane. Additional parameters model the effects of distortion due to the imperfect nature of the camera lens.

Many different models and algorithms have been developed over the years to perform this calibration [8], [9], [10]. We use the method of Zhang [10] as implemented by the Open Source Computer Vision Library, OpenCV [11]. A calibration target consisting of a black-and-white checkerboard pattern of known size is imaged at several different orientations. The corners of the target are extracted with sub-pixel accuracy, and their positions used to find a maximum likelihood estimate for the camera parameters. Once the properties of the left and right cameras have been

independently calculated, point matches from the left and right images can be used for determining the rotation and translation from one camera to the other.

Each camera is modeled as a matrix K that satisifies the following equation, where m is the homogeneous image coordinate $(x, y, 1)^T$, M is the homogeneous world coordinate $(X, Y, Z, 1)^T$, R and t define a rotation and translation, and s is an arbitrary scale factor.

$$sm = K\,[R \mid t]\,M \tag{1}$$

The camera matrix K is composed of the focal length (f_x,f_y) and center of projection (c_x,c_y).

$$K = \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \tag{2}$$

Image distortion is modeled radially as a fourth order polynomial, with two coefficients of radial distortion (κ_1, κ_2) and two coefficients of tangential distortion (ρ_1, ρ_2), where (x_d, y_d) is the distorted image coordinate and (x_n, y_n) is the normalized, undistorted coordinate.

$$\begin{bmatrix} x_d \\ y_d \end{bmatrix} = \begin{bmatrix} x_n \delta x_d^{(r)} + \delta x_d^{(t)} \\ y_n \delta y_d^{(r)} + \delta y_d^{(t)} \end{bmatrix} \tag{3}$$

$$\begin{bmatrix} \delta x_d^{(r)} \\ \delta y_d^{(r)} \end{bmatrix} = \begin{bmatrix} 1 + \kappa_1 r_n^2 + \kappa_2 r_n^4 + \dots \\ 1 + \kappa_1 r_n^2 + \kappa_2 r_n^4 + \dots \end{bmatrix} \tag{4}$$

$$\begin{bmatrix} \delta x_d^{(t)} \\ \delta y_d^{(t)} \end{bmatrix} = \begin{bmatrix} 2\rho_1 x_n y_n + \rho_2 (r_n^2 + 2x_n^2) \\ \rho_1 (r_n^2 + 2y_n^2) + 2\rho_2 x_n y_n \end{bmatrix} \tag{5}$$

Our calibration target consisted of an $8x11$ planar checkerboard pattern with $1/8$-inch grid spacing.

B. Instrument Tracking

Previous work has attempted to separate surgical instruments from the anatomy by use of color histogram matching of standard instruments [5], [6], [7]. While that method is more generalizable to current surgical procedures, it is better suited to the detection of image regions rather than individual features. In order to track the instruments with sufficiently high precision for taking measurements with stereo triangulation, we added high-contrast optical markers to the barrels of the instruments. These markers consisted of an alternating pattern of one-inch long ($25mm$) black and white bars, much like a bar code, aligned with the direction of the instrument barrel. The corners of these bars can be tracked with subpixel resolution, much like those of our calibration target.

Once a pair of features is matched in the left and right camera images, it can be projected into 3D space to a real-world point relative to our cameras. Any image point can be modeled as a ray projected outward from the camera center through the image plane to its three-dimensional location. This ray, when viewed from the perspective of the other camera, is called an epipolar line, and will pass through the same feature point in the second camera's image [12]. Using this knowledge, the pair of rays created by a matched feature in each image should intersect at the true three-dimensional coordinate of that feature. In real images, these rays rarely intersect due to perturbations from image noise and errors in feature detection, but the true location can be approximated by calculating the midpoint of the shortest line segment joining the two rays. We use this method of stereo triangulation to determine the 3D locations of features in the scene.

C. Measurement

The instrument tips are placed alongside a region of interest to measure. The resulting image is thresholded such that the high contrast optical markers are isolated, and the Hough transform extracts the parallel lines of the markers. The end points of one line on each instrument are tracked, and the location of the end of the instrument is calculated by following this vector away from the camera the known distance from the marker to the tip. Once the tips of both instruments are tracked to their 3D locations, then the distance between the two coordinates is calculated as $d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$.

Our test measurements were taken on a peg-board used for training. The known distance between these pegs was compared to the approximate positions calculated using stereo triangulation.

III. RESULTS

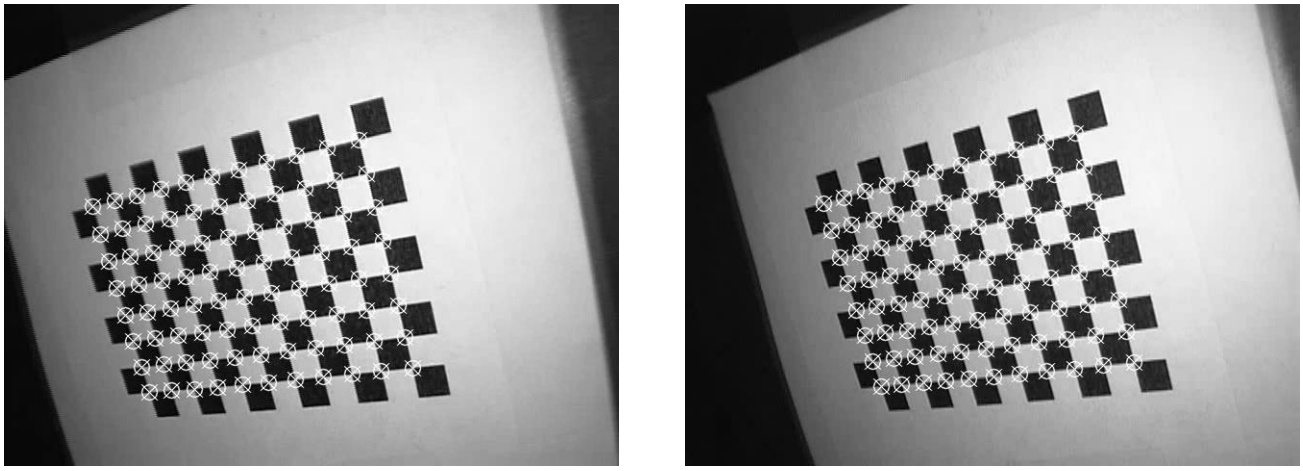
A. Calibration

Our sample data included approximately 3 minutes of video recorded synchronously from the left and right cameras at analog NTSC resolution (Figure 1). From thirty seconds of video of the checkerboard target, twenty "good" images were automatically selected based on criteria of minimal motion, to reduce blur, and a minimum time difference, to reduce the weighting of any particular orientation. The camera calibration routine from the OpenCV library was run over each dataset to determine the camera matrix for the left and right cameras, as well as the transformation from one camera coordinate space to the other. Total runtime for calibration was less than two minutes.

The separation between the left and right cameras was approximately $4.6mm$. The distance for accurate stereo-reconstruction is directly related to this baseline separation and to image noise [12]. Based on simulation at similar noise levels to our images, the working volume for 3D data accurate to within $5mm$ extends to a depth of roughly $150mm$ from the tip of the endoscope [13].

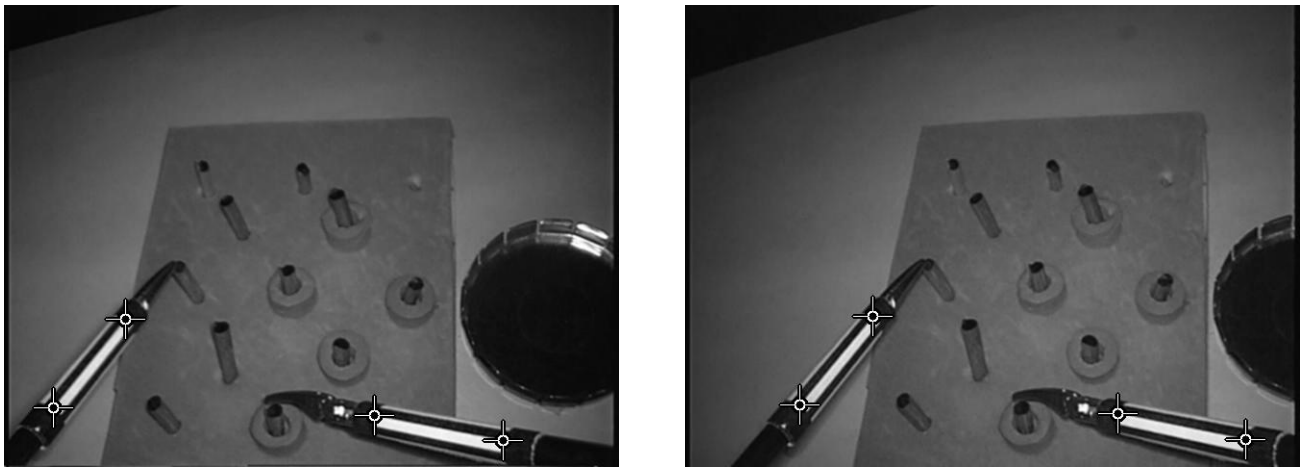
B. Measurement

As shown in Figure 2, the instruments were placed at either end of the region of interest for each measurement. The location of the ends of a single bar in the target pattern were tracked on each instrument. Before the procedure the distance from the end of the pattern to the end of the instrument was measured to be $30mm$ for the right instrument and $25mm$ for the left instrument. The vector from the nearer point on the pattern to the farther point was extended by



(a) (b)

Fig. 1. Synchronized pair of calibration images from the left and right cameras of a stereoscopic endoscope, with detected feature points highlighted.



(a) (b)

Fig. 2. Automatic instrument tracking on a pair of stereo images from an endoscope.

the appropriate distance to approximate the location of the tip of the instrument. This neglects the curved nature of the instrument ends, but adds significantly less than $2.5mm$ error on a $5mm$ surgical instrument.

In total, 16 separate distances were measured over a sample of greater than 500 frames of video. Feature tracking in the video runs at about 15 frames per second (fps) on a 3.0GHz Pentium D workstation, though further optimization or multi-thread processing could bring this closer to the full frame rate of 30 fps. The number of frames per measurement varied from 13 to 53, or between 0.4 and 1.8 seconds of video. The reasonably large sample size allows us to compensate for image noise in any single frame for a more accurate measurement, and to develop a better profile for this noise range and its effects. Ground truth data was obtained with a measuring tape placed between the pegs on our board.

Table I shows each measurement, computed as the mean and standard deviation over the set. Of the sixteen measurements, the means were all within $4mm$ of the true values, and most were within $2mm$. Figure 3 shows these errors as a box plot. These measurements were computed over depths ranging from $95mm$ to $165mm$ from the endoscope.

C. Applications

Accurate anatomical measurements obtained through the scope have immediate utility in some procedures, such as hernia repair. In such a procedure, a mesh is custom trimmed to cover the hernia defect. If the size of the hernia cannot be accurately determined from external palpation, a measuring tape is often inserted through one of the ports and held in place with the instruments to judge the dimensions of the defect. Figure 4 presents a frame of video from an actual anterior abdominal hernia repair showing the use of a tape measure and graspers to measure the area to be repaired. With the addition of rudimentary user interface elements our technique could replace this tape for more accurate measurement without the need to introduce foreign objects into the body and account for their removal afterward.

The restricted view of surgery may preclude the use of obtrusive optical markers, as the entire pattern must be

TABLE I

MEASUREMENTS

Trial	Ground Truth	Mean $\pm$ Std Dev	Frame Count
1	$19mm$	18.65 ± 2.87	48
2	$36mm$	35.21 ± 2.21	34
3	$28mm$	28.45 ± 2.96	53
4	$30mm$	28.97 ± 3.05	39
5	$42mm$	42.07 ± 2.29	48
6	$40mm$	38.74 ± 2.92	45
7	$15mm$	12.00 ± 1.05	40
8	$22mm$	20.00 ± 0.68	30
9	$20mm$	18.92 ± 1.50	24
10	$39mm$	38.64 ± 1.28	52
11	$28mm$	25.10 ± 2.64	27
12	$37mm$	34.68 ± 1.54	13
13	$25mm$	21.28 ± 2.77	33
14	$34mm$	31.20 ± 1.31	27
15	$36mm$	33.27 ± 0.58	18
16	$18mm$	16.01 ± 2.18	46

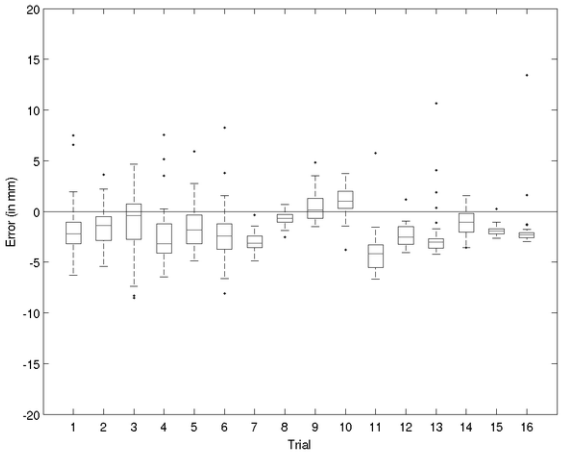


Fig. 3. Box and whiskers plot of the error ranges for each measurement.

visible to project to the instrument tip. Future work will focus on directly tracking surgical instruments without the need for a tracking pattern.

IV. CONCLUSIONS AND FUTURE WORK

We have demonstrated that highly reliable measurements of three-dimensional objects can be quickly made through the lens of a stereoscopic endoscope. These measurements are accurate to within $5mm$ for a useful working volume using our endoscope.

Further refinements of our software will increase the speed and simplicity of calibration and fully automate instrument tracking so that no user intervention is required. The optical markers attached to the instruments allowed for rapid testing of our procedure, but may not be practical in an operating environment. Future work will focus on (1) directly tracking a wide range of standard surgical instruments, (2) efficient capture and analysis of High Definition digital video frames, and (3) the creation of a user interface appropriate for clinical use. For relatively static scenes, it may be possible to extend our stereoscopic techniques to monocular endoscopes by using multiple views at differing orientations.

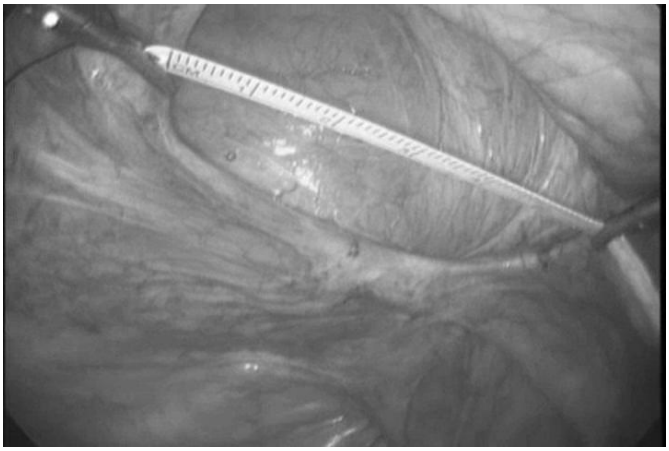


Fig. 4. Example of current technique for laparoscopically measuring a hernia defect. A calibrated stereo-endoscope could perform the same measurement without the tape.

REFERENCES

[1] R. Shahidi, M. R. Bax, J. Calvin R. Maurer, J. A. Johnson, E. P. Wilkinson, B. Wang, J. B. West, M. J. Citardi, K. H. Manwaring, and R. Khadem, "Implementation, calibration and accuracy testing of an image-enhanced endoscopy system," *IEEE Trans. Med. Imag.*, vol. 21, no. 12, pp. 1524–1535, December 2002.

[2] S. Szpala, M. Wierzbicki, G. Guiraudon, and T. M. Peters, "Real-time fusion of endoscopic views with dynamic 3-d cardiac images: A phantom study," *IEEE Trans. Med. Imag.*, vol. 24, no. 9, pp. 1207–1215, September 2005.

[3] H. Fuchs, M. A. Livingston, R. Raskar, D. Colucci, K. Keller, A. State, J. R. Crawford, P. Rademacher, S. H. Drake, and M. Anthony A. Meyer, "Augmented reality visualization for laparoscopic surgery," in *Proceedings of Medical Image Computing and Computer-Assisted Intervention '98 (MICCAI'98)*, 1998.

[4] F. Devernay, F. Mourgues, and E. Coste-Manière, "Towards endoscopic augmented reality for robotically assisted minimally invasive cardiac surgery," in *Proceedings of Medical Imaging and Augmented Reality*, 2001.

[5] D. R. Uecker, C. Lee, Y. F. Wang, and Y. Wang, "Automated instrument tracking in robotically-assisted laparoscopic surgery," *J. of Image Guided Surgery*, vol. 1, 1995.

[6] G.-Q. Wei, K. Arbter, and G. Hirzinger, "Automatic tracking of laparoscopic instruments by color coding," in *CVRMed-MRCAS '97: Proceedings of the First Joint Conference on Computer Vision, Virtual Reality and Robotics in Medicine and Medial Robotics and Computer-Assisted Surgery*. London, UK: Springer-Verlag, 1997, pp. 357–366.

[7] D. R. Uecker and Y. Wang, "A new framework for vision-enabled and robotically-assisted minimally invasive surgery," *Computerized Medical Imaging and Graphics*, vol. 22, pp. 429–437, 1998.

[8] R. Y. Tsai, "A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv cameras and lenses," *IEEE Trans. Robot. Automat.*, vol. 3, no. 4, pp. 323–344, August 1987.

[9] J. Heikkilä and O. Silvén, "A four-step camera calibration procedure with implicit image correction," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'97)*, 1997, pp. 1106–1112.

[10] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 22, no. 11, pp. 1330–1334, November 2000.

[11] Intel Corporation. Open source computer vision library. [Online]. Available: http://sourceforge.net/projects/opencvlibrary/

[12] R. Hartley and A. Zisserman, *Multiple View Geometry (Second Edition)*. Cambridge University Press, 2003.

[13] M. Field, G. Landon, D. Clarke, A. Park, and W. B. Seales, "Evaluation of stereo-endoscopy for image field reconstruction in minimally invasive surgery," submitted for publication in *IEEE Trans. Med. Imag.*