Reconstruction and Enhancement in Monocular Laparoscopic Imagery

Jesus J. Caban *
W. Brent Seales
Computer Science Department
University of Kentucky

Abstract

Constrained minimally-invasive surgical environments create a number of challenges for the surgeon and for automated tools designed to aid in the performance and analysis of complex procedures. The 3D reconstruction of the operative field opens up a number of possibilities for immersive presentation, automated analysis, and post-operative evaluation of surgical procedures.

This paper presents a method for estimating complete 3D information about scope and instrument positioning from monocular imagery. These measurements can be used as the basis for deriving and presenting additional cues during procedures, and can also be used for post-procedure analysis such as objective estimates of high-level performance measures like economy of motion and ergonomic metrics.

1 Introduction

Minimally-invasive surgery (MIS) provides a number of benefits to the patient, including lower risk of infection and swifter recovery times. Minimal invasion is accomplished primarily through the use of a camera (endoscope) and other surgical instruments inserted into the abdominal cavity through small "keyhole" incisions. The surgeon navigates the process by viewing imagery on an external display. By moving the endoscope and instruments, complex surgical tasks can be efficiently accomplished.

In this work we present a method to extract and make explicit information that is implicitly confounded in the imagery. Such information, though valuable as a direct cue, is usually subtle, especially in monocular imagery. Extracting an explicit representation can provide a ready cue or an analytical tool that otherwise would remain subtle and far less useful. In particular, we concentrate on the problem of recovering the 3D position and orientation of instruments within the endoscope's view, as well as the distance of these instruments from the scope, from each other, and from the anatomy.

Providing 3D information is crucial in addressing one of the primary technical and visual obstacles in conducting MIS procedures, which is the lack of an explicit depth cue. Experts become very good at understanding 3D relationships from monocular imagery, which does not make depth explicit but does contain a number of subtle depth cues, such as perspective distortion and scale, expert knowledge of instrument size, shape and relative positioning, and narrow depth of field which provides a focus cue.

We believe that the ability to extract precise depth measurements, including the position and orientation of instruments, scopes and anatomy, can substantially enhance laparoscopic environments of the future. In particular, we envision two immediate uses when depth information can be made explicit for tracked instruments and anatomy: enhanced visualization for the surgical team, and objective performance measures given video of training and simulation cases.

In the case of enhanced visualization, depth cues extracted from monocular imagery can be used to provide alternative views of instrument position, for example. Top views, closing distance, velocity and accelerations between two instruments, and real metric measurements within the operative field to give an accurate sense of the scale of what is potentially a very small operating space are all possible when depth can be extracted from the imagery.

With respect to performance measures, we can formulate motion *signatures* of instruments as a function of time, incorporating distance, orientation, and the derivatives of these values over time. We believe such measures can give a precision to the problem of analyzing the performance of a task completed within a simulation environment or a training box.

We believe that both enhanced visualization and objective performance metrics are valuable. They rely on the extraction of metric 3D information. Stereo scope systems, such as the *da Vinci* System [1], use a stereo endoscope to

* (jesus@netlab.uky.edu) Computer Science Department, University of Kentucky, 2nd Floor Hardymon Building, Lexington KY 40506.

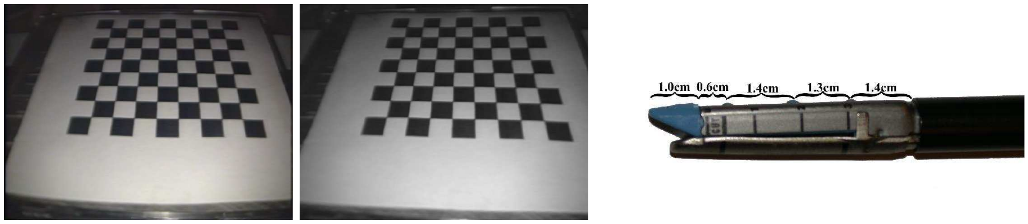

Figure 1: (left) Original image obtained from a 35^o endoscope. (center) Image after undistortion. (right) Stapler instrument with identifiable points and known measurements.

make depth explicit for a viewer using a corresponding stereo viewing system. These systems can also benefit from enhancements and metrics based on depth, which is made even more accessible through the introduction of the stereo scope. One of the crucial contributions of this work is that even when a stereoscopic scope is *not* available, we can obtain certain metric depth measurements. This allows us to apply enhancements and metrics to environments such as trainer boxes, archival procedure video, and current operating areas where it is not practical or even desirable to use specialized hardware like the Da Vinci stereo scope/display.

In the following sections we show how to model and calibrate the camera, and we outline the cue extraction process. We follow this with results from real video that indicate the value of metrics based on extracted cues as we have formulated them. Related work for visual tracking of laparoscopic surgery [3, 2] shows that there continues to be a need for new enhancement techniques, metrics for analysis, and exploration of how best to present extracted cues to the surgical team.

2 The Calibrated Endoscope

In order to formally model the geometry of the endoscope, we assume that the imaging system can be modeled as a pinhole system (i.e., perspective projection). Using this camera model, we apply computer vision methods and algorithms in order to calculate its characteristic geometry and distortion parameters.

First, we compute endoscope parameters and characteristics through a calibration process. According to the pinhole model, the relationship between a 3D point M and its 2D image projection m is given by $sm \approx A[Rt]M$ with

$$A = \begin{pmatrix} \alpha_u & \gamma & u_0 \\ 0 & \alpha_v & v_0 \\ 0 & 0 & 1 \end{pmatrix} \tag{1}$$

where s is an arbitrary scale factor, (R, t) are the extrinsic parameters, and A is a representation of the intrinsic parameters. Extrinsic parameters locate the camera in a 3D world coordinate frame; intrinsic parameters describe internal camera features such as the pixel coordinates of the image center and the focal length. More specifically, A is a 3x3 matrix that relates the pixel coordinate system to the world coordinate system. Contained in A are parameters α_u and α_v, which together represent the focal length and express the total magnification of the imaging system that results from both optics and image sampling. Also contained in A are the parameters u_0 and v_0, which represent the pixel coordinates of the image center.

As a result of the radial curvature of camera lens elements, there is no real lens system that can produce perfect pinhole images. In the case of endoscopes, different viewing angle scopes, which enlarge the field of view, and scopes with wide-angle lenses, cause significant distortion. These distortions can be removed by calculating distortion parameters through optical calibration. After a camera has been calibrated, it is possible to use the camera parameters to resample any image taken by that camera so that its lens distortion is removed from the image. Figure 1 (left and center) shows lens distortion in an image captured from a 35^o endoscope compared with a distortion-free image generated after calibration.

The key assumption that enables depth reconstruction of instruments visible in monocular sequences is knowledge of the shape, size and the metric measurements of visually identifiable fiducials on the instrument. Based on this information, it is possible to track features in imagery and recover the 3D position of each tracked point. From these points, with *a priori* information about the instrument, it is possible to compute the 3D position and orientation of the tip of the instrument. Figure 1 (right) shows a stapler instrument with identifiable marks and known distances between each of the points.

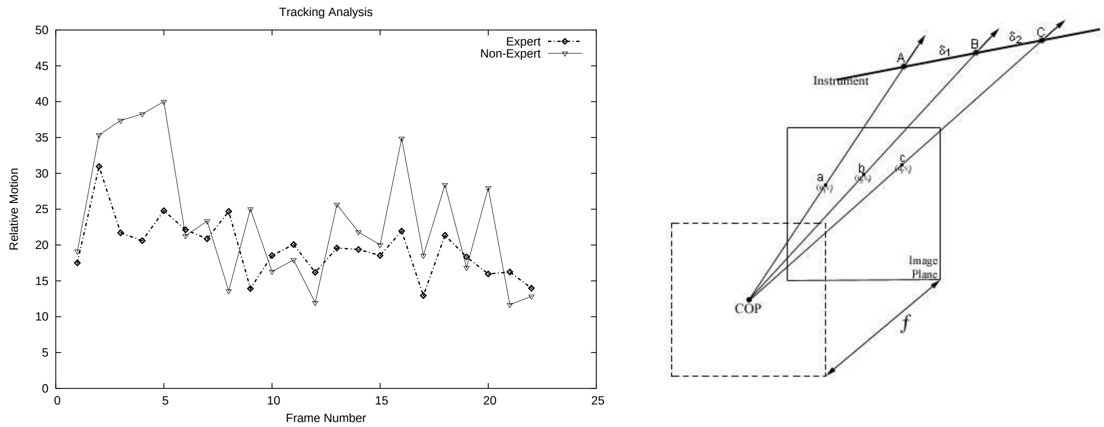

Figure 2: (left) The widely-varying curve shows novice performance; smoother motion is achievable by the expert. (right) The linear constraint leads to a depth solution for known points in monocular imagery.

3 Reconstruction Method

The shape of MIS instruments is almost always linear so that the instrument can be smoothly inserted into ports and manipulated through small incisions. We exploit this fact and as a result simplify the problem of tracking and estimating 3D points that lie on the instrument. By selecting $R = I$ and $t = 0$ in Equation 1, we can compute that $A = z_A A^{-1} a$, $B = z_B A^{-1} b$, and $C = z_C A^{-1} c$, where z_A, z_B and z_C are the unknown depths of the A, B, C points. These equations and constraints lead to a solution [4] for unknown depth wherever the instrument appears in the imagery. These depth values are derived based on the assumption that the instruments are linear, the camera is calibrated (i.e., the projection matrix is available from the off-line calibration process), and the distances between points on the instrument are known *a priori*.

4 Results and Conclusion

We have calibrated various endoscopes (e.g., 0^o and 35^o lenses). After calibration, we used recovered lens distortion estimates to remove lens distortion from images. Using the distortion-corrected images, we tracked the shaft of a stapler instrument (Fig. 2 (right)) in order to recover estimates of the 3D coordinates of points on the instrument. We have found these methods to be very promising as a way to recover 3D cues from monocular data.

As an example, consider the graph in Fig. 2 (left). The two curves on this graph show 3D motion estimates for the stapler instrument over a set of frames. The value plotted as the height of the curve for each frame value is the 3D position of the instrument measured relative to a fixed point. The curve that corresponds to the expert performing the stapling action shows much less relative-motion variation than the curve corresponding to the novice. In this case, economy of motion over a set of frames, evaluated in 3D to capture movement toward and away from the camera, shows how an expert handles the instrument in a way that is measurably and objectively different from the novice.

A number of interesting cues that exist in a subtle, implicit way in monocular laparoscopic imagery can be recovered automatically, enhanced and made explicit. This work shows an example based on the recovery of 3D information relative to the frame of the camera system from points detected on visible instruments as well as points in the operative field tracked over time. These methods do not assume a redesigned camera system, and therefore can be of value for experts using and training with monocular systems today.

References

[1] SunnyBrook Technologies Inc. In *http://www.intuitivesurgical.com*.

[2] R.Raskar et. all H. Fuchs, M. Livingston. Augmented reality visualization for laparoscopic surgery. In *MICCAI 1998*, pages 934–943, Oct 11-13, 1998.

[3] Xiaoli Zhang and Shahram Payandeh. Application of visual tracking for robot-assisted laparoscopic surgery. In *Journal of Robotics Systems*, pages 315–328, Vol. 19, No. 7, 2002.

[4] Z. Zhang. Camera Calibration with One-Dimensional Objects. In *Microsoft Technical Report MSR-TR-2001-120*, August 2002.