Robust Intraoperative US Probe Tracking Using a Monocular Endoscopic Camera

Uditha L. Jayarathne[2], A. Jonathan McLeod[2], Terry M. Peters[1,2], and Elvis C.S. Chen[1]

[1] Robarts Research Institute, London, Ontario, Canada
[2] Western University, London, Ontario, Canada
{ujayarat,jmcleod,tpeters,chene}@robarts.ca

Abstract. In the context of minimally-invasive procedures involving both endoscopic video and ultrasound, we present a vision-based method to track the ultrasound probe using a standard monocular video laparoscopic instrument. This approach requires only cosmetic modification to the ultrasound probe and obviates the need for magnetic tracking of either instrument. We describe an Extended Kalman Filter framework that solves for both the feature correspondence and pose estimation, and is able to track a 3D pattern on the surface of the ultrasound probe in near real-time. The tracking capability is demonstrated by performing an ultrasound calibration of a visually-tracked ultrasound probe, using a standard endoscopic video camera. Ultrasound calibration resulted in a mean TRE of 2.3mm, and comparison with an external optical tracker demonstrated a mean FRE of 4.4mm between the two tracking systems.

1 Introduction

Many procedures can be performed using endoscopy as a viable alternative to open surgery. Such minimally-invasive approaches can reduce recovery time, length of hospital stay, and morbidity. During such procedures, surgeons employ an endoscopic camera to view the organ surface and an endoscopic ultrasound (US) probe to visualize structures within the organ. In a typical surgical configuration, the video and US images are presented separately and in 2D. The surgeon must, therefore, perform spatial reasoning to mentally map the US image onto the video. Furthermore, the 2D nature of these images results in decreased depth perception.

Navigated endoscopy incorporates a spatial tracking device to infer the pose of the US probe relative to the camera, allowing US images to be registered to, and fused with, the video. The fused image is more intuitive, and lowers the cognitive load by eliminating the mental transform between the two images, which may improve hand-eye coordination of the surgeon. Examples of such systems include tumour resections using magnetic [4], robotic [9], photoacoustic [3], and vision-based [12] tracking systems, with Langø et al. [8] providing a comprehensive overview. Systems that employ extrinsic tracking devices increase cost, impact surgical workflow, require additional sterilization, and introduce other

K. Mori et al. (Eds.): MICCAI 2013, Part III, LNCS 8151, pp. 363–370, 2013.

limitations such as line-of-sight and metal interference issues for optical and magnetic tracking solutions respectively. Intrinsic tracking using the endoscopic camera to perform spatial measurements [12] does not incur these limitations, but accuracy, robustness, and the need to modify standard surgical instrument may be of concern. In particular, vision-based tracking in an endoscopic environment may be subject to lighting conditions, erroneous feature detection, and occlusion.

In this paper, a vision-based tracking system using a standard surgical laparoscopic camera is presented. A 3D marker was designed and rigidly attached to the semi-cylindrical back surface of a standard linear laparoscopic US probe, providing a set of features that can be reliably detected in the video sequence. Based on an Extended Kalman Filter framework, the proposed system is capable of tracking the 3D marker in 6 degrees of freedom, in the presence of spurious and/or missing features, even under strong specular lighting conditions. The tracking capability of the proposed system is demonstrated by performing an US probe calibration, along with both visual and quantitative validation procedures.

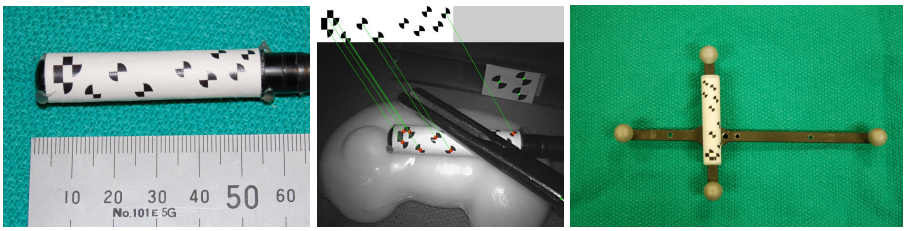
2 Methods and Materials

Given a set of known 3D model points M, and a set of 2D observed features U as imaged by the camera, the process of 2D-to-3D registration can be treated as a Perspective-n-Point (PnP) problem. In general, a PnP solution requires the correspondence between the 2D and 3D features to be established, as well as the optical properties of the imaging system to be known [13].

In practice, spurious features are often detected in the image, and features may be missing due to occlusion or strong specular lighting conditions. Therefore, to ensure robust operation of PnP algorithms, it is advantageous to address both the correspondence and pose estimation problems simultaneously. The proposed method employs an Extended Kalman Filter (EKF) framework [11] to constrain the search for feature correspondence. Given a calibrated camera and a 3D marker, possible poses of the 3D marker with respect to the camera are represented as a Gaussian Mixture Model (GMM). Uncertainty relating to the pose of the marker is propagated from the 3D model space to the 2D image space using EKF equations, constraining the search space for feature correspondence in a sequential fashion. A globally convergent PnP algorithm is applied to further refine the estimation established with at least 4 correct correspondences.

2.1 Hardware Setup

An "X-Corner" fiducial pattern (Fig. 1) was rigidly attached to the curved back-surface of a linear endoscopic US probe (UST-5536-7.5, Aloka, Japan) to serve as the 3D marker. A local 3D coordinate system of the pattern was defined, with the locations of the black-and-white intersections (i.e. X-corners) accurately determined using a measuring microscope (STM6-LM, Olympus, Japan).



(a) (b) (c)

Fig. 1. (a) a 3D X-corner pattern affixed to a linear US probe, (b) the proposed method is able to establish the 2D to 3D correspondence even in the presence of spurious(top right) and missing features (those occluded by surgical grasper), and (c) a surrogate marker designed for validation using an optical tracking system

To locate the X-corners in the video image, a corner detection algorithm [1] specifically designed to detect X-corners was employed. Compared to generic corner and edge detectors [7], this algorithm tends to detect less clutter and is more computationally efficient. Once detected, the locations of X-corners are further refined to sub-pixel accuracy [2]. The camera (Surgical laparoscope, Olympus) was calibrated in a standard fashion using a planar pattern [2]. Both video and US images were captured at an image size of 640×480 pixels (Morphis, Matrox, Canada), and the video was corrected for both tangential and radial distortion.

2.2 Simultaneously Solving for the Pose and Correspondence

The pose of the US probe is represented as a 6D vector: 3 components representing the rotation in Rodrigues' form [2] and 3 representing the translation. It is assumed that the US probe can be freely displaced within the viewing frustum of the endoscopic camera, with a depth ranging from 5cm to 20cm. The orientation of the US probe mimicked realistic surgical conditions. Under these range constraints, pose samples were simulated and a Gaussian Mixture Model (GMM) was learned [5] offline to represent the 6D pose space. The learned GMM was then used to provide the initial pose and correspondence of the US probe by minimizing the following error function [11]:

$$Error(p) = \sum_{(m,u) \in Matches} \|u - Proj(p; m)\| + \tau |NotDetected| \qquad (1)$$

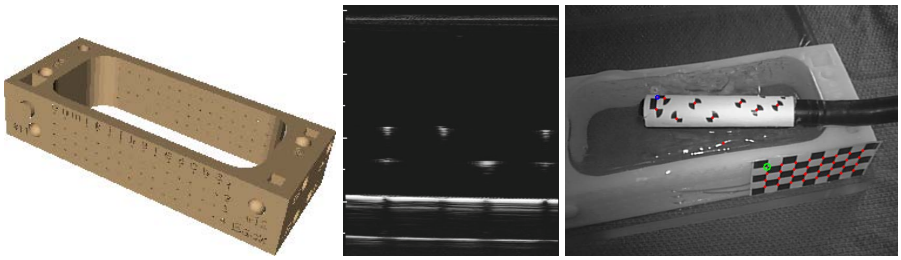
where u is a detected X-corner, $Proj(p; m)$ is the projection of a model point m using the pose p, τ is a penalty term for unmatched points, and $|NotDetected|$ denotes the cardinality of the undetected feature set. Uncertainty in pose space is propagated to image space using EKF equations and used to constrain the matching between the detected X-corners and model points. Once a correspondence is found, the pose and the associated uncertainty are updated using EKF equations to further reduce the search space for subsequent correspondence

matching. This process is repeated until at least 4 correct correspondences are established, by which time the initial pose estimation given by the GMM is updated in response to the found correspondence. This estimated pose is refined by solving the generic PnP problem with a globally convergent algorithm [10].

The GMM, which provides an accurate pose prior but is computationally expensive, is nevertheless used to provide the initial pose estimation for the first video frame. For all subsequent frames, the pose provided by the PnP algorithm at frame i is employed as a motion prior for frame $i+1$. Standard EKF equations are used to propagate the pose and the associated covariance throughout the video sequence.

2.3 Laparoscopic Ultrasound Calibration

The tracked US probe was calibrated [6] using a Z-phantom with 2 line fiducials arranged in a Z pattern. The Z-phantom was filled with polyvinyl chloride-plastisol (PVC) compound, which serves as a clear tissue-mimicking medium (Fig. 2(b)). A rectangular checkerboard pattern was rigidly attached to the planar surface of the phantom, and its pose was determined using OpenCV [2]. Once the homologous features were determined in both the image and camera space, the calibration was solved using the standard Orthogonal Procrustes algorithm.



(a) (b) (c)

Fig. 2. (a) a CAD model of the Z-phantom (designed by PERK `http://perk.cs.queensu.ca`), (b) an US image showing 2 Z-patterns, and (c) pose of the US probe/Z-phantom was determined using the proposed method and OpenCV [2], respectively

2.4 Validation Using Optical Tracker

A surrogate 3D marker, similar to that attached to the US probe, was constructed and rigidly attached to an optical Dynamic Reference Body (DRB) (Fig. 1c). The 3D marker and the optical DRB were carefully coregistered using a calibrated stylus. This setup was simultaneously tracked using both the proposed method and an optical tracking system (OTS) (Vicra, NDI, Canada). A total of 75 measurements were made within the depth range between 92mm to 99mm, typical of an endoscopic intervention.

3 Results

A total of 17 US/video image pairs were acquired for calibration. In each US image, the two fiducials, corresponding to the cross line of the Z-phantom (Fig. 2) were manually identified to obtain a total of 34 homologous points in both the US image and the 3D space. The calibration was accomplished using the Orthogonal Procrustes algorithm and resulted in a mean Fiducial Registration Error (FRE) of 2.2mm with a standard deviation of 0.9mm. A leave-one-out cross validation was performed where one of the 17 image pairs was removed prior to the calibration and the two fiducials on this image were then used as target points. This cross validation was repeated 17 times with different validation images, resulting in a mean Target Registration Error (TRE) of 2.3mm and a standard deviation of 1.0mm. Table 1 summarizes the result.

Table 1. US probe calibration result using a total 17 tracked US images

	mean error (mm)	standard deviation (mm)
Calibration FRE	2.2	0.9
leave-one-out TRE	2.3	1.0

Once US calibration is established, it is used to fuse the US image onto the endoscopic video. For the purpose of visual validation, two phantoms were constructed. The first comprises a spine vertebra semi-submerged into a PVC block; the second, a surgical needle inserted through a block of PVC. Figure 3(b) clearly shows that the outline of the spine forms a continuous contour, verifying that both visual tracking and US calibration are performing as expected. The proposed method is able to track the US probe in the presence of feature occlusion (Fig. 3(c,d)), as well as under ambient and endoscopic lighting conditions (Fig. 3(e,f)). Note that in the fused image, the needle seen in US is in alignment with the ends of the needle protruding from the phantom.

A total of 75 measurements of the surrogate marker were taken simultaneously using the endoscopic camera and the OTS. The tracking accuracy was evaluated at a set of 3D points lying on a virtual US fan created using the calibration obtained from the previous experiment. The 3D location of each pixel in the US fan was determined in both the OTS and the camera coordinate systems, allowing the two tracking systems to be coregistered. The FRE of this registration incorporates tracking errors from the two tracking systems as well as the registration error between the 3D marker and the optical DRB. Assuming the OTS is the gold standard, and the obtained registration is valid over the region of interest, the FRE represents the tracking error of the proposed method propagated to the US fan.

Figure 4 depicts the mean FRE of the camera coregistration using US image pixels as fiducials. The mean FRE exhibits a radial pattern: pixels close to the

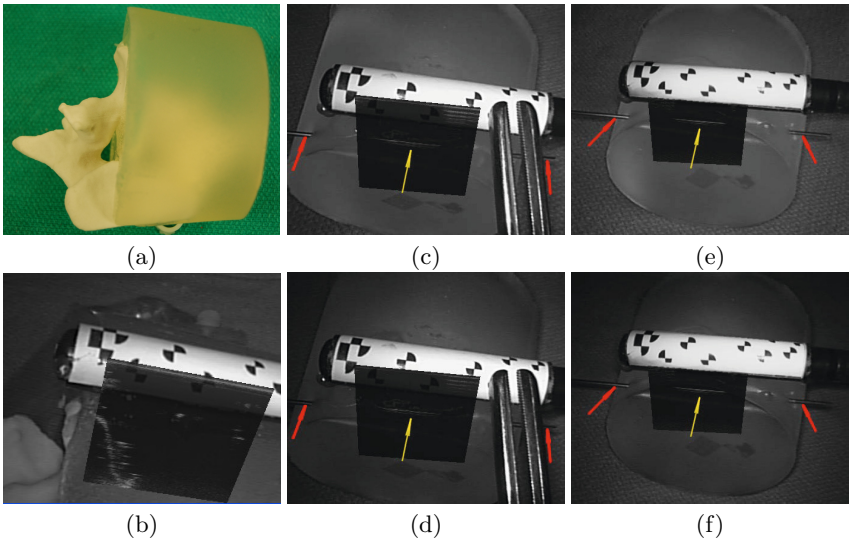


Fig. 3. Visual validation of tracking and US calibration: (a) a spine phantom, (b) US overlay showing a continuous contour of the spine outline under endoscopic lighting conditions. Needle phantom with (c,e) ambient lighting, (c,d) occlusion by surgical tool, (e,f) no occlusion, and (d,f) under endoscopic lighting conditions. Red arrows point to the ends of the needle while the yellow arrow points to its image in US. Note the dramatic improvement in the visual perception of the US beam with respect to the US probe, when the probe occludes the upper part of the US image, compared to the more naïve rendering of simply overlaying the US image on the video scene.

origin of the 3D marker exhibit a smaller mean FRE, which gradually increases as the pixel location is moved further away from the origin. This radial pattern suggests that the main contribution of the tracking error is from the rotational error of pose estimation. Over the acquired 75 measurements, the mean FRE of all pixels was 4.4mm, with a standard deviation of 3.3mm. The per-pixel mean FRE ranges from a minimum of 4.2mm to a maximum of 4.7mm. In our experience, this error is over-estimated due to the coregistration between the checker pattern and the optical DRB. Nevertheless, these results provide an upper-bound on the error introduced by the proposed method. The visual validation results suggest that the actual tracking error is substantially lower.

4 Discussion and Conclusion

A vision-based tracking procedure using a standard surgical monocular camera is proposed. By tracking a 3D pattern that is rigidly attached to a surgical tool, this approach operates under endoscopic lighting conditions and in the presence of occlusion. As with other vision-based systems, the limitation of this method is that it requires a fixed-focus camera that adheres to the pin-hole

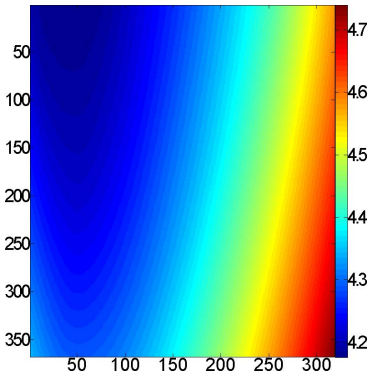


Fig. 4. Per-pixel mean FRE (in mm) of the camera-OTS coregistration shown in US image space. The x and y axis are the US image axis, with the top-left corner being closest to the origin of the 3D marker.

camera model. In contrast to other systems, this approach only requires trivial, cosmetic modification of the tracked object (endo-ultrasound probe).

One advantage of this approach is the ability to track a marker defined on a curved, rather than a planar surface, and can thus be adapted to a wide range of surgical tools. The current implementation tracks X-corners, but any feature that can be reliably detected in a video sequence can be used. Modifications to surgical tools are therefore minimized, and since no extrinsic tracking system is used, additional cost is minimal.

This method was used to calibrate a tracked US probe, and US image overlay gives a visually consistent augmentation. The reported US calibration FRE and TRE are in the order of 2mm which, in our experience, is comparable to results obtained using a magnetic tracking system. An optical tracking system was used to validate the accuracy of the proposed system. Using a surrogate marker to represent that attached to the US probe, simultaneous tracking by both systems was performed, allowing the local coordinate systems of the two trackers to be coregistered. Using the measurements from the OTS as the gold standard, camera tracking proved to be consistent with the OTS. The error of camera tracking exhibits a radial pattern, suggesting that the main contributor of the error is the rotational component of the pose estimation. The reported mean FRE is 4.4mm with a standard deviation of 3.3mm. This FRE is a combined system error including the co-calibration error of the 3D marker and DRB, which provides an upper-bound for the tracking error of the proposed system. Further analysis is warranted to provide a more accurate estimate of this error.

The proposed system currently performs at about 5 frames per second, but can be dramatically improved through more efficient implementation. Future work includes 1) pattern design of the 3D marker to optimize tracking accuracy and range of tracking angle, 2) integration into standard endoscopic procedure,

3) improvements to the algorithm to track multiple surgical instruments, and 4) extending this algorithm to track intrinsic visual features of the surgical tools, eliminating the need for extrinsic 3D markers entirely.

Acknowledgments. This work was supported by Natural Sciences and Engineering Research Council of Canada, Canadian Institutes of Health Research, and Canadian Foundation for Innovation. The authors thank Mohammad Ali Tavallaei and John Moore for their laboratory assistance.

References

1. Bennett, S., Lasenby, J.: ChESS - Quick and Robust Detection of Chess-board Features. Computing Research Repository (CoRR) abs/1301.5491 (2013)
2. Bradski, G.: The OpenCV Library. Dr. Dobb's Journal of Software Tools (2000)
3. Cheng, A., Kang, J.U., Taylor, R.H., Boctor, E.M.: Direct 3D Ultrasound to Video Registration Using Photoacoustic Effect. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part II. LNCS, vol. 7511, pp. 552–559. Springer, Heidelberg (2012)
4. Cheung, C.L., Wedlake, C., Moore, J., Pautler, S.E., Peters, T.M.: Fused Video and Ultrasound Images for Minimally Invasive Partial Nephrectomy: A Phantom Study. In: Jiang, T., Navab, N., Pluim, J.P.W., Viergever, M.A. (eds.) MICCAI 2010, Part III. LNCS, vol. 6363, pp. 408–415. Springer, Heidelberg (2010)
5. Figueiredo, M., Jain, A.: Unsupervised learning of finite mixture models. IEEE Trans. on PAMI 24(3), 381–396 (2002)
6. Gobbi, D.G., Comeau, R.M., Lee, B.K., Peters, T.M.: Correlation of pre-operative MRI and intra-operative 3D ultrasound to measure brain tissue shift. In: Medical Imaging. Proc. SPIE 3982, pp. 77–84 (2000)
7. Harris, C., Stephens, M.: A combined corner and edge detector. In: Proc. of Fourth Alvey Vision Conference, pp. 147–151 (1988)
8. Langø, T., Vijayan, S., Rethy, A., Våpenstad, C., Solberg, O., Mårvik, R., Johnsen, G., Hernes, T.N.: Navigated Laparoscopic ultrasound in abdominal soft tissue surgery: technological overview and perspectives. International Journal of Computer Assisted Radiology and Surgery 7(4), 585–599 (2012)
9. Leven, J., et al.: DaVinci Canvas: A Telerobotic Surgical System with Integrated, Robot-Assisted, Laparoscopic Ultrasound Capability. In: Duncan, J.S., Gerig, G. (eds.) MICCAI 2005. LNCS, vol. 3749, pp. 811–818. Springer, Heidelberg (2005)
10. Lu, C.P., Hager, G.D., Mjolsness, E.: Fast and globally convergent pose estimation from video images. IEEE Trans. on PAMI 22(6), 610–622 (2000)
11. Moreno-Noguer, F., Lepetit, V., Fua, P.: Pose Priors for Simultaneously Solving Alignment and Correspondence. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) ECCV 2008, Part II. LNCS, vol. 5303, pp. 405–418. Springer, Heidelberg (2008)
12. Pratt, P., Di Marco, A., Payne, C., Darzi, A., Yang, G.-Z.: Intraoperative Ultrasound Guidance for Transanal Endoscopic Microsurgery. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012, Part I. LNCS, vol. 7510, pp. 463–470. Springer, Heidelberg (2012)
13. Wu, Y., Hu, Z.: PnP Problem Revisited. Journal of Mathematical Imaging and Vision 24(1), 131–141 (2006)