This paper was published in the Proceedings of Image and Vision Computing New Zealand (IVCNZ)
B. McCane (Ed.), Dunedin, November 2005. ISBN 0-473-10523-3.

Video Understanding for Laparoscopic Surgery: Instrument Tracking

S. J. McKenna[1], H. Nait Charif[2], and T. Frank[3]

[1]Division of Applied Computing, University of Dundee,
Dundee DD1 4HN, Scotland.

[2]Department of Physics, Royal Holloway, University of London,
Egham, Surrey, TW20 0EX, England.

[3]Department of Surgery and Molecular Oncology, Ninewells Hospital,
University of Dundee, Dundee DD1 9SY, Scotland.

Email: stephen@computing.dundee.ac.uk

Abstract

Endoscopic surgery routinely produces video data. Automated understanding of these data has applications both online (visual servoing for telemanipulation) and offline (procedure analysis, annotation and content-based retrieval). As a step towards such applications, a method for tracking surgical instruments in endoscopic video without modifications to instruments or endoscopic equipment is described. The shape model used is flexible enough to track a variety of surgical instruments. Colour distributions are used to compute measurements that drive particle filter-based tracking. The method is demonstrated using routinely obtained video from laparoscopic cholecystectomy.

Keywords: Endoscopic surgery, object tracking, particle filter

1 Introduction

A routinely performed endoscopic surgical procedure is laparoscopic cholecystectomy (gall bladder removal). The video endoscope is inserted through a small incision near the navel. The abdomen is inflated with carbon dioxide to provide space for the surgery to be viewed and performed. Various instruments are inserted through further entry ports. These enable the surgeon to pick up the gallbladder, move intestines around and to safely dissect and remove the gallbladder and stones. Further instruments are used to apply clips to the dissected gallbladder artery and bile duct. Laparoscopic cholecystectomy typically takes between 30 and 60 minutes.

Clearly, endoscopic surgery routinely produces enormous amounts of image sequence data, although much of it is not currently recorded. These data are usually, though not always, monoscopic. We are interested in using computer vision for video understanding in this domain.

An online application is visual servoing for an endoscopic robot. It is common practice for a human assistant to hold and move the endoscope during surgery in order that appropriate views are obtained for the surgeon. This is costly and has a number of limitations. For example, suitably trained assistants can be a scarce resource other than in teaching hospitals. Furthermore, the human operator is likely to become fatigued, may introduce motion tremor and is open to errors of communication with the surgeon. Telemanipulation, in which a robot holds and moves the endoscope based on a combination of automatic video understanding and surgeon control, provides an alternative.

Potential offline applications of endoscopic video understanding are indexing, retrieval and analysis of surgery. Despite widespread interest in content-based visual information retrieval, tools for retrieval of video specialised to the medical domain do not yet exist [1]. In particular, there are no tools available to automatically annotate endoscopic surgery videos. Content-based access of these videos would benefit large domains of teaching and research, facilitating training and outcomes studies.

As a step towards such applications, this paper presents a method for tracking surgical instruments in endoscopic video. The resulting track data could be used to help infer the desired field of view for servoing a robot or for modelling and recognising surgical actions and activities for video annotation. The aim here is to perform tracking without modifications to surgical instruments or endoscopic equipment. The video data are acquired as normal during surgery using a standard endoscope and routinely used surgical instruments. In particular, the use of special visual markers on the surgical instruments is avoided. This means that the method developed is applicable to pre-existing video data which will have been acquired without such markers. Furthermore, requiring the use of visual markers has other disadvantages. Painted markers would present a biological hazard. Tape-based markers can be awkward to fit and problematic if the instrument is to be introduced down a port with very little radial clearance. Moulded features would not be retrofittable, and etched or shot blasted markers would not yield high colour contrast.

The remainder of the paper is organised as follows. Section 2 reviews previous research on endoscope control and instrument tracking. Section 3 describes the tracking method. Section 4 presents results from video of laparoscopic cholecystectomy. Finally, conclusions and directions for future work are given in Section 5.

2 Previous Work on Endoscope Telemanipulation and Instrument Tracking

Various command interfaces to enable the surgeon to control the endoscope directly have been developed. Systems have been based on speech (AESOP, Computer Motion (discontinued); [2]), glove control (Zeus, Computer Motion) and head movement (EndoAssist, Armstrong Healthcare Ltd.) for example. The latter used a wireless head-mounted device to enable camera pan, tilt and zoom control in response to small head movements in combination with a footswitch. There is evidence that this can reduce operation times [3]. A force sensor was used for safety to deactivate movement. Nishikawa et al. [4] investigated a computer vision-based face tracker as a means of allowing the surgeon to control the laparoscope.

Previous attempts at 'self-guided' telemanipulation systems (visual servoing) have also been reported. The idea here is that a robot automatically controls the endoscope pose and zoom parameters. This has to date consisted of attempts to centre the field of view near the tip(s) of relevant surgical instruments. Several research groups used instruments with special coloured or patterned markers on them and attempted to detect the instruments' positions [5, 6, 7, 8]. The image analysis methods used were 'bottom-up', typically consisting of colour thresholding and subsequent binary image analysis. These systems showed preliminary success but were not reliable enough. For example, simple colour thresholding has problems due to inter-reflections, varying illuminant position and specular highlights. Zhang and Payandeh [9] described some simplified endoscope calibration methods and proposed tracking markers on the instrument tips in monochromatic images to recover robot control parameters including zoom. The calibration methods aimed to correct for barrel distortion due to the wide-angle lenses used as well as determining other intrinsic (focal length, image centre) and extrinsic (position and orientation relative to robot) parameters. Asari et al. [10] also considered distortion correction for endoscopic images. A related problem of bringing the instruments into the endoscope's field of view was tackled by Krupa et al. [11] who used lasers fitted to the instruments to project a structured light pattern which was then automatically analysed to guide the endoscope.

Roubert [12] attempted to detect instruments without markers using frame differencing and least squares line fitting but the lack of strong models and a tracking framework meant that performance was not robust. Computer Motion Inc. also investigated vision-based control [13, 14] based on colour cues, low-level grouping and simple temporal filtering.

3 Method

A pixel's colour values provide a strong discriminatory cue for whether or not it is occupied by a surgical instrument. In particular, given a pixel's RGB values, denoted γ, and ignoring all other data, the posterior probability of a surgical instrument at that pixel is given by Bayes' rule:

$$P(\iota|\gamma) = \frac{P(\gamma|\iota)P(\iota)}{P(\gamma|\iota)P(\iota) + P(\gamma|\neg\iota)(1 - P(\iota))}$$

where ι and $\neg\iota$ denote the instrument and non-instrument classes, respectively. Figure 1 shows an image of such posterior probabilities computed using $P(\iota) = 0.3$ and likelihoods, $P(\gamma|\iota)$ and $P(\gamma|\neg\iota)$, estimated as RGB histograms from manual segmentations of four images selected at random from other image sequences of this surgery. Whilst this provides a strong local cue,

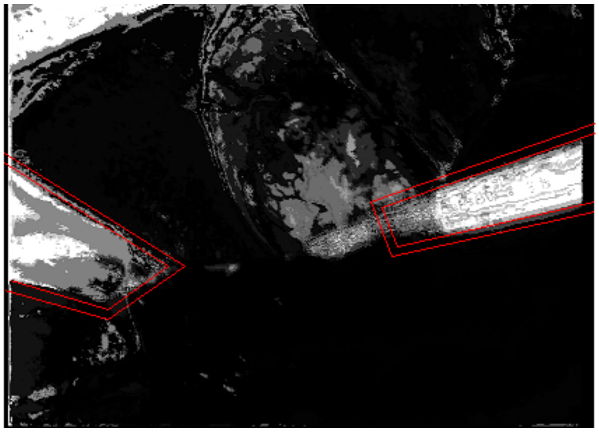

Figure 1: An image of instrument posterior probabilities computed pixelwise using RGB histograms. Hypotheses for two instruments are overlaid along with their adjacent background region contours.

using it to drive low-level grouping without shape constraints will clearly be insufficient.

A variety of surgical instrument designs are used and these designs will be adapted over time. Therefore, rather than use models specific to particular instruments, a middle ground is adopted here by using a generic model with two-dimensional shape constraints that are satisfied by the images of the instruments to be tracked. The instrument state model has six degrees of freedom and specifies three line segments, referred to as the side segments and the tip segment. The side segments each have an endpoint at the image boundary and they are constrained to be approximately parallel. The angles between the tip segment and the side segments are constrained to lie between $75°$ and $105°$.

Figure 1 shows two hypothesised instruments overlaid on a pixel colour posterior probability image. The inner contours represent the hypothesised instrument boundaries. The regions enclosed between the inner and outer contours are referred to as the instruments' adjacent backgrounds. Let $\bar{P}_{\mathcal{R}}$ and $\bar{P}_{\mathcal{B}}$ denote the average pixel posterior probability within the instrument region and the adjacent background region, respectively. Intuitively, a good hypothesis should have large $\bar{P}_{\mathcal{R}}$ and small $\bar{P}_{\mathcal{B}}$. A reasonable approximation to the likelihood of the state is $\bar{P}_{\mathcal{R}}(1 - \bar{P}_{\mathcal{B}})$.

Instrument tracking was performed using a particle filter to propagate an estimate of the state posterior over time. Specifically, iterated likelihood weighting [15] with zero-mean Gaussian dynamics ($\sigma = 20$ pixels) was used. This method combines sampling importance resampling with local likelihood search.

4 Results

Figures 2 and 3 show example frames from two sequences in which instruments are tracked. Overlaid for each tracker is the mean of its 10 most strongly weighted particles. The images had 720×576 pixels and each sequence was acquired at a frame rate of $24Hz$.

In order to provide a quantitative indication of accuracy, 'ground-truth' data were manually annotated for the two main instruments in the subsequence represented in Figure 2 by clicking points on the sides of the instruments in each frame and fitting lines through them in order to give an instrument orientation in the image plane. The mean absolute difference between the orientation recovered by the tracker and the manually annotated orientation was computed. For the instrument on the right side of the images, it was $5.2°$. The instrument on the left side had a larger mean absolute difference of $11°$ due to the relatively small fraction of its length that was visible.

5 Summary and Future Work

A method for tracking surgical instruments in monocular endoscopic video was presented as a step towards automated video understanding in this domain. Whereas previous research has tended to focus on tracking instruments with special markers, this paper has demonstrated the feasibility of tracking unmodified instruments. The results presented were on subsequences extracted from video of laparoscopic cholecystectomy.

Robust tracking of entire surgical procedures will require further work. The current method needs to be extended to cope with situtions in which significant partial occlusion by tissue and other instruments occurs. Other challenges include abrupt motion and periods in which smoke partially obscures the view. Further work is also needed to determine the extent to which reliable $3D$ information can be extracted in order, for example, to perform visual servoing.

6 Acknowledgements

The authors are grateful to Dr Stuart Brown for assistance in obtaining the video and for helpful discussions.

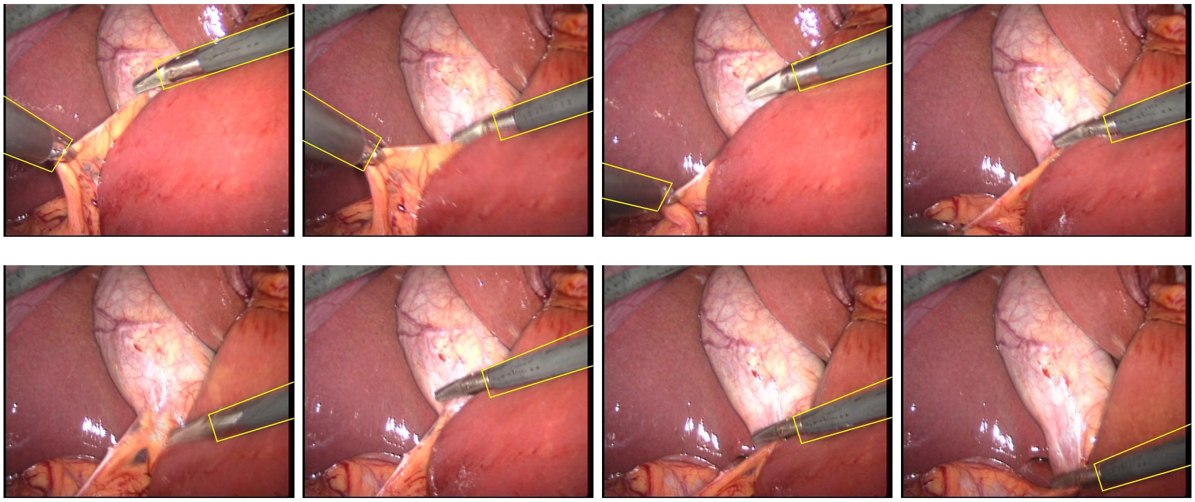

Figure 2: Tracking result showing example frames at intervals of 20 frames.

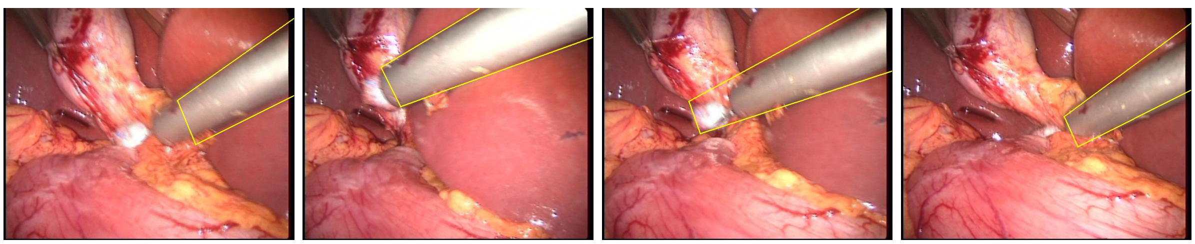

Figure 3: Tracking result showing example frames (786, 798, 817 and 835).

References

[1] H. Muller, N. Michoux, D. Bandon, and A. Geissbuhler, "A review of content-based image retrieval systems in medical applications - clinical benefits and future directions," *International Journal of Medical Informatics*, vol. 73, pp. 1–23, 2004.

[2] V. F. Munoz, J. Gomez de Gabriel, J. Fernandez-Lozano, I. Garcia-Morales, R. Molina-Mesa, C. Perez-del Pulgar, J. Seron-Barba, and M. Azouaghe, "Design and control of a robotic assistant for laparoscopic surgery," in *Int. Symp. Intelligent Robotic Systems (SIRS)*, (Toulouse), 2001.

[3] S. Aiono, M. Gilbert, B. Soin, P. A. Finlay, and A. Gordon, "Controlled trial of the introduction of a robotic camera assistant (EndoAssist^{TM}) for laparoscopic cholecystectomy," in *11th Annual Scientific Meeting of the Society for Minimally Invasive Therapy*, (Boston, USA), 1999.

[4] A. Nishikawa, T. Hosoi, K. Koara, D. Negoro, A. Hikita, S. Asano, F. Miyazaki, M. Sekimoto, Y. Miyake, M. Yasui, and M. Monden, "Real-time visual tracking of the surgeon's face for laparoscopic surgery," in *MICCAI*, vol. LNCS 2208, pp. 9–16, 2001.

[5] A. Casals, J. Amat, and E. Laporte, "Automatic guidance of an assistant robot in laparoscopic surgery," in *IEEE Int. Conf. Robotics and Automation*, vol. 1, pp. 895–900, 1996.

[6] K. Omote, H. Feussner, A. Ungeheuer, K. Arbter, and G.-Q. Wei, "Self-guided robotic camera control for laparoscopic surgery compared with human camera control," *American Journal of Surgery*, vol. 177, no. 4, pp. 321–324, 1999.

[7] A. Nishikawa, S. Asano, R. Fujita, T. Yohda, F. Miyazaki, M. Sekimoto, M. Yasui, S. Takiguchi, and M. Monden, "Robust visual tracking of multiple surgical instruments for laparoscopic surgery," in *Computer Assisted Radiology and Surgery (CARS)*, (London), 2003.

[8] G. Wei, K. Arbter, and G. Hirzinger, "Real time visual servoing for laparoscopic surgery: Controlling robot motion with color image segmentation," *IEEE Engineering in*

Medicine and Biology Magazine, vol. 16, no. 1, pp. 40–45, 1997.

[9] X. Zhang and S. Payandeh, "Application of visual tracking for robot-assisted laparoscopic surgery," *Journal of Robotics Systems*, vol. 19, no. 7, pp. 315–328, 2002.

[10] K. V. Asari, S. Kumar, and D. Rashakrishnan, "Technique of distortion correction in endoscopic images using a polynomial expansion," *Medical and Biological Engineering and Computing*, vol. 37, no. 1, pp. 8–12, 1999.

[11] A. Krupa, J. Gangloff, C. Doignon, M. de Mathelin, G. Morel, J. Leroy, L. Soler, and J. Marescaux, "Autonomous 3-D positioning of surgical instruments in robotized laparoscopic surgery using visual servoing," *IEEE Trans. Robotics and Automation*, vol. 19, no. 5, pp. 842–853, 2003.

[12] J. Roubert, "Automatic guidance of a laparoscope using computer vision," M.Sc. Thesis, Technical University of Catalonia, 2002.

[13] D. R. Uecker, C. Lee, Y. F. Wang, and Y. Wang, "Automated instrument tracking in robotically assisted laparoscopic surgery," *J. Image Guided Surgery*, vol. 1, no. 6, pp. 308–325, 1995.

[14] Y.-F. Wang, 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, no. 6, pp. 429–437, 1998.

[15] H. Nait-Charif and S. J. McKenna, "Tracking poorly modelled motion using particle filters with iterated likelihood weighting," in *Asian Conference on Computer Vision*, (Jeju Island, Korea), pp. 156–161, 2004.