

A Novel Interaction Method for Laparoscopic Surgery Training

Andrzej Wytyczak-Partyka †, Jan Nikodem †

Ryszard Klempous †, *Senior member, IEEE*, and Jerzy Rozenblit‡, *Senior member, IEEE*

†Wroclaw University of Technology, Wroclaw, Poland

‡ University of Arizona, Tucson, Arizona, USA

iapart@gmail.com, {jan.nikodem, ryszard.klempous}@pwr.wroc.pl, jr@ece.arizona.edu

Abstract—In the following paper a system for laparoscopic surgery training is presented. The main contribution of the paper is a novel approach to training, where a 3D model of the training object serves as a base of the interaction between the trainee and the system. A thorough state of the art is analyzed, later the main components of the system are described, as well as an outline of the applied 3D recovery algorithm.

Keywords—Laparoscopy, training, interface, structure-from-motion, minimally-invasive surgery.

I. INTRODUCTION

LAPAROSCOPIC surgery brings significant benefits into the healing process and therefore an increasing interest in it is observed. Great benefits over traditional surgery include: limited scarring, reduction in pain and recovery time, leading to a smaller risk of complications. Study conducted by Hansen et al. [1] shows that patients who have undergone laparoscopic appendectomy had five times fewer wound infections, two times shorter discharge time and fewer of them required narcotic analgesia. On the other hand, there is a number of downsides, for instance the surgeon's perception both haptic and visual - is very limited, which extends the procedure time (in the open appendectomy case 63 as opposed to 40 minutes in laparoscopic appendectomy [1]) and the likelihood of human error. Also investment in expensive instruments and a very long training period are required. Surgical training should be modular, where each module should focus on the development of certain behaviours - knowledge-based, rule-based and skill-based. This paper will focus on a system designed particularly for the skill-based behaviour training.

The primary skills that have to be developed during the training involve - depth perception, dexterity and hand-eye coordination. Traditionally, according to the Surgical Papyrus [2] since the times of ancient Egypt, surgical training has always followed the model of master and apprentice and involved mentorship in real-life clinical cases, where the apprentice would gain skills and experience from his teacher. That approach to surgical education hasn't changed significantly since ancient times. It is also worth noting that the training was strictly dependent on the availability of a patient and tied with the course of patient care [3], [4], [5].

Since some of the surgical skills do not strictly require practicing them in a real-life clinical situation - it is desirable to enhance them in a safe environment, without any risk to patients. It is also important to note, that the operating room is not the best learning environment

because of factors such as stress, time constraints and costs that have a negative impact on the learning process.

On the other hand, reference [6] shows, that surgical simulation has positive impact on improvement of psychomotoric skills, i.e. the gallbladder dissection procedure was 29% faster for residents trained on simulators, also accidental injuries during the procedure were 5 times less frequent in that group. Similar results in a cholecystectomy procedure are shown in [7]. Therefore a number of surgical simulators, both physical-model-based and software-based, have been developed, all allowing the trainees to safely master the basics skillset.

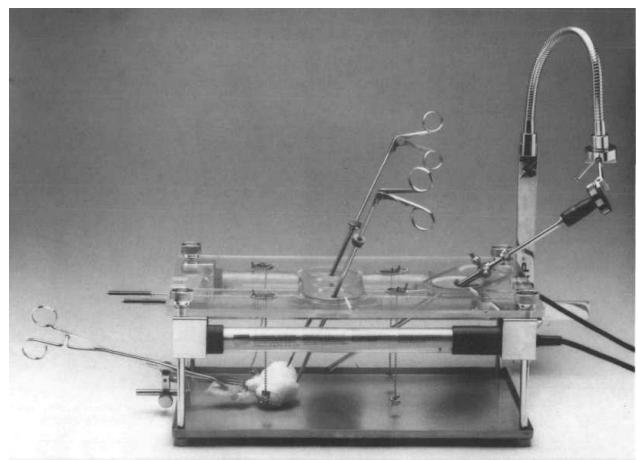


Fig. 1. Karl Semm's pelvi-trainer [8].

One of the first simulation devices for laparoscopic training was the Pelvi-trainer, designed by Karl Semm in 1986 [8]. The original concept consisted of a transparent box, where organs were put. The box contained several holes to introduce the instruments and the camera, Figure 1. The training method proposed by Semm was gradual: at first the trainee would only learn to use the instruments, without the use of the endoscope, secondly the endoscope would be introduced, while the trainee would be still allowed to occasionally look at the organ through the transparent walls of the box, and finally the box would be covered with a cloth to obscure vision. Since 1986 the concept of a pelvi-trainer has influenced many simulators and is currently used in many training programs.

Semm's pelvi-trainer is a typical example of a skill-based training device, that focuses on development of dexterity, depth perception and hand-eye coordination. There are several simulators that serve for knowledge- and/or

behaviour-based training as well as skill-based. One of the first systems built for that purpose is the Karlsruhe Endoscopic Surgery Trainer and the KISMET software [9], [10]. It is a complete training system with built in scenarios for practicing procedures like laparoscopic cholecystectomy. There are several products similar to the Karlsruhe Trainer, for instance the MISTELS system [11], or the LapMentor [12]. A lot of effort has been put in those systems to reproduce details such as graphics, organ deformations and haptic feedback. Needless to say those systems are very expensive.

One skill-based approach that combines a simple pelvi-trainer with a high-tech system is the Virtual Assistant Surgical Training (VAST) system, developed at the University of Arizona [13]. It is comprised of a pelvi-trainer and a computer system. The computer, through a magnetic position sensor, collects data about the instrument's tip position, which is used to rate the trainee's performance in a certain exercise, based on time, path length and accuracy. The individual's progress can therefore be precisely measured and monitored.

The approach represented by VAST is especially interesting and will serve as a basis for the further described training system which incrementally adds to the VAST trainer.

II. DESCRIPTION OF THE SYSTEM

The purpose of the proposed system is to aid a trainee in the development of basic skills, as dexterity, hand-eye coordination and depth perception. The novel way of interaction, and the main contribution of this paper, allows the trainee to develop the skill of avoiding certain regions where the appearance of an instrument might inflict damage to the patient.

A. System outline

The system resembles the VAST prototype trainer [13] and is comprised of a standard pelvi-trainer setting an endoscopic camera and two instruments - and a computer. The instruments have an embedded position sensor. The video output of the camera, as well as the position sensors, are connected to the computer, where additional processing occurs, as in Figure 2. There are two results of the processing. First - each exercise performed by a trainee is scored, and thus the performance can be analysed, and secondly the trainee is optionally informed by an auditory signal coming from a speaker, whenever he approaches the hazardous area with the instruments.

The score achieved in the exercises is calculated by the computer and based on the information collected from position sensors and the video camera.

Initially, before the exercise can start, and after a new model is introduced to the pelvi-trainer a 3D representation of the model is built, from images collected with the endoscopic camera. Secondly a hazardous region can be selected, the trainee has the opportunity to see and understand where the region is on the model, so it can later be avoided, during the course of the exercise. This can be done individually for each exercise to introduce more difficulty into the exercises. Later the coordinates of

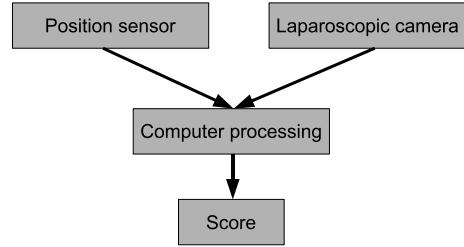


Fig. 2. Flow of the proposed system.

the instruments obtained from the position sensors can be used to determine the location of the instrument in the reconstructed environment.

The 3D structure of the viewed scene, recovered in the process of the system's work, is at this point not intended strictly for visualization - the trainee observes the image as it comes from the endoscopic camera, but rather for the purpose of determining the hazardous regions within the 3D model. Through fusion of the 3D reconstructed model and sensor data the coordinates collected from the position sensor located on the tip of the instrument can be used to determine its location and proximity to the hazardous area. If the position sensor indicates appearance of the instrument in a hazardous region an auditory signal is produced and a penalty score applied.

B. Training model

The training is based on simple tasks that trainees have to complete within a specified time. Several types of exercises have been proposed by others [14] [15], the common goal is to practice dexterity, coordination and depth perception. Example exercises are :

- 1) Knot tying
- 2) Cutting and suturing
- 3) Picking up objects
- 4) Touching different points on a model

Each exercise is associated with a different physical model which is placed in the pelvi-trainer's box. The trainee's score is calculated with respect to :

- 1) Elapsed time
- 2) Length of the path of the instrument tip
- 3) Accuracy

It is proposed that another factor is introduced and that the score is significantly decreased upon hitting a hazardous region, defined at the beginning of each exercise.

III. PROCESSING

The following paragraph describes a method for recovering the 3D structure of the scene, which is the initial step in the work of the proposed system. It also discusses the method of key-frame selection as one of the problems in the structure-from-motion approach.

During laparoscopic procedures it is natural that after introducing the camera into the body, the surgeon performs a series of movements with the camera to find the best viewing point for performing the procedure, and also to discover any potential abnormalities in the viewed organs. It is therefore natural that a video sequence containing

images of the operating field can be obtained in that manner also during the training procedure and serve as a source of images for a structure from motion 3D geometry recovery algorithm.

A. Structure from motion

Since the early 1980's there has been a lot of research in the field of recovering 3D structure from camera image sequences [16], [17]. Current state of the art algorithms [18], [19] perform the task without previous camera calibration and with the only constraints on the image sequence that a certain number of corresponding feature points between frames can be established and that a sufficient baseline (distance between the camera locations) exists. It has been proven [18] that the point-correspondences, along with several constraints on the camera, can lead to a metric reconstruction of the imaged scene.

In general the reconstruction process consists of: (a) selecting potential point matches between the frames of the sequence, (b) estimating the 3D geometry, (c) refinement. Since the whole process is based on point correspondences it is very important that the first step is performed carefully. Therefore potential matches, (m, m') , are selected from a set of points that contains only interest points that significantly differentiate from their neighborhood and can be matched with their corresponding points in other views. The set of interest points is populated through a Harris' corner detection procedure [20], later the matches can be found using a similarity measure, i.e. cross-correlation. It is important that the usage of a video sequence limits the search for matches to a certain subwindow of the image, because the camera movement between the frames (baseline) is small and therefore feature points stay in a several-pixel range.

After the feature matching has concluded two initial views are selected that will serve as a base for further sequential structure refinement. The criteria of selecting such views are:

- maximization of the number of features matched between the views,
- sufficient baseline between the views, so the initial 3D estimate can be properly performed.

While the first criterion is easy to fulfill, based on the results of feature matching, it is significantly harder to determine if a view meets the second one, which is especially important in images from a video sequence, where the baselines are usually very small.

The simplest approach would be to limit the frame sequence and select only every k -th frame, where k depends on the framerate of the video. In fact, basing on the constraint that the camera moves are rather smooth and slow and the imaged object itself remains still, it is suggested to perform such a reduction of the input video sequence, F_{in} , by reducing the framerate, fps , to 2 frames per second.

$$F_{in} = \{1, 2, \dots, n\},$$

$$F_{new} = \{1, k, 2k, \dots\},$$

$$k = \left[\frac{fps}{2} \right].$$

Such a reduction will majorly decrease the computational effort involved with the feature point selection. A finer approach, that also allows to ensure the constraint of general motion of the camera is to use the Geometric Robust Information Criterion (GRIC) [21], to check if the set of two-view point correspondences can be better described by a homography or by epipolar geometry. The homography model scores better with GRIC for small baselines, while the fundamental matrix model for bigger camera distances. In any case the matching points should lie on an epipolar line and satisfy the following equation:

$$m'^T F m = 0, \quad (1)$$

where m and m' are row-vectors containing the homogenous coordinates of a matched point pair and F is the fundamental matrix relating the two views. Equation 1 can be further used to find additional point correspondences. This approach is used in [18], where the first key frame is always assumed to be the first frame of the video sequence, and the second is selected when the epipolar geometry model overtakes the homography model (which guarantees a certain baseline) and where the number of matched features stays over a certain threshold. Another similar method based on comparing the reprojection errors of the homography and epipolar geometry models has been given in [22], while a more sophisticated and optimized approach to keyframe selection has been presented in [23].

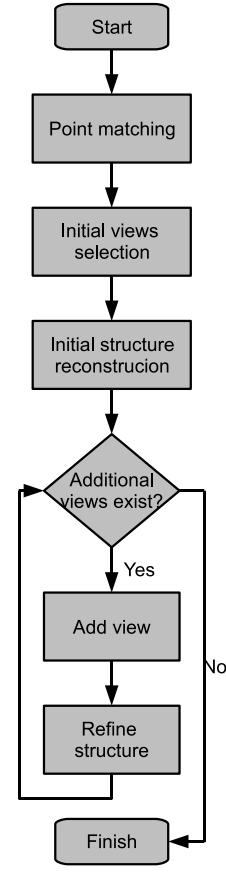


Fig. 3. Structure from motion algorithm sparse surface model reconstruction.

Figure 3. illustrates the process of sequential refinement of the recovered structure. After the initial two views have been selected the images may be further processed for the retrieval of the 3D geometry, next additional views are added, selected in a similar manner as in the initial step. The recovery of the 3D geometry is based on the normalized 8-point algorithm and a triangulation method described in [19]. The 8-point algorithm is used to compute the fundamental matrix, F , and the epipoles (e, e'), from a set of point correspondences (m, m') between 2 images by solving (1). The F matrix can be then used to calculate the camera matrices, (P, P') , which are needed for the triangulation step.

$$P = [I_{3 \times 3} \mid 0], \quad (2)$$

$$P' = [[e']]_x F | e'], \quad (3)$$

where $[e]_x$ is a skew symmetric matrix of the form:

$$[a]_x = \begin{bmatrix} 0 & -a_3 & a_2 \\ a_3 & 0 & -a_1 \\ -a_2 & a_1 & 0 \end{bmatrix}, \quad (4)$$

and linked with vector product in the following manner:

$$a \times b = [a]_x b. \quad (5)$$

It is now clear how important careful selection of point matches is. To ensure that the set of matches doesn't contain any invalid pairs a RANSAC fitting of the fundamental matrix is used, following [19],[18].

Finally the 3D coordinates of the points can be retrieved by solving the following equations :

$$m = PM, \quad (6)$$

$$m' = P'M, \quad (7)$$

where M are the real-world coordinates of a point, and (m, m') is a pair of matching images of the point M .

The resulting surface model is sparse and has to undergo dense disparity matching to become geometrically correct and to be used in the process of selecting hazardous regions.

IV. DISCUSSION

The goal of this paper has been to present a laparoscopic surgery training system with a new approach to interaction between the trainee and the system. It has been proposed to build a 3D representation of the training model used in an exercise, select hazardous areas which should be avoided by the trainee and finally score the trainee's performance appropriately. An approach to one of the problems arising in structure-from-motion 3D geometry recovery - key frame selection - has also been selected.

It has been argued [23], that the selected key-frame selection algorithm, following Pollefeys [18], is not optimal, and a better approach has been presented. The benefits of the approach presented in [23] for the proposed system should be evaluated.

Knowledge of the 3D model and geometry of the scene can be further used to visually augment additional information into video from the endoscopic. Similar systems, although with different applications, have been presented

[18]. It is interesting to apply the augmented-reality approach to laparoscopic training and examine it's usability in the Operating Rooms. The presented system can serve as a basis for such augmented-reality trainer.

REFERENCES

- [1] J. Hansen, "Laparoscopic versus Open Appendectomy: Prospective Randomized Trial," *World Journal of Surgery*, vol. 20, no. 1, pp. 17–21, 1996.
- [2] J. Breasted, "The Edwin Smith Surgical Papyrus," 1991.
- [3] P. Gorman, A. Meier, and T. Krummel, "Computer-assisted training and learning in surgery," *Computer Aided Surgery*, vol. 5, no. 2, pp. 120–130, 2000.
- [4] P. Cosman, P. Cregan, C. Martin, and J. Cartmill, "Virtual reality simulators: Current status in acquisition and assessment of surgical skills," *ANZ Journal of Surgery*, vol. 72, no. 1, pp. 30–34, 2002.
- [5] R. Kneebone, "Simulation in surgical training: educational issues and practical implications," *Medical Education*, vol. 37, no. 3, pp. 267–277, 2003.
- [6] N. Seymour, A. Gallagher, S. Roman, M. O'Brien, V. Bansal, D. Andersen, and R. Satava, "Virtual reality training improves operating room performance: results of a randomized, double-blinded study," *Ann Surg*, vol. 236, no. 4, pp. 458–63, 2002.
- [7] T. Grantcharov, V. Kristiansen, J. Bendix, L. Bardram, J. Rosenberg, and P. Funch-Jensen, "Randomized clinical trial of virtual reality simulation for laparoscopic skills training," *British Journal of Surgery*, vol. 91, no. 2, pp. 146–150, 2004.
- [8] K. Semm, "Pelvi-trainer, a training device in operative pelviscopy for teaching endoscopic ligation and suture technics," *Geburtshilfe und Frauenheilkunde*, vol. 46, no. 1, pp. 60–2, 1986.
- [9] U. Kühnapfel, H. Çakmak, and H. Maas, "3D Modeling for Endoscopic Surgery," *Proceedings of the IEEE Symposium on Simulation*, pp. 22–32, 1999.
- [10] U. Kühnapfel, H. Krumm, C. Kuhn, M. Hübner, and B. Neisius, "Endosurgery Simulations with KISMET: A flexible tool for Surgical Instrument Design, Operation Room Planning and VR Technology based Abdominal Surgery Training," *Proc. Virtual reality World*, vol. 95, pp. 165–171, 1995.
- [11] S. Fraser, D. Klassen, L. Feldman, G. Ghitulescu, D. Stanbridge, and G. Fried, "Evaluating laparoscopic skills," *Surgical Endoscopy*, vol. 17, no. 6, pp. 964–967, 2003.
- [12] Simbionix Ltd., <http://www.simbionix.com>, 2006.
- [13] C. Feng, J. Rozenblit, and A. Hamilton, "A Hybrid View in a Laparoscopic Surgery Training System," *Proceedings of the 14th Annual IEEE International Conference and Workshops on the Engineering of Computer-Based Systems*, pp. 339–348, 2007.
- [14] A. Derossis, G. Fried, M. Abrahamowicz, H. Sigman, J. Barkun, and J. Meakins, "Development of a model for training and evaluation of laparoscopic skills," *Am J Surg*, vol. 175, no. 6, pp. 482–7, 1998.
- [15] G. Fried, L. Feldman, M. Vassiliou, S. Fraser, D. Stanbridge, G. Ghitulescu, and C. Andrew, "Proving the value of simulation in laparoscopic surgery," *Ann Surg*, vol. 240, no. 3, pp. 518–528, 2004.
- [16] B. Horn and B. Schunck, "Determining Optical Flow," *Artificial Intelligence*, vol. 17, no. 1-3, pp. 185–203, 1981.
- [17] S. Ullman and E. Hildreth, "The measurement of visual motion," *Physical and Biological Processing of Images* Eds OJ Braddick, AC Sleigh (New York: Springer) pp. 154–176, 1983.
- [18] M. Pollefeys, L. Van Gool, M. Vergauwen, F. Verbiest, K. Cornelis, J. Tops, and R. Koch, "Visual Modeling with a Hand-Held Camera," *International Journal of Computer Vision*, vol. 59, no. 3, pp. 207–232, 2004.
- [19] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2003.
- [20] C. Harris and M. Stephens, "A combined corner and edge detector," *Alvey Vision Conference*, vol. 15, p. 50, 1988.
- [21] P. Torr, A. Fitzgibbon, and A. Zisserman, "The Problem of Degeneracy in Structure and Motion Recovery from Uncalibrated Image Sequences," *International Journal of Computer Vision*, vol. 32, no. 1, pp. 27–44, 1999.
- [22] S. Gibson, J. Cook, T. Howard, R. Hubbold, and D. Oram, "Accurate camera calibration for off-line, video-based augmented reality," *Mixed and Augmented Reality, 2002. ISMAR 2002. Proceedings. International Symposium on*, pp. 37–46, 2002.
- [23] T. Thormahlen, H. Broszio, and A. Weissenfeld, "Keyframe selection for camera motion and structure estimation from multiple views," *Proceedings of European Conference on Computer Vision*, pp. 523–535, 2004.