

# Light Fields for Minimal Invasive Surgery Using an Endoscope Positioning Robot

F. Vogt<sup>1</sup>, S. Krüger<sup>2</sup>, J. Schmidt<sup>1</sup>, D. Paulus<sup>3</sup>, H. Niemann<sup>1</sup>, W. Hohenberger<sup>2</sup>, C. H. Schick<sup>2</sup>

<sup>1</sup>Chair for Pattern Recognition, University of Erlangen-Nuremberg, Erlangen, Germany

<sup>2</sup>Department of Surgery, University of Erlangen-Nuremberg, Erlangen, Germany

<sup>3</sup>Institute for Computational Visualistics, University of Koblenz-Landau, Koblenz, Germany

## Summary

**Objectives:** To generate a fast and robust 3-D visualization of the operation site during minimal invasive surgery.

**Methods:** Light fields are used to model and visualize the 3-D operation site during minimal invasive surgery. An endoscope positioning robot provides the position and orientation of the endoscope. The a priori unknown transformation from the endoscope plug to the endoscope tip (hand-eye transformation) can either be determined by a three-step algorithm, which includes measuring the endoscope length by hand or by using an automatic hand-eye calibration algorithm. Both methods are described in this paper and their respective computation times and accuracies are compared.

**Results:** Light fields were generated during real operations and in the laboratory. The comparison of the two methods to determine the unknown hand-eye transformation was done in the laboratory. The results which are being presented in this paper are: rendered images from the generated light fields, the calculated extrinsic camera parameters and their accuracies with respect to the applied hand-eye calibration method, and computation times.

**Conclusion:** Using an endoscope positioning robot and knowing the hand-eye transformation, the fast and robust generation of light fields for minimal invasive surgery is possible.

## Keywords

Light field, hand-eye calibration, minimally invasive surgical procedures, endoscopy, 3-D visualization

Methods Inf Med 2004; 43: 403–8

## 1. Introduction

Development in the field of surgical operations is moving towards techniques called minimal invasive operation. These traumatize the patient less. The time of convalescence and complications after the operation are reduced. During the operation, the surgical instruments and the endoscope are introduced into the patient through small approaches, so-called trocars. Only very small incisions are needed (1-2 cm). The surgeon works without direct visual contact to the operation area since the camera image of the operation site (e.g., abdomen or thoracic cavity) is displayed on a video monitor for visual feedback (see also Fig. 1).

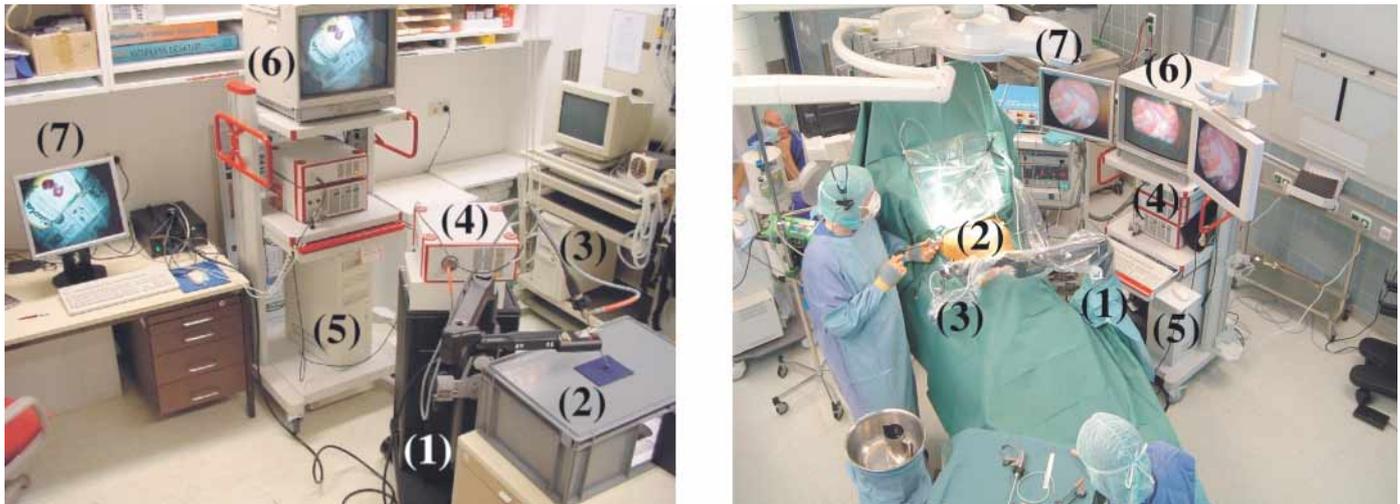
During the course of an operation, image quality may be low due to degradations of the endoscopic images, e.g. the cutting techniques using high frequency diathermia lead to smoke and small flying particles. If blood covers large areas of the visible field, the image gets reddish and the ability to discriminate different tissue types is reduced. Navigation and vision are more difficult compared to a conventional operation with a large incision: the 3-D impression is lost due to the visual feedback on a monitor and the camera image usually only shows a part of the operation site.

Currently, only hardware-based image enhancement methods are used (e.g., white balancing, i.e. telling the camera what “white” is). These are usually only used once at the beginning of an operation. Apart from the reduction of image degradations in endoscopic images [1-3], a 3-D visualization of the operation site would be very helpful for the surgeon. Thus, the operation area can be observed without

moving the endoscope, or additional information could be displayed in the 3-D visualization. The generation of a realistic 3-D surface model using endoscopic images of the colon is described in [4]. This model is obtained by point tracking and texture mapping (putting the image onto the 3-D surface) of the image data to the calculated triangular mesh of the surface (approximation of a surface by triangles where the corners of the triangles are the 3-D points).

Modeling 3-D scenes using *light fields* [5] allows the generation of new views of a scene, even views of a scene the real camera has never seen. Light fields are therefore suitable to provide a 3-D visualization of the operation site. A light field consists of extrinsic and intrinsic camera parameters, namely, pose (i.e. position and orientation) and focal length and principal point (i.e., the intersection of the optical axis with the sensor chip), respectively, and also the image data at each camera pose. If an appropriate parameterization is used, views of the scene can be rendered in real-time. This contribution focuses on the techniques required for generating light fields during minimal invasive operations. The authors do not know of any research group working on the special problem of endoscopic light fields.

Applying structure-from-motion techniques, the generation of a light field is possible based only on the video source, where the extrinsic and intrinsic camera parameters are estimated by point tracking, self-calibration (i.e., estimating the intrinsic and extrinsic camera parameters), and 3-D reconstruction (i.e., estimating 3-D points of the surface) [5-7]. The application of these techniques to endoscopic sequences is very



**Fig. 1** Experimental setup, laboratory (left image) and surgery (right image): (1) AESOP 3000, (2) patient, (3) camera head and endoscope, (4) light source, (5) PC, (6) video-endoscopic system (original image), (7) second monitor (processed image/light field)

difficult and only possible under certain prerequisites: no movement inside the scene, smooth camera movement during recording of the image sequence, structure in the scene for point tracking, and good illumination conditions. The algorithm fails if the prerequisites are not fulfilled. As our goal is the generation of a light field of an endoscopic sequence, an alternative way of determining the camera parameters is needed. Besides optical or magnetic tracking [8, 9] an endoscope positioning robot like the AESOP 3000, which is presently available in many operating rooms, allows the robust calculation of the extrinsic camera parameters during minimal invasive surgery. The advantages of AESOP 3000 compared to the alternatives are: no interferences with metallic objects like the operation table or surgical instruments (a problem for magnetic tracking) and no “line of sight” requirement, i.e. the tracking system has to “see” the tracked object (needed for optical tracking). Furthermore, the robot arm possibly is already installed in the operation room (as in our case) and reduces the number of persons required for a minimal invasive operation. Only the extrinsic parameters can be determined with AESOP 3000 (or with any other tracking system). The intrinsic camera parameters can be estimated using a calibration pattern (an object with known 3-D geometry). Knowing the kinematics of the robot arm,

the problem of determining the extrinsic camera parameters reduces to the problem of determining the a priori unknown transformation from the endoscope plug to the endoscope tip (so-called hand-eye calibration).

In this contribution, we describe and compare two techniques for hand-eye calibration of the robot arm AESOP 3000: a semi-automatic and a fully automatic approach. Based on a defined data set (here, an image sequence of a calibration pattern), a fully automatic approach does not require any human interaction whereas a semi-automatic approach consists of one or more steps where human interaction is usually used for hand-eye calibration.

The experiments for the comparison between the two techniques were made using a simulated endoscopic setup in the laboratory. The same setup is also used before an operation to determine the hand-eye transformation. Knowing the hand-eye transformation, light fields can be generated, which provide a 3-D visualization of the operation site. Sixteen light fields were generated in the laboratory, three during a minimally invasive operation. Section 2 addresses the problem of hand-eye calibration; Section 3 summarizes the experiments and results. This article closes with a discussion of the results and a conclusion (Sections 4 and 5).

## 2. Methods

The Computer Motion Inc. robot arm AESOP 3000 used (see Fig. 1) has 7 degrees of freedom (3 active and 2 passive joints, 1 manually adjustable joint, and 1 active translation). A PC grabs the images from the endoscope directly from the endoscopic camera with a frame grabber card. Before and after grabbing an image, the 7 degrees of freedom of the robot arm (one length and six angular values) are read out from the robot arm controller through a serial interface and stored. The following subsections describe the process of calculating the camera pose from these 7 values. The setup for the experiments in the laboratory and in the operation room is shown in Figure 1.

First, a calibration pattern is used to determine the intrinsic camera parameters [10]. Radial and tangential distortion coefficients are also calculated. They are used for correcting the highly distorted endoscopic images. The distortion is caused by the small focal length of the optics used.

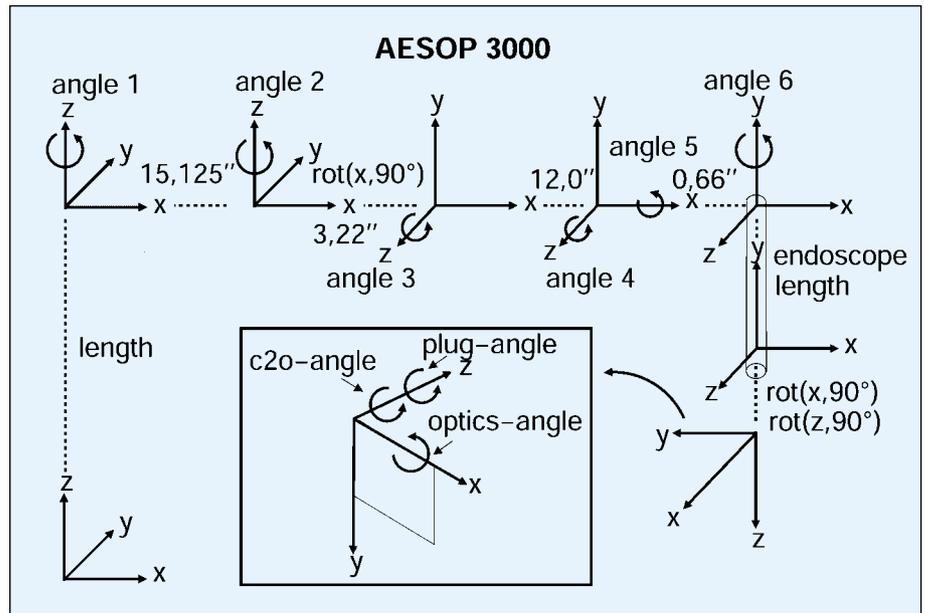
The mean values of the length and the six angular values of the robot arm before and after grabbing the image are used. Applying the known transformations of the robot kinematics (see Fig. 2), the pose of the endoscope plug can be calculated, but not the required pose of the endoscope tip (the „real“ camera position). Since the op-

tics are mounted in a new position into the plug for each surgery, the a priori unknown transformation from the endoscope plug to the endoscope tip has to be determined before each operation. This problem is also known as hand-eye calibration. In the following two subsections, two solutions are presented. With each solution, the real camera pose for each image relative to the robot coordinate system can be calculated.

### 2.1 Single Step, Semi-automatic Approach

This approach extends the kinematics to describe the unknown transformation, which leads to four unknown values that have to be determined (see Fig. 2): the length of the endoscope (endoscope length), the angle between camera head and optics (c2o-angle), the angle inside the endoscope plug (plug-angle), and the angle of the optics (optics-angle). Only the angle of the optics is known, which in our case is 30°. The length of the endoscope is measured by hand. For the c2o-angle, a notch at the optics border is detected with a simple threshold technique (see Fig. 3). The notch is defined as the point with the largest distance to the middle point of the contour of the optics border.

Now only one unknown value, the plug-angle, is left. The idea for its determination is to calculate the relative movement between two images using a calibration pattern and adjust the plug-angle such that the relative movement calculated by the kinematics equals the real one. Let  $C_1$  and



**Fig. 2** Kinematics of the robot arm AESOP 3000: 7 degrees of freedom (length, angles 1, ..., 6) for the transformation from the robot origin to the endoscope plug; four a priori unknown but fixed values for the transformation from the endoscope plug to the endoscope tip (endoscope length, c2o-, plug- and optics-angle). c2o-angle is the angle between the camera head and the endoscope optics, plug-angle is the angle of the endoscope optics inside the endoscope plug, and optics-angle is the side-view angle of the optics.

$C_2$  be the rotation matrices obtained from the calibration pattern,  $t_{C1}$  and  $t_{C2}$  the corresponding translation vectors, and  $A_1, A_2, t_{A1}$ , and  $t_{A2}$  the parameters obtained from AESOP. The positions of the second camera relative to the first, denoted by  $t_{C12}$  and  $t_{A12}$ , are calculated as follows (the translation vector of the second camera is transformed into the coordinate system of the first camera):

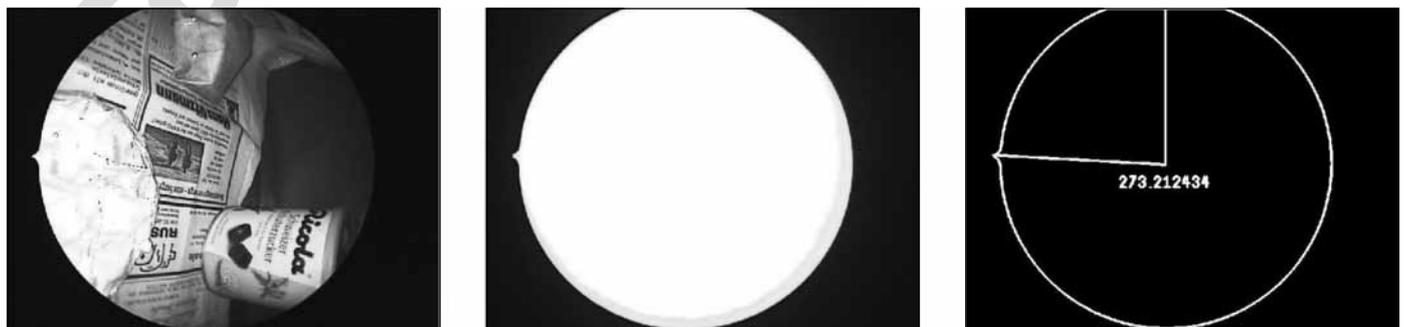
$$t_{C12} = C_1^T (t_{C2} - t_{C1}) \tag{1}$$

$$t_{A12} = A_1^T (t_{A2} - t_{A1}) \tag{2}$$

The 3-D angle  $\angle (t_{C12}, t_{A12})$  is used as a one-dimensional similarity measure that is optimized. Because the plug-angle is in the range  $[0^\circ, 360^\circ]$ , a loop with predefined accuracy over the range is sufficient to determine the value which minimizes  $\angle (t_{C12}, t_{A12})$ .

### 2.2 Automatic Hand-eye Calibration Approach

This approach extends the hand-eye calibration algorithm described in [11] by a



**Fig. 3** Determination of the angle between camera and optics (c2o-angle): original (left image), white paper (middle image), and calculated angle, detected notch and contour of the optics border (right image)

data selection step because the algorithm is very sensitive to the data used. The algorithm is shortly described first, followed by our method for data selection.

The problem of hand-eye calibration is shown in the following commutative diagram; robot arm poses are denoted by  $\mathbf{A}$ , camera poses by  $\mathbf{C}$  at two time steps  $i$  and  $k$ . The unknown hand-eye transformation is expressed by  $\mathbf{R}_{\text{HE}}$  and  $\mathbf{t}_{\text{HE}}$ :

$$\begin{array}{ccc} & \mathbf{R}_{\text{HE}}, \mathbf{t}_{\text{HE}} & \\ & \rightarrow & \\ \mathbf{A}_k & & \mathbf{C}_k \\ \uparrow & & \uparrow \\ \mathbf{R}_{\text{Aik}}, \mathbf{t}_{\text{Aik}} & & \mathbf{R}_{\text{Cik}}, \mathbf{t}_{\text{Cik}} \quad (3) \\ & \mathbf{R}_{\text{HE}}, \mathbf{t}_{\text{HE}} & \\ & \rightarrow & \\ \mathbf{A}_i & & \mathbf{C}_i \end{array}$$

The hand-eye parameters  $\mathbf{R}_{\text{HE}}$  and  $\mathbf{t}_{\text{HE}}$  can be recovered from the diagram above using the commutative law (3):

$$\mathbf{R}_{\text{HE}} \mathbf{R}_{\text{Aik}} = \mathbf{R}_{\text{Cik}} \mathbf{R}_{\text{HE}} \quad (4)$$

$$(\mathbf{I}_3 - \mathbf{R}_{\text{Cik}}) \mathbf{t}_{\text{HE}} = \mathbf{t}_{\text{Cik}} - \mathbf{R}_{\text{HE}} \mathbf{t}_{\text{Aik}} \quad (5)$$

where  $\mathbf{I}_3$  is the  $3 \times 3$  identity matrix.

The classical way is to first solve (4) for  $\mathbf{R}_{\text{HE}}$ , and then (5) for  $\mathbf{t}_{\text{HE}}$ . Daniilidis [11] is the first who presented a linear algorithm for simultaneous computation of the hand-eye parameters.

Note that at least two motions of the robot arm/camera with different rotation axes are necessary for reconstructing the rigid hand-eye transformation. Hence, it is usually suboptimal to process the arm/camera positions in their temporal order. It is much

better to select the data so that those relative movements are used for calibration that actually fulfil the requirement above. As an optimality criterion we propose to use the scalar product between well-defined rotation axes of two camera movements (for rotations very similar to the identity the rotation axis is not well-defined). Let  $\mathbf{a}_{ij}$  and  $\mathbf{a}_{kl}$  be the normalized rotation axes of two relative movements from frame  $i$  to  $j$  and from  $k$  to  $l$ , respectively. Then

$$s_{ij,kl} = |\mathbf{a}_{ij}^T \mathbf{a}_{kl}| \quad (6)$$

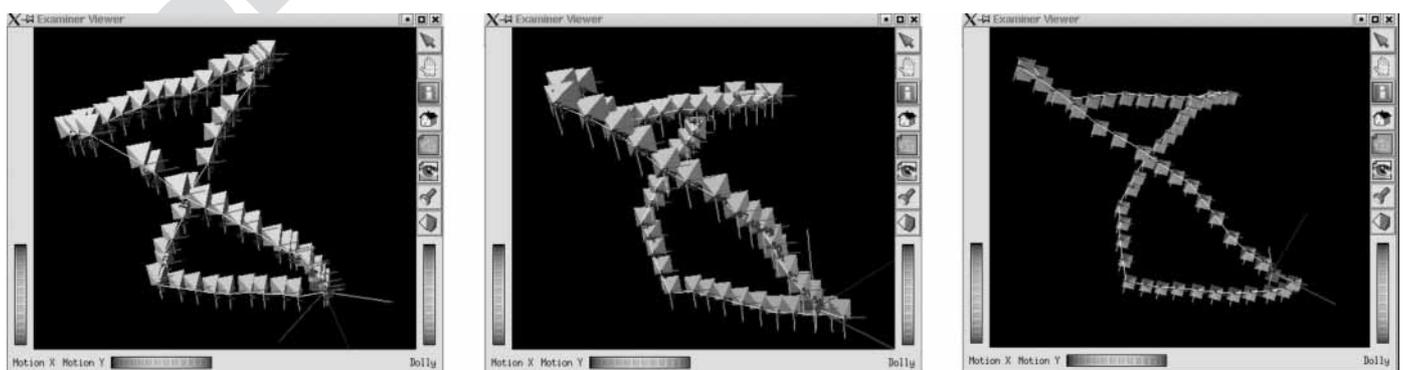
gives a value of one for parallel rotation axes, and zero for orthogonal axes. The camera movements with orthogonal rotation axes are the best suited ones for hand-eye calibration. Note that the camera and not the robot arm data should be used at this point, since the camera was calibrated accurately using a calibration pattern, while the data provided by the robot is still corrupted by outliers and hence unreliable.

### 3. Results

For the simulated endoscopic setup (cf. Fig. 1), one half of the “patient”, a box with an opening for the endoscope, was inlaid with newspaper, the other half with printed color images from real operations (cholecystectomies). To compare the two methods for hand-eye calibration, two image sequences of a calibration pattern were captured (ALF014 and ALF040). Figure 4

displays the camera poses of the sequence ALF014 compared to those calculated by using the two-hand eye calibration methods. The intrinsic camera parameters were calculated in advance from 10 images of a calibration pattern. The extrinsic camera parameters of the two sequences obtained by a camera calibration algorithm were defined as ground truth data, because in our case no other ground truth data can be acquired. The mean errors per frame for these two sequences are shown in Table 1. Translation errors range from 0 to 4 mm, rotation errors range from 0 to  $1.5^\circ$  per axis for relative movements from image to image (for both methods). As can be seen, the accuracy of the two approaches is comparable. Since the accuracy of the automatic approach is very sensitive to the data used and the overall time needed for obtaining the unknown transformation is about the same for both methods ( $\approx 10$  minutes, including capturing of the needed calibration images), the experiments presented in the following were done with the three-step semi-automatic method.

Sixteen light fields of artificial objects (e.g., candy can, newspaper ball) under OR-realistic conditions, and 3 light fields during real minimal invasive surgeries were generated (cf. Fig. 1). A rendered image of the light field generated from the laboratory sequence ALF029 can be found in Figure 5. Figure 6 shows the camera poses and a rendered image from one of the light fields obtained during minimal invasive surgery of the thoracic cavity (ALF051). The camera pose for the rendered images



**Fig. 4** Sequence ALF014: extrinsic camera parameters (the pyramids represent the camera poses), determined by camera calibration (left image), single step (middle image) and automatic hand-eye calibration (right image)

was selected arbitrarily between the original camera poses. The 3-D effect of the visualization is only visible when the camera pose is moved interactively during the rendering. The generation of a typical light field (not including hand-eye calibration) lasts 3-5 minutes; Table 2 summarizes the computation times of selected light fields.

## 4. Discussion

The single step, semi-automatic approach is more robust than the automatic approach, which is very sensitive to the input data (movement pairs) provided. Therefore, we currently prefer the single step approach.

In contrast to the conventional method of generating light fields by point tracking, self-calibration, and 3-D reconstruction [6],

**Table 1** Mean error per frame for single step and automatic method, once with best movement pairs, once for pairs in temporal order. For the Euler angles ( $\alpha$ ,  $\beta$ ,  $\gamma$ ), the error is given in degrees and for the translation in mm.

Sequence / Method	$\alpha$	$\beta$	$\gamma$	Transl.
ALF014, single step	0.29	0.28	0.25	0.67
ALF014, automatic, best 30% (= 36,532 pairs)	0.22	0.38	0.24	0.89
ALF014, automatic, temporal order	0.94	0.73	0.68	10.7
ALF040, single step	0.26	0.26	0.35	0.91
ALF040, automatic, best 30% (= 1,721 pairs)	0.20	0.26	0.29	1.10
ALF040, automatic, temporal order	0.50	0.83	1.5	10.8

the presented method can be applied to endoscopic sequences and is very fast. The quality of the generated light fields depends mainly on the correctness of the

determined hand-eye transformation and on the accuracy of the robot arm. A problem that was not addressed in this paper is the inaccuracy of the robot arm due to the



**Fig. 5** Sequence ALF029: original endoscopic image (left image), rendered image from the generated light field (middle image), calculated extrinsic camera parameters (right image, the pyramids represent the camera poses)



**Fig. 6** Sequence ALF051: original (left image), rendered image from the generated light field (middle image), calculated extrinsic camera parameters (right image, the pyramids represent the camera poses)

Sequence	#Images	Light field generation	LF-type
ALF027	16	01:26 min	A
ALF014	54	04:26 min	A
ALF040	100	05:45 min	A
ALF029	143	05:10 min	A
ALF028	236	05:20 min	A
ALF051	45	03:04 min	AR

**Table 2**  
Computation times (Pentium IV, 2.4 GHz): LF-types: A = AESOP in laboratory, AR = AESOP in real surgery; image size 512 × 512.

manufacturing. According to Computer Motion Inc., the relative error of the position of the endoscope plug is about 1%.

The quality of the rendered images decreases if the distance of the objects from the endoscope becomes shorter. This is due to the inherent assumption of a plane as the geometry of the scene, since currently no depth information is available (it is a part of future work). The lack of depth information is the main disadvantage compared to the conventional method, where 3-D points of the scene are reconstructed.

## 5. Conclusions

Two methods were presented to determine the a priori unknown hand-eye transformation from the endoscope plug to the real camera pose for the robot arm AESOP 3000. Both methods are comparable with respect to computation time and accuracy, therefore the decision which method is going to be used in the future will depend on which one is better suited for usage in the operation room.

If an endoscope positioning robot is available in the operation room, the generation of light fields for minimal invasive surgery with the technique described in this work is fast (3-5 minutes, depending on the number of images per sequence) and robust (i.e. without adjusting scene-dependent parameters). Three to five minutes is a reasonable time for the surgeon to perform other tasks or wait while the 3-D visualization is generated, i.e. the algorithms are fast

enough to be used during an operation. The test of our system during a real minimal invasive endoscopic operation was very positive: after determining the hand-eye transformation with the semi-automatic method, the (automatic) generation of light fields was possible and the surgeon could navigate through the 3-D visualization of the operation site. With our method, the surgeon gets back the lost 3-D impression of the operation site during minimal invasive operations.

To increase the quality of the light field visualization in the future, the integration of information about the geometry of the surface of the scene (depth maps) is possible. Surface geometry will also be the basis of our approach for the fusion of the visualization with CT or MRT data.

### Acknowledgments

This work was partly funded by Deutsche Forschungsgemeinschaft (DFG) under grant SFB 603/TP B6. Only the authors are responsible for the content. The authors express their thanks to Christian Vogelgsang, who provided the graphics and rendering part. His work is also supported by DFG under grant SFB 603/TP C2.

## References

- Vogt F, Klimowicz C, Paulus D, Hohenberger W, Niemann H, Schick CH. Bildverarbeitung in der Endoskopie des Bauchraums. In: Handels H, Horsch A, Lehmann T, Meinzer HP (eds). 5. Workshop Bildverarbeitung für die Medizin; 2001 (October); Lübeck, Germany. Berlin: Springer; 2001; pp. 320-4.
- Palm C, Lehmann T, Spitzer K. Bestimmung der Lichtquellenfarbe bei der Endoskopie makrotexturierter Oberflächen des Kehlkopfs. In: Franke KH (ed). 5. Workshop Farbbild-

verarbeitung; 1999. Ilmenau, Germany. Ilmenau: Ilmenau Schriftenreihe des Zentrums für Bild- und Signalverarbeitung e.V. Ilmenau; pp. 3-10.

- Vogt F, Paulus D, Niemann H. Highlight Substitution in Light Fields. In: Proceedings of the IEEE International Conference on Image Processing (ICIP), 2002 (September); Rochester, USA. IEEE Computer Society Press; pp. 637-40.
- Thormählen T, Broszio H, Meier PN. Automatische 3D-Rekonstruktion aus endoskopischen Bildfolgen. In: Meiler et al. [12]; pp. 207-10.
- Levoy M, Hanrahan P. Light Field Rendering. Computer Graphics Proceedings, Annual Conference Series (Proc. SIGGRAPH); 1996; pp. 31-42.
- Heigl B, Koch R, Pollefeys M, Denzler J, Gool LV. Plenoptic Modeling and Rendering from Image Sequences Taken by a Hand-Held Camera. In: Förstner W, Buhmann J, Faber A, Faber P (eds). Mustererkennung; 1999 (September); Heidelberg: Springer; 2001; pp. 94-101.
- Pollefeys M. Self-Calibration and Metric 3D Reconstruction from Uncalibrated Image Sequences. Katholieke Universiteit Leuven, Belgium; May 1999.
- Holl K, Wies W, Wurm G. Clinical Benefit of Image Guidance in Neurosurgery. In: Lempke HU, Vannier MW, Inamura K, Farman AG, Doi K (eds). Computer Assisted Radiology and Surgery, Proceedings of the 15th International Congress and Exhibition CARS, 2001; Amsterdam: Elsevier Science B.V. 2001; pp. 99-101.
- Scheuering M, Rezk-Salama C, Barfuss H, Barth K, Schneider A, Greiner G et al. Intraoperative Augmented Reality with Magnetic Navigation and Multi-texture Based Volume Rendering for Minimally Invasive Surgery. Rechner- und Sensorgestützte Chirurgie 2001; pp. 83-91.
- Tsai RY. A Versatile Camera Calibration Technique for High-Accuracy 3D Machine Vision Metrology using Off-the-Shelf TV Cameras and Lenses. IEEE Journal of Robotics and Automation 1987; Ra-3 (3): 323-44.
- Daniilidis K. Hand-Eye Calibration Using Dual Quaternions. International Journal of Robotics Research 1999; 18: 286-98.
- Meiler M, Handels H, Kruggel F, Lehmann T, Saupe D (eds). 6. Workshop Bildverarbeitung für die Medizin 2002, Leipzig 2002. Berlin: Springer; 2002.

### Correspondence to:

Dipl.-Inf. Florian Vogt  
Lehrstuhl für Mustererkennung (Informatik 5)  
Universität Erlangen-Nürnberg  
Martensstraße 3  
91058 Erlangen  
Germany  
E-mail: vogt@informatik.uni-erlangen.de