Calibration–Free Hand–Eye Calibration: A Structure–from–Motion Approach

Jochen Schmidt*, Florian Vogt**, and Heinrich Niemann

Lehrstuhl für Mustererkennung, Universität Erlangen-Nürnberg
Martensstr. 3, 91058 Erlangen, Germany
{jschmidt,vogt,niemann}@informatik.uni-erlangen.de

Abstract. The paper presents an extended hand-eye calibration approach that, in contrast to the standard method, does not require a calibration pattern for determining camera position and orientation. Instead, a structure-from-motion algorithm is applied for obtaining the eye-data that is necessary for computing the unknown hand-eye transformation. Different ways of extending the standard algorithm are presented, which mainly involves the estimation of a scale factor in addition to rotation and translation. The proposed methods are experimentally compared using data obtained from an optical tracking system that determines the pose of an endoscopic camera. The approach is of special interest in our clinical setup, as the usage of an unsterile calibration pattern is difficult in a sterile environment.

1 Introduction

Hand-eye calibration algorithms [9, 10, 7, 5] solve the following problem that originated in the robotics community: Given a robot arm and a camera mounted on that arm, compute the rigid transformation from arm to camera (hand-eye transformation). Knowledge of this transformation is necessary, because the pose of the robot arm is usually provided by the robot itself, while the pose of the camera is unknown but needed for visual guidance of the arm. However, if the hand-eye transformation is known the camera pose can be computed directly from the pose data provided by the robot.

Usually, the camera (*eye*) poses are computed using a calibration pattern and standard camera calibration techniques. In contrast to that, a method for hand-eye calibration is presented in this paper, where no calibration pattern is needed. Instead, the camera poses are obtained solely from an image sequence recorded using a hand-held camera by applying structure-from-motion methods.

Hand-eye calibration is also interesting for applications that are not directly related to robotics, but where similar problems arise. Instead of a robot we used an optical tracking system that provides hand data, and a camera, where the camera poses (*eye*) are computed using a calibration pattern for standard hand-eye calibration, and structure-from-motion for the extended hand-eye calibration described in this paper. The camera may in general be an arbitrary hand-held video camera. For our application—the reconstruction of high-quality medical light fields [11]—we used an endoscope with a rigidly

* This work was partially funded by the European Commission 5th IST Programme - Project VAMPIRE. Only the authors are responsible for the content.

** This work was partially funded by the Deutsche Forschungsgemeinschaft (DFG) under grant SFB 603/TP B6. Only the authors are responsible for the content.

mounted CCD camera. The endoscope is moved by hand, its pose is determined by the optical tracking system. More details on this system will follow in the experiments section. The hand-eye transformation has to be estimated every time when the camera head is mounted anew on the endoscope optics, which is done before each operation because the endoscope has to be sterilized. This requires an algorithm that works automatically and fast with a minimum of human interaction.

The paper is structured as follows: After an introduction to hand-eye calibration in Sect. 2 the structure-from-motion algorithm will be described (Sect. 3). Section 4 shows the modifications to the hand-eye calibration equations that are necessary when using structure-from-motion instead of standard camera calibration. Experimental results are presented in Sect. 5.

2 Hand-Eye Calibration

The first hand-eye calibration methods were published by Tsai and Lenz [10], and Shiu and Ahmad [9], where the latter formulated the hand-eye calibration problem as a matrix equation of the form

$$\boldsymbol{T}_{\mathrm{E}ij}\boldsymbol{T}_{\mathrm{HE}} = \boldsymbol{T}_{\mathrm{HE}}\boldsymbol{T}_{\mathrm{H}ij}, \quad \boldsymbol{T}_{\chi} = \begin{pmatrix} \boldsymbol{R}_{\chi} & \boldsymbol{t}_{\chi} \\ \boldsymbol{0}_3^{\mathrm{T}} & 1 \end{pmatrix}, \chi \in \{\mathrm{H}ij, \mathrm{E}ij, \mathrm{HE}\} \quad . \tag{1}$$

$\boldsymbol{T}_{\mathrm{H}ij}$ is the robot arm (*hand*) movement from time step i to j, $\boldsymbol{T}_{\mathrm{E}ij}$ the camera (*eye*) movement, and $\boldsymbol{T}_{\mathrm{HE}}$ is the unknown hand-eye transformation, i. e. the transformation from gripper to camera. All transformations $\boldsymbol{T}_{\chi}$ are described by a 3×3 rotation matrix $\boldsymbol{R}_{\chi}$ and a 3-D translation vector $\boldsymbol{t}_{\chi}$. Equation (1) can be directly derived from the following diagram:

$$\begin{array}{ccc} \boldsymbol{H}_j & \xrightarrow{\boldsymbol{T}_{\mathrm{HE}}} & \boldsymbol{E}_j \\ {\scriptstyle \boldsymbol{T}_{\mathrm{H}ij}}\uparrow & & \uparrow{\scriptstyle \boldsymbol{T}_{\mathrm{E}ij}} \\ \boldsymbol{H}_i & \xrightarrow{\boldsymbol{T}_{\mathrm{HE}}} & \boldsymbol{E}_i \end{array} \tag{2}$$

$\boldsymbol{H}_i$ and $\boldsymbol{H}_j$ denote the gripper poses, $\boldsymbol{E}_i$ and $\boldsymbol{E}_j$ the camera poses at times i, j. The usual way to solve (1) is to split it into two separate equations, one that contains only rotation, and a second one that contains rotation and translation:

$$\boldsymbol{R}_{\mathrm{HE}}\boldsymbol{R}_{\mathrm{H}ij} = \boldsymbol{R}_{\mathrm{E}ij}\boldsymbol{R}_{\mathrm{HE}}, \qquad (\boldsymbol{I}_{3\times 3} - \boldsymbol{R}_{\mathrm{E}ij})\boldsymbol{t}_{\mathrm{HE}} = \boldsymbol{t}_{\mathrm{E}ij} - \boldsymbol{R}_{\mathrm{HE}}\boldsymbol{t}_{\mathrm{H}ij} \quad . \tag{3}$$

Thus, the rotational part of the hand-eye transformation can be determined first, and, after inserting it into the second equation, the translational part can be computed. This is the way hand-eye calibration is done, e. g., in [9, 10, 3]. Different parameterizations of rotation have been applied. The original works of [9, 10] use the axis/angle representation, quaternions were used by [3, 7], and dual quaternions were introduced by [5]. In contrast to the former approaches, it was suggested in [2] that rotation and translation should be solved for simultaneously and not separately. This approach is also followed in [7], where a non-linear optimization of rotation and translation is done.

3 Structure-from-Motion

The usual way to obtain the camera poses $\boldsymbol{E}_i$ is to capture images of a calibration pattern and apply standard camera calibration techniques [12]. In contrast to that our

approach is capable of using an image sequence without a pattern and an uncalibrated camera. By applying a structure-from-motion approach, the camera motion (and therefore the eye poses $\boldsymbol{E}_i$) can be computed for each recorded image. Basically, it is possible to use any algorithm that results in camera poses, as the following computation steps do not rely on the actual method used. The approach applied in this paper is based on the work of [6]; it will be outlined in the following.

The algorithm starts with establishing 2-D point correspondences between images. Each detected feature point has to be tracked over a sequence of images to allow a 3-D reconstruction. Here, a modified more robust and faster version of the Tomasi-Kanade-Shi tracker is used that can also deal with illumination changes [13].

After point features are tracked, the actual 3-D reconstruction step starts. First, an initial reconstruction is computed using the paraperspective factorization algorithm on a subset of the images, since all features have to be visible in all images. The reconstructed affine cameras are now converted to perspective ones by assuming a reasonable value for focal length and by choosing the center of the image as the principal point. These perspective cameras can be used as an initialization for a non-linear optimization step, where camera matrices and 3-D points are optimized alternatingly.

The initial sequence is now extended by performing the following steps for each frame that is to be added: First, 3-D scene points are triangulated from feature points that are visible in the new image using already reconstructed camera matrices. This way it is possible to use the triangulated points as calibration points and apply standard camera calibration techniques. In fact, since differences from one camera pose to the next will usually be small, it is sufficient in practice to skip the linear standard calibration methods and initialize the new camera pose with the parameters of the neighboring one. Non-linear optimization of this camera will yield the desired result. These two steps are repeated until all frames are processed. Optionally, the whole reconstruction can be optimized non-linearly by a final bundle-adjustment step.

The result is a reconstruction of the 3-D scene points as well as the extrinsic and intrinsic camera parameters for each recorded image. Note, however, that the reconstruction is only unique up to a similarity transformation, i. e., the world coordinate system can be chosen arbitrarily, and the scale of the reconstruction is unknown. While the choice of the world coordinate system is exactly the problem that is solved with standard hand-eye calibration, the unknown scale factor has to be estimated additionally. This topic will now be addressed.

4 Extended Hand-Eye Calibration

The drawback of using structure-from-motion instead of camera calibration is that the scaling factor mentioned above has to be estimated in addition to rotation and translation during hand-eye calibration; the modified method will be called *extended hand-eye calibration* in the following.

In [1] a structure-from-motion based hand-eye calibration approach was presented already, where the scaling factor has been integrated into the standard equations (3). The main drawback of that method is that the orthogonality of the rotation matrix $\boldsymbol{R}_{\mathrm{HE}}$ is not guaranteed by the extended equations, but has to be enforced afterwards using the SVD.

4.1 Extension of Basic Equations

The straight-forward method is to extend (3) by a scaling factor s_{HE}, resulting in:

$$\boldsymbol{R}_{\mathrm{HE}}\boldsymbol{R}_{\mathrm{H}ij} = \boldsymbol{R}_{\mathrm{E}ij}\boldsymbol{R}_{\mathrm{HE}} \quad , \tag{4}$$

$$(\boldsymbol{I}_{3\times 3} - \boldsymbol{R}_{\mathrm{E}ij})\boldsymbol{t}_{\mathrm{HE}} = \boldsymbol{t}_{\mathrm{E}ij} - s_{\mathrm{HE}}\boldsymbol{R}_{\mathrm{HE}}\boldsymbol{t}_{\mathrm{H}ij} \quad . \tag{5}$$

It can be observed that the rotational equation in (3) and eq. (4) are the same, i. e., the scale factor has no influence on the computation of rotation. Therefore, the rotation can be obtained by standard methods, e. g., using the quaternion representation of rotations [3, 7], which guarantees that the resulting matrix actually is a rotation. Equation (5), however, contains translation and scale, and can be formulated as a linear system of equations as follows:

$$\left((\boldsymbol{I}_{3\times 3} - \boldsymbol{R}_{\mathrm{E}ij})\ \boldsymbol{R}_{\mathrm{HE}}\boldsymbol{t}_{\mathrm{H}ij}\right)\begin{pmatrix}\boldsymbol{t}_{\mathrm{HE}}\\ s_{\mathrm{HE}}\end{pmatrix} = \boldsymbol{t}_{\mathrm{E}ij} \quad . \tag{6}$$

This method of extended hand-eye calibration has the advantage that all equation systems are linear, but the disadvantage that one has to solve for rotation first, and then for translation and scale.

The equations (4) and (5) can be used to formulate an objective function $f(\cdot)$ for non-linear optimization, which is based on the objective function for standard hand-eye calibration proposed by [7]:

$$\begin{aligned} f(\boldsymbol{q}_{\mathrm{HE}}, \boldsymbol{t}_{\mathrm{HE}}, s_{\mathrm{HE}}) = &\sum_{i=1}^{N_{\mathrm{rel}}} \left\|\boldsymbol{q}_{\mathrm{E}i} - \boldsymbol{q}_{\mathrm{HE}}\boldsymbol{q}_{\mathrm{H}i}\boldsymbol{q}_{\mathrm{HE}}^{*}\right\|^2 + \\ &\sum_{i=1}^{N_{\mathrm{rel}}} \left\|Q\left((\boldsymbol{I}_{3\times 3} - \boldsymbol{R}_{\mathrm{E}i})\,\boldsymbol{t}_{\mathrm{HE}} - \boldsymbol{t}_{\mathrm{E}i}\right) + \boldsymbol{q}_{\mathrm{HE}}Q(s_{\mathrm{HE}}\boldsymbol{t}_{\mathrm{H}i})\boldsymbol{q}_{\mathrm{HE}}^{*}\right\|^2 + \lambda\left(1 - \boldsymbol{q}_{\mathrm{HE}}\boldsymbol{q}_{\mathrm{HE}}^{*}\right)^2 . \end{aligned} \tag{7}$$

where $\boldsymbol{q}_{\mathrm{HE}}$ is the quaternion used for parameterization of the rotation matrix $\boldsymbol{R}_{\mathrm{HE}}$ and λ is a regularization factor (e. g., $\lambda = 2 \cdot 10^6$) that penalizes deviations of the quaternion $\boldsymbol{q}_{\mathrm{HE}}$ from norm one and thus implements the norm one constraint. The function $Q(\cdot)$ maps a 3-D vector to a purely imaginary quaternion: $Q(\boldsymbol{x}) = 0 + x_1\mathrm{i} + x_2\mathrm{j} + x_3\mathrm{k}$, where $\boldsymbol{x} = \begin{pmatrix}x_1 & x_2 & x_3\end{pmatrix}^{\mathrm{T}}$. The single terms of (7) can be derived directly from the hand-eye equations (4) and (5): The first summand equals (4) in quaternion notation. The second one is derived from (5) by reformulating the multiplication of the rotation matrix $\boldsymbol{R}_{\mathrm{HE}}$ and the translation vector of the relative movement of the left camera using quaternions.

4.2 Extension of the Dual Quaternion Algorithm

This section shows how the estimation of rotation, translation, and scale can be formulated using dual quaternions. As quaternions are a representation for 3-D rotations, dual quaternions treat rotations *and* translations in a unified way.

Dual numbers were proposed by Clifford in the 19$^{\mathrm{th}}$ century [4]. They are defined by $\widetilde{z} = a + \varepsilon b$, where $\varepsilon^2 = 0$. When using vectors for a and b instead of real numbers, the result is a dual vector.

A dual quaternion $\widetilde{\boldsymbol{q}}$ is defined as a quaternion, where the real and imaginary parts are dual numbers instead of real ones, or equivalently as a dual vector where the dual

and the non-dual part are quaternions: $\widetilde{\boldsymbol{q}} = \boldsymbol{q}_{\mathrm{nd}} + \varepsilon \boldsymbol{q}_{\mathrm{d}}$. Just as unit quaternions represent rotations, unit dual quaternions contain rotation and translation [5]. In the dual quaternion representation of $\boldsymbol{R}$ and $\boldsymbol{t}$, the non-dual part $\boldsymbol{q}_{\mathrm{nd}}$ is the well-known quaternion representation of $\boldsymbol{R}$, and the dual part is given by

$$\boldsymbol{q}_{\mathrm{d}} = \frac{1}{2}\boldsymbol{t}_{\mathrm{q}}\boldsymbol{q}_{\mathrm{nd}}, \quad \boldsymbol{t}_{\mathrm{q}} = (0, \boldsymbol{t}) \quad , \tag{8}$$

where $\boldsymbol{t}_{\mathrm{q}}$ is a purely imaginary quaternion defined by the translation vector $\boldsymbol{t}$. A dual quaternion formulation of hand-eye calibration was introduced by [5]. We will now show how to integrate scale into the dual quaternion formulation by using non-unit dual quaternions, which results in a unified representation of similarity transformations.

For this purpose a dual quaternion $\widetilde{\boldsymbol{q}}_{s\mathrm{HE}}$ containing all these parameters is introduced, which is defined by:

$$\widetilde{\boldsymbol{q}}_{s\mathrm{HE}} = \boldsymbol{q}_{s\mathrm{HEnd}} + \varepsilon \boldsymbol{q}_{s\mathrm{HEd}} = s_{\mathrm{HE}}\boldsymbol{q}_{\mathrm{HE}} + \varepsilon\frac{1}{2}\boldsymbol{t}_{\mathrm{HEq}}\boldsymbol{q}_{\mathrm{HE}} \quad . \tag{9}$$

A dual quaternion has eight elements, but for rotation, translation, and scale only seven degrees of freedom are necessary. The norm of a dual quaternion is in general a *dual* number with non-negative real part given by:

$$|\widetilde{\boldsymbol{q}}|^2 = \widetilde{\boldsymbol{q}}\widetilde{\boldsymbol{q}}^* = \boldsymbol{q}_{\mathrm{nd}}\boldsymbol{q}_{\mathrm{nd}}^* + \varepsilon(\boldsymbol{q}_{\mathrm{nd}}\boldsymbol{q}_{\mathrm{d}}^* + \boldsymbol{q}_{\mathrm{d}}\boldsymbol{q}_{\mathrm{nd}}^*) \quad . \tag{10}$$

When the dual quaternion as defined in (9) is used, the scale is actually modeled as the norm of $\widetilde{\boldsymbol{q}}_{s\mathrm{HE}}$:

$$|\widetilde{\boldsymbol{q}}_{s\mathrm{HE}}|^2 = {s_{\mathrm{HE}}}^2 + \varepsilon 0 \quad \Leftrightarrow \quad |\widetilde{\boldsymbol{q}}_{s\mathrm{HE}}| = s_{\mathrm{HE}} \quad . \tag{11}$$

Since the scale factor will always be a positive real number, the dual part of the norm has to be zero. Therefore, one degree of freedom is lost, and we get an additional constraint that is given by:

$$\boldsymbol{q}_{s\mathrm{HEnd}}\boldsymbol{q}_{s\mathrm{HEd}}^* + \boldsymbol{q}_{s\mathrm{HEd}}\boldsymbol{q}_{s\mathrm{HEnd}}^* = 0 \quad . \tag{12}$$

Using (9), the extended hand-eye calibration problem (cf. [5] for the standard formulation) solving for scale, rotation, and translation can be formulated as:

$$\boldsymbol{q}_{\mathrm{End}}\boldsymbol{q}_{s\mathrm{HEnd}} = \boldsymbol{q}_{s\mathrm{HEnd}}\boldsymbol{q}_{\mathrm{Hnd}} \quad , \tag{13}$$

$$\boldsymbol{q}_{\mathrm{End}}\boldsymbol{q}_{s\mathrm{HEd}} + \frac{1}{s_{\mathrm{HE}}}\boldsymbol{q}_{\mathrm{Ed}}\boldsymbol{q}_{s\mathrm{HEnd}} = \boldsymbol{q}_{s\mathrm{HEnd}}\boldsymbol{q}_{\mathrm{Hd}} + \boldsymbol{q}_{s\mathrm{HEd}}\boldsymbol{q}_{\mathrm{Hnd}} \quad . \tag{14}$$

The indices ij that indicate a relative movement from frame i to frame j have been omitted for reasons of simplicity. Note that s_{HE} is not an additional independent parameter as in the previous section, but the norm of the dual quaternion $\widetilde{\boldsymbol{q}}_{s\mathrm{HE}}$ (cf. (11)).

It can be observed that (14) is a non-linear equation; an objective function $f'(\cdot)$ for non-linear optimization will look as follows:

$$\begin{aligned} f'(\boldsymbol{q}_{s\mathrm{HEnd}}, \boldsymbol{q}_{s\mathrm{HEd}}) = &\sum_{i=1}^{N_{\mathrm{rel}}} \left\| \boldsymbol{q}_{\mathrm{End}\,i}\boldsymbol{q}_{s\mathrm{HEnd}} - \boldsymbol{q}_{s\mathrm{HEnd}}\boldsymbol{q}_{\mathrm{Hnd}\,i} \right\|^2 + \\ &\sum_{i=1}^{N_{\mathrm{rel}}} \left\| \boldsymbol{q}_{\mathrm{End}\,i}\boldsymbol{q}_{s\mathrm{HEd}} + \frac{1}{\sqrt{\boldsymbol{q}_{s\mathrm{HEnd}}\boldsymbol{q}_{s\mathrm{HEnd}}^*}}\boldsymbol{q}_{\mathrm{Ed}\,i}\boldsymbol{q}_{s\mathrm{HEnd}} - \boldsymbol{q}_{s\mathrm{HEnd}}\boldsymbol{q}_{\mathrm{Hd}\,i} + \boldsymbol{q}_{s\mathrm{HEd}}\boldsymbol{q}_{\mathrm{Hnd}\,i} \right\|^2 + \\ &\lambda \left(\boldsymbol{q}_{s\mathrm{HEnd}}\boldsymbol{q}_{s\mathrm{HEd}}^* + \boldsymbol{q}_{s\mathrm{HEd}}\boldsymbol{q}_{s\mathrm{HEnd}}^*\right)^2 \quad . \end{aligned} \tag{15}$$

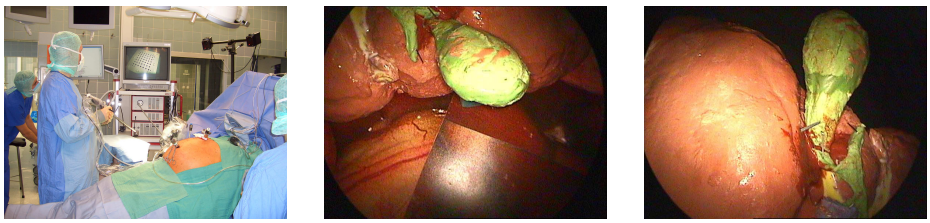

Fig. 1. Optical tracking system (left) and one image of the sequences *ART1* (middle) and *ART2* (right) showing a silicon liver/gall-bladder model that were used for structure-from-motion based 3-D reconstruction.

This objective function estimates rotation, translation, and scale, which are all encoded in the dual and non-dual parts of $\widetilde{\boldsymbol{q}}_{s\mathrm{HE}}$. The first summand is derived from (13), the second one from (14). As before, a regularization term enforces the constraint (12) on the norm of the dual quaternion $\widetilde{\boldsymbol{q}}_{s\mathrm{HE}}$, which has to be a real number. In this case, eight parameters with only seven degrees of freedom are optimized.

5 Experiments

In the following we present an experimental evaluation of the extended hand-eye calibration methods. The data were acquired using an endoscope with a camera mounted on it (the *eye*), which was moved by hand. An optical tracking system (cf. Fig. 1, left) provides pose data of a so-called target (the *hand*) that is fixed to the endoscope. The infrared optical tracking system smARTtrack1 by Advanced Realtime GmbH is employed. It is a typical optical tracking system consisting of two (or more) cameras and a target that is tracked. The target is built from markers that can easily be identified in the images captured by the cameras. In our case spheres with a retro-reflective surface are used. Infrared light simplifies marker identification. The 3-D position of each visible marker is calculated by the tracking system. The knowledge of the geometry of the target then allows to calculate its pose.

Instead of using consecutive movements for calibration, we applied the vector quantization based data selection method proposed in [8], which leads to more accurate results.

Since no ground truth is available when calibrating real data, we cannot give errors between the real hand-eye transformation and the computed one. It is desirable, however, that an error measure is available which rates the quality of the resulting transformation. Therefore, the following error measure is used: After applying the computed hand-eye transformation on the *hand* data, we get an estimate of the *eye* movements $\mathcal{E}'$. This estimated movement can now be compared to the original *eye* movement $\mathcal{E}$, which has been obtained by structure-from-motion: If the hand-eye transformation is correct, the relative movements between single camera positions are equal in $\mathcal{E}$ and $\mathcal{E}'$. The errors are computed by averaging over a set of randomly selected relative movements.

Table 1 shows residual errors in translation and rotation as well as the computation times for hand-eye calibration on a Linux PC (Athlon XP2600+) including data selection, but not feature tracking and 3-D reconstruction. The latter steps are the same for all methods, and take approximately 90 sec for tracking and 200 sec for 3-D reconstruction. The values shown are relative and absolute residuals for rotation, and relative errors for the norm of the translation vector of relative movements. Absolute errors for translation

Table 1. Mean errors in rotation and translation of relative *eye* movements computed with different hand-eye calibration methods using structure-from-motion as a basis.

Data Set	Method	Translation	Rotation		Time
	DQ, scale sep.	22.3%	0.191°	3.75%	310 msec
	Hor., scale sep.	13.0%	0.179°	3.72%	2090 msec
ART1	non-lin., eq. (7)	17.8%	0.191°	3.75%	2350 msec
	non-lin., eq. (15)	44.6%	0.191°	3.75%	756 msec
	Andreff	13.2%	0.172°	3.60%	309 msec
	DQ, scale sep.	20.7%	0.290°	7.25%	433 msec
	Hor., scale sep.	18.7%	0.266°	7.00%	1590 msec
ART2	non-lin., eq. (7)	18.4%	0.290°	7.25%	1770 msec
	non-lin., eq. (15)	20.7%	0.290°	7.25%	482 msec
	Andreff	19.7%	0.272°	7.10%	434 msec

are not given, as these are highly dependent on the estimated scale factor and therefore cannot be compared directly, whereas absolute rotational residuals are independent of scale. We show the results for two data sets, namely *ART1* (190 images) and *ART2* (200 images). After feature tracking and 3-D reconstruction, different hand-eye calibration methods have been evaluated; in all cases the reconstructed camera movement has been used as eye-data. The results shown in Table 1 were computed as follows:

DQ, scale sep.: Here, the scale factor was estimated first by solving (4) and (5). After scaling the eye-reconstruction appropriately, rotation and translation were re-estimated using a standard hand-eye calibration method, namely the linear dual quaternion algorithm of [5].

Hor., scale sep.: The same as *DQ, scale sep.*, i. e., scale and rotation/translation were computed separately. Instead of dual quaternions the non-linear method proposed by [7] was used for hand-eye calibration.

non-lin., eq. (7)**/**(15)**:** Here, the non-linear objective functions (7), (15) were used, which were initialized with the result of *DQ, scale sep*. After non-linear optimization of rotation, translation, and scale, rotation and translation were re-estimated using the linear dual quaternion method, which results in a more accurate hand-eye transformation compared to non-linear optimization alone.

Andreff: This is the result of the hand-eye calibration method proposed by [1].

The relative residual errors obtained using standard hand-eye calibration and a calibration pattern are 4.20% (transl.) and 0.725% (rot.) for the configuration similar to *ART2* and 5.39% (transl.) and 1.09% (rot.) for the configuration similar to *ART1*. The deviations to the hand-eye transformation computed this way compared to the extended approach using structure-from-motion (depending on the method) are 15% to 16% in rotation and about 35% for translation (*ART2*), and 9% to 11% in rotation and 26% to 32% in translation (*ART1*).

6 Conclusion

We presented methods for an extended hand-eye calibration, which allow to compute the hand-eye transformation without the necessity for using a calibration pattern in order to obtain the camera (eye) poses. Instead, these are computed using feature tracking

and a structure-from-motion approach, which makes the extension of standard hand-eye calibration necessary since in addition to rotation and translation a scale factor has to be estimated. Different ways of extending these equations have been presented and compared. The main result is that the estimation of the hand-eye transformation is feasible without a calibration pattern. Of course, one could not expect to obtain results as accurate as with standard calibration; depending on the application, however, the advantages of the extended method may outweigh this drawback. This is especially true for the clinical setup that we have in mind, as hand-eye calibration has to be performed before each operation. The usage of an unsterile calibration pattern in combination with a sterile endoscope and a surgeon working under sterile conditions is difficult in practice, and can be completely circumvented when using the methods proposed here.

References

1. N. Andreff, R. Horaud, and B. Espiau. Robot Hand-Eye Calibration Using Structure from Motion. *Int. Journal of Robotics Research*, 20:228–248, 2001.
2. H. Chen. A Screw Motion Approach to Uniqueness Analysis of Head–Eye Geometry. In *Proc. of CVPR*, pages 145–151, Maui, Hawaii, 1991.
3. J. C. K. Chou and M. Kamel. Finding the Position and Orientation of a Sensor on a Robot Manipulator Using Quaternions. *Int. Journal of Robotics Research*, 10(3):240–254, 1991.
4. W. Clifford. Preliminary Sketch of Bi-quaternions. *Proc. of the London Mathematical Society*, 4:381–395, 1873.
5. K. Daniilidis. Hand-Eye Calibration Using Dual Quaternions. *Int. Journal of Robotics Research*, 18:286–298, 1999.
6. B. Heigl. *Plenoptic Scene Modeling from Uncalibrated Image Sequences*. ibidem-Verlag Stuttgart, 2004.
7. R. Horaud and F. Dornaika. Hand-Eye Calibration. *Int. Journal of Robotics Research*, 14(3):195–210, 1995.
8. J. Schmidt, F. Vogt, and H. Niemann. Vector Quantization Based Data Selection for Hand-Eye Calibration. In B. Girod, M. Magnor, and H.-P. Seidel, editors, *Vision, Modeling, and Visualization 2004*, pages 21–28, Stanford, USA, 2004. Aka / IOS Press, Berlin, Amsterdam.
9. Y. Shiu and S. Ahmad. Calibration of Wrist Mounted Robotic Sensors by Solving Homogeneous Transform Equations of the Form $AX = XB$. *IEEE Trans. on Robotics and Automation*, 5(1):16–29, 1989.
10. R. Y. Tsai and R. K. Lenz. A New Technique for Fully Autonomous and Efficient 3D Robotics Hand/Eye Calibration. *IEEE Trans. on Robotics and Automation*, 5(3):345–358, 1989.
11. F. Vogt, S. Krüger, J. Schmidt, D. Paulus, H. Niemann, W. Hohenberger, and C. H. Schick. Light Fields for Minimal Invasive Surgery Using an Endoscope Positioning Robot. *Methods of Information in Medicine*, 43(4):403–408, 2004.
12. Z. Zhang. A Flexible New Technique for Camera Calibration. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 22(11), 2000.
13. T. Zinßer, Ch. Gräßl, and H. Niemann. Efficient Feature Tracking for Long Video Sequences. In C. E. Rasmussen, H. H. Bülthoff, M. A. Giese, and B. Schölkopf, editors, *Pattern Recognition, 26th DAGM Symposium*, volume 3175 of *Lecture Notes in Computer Science*, pages 326–333. Springer-Verlag, Berlin, Heidelberg, New York, 2004.