

Comparing calibration approaches for 3D ultrasound probes

C. Bergmeir · M. Seitel · C. Frank · R. De Simone ·
H.-P. Meinzer · I. Wolf

Received: 10 January 2008 / Accepted: 14 September 2008 / Published online: 4 November 2008
© CARS 2008

Abstract

Objective By adding a tracking sensor to a 3D ultrasound (US) probe and thus locating the probe in space, new applications within the fields of image guided surgery and radiation therapy are possible. To locate the US volume in space, a calibration is necessary to determine the mathematical transformation for mapping points from the tracking coordinate system to the US image coordinate system. We present a comprehensive comparison of two different approaches to perform this calibration for 3D US.

Methods For both approaches a phantom is scanned and located in the images by means of segmentation and registration techniques. Calibration is then performed by either relating the tracked phantom's (TP) spatial location to the calibration scans, or by solely correlating scans taken from multiple perspectives when using hand-eye calibration methods (HE). Depending on which approach is utilized, a minimum of one or three images, respectively, need to be acquired for the calibration process.

Results We evaluated both approaches for calibration and reconstruction precision. Regarding the latter, the performed tests led to mean target localization errors of 3.5 mm (HE) and 3.3 mm (TP) for real data, and of 1.4 mm (HE) and 0.9 mm (TP) for simulated data.

Conclusion Our results indicate that taking additional scans leads to a significant improvement in the calibration. Furthermore, the obtained calibration and reconstruction precisions suggest the use of a TP.

Introduction

Ultrasound (US) is a well-established imaging modality with many practical applications in medical diagnosis and therapy. As a low-cost, radiationless, and real-time technology applicable both at the patient's bed-side and in the operation room, US is the modality of choice for many situations. The recent advent of 3D US probes brought forth a whole field of new and promising applications.

Employing real-time US images to facilitate monitoring and navigation, e.g. during surgery or radiation therapy [5, 10], requires their spatial location to be known within a frame of reference. The ultrasound probe's location can be determined by providing its body with tracking sensors. To transform tracking coordinates into image coordinates, a calibration needs to be performed.

Spatial US calibration generally involves a phantom, an object with known geometrical features which is clearly visible in US scans. The calibration is done by scanning the phantom with the US probe and then determining its location within the scanned images by means of segmentation and registration methods. From these data, the calibration transform can be calculated subsequently using two different approaches, which we refer to as the hand-eye (HE) and the tracked phantom (TP) approach.

With the HE approach, at least three acquisitions taken from distinct perspectives are necessary [7] to calculate the calibration transform with methods adopted from HE calibration techniques commonly used in robotics. The HE

C. Bergmeir (✉) · M. Seitel · H.-P. Meinzer · I. Wolf (✉)
Division of Medical and Biological Informatics,
German Cancer Research Center, INF 280,
69120 Heidelberg, Germany
e-mail: c.bergmeir@dkfz.de; christoph.bergmeir@gmail.com

I. Wolf
e-mail: i.wolf@dkfz.de

C. Frank · R. De Simone
Department of Cardiac Surgery, University Hospital,
Heidelberg, Germany

calibration problem can be solved for rotation and translation considered either separately or simultaneously, and both linear as well as non-linear algorithms exist. Solving for rotation and translation separately leads to the disadvantage that errors in rotation computation propagate to translation [23]. The first publications addressing HE calibration solve for rotation and translation separately. Tsai and Lenz [25] use axis and angle representation of the rotation, Chou and Kamel [8] use quaternions. Chen [7] is the first to solve for translation and rotation simultaneously, using screw theory. Horaud and Dornaika [12] use quaternions and solve the equation system simultaneously using non-linear optimization. Daniilidis [9] uses dual quaternions, an algebraic representation of the screw theory. Andreff et al. [2] are able to solve the problem linearly and simultaneously by increasing the dimension of the rotational aspect.

If in addition to the probe the phantom is tracked as well, a single acquisition is sufficient for calibration with the TP approach. This method requires tracking sensors to be mounted with high precision onto the phantom, for which purpose it usually gets measured by a precalibrated tracking pointer. If the pointer directly serves as the phantom, the problem of accurately defining the phantom geometry is avoided, as this already has to be done by the manufacturer.

While the calibration of 2D and of 3D US probes share a common purpose, the methods employed for both cases have rather distinct characteristics. On the one hand, 3D probes yield better redundancy by providing a larger amount of data, and the detection of the phantom and determination of its original geometry are simplified in comparison to 2D probes. No compounding is necessary to depict the entire phantom, and thus only few scans need to be taken. On the other hand, 3D probes have a lower spatial resolution. When applying 2D calibration procedures, it is not trivial to integrate the additional information provided by 3D probes. At worst, it remains unused, leading to an unnecessarily laborious procedure. Furthermore, modern 3D ultrasound systems usually allow for digital data transfer, which implicitly defines several characteristics of the images, such as image spacing.

Many calibration methods for 2D US probes can be found in the literature. A review is given by Mercier et al. [18]. Common methods are single-point calibration methods (as developed by, e.g. Treece et al. [24]), and methods which use Z shaped wire phantoms (see, e.g. Lindseth et al. [16]). Based on these, cross-wire calibration methods were developed, where the probe is aligned to a pre-defined point visible in the image. Carr [6] and Prager et al. [21] use a phantom consisting of three orthogonal nylon wires, which gives a straightforward definition of the phantom coordinate system. Furthermore, Prager et al. compare this type of phantom to a cross-wire phantom and a single-wall phantom. The latter consists of a thin membrane or the bottom of a water tank, which leads to lines in the 2D images. In this

way, more redundancy than within single-point calibration is achieved, and a more reliable, potentially automatic feature localization is possible. A further enhancement to the single-wall phantom is the Cambridge phantom [21], which avoids the problem of lines getting blurred at oblique angles due to the finite width of the US beam.

For 3D probe calibration, however, only few publications exist. Poon and Rohling [20] compare three different phantoms, namely an IXI-shaped wire phantom, a cube phantom, and a stylus. All phantoms are tracked. In calibration reproducibility tests, they use different calibrations to transform the point located in the farthest corner of the US volume, and compute pairwise differences between the transformed points. For this differences, they compute mean values of 1.50, 1.16, and 5.13 mm from 45 trials each. In point accuracy tests, they acquire images of a tracking stylus and compute the distance between its position measured by the tracking system and by the calibrated probe. For this measure, their methods show root mean square errors of 2.15, 4.91, and 2.36 mm. They conclude that the nylon wire phantom performs best. Bouchet et al. [5] use a phantom consisting of 39 nylon wires which is molded in a tissue-equivalent anechoic medium, and track the phantom optically. During target localization experiments, they locate sphere fiducials using a CT scan and a tracked 3D probe. Within these experiments, they measure maximum errors of 0.4 mm in the elevational, 1.5 mm in the lateral, and 1.1 mm in the axial direction, leading to a total target localization error of 1.9 mm. Hastenteufel et al. [11] calibrate the probe using a metallic cross phantom. They only evaluate their registration procedure, i.e. the localization of their phantom in the images, but not the whole calibration procedure. Boctor et al. [4] present a calibration system which needs no phantom at all, and can be used both with 2D and 3D probes. They use Andreff's algorithm to do HE calibration without stating accuracy results regarding 3D probes. Lange and Eulenstein [15] perform calibration both with TP and HE calibration methods. A magnetic tracking system, a swept-volume probe and a commercial phantom are used. During calibration precision experiments, ten calibrations are used to map test points to a global coordinate system. Root mean square errors of 1.8 mm for the TP and 2.2 mm for the HE calibration algorithm are given. Also, reconstruction precision is computed by locating a point feature in images acquired from five different directions and mapping these locations to a global coordinate system. Root mean square errors of 3.4 and 3.3 mm are reported. They conclude that optical tracking (datasets are merely simulated for this purpose) and a nylon wire phantom should be evaluated as well.

To the best of our knowledge, we are the first to give a detailed comparison of calibration using a TP and calibration with HE methods for 3D probes. Making use of Daniilidis' algorithm for HE calibration, which shows several advan-

tages in robustness and precision compared to other work [9], we use a nylon wire phantom and optical tracking.

Materials and methods

Experimental setup

To perform calibration, datasets consisting of US volumes showing the phantom and related tracking data need to be acquired.

Combination of tracking and ultrasound system

We use a SONOS 7500 ultrasound system with the ultrasound probe X4 (both Philips Medical Systems). The probe is tracked by an optical system (Polaris, Northern Digital Inc.). For this purpose, four passive markers are mounted on a rigid plate which is attached to the probe using adhesive tape (see Fig. 1). The X4 probe is a 2D array probe with a frequency of 2–4 MHz.

Phantom

An orthogonal cross phantom has the advantage of intuitively defining a phantom coordinate system. The shortcoming of ambiguity with respect to registration of the cross in the images is eliminated by our phantom through additional wires.

Hence our phantom consists of five nylon wires mounted to a rack manufactured of acrylic glass. The relevant cubic volume of the phantom has an edge length of 260 mm. Three of the wires form an orthogonal cross, the other two are placed such that the cross becomes non-ambiguous with respect to rotation transforms, i.e. one parallel to the x -axis with a distance of 22 mm, the other subtending the x - and the y -axis by an angle of 45° with a distance of 16 mm from the origin. The nylon wires consist of drilled fibres (to enhance visibility in the US images by the more inhomogeneous material) and have a diameter of 0.9 mm.

To track the phantom, four markers of the optical tracking system are attached to its surface. The phantom is assembled



Fig. 1 Passive markers of the tracking system are attached to the ultrasound probe

by our on-site workshop with a drilling accuracy of 0.1 mm. Furthermore, the location of the wires is controlled by means of a CT scan, so that the phantom coordinate system can be defined accurately (see Fig. 2).

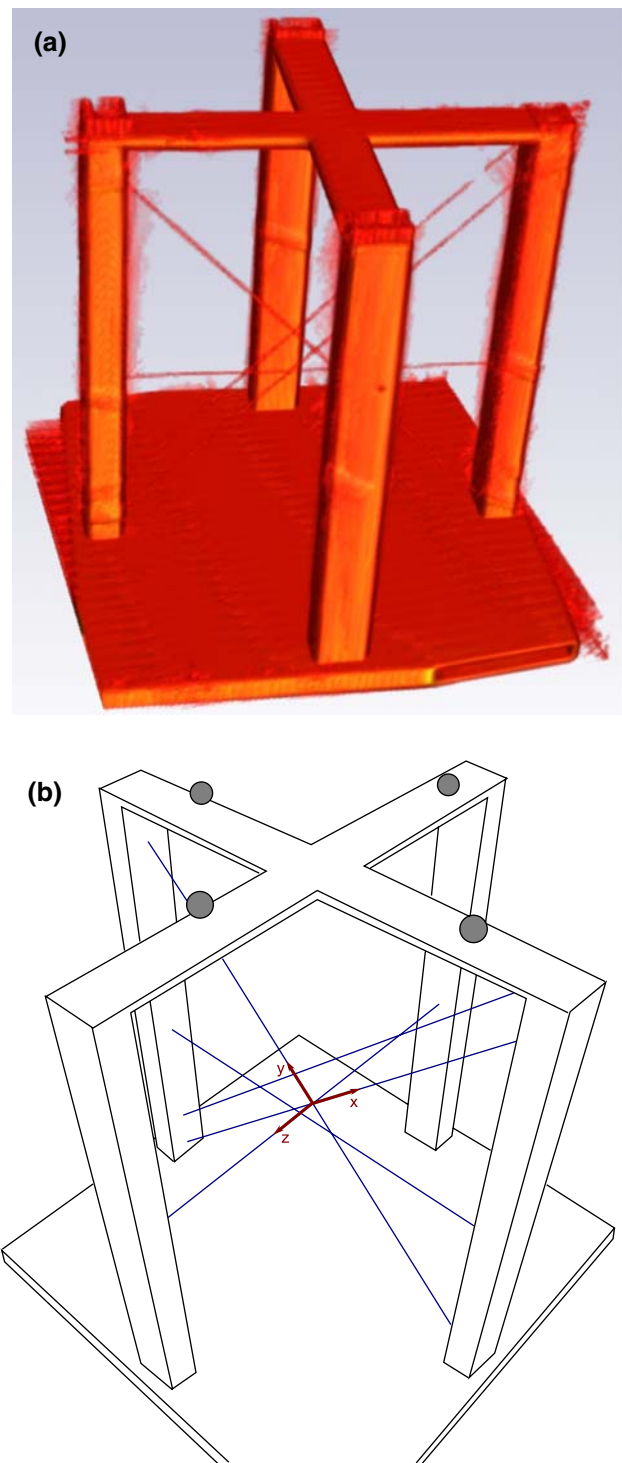


Fig. 2 To verify the geometrical parameters of the phantom, a CT scan was performed, and a three dimensional visualization was generated (a). With these data, the phantom coordinate system could be defined accurately in relation to the passive tracking markers, illustrated by (b)

Image acquisition

The phantom is immersed in a water bath inside a rigid plastic box. The probe is mounted to a flexible arm, which both makes an easy positioning of the probe possible, and ensures that the probe remains fixed during image acquisition. With this setup, datasets consisting of ultrasound images and corresponding tracking data can be acquired from different directions. All acquired volumes have a size of $160 \times 144 \times 208$ voxels with scaling factors of $s_x = 1.24$, $s_y = 1.26$ and $s_z = 0.80$ (scaling factors in mm/voxel). During image generation the position is continuously recorded with a frequency of 20 Hz. We define the imaging duration as two seconds (which ensures that the ultrasound system has completed the entire acquisition process), within which the probe is tracked and must not be moved. Hence, we obtain approximately 40 tracking measurements per acquisition, which are used to compute an average position and to detect irregularities during image acquisition. The probe always has a distance of approximately 6 cm from the center of the cross. The rack allows for taking scans from four distinct perspectives (see Fig. 2). For every calibration, three scans are taken from each perspective, holding the probe in different orientations. As the stereo cameras of the tracking system are placed directly above the experimental setup, the probe cannot be rotated completely freely, as the sensors need a direct line of sight to the cameras. Thus, we define the first orientation to be horizontal, with the tracking sensors pointing directly to the tracking system. The other two scans are acquired by rotating the probe by $+45^\circ$ and -45° about its principal axis, respectively (see Fig. 3). After the acquisition of each set of 12 images, both the phantom and the stereocamera system are repositioned to assure independence of the trials.

Image processing

To locate the phantom in the images, after some preprocessing, the phantom is segmented and afterwards registered with a computer generated model.

Preprocessing

During the image acquisition process, the US machine automatically performs envelope detection and logarithmic compression of the signal. Time gain compensation is controlled by slide potentiometers. In our experiments, we adjusted the gain compensation slides such that the resulting image had a reasonably homogenous distribution of gray level over depth. Making use of the built-in digital interface, the resulting grayscale intensity images are transferred via compact discs to an external computer running the calibration software.

Since the speed of sound in water at room temperature differs from the average speed of sound in human tissue, the

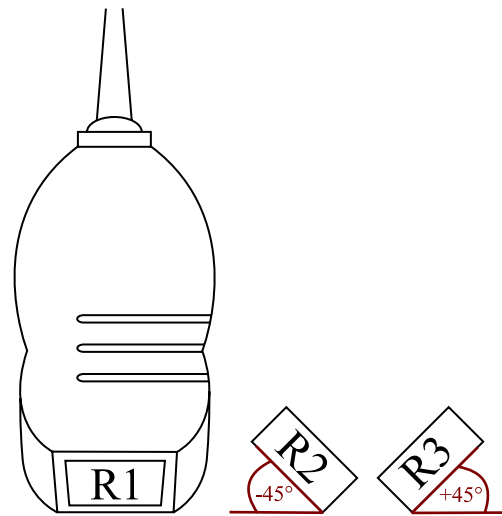


Fig. 3 Orientations (R1–R3) of the probe for acquiring scans from different positions

ultrasound volume datasets are scaled with a correction factor of 1.04 (compare [18, p. 458]) in a first processing step. The center for the scaling is determined manually from a sample dataset.

Segmentation

Afterwards, we segment the phantom in the image using a region growing algorithm (which is a common method in image segmentation, for a detailed description see, e.g. Adams and Bischof [1]). Besides the 3D US image, the algorithm we use has two inputs, the marker mr and the mask mk , which are real values in the interval $[0, 1]$. We denote by \max the maximum intensity value occurring in the image. Initially, all voxels v whose intensity $I(v)$ lies in the interval $[mr * \max, \max]$ are added to the segmentation (they serve as seed points). Then, iteratively voxels w adjacent to the segmentation are added if their intensity $I(w)$ lies in $[mk * \max, \max]$. The algorithm terminates if no more voxels are added to the segmentation during an iteration.

Selecting suitable values for the input parameters depends highly on the settings of the US machine (in particular time gain compensation). In our experiments mr was defined as 0.95, and mk as 0.8.

Registration

After segmentation, we register a computer generated model of the cross to the image in two steps.

Registration by distance minimization At first, we generate two point sets (each containing 400 points) by random sampling of the segmentation and the model. In the following we

call the point set generated from the segmentation the fixed point set, the one generated from the model the moving point set. Then, the iterative closest point algorithm (ICP), proposed by Besl and McKay [3], is applied for registering the two point sets. They define the Euclidean distance d between two points $p_1 = (x_1, y_1, z_1)$ and $p_2 = (x_2, y_2, z_2)$ as

$$d(p_1, p_2) = \|p_1 - p_2\| = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}.$$

Furthermore, the distance D of a point p to a point set $P = \{p_1, \dots, p_n\}$ is defined as

$$D(p, P) = \min_{i \in \{1, \dots, n\}} d(p, p_i).$$

Then, the algorithm minimizes the squares of the distances between the two point sets. For this purpose, a vector is computed that contains for every point in the moving point set the distance D to the fixed point set. The Levenberg–Marquardt algorithm [19] is then used to find a rigid transform that maps the moving point set in a way that this vector becomes minimal. Finally, we position the model using this transform.

Registration by grayscale value maximization To perform the second registration step, a point set $P = \{p_1, \dots, p_n\}$ (in our experiments $n = 500$) is generated by random sampling of the model, which is already initially positioned by the first registration step. Every point in P gets assigned a grayscale value g from the corresponding voxel in the image. Then, a vector of the differences of the grayscale value maximum \max (in our case $\max = 255$) and the grayscale values g_i is computed. The components of the vector are then squared and added up, to define a value s :

$$s = \sum_{i=1, \dots, n} (\max - g_i)^2.$$

Then a gradient descent optimization is used to iteratively compute s and find a transform that maps the point set to a location such that s becomes minimal (see Fig. 4). This processing seems feasible since the ultrasound images depict the wires in a blurry way. With the additional grayscale value information, which was not taken into account in the first registration step, the model is repositioned to the brightest areas of the images, which typically appear in the middle of the wires.

Calibration

As illustrated by Fig. 6, the transforms S , P , R , and X are relevant during calibration, where S transforms from tracking coordinates to probe sensor coordinates, P transforms from tracking coordinates to phantom coordinates, R from image coordinates to phantom coordinates, and X from image coordinates to probe sensor coordinates. By the tracking system

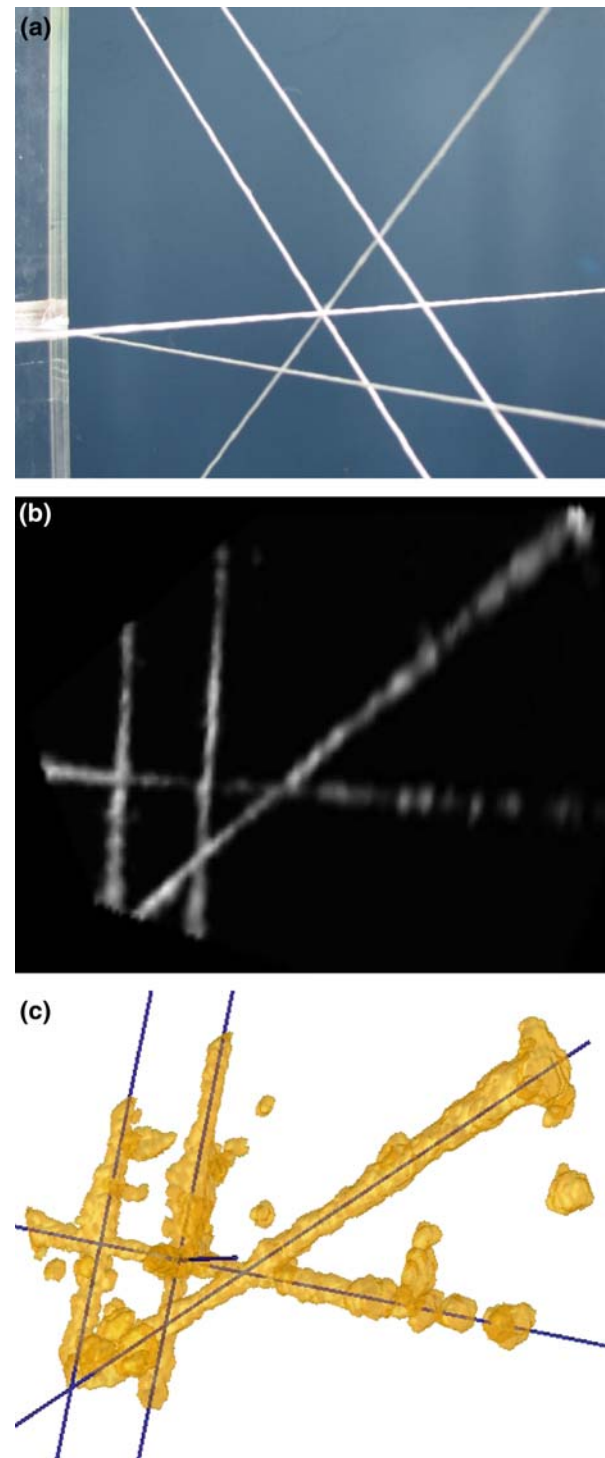


Fig. 4 Processing steps of phantom detection. **a** Nylon wire phantom mounted in a rack (left in the image). **b** Selected slice of the ultrasound volume, showing the wires. **c** Registration result of the computer model and the ultrasound image (visualized by the segmentation which is used during the first registration step)

and the known geometry of the marker plate, the transforms S and P are known. The registration computes R , and X is the unknown calibration matrix.

Two different calibration approaches are used, the HE and the TP approach. For the HE approach, two different methods are evaluated. The 12 images directly serve as input for the corresponding methods to obtain a single calibration. For the TP approach, a calibration transform is computed for every single image. The resulting 12 calibrations are used for averaging.

Notation

Matrices are represented by uppercase italic boldface letters (e.g. \mathbf{M}), vectors by lowercase boldface letters (e.g. \mathbf{v}), and quaternions in a sans serif style (e.g. \mathbf{q}). A rigid transform (e.g. \mathbf{X}) is given by homogeneous coordinates:

$$\mathbf{X} = \begin{pmatrix} \mathbf{R}_x & \mathbf{t}_x \\ 0 & 0 & 0 & 1 \end{pmatrix},$$

where \mathbf{R}_x is a 3×3 rotation matrix and \mathbf{t}_x a translation vector. As a rotation can be represented both by rotation matrices and quaternions, we denote transforms by the combination of a translation vector \mathbf{t} and a rotation either represented by \mathbf{R} or by \mathbf{q} .

Hand–eye calibration (HE)

Without the phantom being tracked, at least three images acquired with the tracked probe are required. With the transforms $\mathbf{A} = \mathbf{R}_1\mathbf{R}_2^{-1}$ and $\mathbf{B} = \mathbf{S}_1^{-1}\mathbf{S}_2$ (see Fig. 5b), which describe the movement (i.e. modification of position) of the image and the sensor, respectively, we can formulate the equation

$$\mathbf{A}\mathbf{X} = \mathbf{X}\mathbf{B} \quad (1)$$

for each movement taken into account. Having at least two such equations, we can compute \mathbf{X} . We use two different algorithms to solve for the HE calibration problem. The first, basic algorithm solves first for rotation and afterwards for translation. The second algorithm is the dual quaternion approach proposed by Daniilidis [9].

Separate computation of translation (TS) Assuming that the transforms are given in homogeneous coordinates, the separation of equation 1 into its translational and its rotational part is given by the following two equations:

$$\mathbf{R}_a\mathbf{R}_x = \mathbf{R}_x\mathbf{R}_b, \quad (2)$$

$$\mathbf{R}_a\mathbf{t}_x + \mathbf{U}_a\mathbf{s} = \mathbf{R}_x\mathbf{t}_b + \mathbf{t}_x, \quad (3)$$

where $\mathbf{U}_a\mathbf{s} = \mathbf{t}_a$. So \mathbf{t}_a is not computed directly, but is derived by multiplying a 3×3 diagonal matrix \mathbf{U}_a with the scalar factors $\mathbf{s} = (s_1, s_2, s_3)$. In our setup, the scalar factors are provided as spacing in the image data, and thus could be used

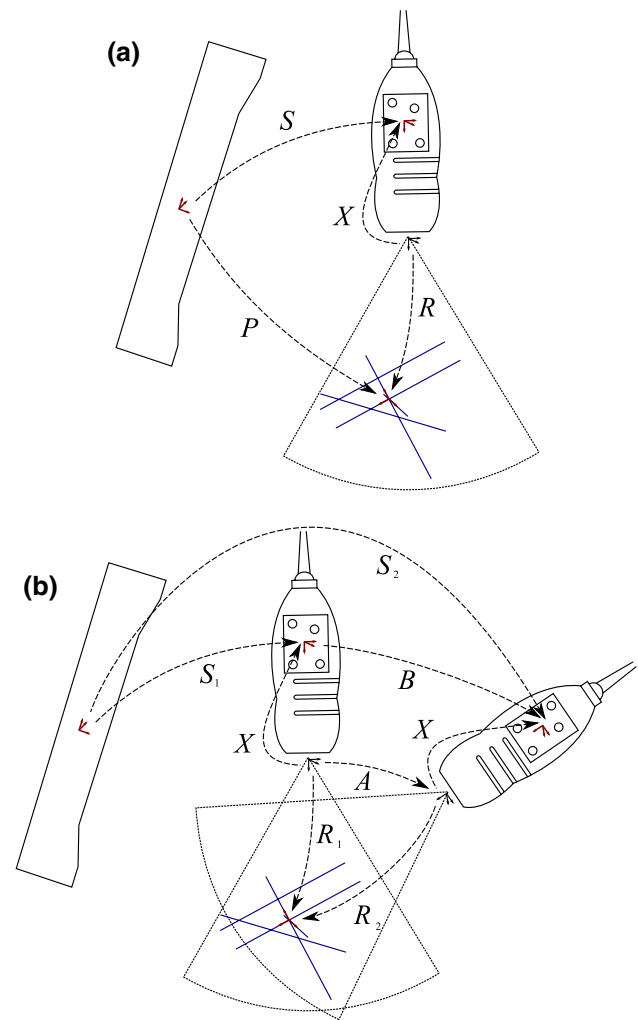


Fig. 5 Transforms involved in the calibration process with tracked phantom (a) and with hand–eye calibration (b)

for quality control of the system, e.g. by discarding solutions whose spacing lies outside a given tolerance.

Furthermore, we will use the linear operator vec and the Kronecker product \otimes in analogy to Andreff et al. [2]. Given an $m \times n$ matrix \mathbf{M} , vec reorders \mathbf{M} to an mn vector:

$$\text{vec}(\mathbf{M}) = (M_{11}, \dots, M_{1n}, M_{21}, \dots, M_{mn})^T.$$

With an additionally given $o \times p$ matrix \mathbf{N} , the Kronecker product is defined by

$$\mathbf{M} \otimes \mathbf{N} = \begin{pmatrix} M_{11}\mathbf{N} & \dots & M_{1n}\mathbf{N} \\ \vdots & \ddots & \vdots \\ M_{m1}\mathbf{N} & \dots & M_{mn}\mathbf{N} \end{pmatrix}.$$

Furthermore, we denote by \mathbf{I}_n an n -dimensional identity matrix. According to Andreff et al. [2], the rotational part \mathbf{R}_x of the calibration can be computed by solving

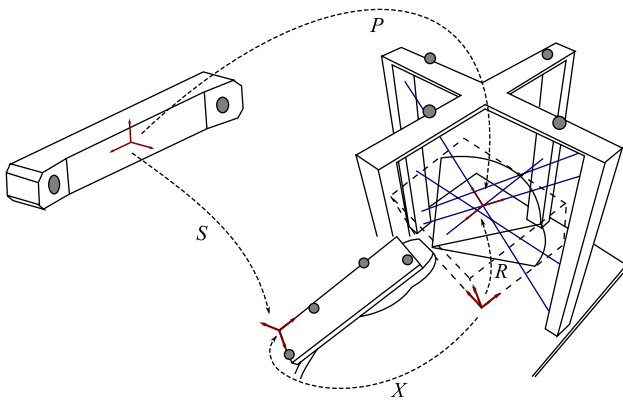


Fig. 6 The relevant transforms in our calibration setup. The transforms S and P are known from the tracking system and the known geometry of the marker plate, R is computed by the registration process. The unknown calibration matrix is X . With the help of a given X , S can be computed in dataset simulation

$$\begin{pmatrix} I_9 - R_{a1} \otimes R_{b1} \\ \vdots \\ I_9 - R_{an} \otimes R_{bn} \end{pmatrix} \text{vec}(R_x) = 0. \quad (4)$$

Afterwards, the translational part t_x can be computed by solving

$$\begin{pmatrix} I_3 - R_{a1} - U_{a1} \\ \vdots \\ I_3 - R_{an} - U_{an} \end{pmatrix} \begin{pmatrix} t_x \\ s \end{pmatrix} = \begin{pmatrix} -R_x t_{b1} \\ \vdots \\ -R_x t_{bn} \end{pmatrix}. \quad (5)$$

We solve both these systems of equations with a Singular Value Decomposition (SVD).

Dual quaternion approach (DQ) Quaternions are a common method for representing rotations in space, for their definition see, e.g. Chou and Kamel [8]. Dual numbers are also common in the field of HE calibration [9, 22]. A dual number is defined by

$$\tilde{z} = a + \varepsilon b \quad \text{with } \varepsilon^2 = 0, \quad (6)$$

where a is the non-dual part and b the dual part. With multiplication and addition the dual numbers form an Abelian ring.

A dual quaternion can be defined as a quaternion whose real- and imaginary parts consist of dual numbers, or equivalently as a dual number, whose dual and non-dual parts consist of quaternions:

$$\tilde{q} = q + \varepsilon q'. \quad (7)$$

A transform (q, t) can be converted to a dual quaternion by:

$$\tilde{q} = q + \varepsilon \frac{1}{2} tq, \quad (8)$$

where t is a quaternion with real part equal to zero and imaginary part equal to the translation vector t .

Using dual quaternions equation 1 can be written as:

$$\tilde{a} = \tilde{x} \tilde{b} \tilde{x}, \quad (9)$$

where \tilde{x} is the conjugate of \tilde{x} (because we use unit dual quaternions, the inverse of \tilde{x} can be obtained by computing its conjugate [9]). Again, there is one such equation for every movement. Daniilidis [9] uses this representation of the HE equation to formulate a linear system of equations which has to be solved non-trivially for x and x' :

$$T \begin{pmatrix} x \\ x' \end{pmatrix} = 0. \quad (10)$$

T is a $6n \times 8$ matrix (with n being the number of movements) whose components are obtained directly from the components of all involved dual quaternions \tilde{a} and \tilde{b} . Equation 10 can be solved using SVD. For the detailed algorithm, we refer to Daniilidis [9] and Schmidt et al. [22].

Tracked phantom approach (TP)

With the phantom being tracked, besides S and R , also P is known, so that X can be computed by

$$X = RP^{-1}S \quad (11)$$

in each dataset (see Fig. 5a). Given the 12 acquired datasets we compute an average transform using dual quaternion linear blending (DLB), as proposed by Kavan et al. [13]. Given the rigid calibration transforms as dual quaternions $\tilde{x}_1, \dots, \tilde{x}_n$, an average transform \tilde{x}_a can be defined as

$$\tilde{x}_a = DLB(\tilde{x}_1, \dots, \tilde{x}_n) = \frac{\tilde{x}_s}{\|\tilde{x}_s\|} = \frac{\tilde{x}_1 + \dots + \tilde{x}_n}{\|\tilde{x}_1 + \dots + \tilde{x}_n\|}.$$

Using the definition of the dual quaternion norm and dual number arithmetics (see [13]), \tilde{x}_a can be computed by

$$\tilde{x}_a = \frac{x_s}{\|x_s\|} + \varepsilon \left(x'_s - x_s \frac{\langle x_s, x'_s \rangle}{\|x_s\|^2} \right) \frac{1}{\|x_s\|},$$

where $\langle x_s, x'_s \rangle$ denotes to the scalar product of the non-dual and the dual part of \tilde{x}_s .

Evaluational methodology

To perform the evaluation, we simulate datasets and define different kinds of errors and evaluation measurements.

Dataset simulation

In order to test parts of the system, datasets are simulated. By defining a calibration transform X , a phantom transform

P and registration transforms R_1, \dots, R_n , we can compute probe positions S_1, \dots, S_n by

$$S_i = PR_i^{-1}X \quad \text{with } i = 1, \dots, n. \quad (12)$$

During our experiments, transforms obtained from a real calibration are used, so that the newly computed probe positions lead to datasets similar to the experimentally obtained ones, but without containing noise. Furthermore, the calibration transform is already known, and can be used as a gold standard. By artificially adding noise to these simulated datasets afterwards, the behavior of system parts or the whole system can be analyzed systematically. We use the anisotropic noise model simulating the Polaris system proposed by Khadem et al. [14] and Ma et al. [17]. It assumes that the accuracy of the tracking system is three times less in the direction of sight than in the other two directions. We add noise with a standard deviation of 0.1 mm in x and y direction, and 0.3 mm in z direction. The noise added to the registration data is isotropically distributed, with a standard deviation of 0.15 mm in the translation parameters and 0.35° in the axis angles (as empirically acquired through experiments, see “Results-Registration” section).

Calibration evaluation

Within the experiments, calibration is performed several times. Calibration precision indicates the relative error within these calibrations, i.e. the variation within the results. It can be used to get information about the reliability of the method. When using simulated datasets, the correct calibration is known, so that the error between the computed calibrations and the known calibration can be determined, which leads to a calibration accuracy measurement.

Reconstruction precision tests

We define the transforms T_i , which map points from the phantom coordinate system to the world coordinate system by

$$T_i = R_i^{-1}XS_i^{-1} \quad \text{with } i = 1, \dots, n. \quad (13)$$

Here, X is the calibration and $i = 1, \dots, n$ refers to acquisitions not used for the calibration that yielded X . Within these acquisitions, the tracking system and the phantom are not moved, so that for all n datasets the transforms T_1, \dots, T_n in theory are equal.

We define a point set P , which contains all points whose x , y , and z coordinates lie in the set $\{-25, -20, \dots, 20, 25\}$ mm. Thus, P contains 1,331 points in total.

We map the points in P with the transforms T_i . Then we compute pairwise (i.e. the points transformed by T_i paired with the points transformed by another transform T_j) the distances between corresponding points. T_i and T_j are chosen

in a way that they represent acquisitions from different positions and with different orientations. The average D_a over these distances is computed.

To avoid using the same probe position transforms and registration transforms for calibration and for reconstruction precision evaluation, we apply this process using a leave-one-out method, i.e. D_a is computed for all experimentally acquired datasets but the ones used for computing the current calibration transform (both the phantom and the stereo camera system are repositioned after each calibration).

Since the evaluation is based on pre-computed registrations instead of using an entirely separate evaluation methodology, such as detecting predefined points in a special evaluation phantom, unrecognized systematic errors in the registration part may potentially influence the evaluation results. Though our method thus computes a precision result, systematic registration errors can be easily detected. We verified the registrations visually by comparing the registered model of the phantom with the image data in cross-sections and in 3D. Thus, performing the evaluation using a different phantom would presumably lead to only minor knowledge gain.

Error measurement

We specify the translational error as the length of an error vector, i.e. $\|t_e - t\|$. When computing calibration precision, t_e is defined by the translational part of X , which is known in simulation experiments and has to be estimated through averaging in real experiments. We specify the rotational error (in analogy to the translational part) using the quaternion norm $\|q_e - q\|$. According to Andreff et al. [2], this measure is directly related to the angle α of the residual rotation between these two orientations by $\|q_e - q\| = 2 - 2 \cos \frac{\alpha}{2}$.

Results

Registration

We acquired 50 images of the phantom with the probe being always in the same position and registered the phantom in the images. The resulting variation is the sum of noise arising from the ultrasound system and the registration method. Results are shown in Table 1.

Calibration

The calibration was performed 12 times, so that 144 datasets were generated in total. We used 4, 8, and all 12 images, for calibration with the three methods DQ, TP, and TS. The calibration precision was evaluated; Table 2 shows standard deviations for the six parameters of the calibration trans-

Table 1 Results of registration precision tests on real data: standard deviations of the parameters defining the registration transform, computed over 50 trials

Axis	x	y	z
Translation (mm)	0.17	0.14	0.15
Rotation (°)	0.34	0.49	0.33

Table 2 Results of calibration precision tests on real data: standard deviations of the parameters defining the calibration transform \mathbf{X} , using 12 images per calibration and calculated over 12 calibrations in total

Axis	TS	DQ	TP
Translation (mm)			
x	1.34	1.26	0.48
y	2.27	1.46	0.35
z	3.33	0.88	0.23
Rotation (°)			
x	0.74	0.93	0.31
y	0.35	0.33	0.28
z	0.85	1.20	0.79

Table 3 Results of calibration precision tests on real data: minimum, maximum, and mean of the relative translation and rotation errors for calibrating with 4, 8, and 12 images

Images	TS			DQ			TP		
	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean
Translation (mm)									
04	4.59	43.20	17.08	0.69	6.10	4.29	0.48	1.60	0.95
08	1.73	8.54	4.11	1.27	3.39	2.31	0.21	1.24	0.66
12	0.85	6.47	3.54	0.65	3.50	1.84	0.21	1.12	0.56
Rotation (quat. norm)									
04	0.005	0.033	0.018	0.006	0.052	0.026	0.003	0.013	0.007
08	0.005	0.017	0.011	0.005	0.023	0.013	0.001	0.017	0.007
12	0.004	0.017	0.009	0.005	0.023	0.012	0.002	0.017	0.006

form \mathbf{X} , when using 12 images. The minimum, maximum and mean errors for all number of images are shown in Table 3.

Then, one of the computed calibrations was used to simulate datasets out of the 144 original ones using the method described in “Dataset simulation” section. Calibration was performed with these simulated datasets afterwards, and calibration accuracy (as the correct calibration is known from simulation) was computed. Results are shown in Table 4.

Reconstruction

For reconstruction precision evaluation we performed leave-one-out tests as described in “Reconstruction precision tests” section. As we calibrated 12 times with 12 images each, we

Table 4 Results of calibration accuracy tests on simulated data

Images	TS			DQ			TP		
	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean
Translation (mm)									
04	1.24	13.09	5.47	0.80	15.90	3.83	0.09	1.81	0.64
08	0.48	6.39	2.86	0.41	3.57	1.58	0.09	1.56	0.63
12	0.23	6.26	2.19	0.39	2.61	1.11	0.10	1.76	0.50
Rotation (quat. norm)									
04	0.002	0.032	0.010	0.011	0.124	0.029	0.002	0.017	0.006
08	0.002	0.027	0.012	0.003	0.018	0.010	0.002	0.012	0.006
12	0.000	0.021	0.008	0.002	0.021	0.008	0.001	0.013	0.004

As the correct calibration transform is known, the absolute error to this transform can be computed

Table 5 Results of reconstruction precision tests on real and simulated data: mean target localization errors in (mm)

	TS	DQ	TP
Real	5.9	3.5	3.3
Simulated	3.1	1.4	0.9

used every calibration to compute D_a on the other 11 calibration datasets. Hence, 792 values for D_a were computed. The results both for real and simulated data are given in Table 5. Boxplots of the results are shown in Fig. 7.

Discussion

Our results are comparable to those obtained by Poon and Rohling [20] and by Bouchet et al. [5] with nylon wire phantoms. Poon and Rohling [20] give standard deviations of 0.26, 0.20, and 0.84 mm for the translational and 0.18°, 0.64°, and 0.28° for the rotational components of the calibration transform. Bouchet et al. [5] report standard deviations of 0.2, 0.3, 0.2 mm, 0.2°, 0.4°, and 0.1° for the transform parameters. These results are similar to ours, namely 0.48, 0.35, 0.23 mm, 0.31°, 0.28°, and 0.79° (given in Table 2). We suppose that by using more nylon wires and therefore obtaining a higher redundancy of features, Bouchet et al. achieve slightly better results.

Lange and Eulenstein [15], by applying TP and HE calibrations using a commercial phantom, obtain similar results as well. They report a root mean square error for the reconstruction precision of 3.4 mm during TP calibration and 3.3 mm during HE calibration. With comparable experiments, we obtained mean target localization errors of 3.5 mm using the DQ method and, using the TP method, 3.3 mm (see Table 5).

When comparing the two HE calibration methods, the DQ method leads to better stability regarding the translation,

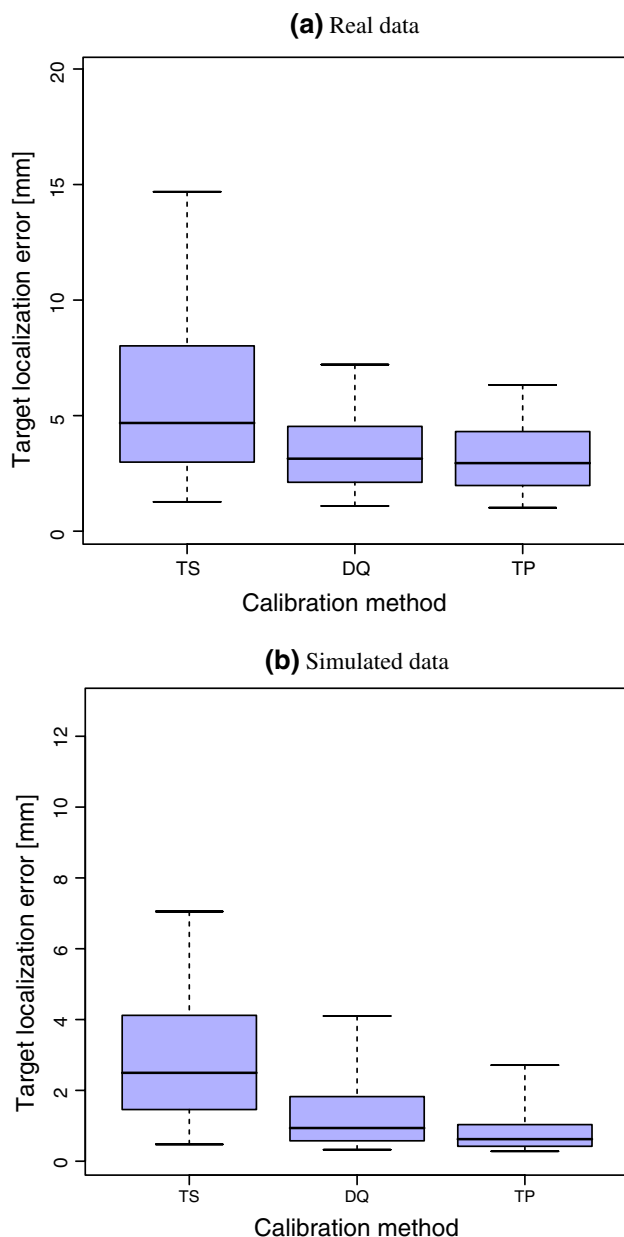


Fig. 7 Results of reconstruction precision tests on real and simulated data: mean target localization errors. Boxes indicate the 25 and 75% quantiles, whiskers the 5 and 95% quantiles; lines inside the boxes show the median

but to worse stability (especially when only few datasets are used) for rotation (see Tables 3, 4). We assume that this is due to the fact that the TS method solves for rotation first and afterwards for translation, whereas the DQ method solves for both at the same time. In the TS method, the rotational error propagates to the translation. Therefore, if HE calibration is necessary, the DQ method should be used.

In our experiments, we used 4, 8, and 12 images for calibration. As Table 3 indicates, with all methods, a higher calibration precision (and therefore a more reliable calibration)

was obtained when using 12 images. Tracking the phantom led to a more straightforward computation and to better results in calibration precision, which is shown by the results for real data with a mean translation error of 0.56 mm for the TP method, in contrast to 3.54 mm (TS) and 1.84 mm (DQ). Concerning reconstruction precision, we obtained slightly better results using the TP method as well, i.e. 3.3 mm (TP) versus 5.9 mm (TS) and 3.5 mm (DQ). However, the boxplot (see Fig. 7a) indicates that in our reconstruction precision evaluation, no significant difference can be found between the TP and the DQ method.

Our simulation experiments support these findings that the TP approach performs better than the HE methods, and that within the HE methods the DQ method performs better regarding stability in the translation parameters, and slightly worse regarding rotation parameters. During reconstruction precision experiments, the TP method shows a superiority over the other methods (see Fig. 7b).

We conclude that when tracking the phantom, the algorithm is provided with more information and thus yields better results. According to these findings, HE calibration should only be used if tracking of the phantom is not possible (e.g. if no phantom is used at all, as Boctor et al. [4] propose).

The presented system could be used in order to solve different clinical problems. In future work we will identify one of them, in order to both make use of the system and to further examine its accuracy and usability in the context of a concrete application.

References

- Adams R, Bischof L (1994) Seeded region growing. *IEEE Trans Pattern Anal Mach Intell* 16(6):641–647
- Andreff N, Horaud RP, Espiau B (2001) Robot hand-eye calibration using structure from motion. *Int J Rob Res* 20(3):228–248
- Besl PJ, McKay HD (1992) A method for registration of 3-d shapes. *IEEE Trans Pattern Anal Mach Intell* 14(2):239–256
- Boctor EM, Iordachita I, Fichtinger G, Hager GD (2006) Ultrasound self-calibration. *Proc SPIE* 6141:784–795
- Bouchet LG, Meeks SL, Goodchild G, Bova FJ, Buatti JM, Friedman WA (2001) Calibration of three-dimensional ultrasound images for image-guided radiation therapy. *Phys Med Biol* 46: 559–577
- Carr J (1996) Surface reconstruction in 3D medical imaging. Ph.D. thesis, University of Canterbury, Christchurch, New Zealand
- Chen HH (1991) A screw motion approach to uniqueness analysis of head-eye geometry. In: *Proceedings CVPR, IEEE Computer Society Conference*, pp 145–151
- Chou JCK, Kamel M (1991) Finding the position and orientation of a sensor on a robot manipulator using quaternions. *Int J Rob Res* 10(3):240–254
- Daniilidis K (1999) Hand-eye calibration using dual quaternions. *Int J Rob Res* 18:286–298
- Hassenpflug P, Vetter M, Wolf I, Thorn M, Grenacher L, Richter GM, Lamade W, Uhl W, Buechler MW, Meinzer H-P (2003) Generation of attributed relational vessel-graphs from

- three-dimensional freehand ultrasound for intraoperative registration in image-guided liver surgery. *Proc SPIE* 5029:222–230
11. Hastenteufel M, Mottl-Link S, Wolf I, de Simone R, Meinzer H-P (2003) A method for the calibration of 3D ultrasound transducers. *Proc SPIE* 5029:231–238
 12. Horaud R, Dornaika F (1995) Hand-eye calibration. *Int J Rob Res* 14:195–210
 13. Kavan L, Collins S, O'Sullivan C, Zara J (2006) Dual Quaternions for Rigid Transformation Blending. Technical Report TCD-CS-2006-46, Trinity College, Dublin
 14. Khadem R, Yeh CC, Sadeghi-Tehrani M, Bax MR, Johnson JA, Welch JN, Wilkinson EP, Shahidi R (2000) Comparative tracking error analysis of five different optical tracking systems. *Comput Aided Surg* 5(2):98–107
 15. Lange T, Eulenstein S (2002) Calibration of swept-volume 3D ultrasound. In: *Proceedings of Medical Image Understanding and Analysis 2002*, The University of Portsmouth
 16. Lindseth F, Tangen GA, Lango T, Bang J (2003) Probe calibration for freehand 3-d ultrasound. *Ultrasound Med Biol* 29(11):1607–1623
 17. Ma B, Moghari MH, Ellis RE, Abolmaesumi P (2007) On fiducial target registration error in the presence of anisotropic noise. In: *Proc MICCAI*, vol 4792. Springer, Berlin, pp 628–635
 18. Mercier L, Lango T, Lindseth F, Collins DL (2005) A review of calibration techniques for freehand 3D ultrasound systems. *Ultrasound Med Biol* 31(4):449–471
 19. Moré JJ (1978) The Levenberg-Marquardt algorithm: implementation and theory. In: Watson GA (ed) *Numerical analysis: proceedings of the Biennial Conference held at Dundee, June 28–July 1, 1977*. Lecture Notes in Mathematics, vol 630. Springer, Berlin, pp 104–116 (of xii + 199)
 20. Poon TC, Rohling RN (2005) Comparison of calibration methods for spatial tracking of a 3D ultrasound probe. *Ultrasound Med Biol* 31(8):1095–1108
 21. Prager RW, Rohling RN, Gee AH, Berman L (1998) Rapid calibration for 3-d freehand ultrasound. *Ultrasound Med Biol* 24(6):855–869
 22. Schmidt J, Vogt F, Niemann H (2005) Robust hand-eye calibration of an endoscopic surgery robot using dual quaternions. In: *Pattern Recognition, 27th DAGM Symposium*, vol 3663. Springer, Berlin, pp 548–556
 23. Strobl KH, Hirzinger (2006) Optimal hand-eye calibration. In: *Proc IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2006)*, Beijing, China, pp 4647–4653
 24. Treece GM, Gee AH, Prager RW, Cash CJC, Berman LH (2003) High-definition freehand 3-d ultrasound. *Ultrasound Med Biol* 29(4):529–546
 25. Tsai RY, Lenz RK (1989) A new technique for fully autonomous and efficient 3D robotics hand/eye calibration. *IEEE Trans Rob Autom* 5:345–358