The International Journal of **Medical Robotics** and **Computer Assisted Surgery**

THE INTERNATIONAL JOURNAL OF MEDICAL ROBOTICS AND COMPUTER ASSISTED SURGERY
Int J Med Robotics Comput Assist Surg (2013)
Published online in Wiley Online Library (wileyonlinelibrary.com) **DOI:** 10.1002/rcs.1478

ORIGINAL ARTICLE

Hand–eye and radial distortion calibration for rigid endoscopes

Abed Malti[1]*
Joao Pedro Barreto[2]

[1]*ALCoV-ISIT, UMR 6284 CNRS/ Université d'Auvergne, Clermont-Ferrand, France*

[2]*ISR–Coimbra, Faculty of Science and Technology, University of Coimbra, Coimbra, Portugal*

*Correspondence to: A. Malti, ALCoV-ISIT, UMR 6284 CNRS/ Université d'Auvergne, 28 Place Henri Dunant, Clermont-Ferrand, France.
E-mail: abed.malti@gmail.com

Accepted: 26 November 2012

Abstract

Background In this paper, we propose a non-linear calibration method for hand–eye system equipped with a camera undergoing radial distortion as the rigid endoscope. Whereas classic methods propose either a separated estimation of the camera intrinsics and the hand–eye transform or a mixed non-linear estimation of both hand–eye and camera intrinsics assuming a pin-hole model, the proposed approach enables a simultaneous refinement of the hand–eye and the camera parameters including the distortion factor with only three frames of the calibrated pattern.

Methods Our approach relies on three steps: (i) linear initial estimates of hand–eye and radial distortion with minimum number of frames: one single image to estimate the radial distortion and three frames to estimate the initial hand–eye transform, (ii) we propose to express the camera extrinsic with respect to hand–eye and world–grid transforms and (iii) we run bundle adjustment on the reprojection error with respect to the distortion parameters, the camera intrinsics and the hand–eye transform.

Results Our method is quantitatively compared with state-of-the-art linear and non-linear methods. We show that our method provides a 3D reconstruction error of approximately 5% of the size of the 3D shape.

Conclusions Our experimental results show the effectiveness of simultaneously estimating hand–eye and distortion parameters for 3D reconstruction.

Keywords hand–eye calibration; radial distortion; endoscopy; 3D reconstruction

Introduction

Developing a vision system for 3D registration and 3D scene reconstruction requires precise calibration of the used camera. Besides knowing the intrinsics (1), the problem is further constrained by tracking the poses of the camera body using an external sensor (opto-tracker, kinematic readings etc). To register pre-acquired scene model with current views, the 3D motion of the camera reference frame needs to be recovered. Because the camera body and the camera reference frame are not coincident, the rigid displacement between them needs to be estimated.

Determining such displacement is known as the hand–eye calibration. This problem initially arose in robotics when a camera was mounted on a robot to measure 2D and 3D geometric relationships among different viewed objects (2). It has been applied in several contexts: for instance, in sensor-based motion planning (3) to automatically determine the optimal positions of the sensor so that all the desired features can be viewed while taking care of

problems of occlusion, depth of focus, field of view and so on. During the past few years, rigid endoscopes have raised much interest in the context of minimal invasive surgery (4,5) and inspection of manufactured devices in industry.[1] An accurate estimate of such a rigid displacement between the rigid endoscope body and the camera frame is important to register the pre-operative models of organs or computer-aided design models of devices with current views. The hand–eye problem in the case of a rigid endoscope is shown in Figure 1.

In a first glance, the hand–eye calibration consists on solving for $X \in SE(3)$ the following equation:

$$AX = XB \tag{1}$$

With $A \in SE(3)$ represents the motion of the camera reference frame and $B \in SE(3)$ encodes the motion of the rigid body to which the camera is attached. It has been proven by Tsai and Lenz (6) that at least two motions with different rotation axes are needed to solve this equation. It can be observed that the rotation involved by the camera motion A is the same as the one involved by the effector motion B. This intrinsic property of the problem is highly relevant to balance the estimation of the rotation part of X in the presence of noise. Since the end of the 1980s, the resolution of such a linear equation has been addressed in a set of extensive works. In contrast, in the work by Horaud (7), a non-linear approach is proposed to solve this equation. It is proven that the hand–eye estimation is not independent from camera intrinsics and that estimating all parameters together improve the hand–eye estimates within the camera intrinsics. Moreover, in the presence of lens distortion (such as in the case of a rigid endoscope), the resolution of equation (1) is not sufficient for an accurate 3D registration and reconstruction as will be proven experimentally in this paper.

The hand–eye calibration applied to the context of endoscopy has specificities: (i) the translation components of the motions A and B are usually small (below 10 cm) because of the close range image characteristics of the device and (ii) the calibration has to be performed by a non-expert and several times when the boroscope (the rigid tube) has to be changed, which requires the method to be robust with a minimum number of motions.

A unified method for simultaneous calibration of hand–eye and camera intrinsics undergoing radial distortion is proposed. The problem is formulated as non-linear optimization problem with minimum number of motions. Our approach relies on three steps: (i) linear initial estimates of hand–eye and radial distortion with minimum number of frames: one single image to estimate the radial distortion and three frames to estimate the initial hand–eye transform, (ii) to express the camera extrinsic with respect to hand–eye and world–grid transforms is proposed and (iii) bundle adjustment (8) on the reprojection error is run with respect to the distortion parameters, the camera intrinsics and the hand–eye transform.

Our former paper (9) is extended in this paper, where we proposed a linear hand–eye calibration using dual quaternion

[1]http://visionscope.moonfruit.com/

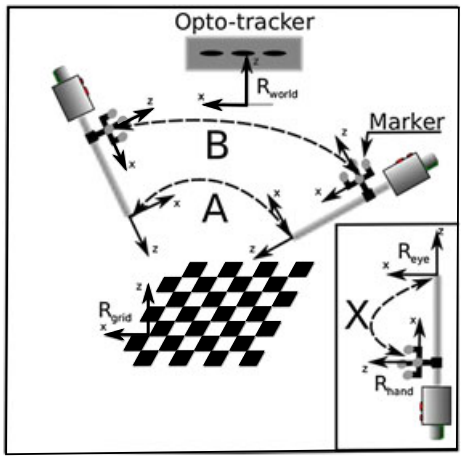

Figure 1. The hand–eye problem in the case of a rigid endoscope. The linear formulation of the hand–eye problem gives rise to the well-known conjugation equation in SE(3): AX = XB. $A \in SE(3)$ is the motion of the camera reference frame computed, thanks to the calibration of the camera with the help of a planar chess grid. $B \in SE(3)$ encodes the motion of the endoscope's body tracked by an external sensor (here an opto-tracker)

by separately estimating rotation and translation. This version is self-contained and explains in detail this previous contribution. The specific contributions of this paper include (i) the non-linear formalization of the simultaneous hand–eye and radial distortion calibration, (ii) experimental results on synthetic data where we show the improvement of our non-linear method above state-of-the-art methods and (iii) real-world experimental results including 3D reconstruction of 3D grid pattern, ex vivo organ and electronically manufactured device. These real experimental results are compared with both state-of-the-art methods for hand–eye calibration and stereo 3D reconstruction.

Paper organization

The related work and our contribution are presented in the second section. The hand–eye problem is presented in the third section. The classic linear solving methods are presented in the fourth section. Our proposed linear approach is presented in the fifth section. An overview of classic non-linear solving methods is presented sixth section. Our proposed non-linear unified approach is presented in the seventh section. The effectiveness of our unified non-linear method using synthetic data is shown in the eighth section. Real experimental results with comparison with ground truth and state-of-the-art non-linear method are presented in the ninth section.

Notations

Quaternions are represented by lowercase bold font (e.g. $\mathbf{q}$). Matrices are denoted by uppercase sans serif font (e.g. A). Vectors providing a direction in 3D are represented using plain lowercase topped by an arrow (e.g. $\vec{l}$). For convenience, and given two 3×1 vectors $\vec{l}$

DOI: 10.1002/rcs

and $\vec{m}$, the dot product is indicated either using $\langle .,. \rangle$ or sing regular matrix/vector multiplication (e.g. $\langle \vec{l}, \vec{m} \rangle = \vec{l}^T \vec{m}$), and the cross-product is carried either using the symbol $\times$ or using the skew symmetric matrix (e.g. $\vec{l} \times \vec{m} = [\vec{l}] \times \vec{m}$). For the sake of simplicity, the vector norm 2 in either $\mathbb{R}^3$ or $\mathbb{R}^4$ is denoted by $||.||_2$. In both cases, the concerned space will be mentioned.

Related work

Several methods exist in the literature to solve the hand–eye calibration problem. It is difficult to provide an exhaustive list of all these works (22). In the focus of our contribution, the different approaches are split into two categories: linear and non-linear methods. The linear methods include subsequently either separated or simultaneous estimation of the rotation and translation.

In the linear separated estimation (6,10–13), first the rotation and then the translation are estimated. Quaternions to solve the rotation have been used by Chou and Kamel (11,12), whereas the rotation in the rotation group SO(3) is solved by Tsai and Lenz (6) and Park and Martin (14). A linear method to separately estimate both hand–eye and world–grid transforms has been proposed by Zhuang *et al.* (15). In the linear simultaneous estimation, the idea is to simultaneously solve for the translation and rotation by using a screw representation of the problem. The most famous method is proposed by Dannilidis (16), where dual quaternions are used to encode the screw representation of the hand–eye problem. A conformal algebra formulation that is similar to dual quaternion formulation is used by Rivera-Rovelo *et al.* (17). A method for data and motion selection has been provided by Schmidt (18) to improve the simultaneous estimation.

In the non-linear methods, the hand–eye transform is estimated by minimizing the norm 2: $\| AX - XB \|$ as was proposed by Dornaika and Horaud (19) to estimate both the hand–eye and the world–grid transforms. A robust metric to minimize has been proposed by Strobl *et al.* (20) and has been used for camera with narrow field of view (21). It has been shown by Horaud and Dornaika (7) that camera intrinsics are not independent from hand–eye parameters. Both camera and hand–eye parameters were simultaneously estimated in a global iterative optimization. In the presence of lens distortion, no method for simultaneous hand–eye and camera intrinsic estimation has been proposed. Moreover, our new formulations allow us to obtain accurate 3D reconstruction with less number of motion than classic linear or non-linear approaches.

Contribution

The contribution of this paper is threefold: (i) a robust linear method is proposed to estimate the hand–eye transform by using a separated estimation of translation and rotation with dual quaternions. The dual quaternion formulation of the problem is used in our linear formulation because it offers a stable representation of the translation as a shift along the rotation axis. The rotation and the translation are estimated separately for mainly two reasons: (1) the rotation part is fully self-described in SO(3) and (2) the noise in translation badly affects the rotation in a joint estimation. (ii) A generic formulation of a unified optimization criterion including hand–eye and camera intrinsics with radial distortion is proposed. This generic formulation uses a minimal number of parameters and can be extended to any distortion model. It is embedded in the reprojection error that is optimized using bundle adjustment (8) and (iii) the application to endoscopy through synthetic and real extensive experimental results: 3D reconstruction of a 3D grid pattern, an ex vivo organ and an electronic cardboard. The method, although validated for hand–eye and distortion calibration for rigid endoscopes, can be easily used in any other robotic vision contexts.

Problem description

Let ${}^{w}T_{h}(t)$ be the SE(3) transformation T linking the hand frame to the world/opto-tracker frame at an instant t, see Figure 2. ${}^{w}T_{h}$ is a sensor measure information that can result either from encoder readings if the camera is mounted upon an articulated robot or from an external opto-sensor tracking the camera body poses. Let again ${}^{e}T_{g}(t)$ be the SE(3) transformation relating the grid/object frame to the camera frame at an instant t. ${}^{e}T_{g}$ stems from the absolute extrinsic parameters of the camera calibration process realized beforehand. ${}^{e}T_{h}$ and ${}^{w}T_{g}$ are two unknown constant transforms that stand, respectively, for the hand-to-camera/eye and the grid-to-world frames. At an instant pose t of the hand camera rigid body, the loop of rigid transforms linking the different system frames can be written as

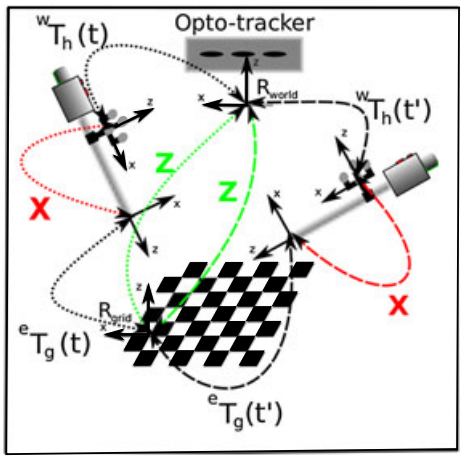

Figure 2. The different transforms between frames. ${}^{w}T_{h}(t)$ is the SE(3) transformation linking the hand frame to the world/opto-tracker frame. ${}^{e}T_{g}(t)$ is the SE(3) transformation relating the grid/object frame to the camera frame. X and Z are two constant unknown transforms

DOI: 10.1002/rcs

$$({}^{g}\mathrm{T}_{e}(t))({}^{e}\mathrm{T}_{h})\left({}^{h}\mathrm{T}_{w}(t)\right)\left({}^{w}\mathrm{T}_{g}\right) = I_4 \tag{2}$$

By using the fact that ${}^{w}\mathrm{T}_{g}$ is constant, the movement of the hand from an instant t to an instant t' gives rise to an equality that depends on the hand–eye transform:

$$({}^{g}\mathrm{T}_{e}(t))({}^{e}\mathrm{T}_{h})\left({}^{h}\mathrm{T}_{w}(t)\right) = ({}^{g}\mathrm{T}_{e}(t'))({}^{e}\mathrm{T}_{h})\left({}^{h}\mathrm{T}_{w}(t')\right) \tag{3}$$

and thus,

$$\mathrm{AX}=\mathrm{XB} \tag{4}$$

where X is the unknown hand–eye transform ${}^{e}\mathrm{T}_{h}$, $\mathrm{A} = {}^{e}\mathrm{T}_{g}(t'){}^{g}\mathrm{T}_{e}(t)$ stands for the camera motion and $\mathrm{B} = {}^{h}\mathrm{T}_{w}(t'){}^{w}\mathrm{T}_{h}(t)$ represents the hand motion. Writing equation (4) as $\mathrm{A} = \mathrm{XBX}^{-1}$ shows up that A and B are the same rigid transformations assessed in different frames of reference. In the same fashion the, second constant transform $Z = {}^{w}T_{g}$ can be expressed in a conjugation equation independently of the hand–eye transform as

$$\mathrm{CZ} = \mathrm{ZD} \tag{5}$$

with $\mathrm{C} = {}^{w}\mathrm{T}_{h}(t){}^{h}\mathrm{T}_{w}(t')$ and $\mathrm{D} = {}^{g}\mathrm{T}_{e}(t){}^{e}\mathrm{T}_{g}(t')$. In the classic solving approaches, the transform Z has received few interest (19,20) because it has no direct influence in the camera–hand–world system. In our formulation, this transform is used to constrain the minimization criterion as is described in the seventh section.

Classic linear solving of X

Separated solving

For one motion of the camera body, the hand–eye formulation of equation (4) can be rewritten as

$$\begin{pmatrix} \mathrm{R}_{\mathrm{A}} & \vec{t}_{\mathrm{A}} \\ 0_{3\times3} & 1 \end{pmatrix}\begin{pmatrix} \mathrm{R}_{\mathrm{X}} & \vec{t}_{\mathrm{X}} \\ 0_{3\times3} & 1 \end{pmatrix} = \begin{pmatrix} \mathrm{R}_{\mathrm{X}} & \vec{t}_{\mathrm{X}} \\ 0_{3\times3} & 1 \end{pmatrix}\begin{pmatrix} \mathrm{R}_{\mathrm{B}} & \vec{t}_{\mathrm{B}} \\ 0_{3\times3} & 1 \end{pmatrix} \tag{6}$$

and then,

$$\begin{cases} \mathrm{R}_{A}\mathrm{R}_{X} = \mathrm{R}_{X}\mathrm{R}_{B} \\ (\mathrm{R}_{A} - \mathrm{I}_3)\,\vec{t}_{\mathrm{X}} = \mathrm{R}_{X}\,\vec{t}_{\mathrm{B}} - \vec{t}_{\mathrm{A}} \end{cases} \tag{7}$$

Solving for rotation

Considering only the rotation part, it comes that

$$\mathrm{R}_{A}\mathrm{R}_{X} = \mathrm{R}_{X}\mathrm{R}_{B} \tag{8}$$

Let **a**, **b** and **q** be the quaternions associated with R_A, R_B and R_X, respectively. By using the quaternion multiplication, it follows that

$$\mathbf{a}.\mathbf{q} = \mathbf{q}.\mathbf{b} \tag{9}$$

By using the result of equation (A.3) (cf. Appendix A) and performing some algebraic operations on equation (9), it is concluded that

$$\underbrace{\begin{pmatrix} a_0 - b_0 & -\left(\vec{a} - \vec{b}\right)^{\mathrm{T}} \\ \vec{a} - \vec{b} & \left[\vec{a} + \vec{b}\right]_{\times} + (a_0 - b_0)\mathrm{I}_3 \end{pmatrix}}_{\mathrm{K}(\mathbf{a},\ \mathbf{b})}\mathbf{q} = 0 \tag{10}$$

with I_3 being the 3×3 identity matrix. The quaternion equation can be written in matrix form as

$$\mathrm{K}(\mathbf{a},\mathbf{b})\mathbf{q} = 0 \tag{11}$$

where K(**a**,**b**) is a 4×4 matrix defined for each pair of quaternions (**a**, **b**).

Equation (8) is a particular case of a similarity transformation. It is well known that it exists a solution X iff the trace of R_{A} is equal to the trace of R_{B} (rotations with the same angle θ but different axis). This means that the scalar part of **a** and **b** are equal ($a_0 = b_0$) and then, the vector parts have the same modulus $\left(\|\vec{a}\| = \|\vec{b}\|\right)$.

In general, equation (11) only admits the trivial solution because K(.,.) is full rank. For the case of **a** and **b** being two rotations with the same angle, matrix K(.,.) becomes of rank 3, and there is a valid quaternion solution **q** different from zero. The existence of a solution can be understood geometrically. Because ($a_0 = b_0$) and $\left(\|\vec{a}\| = \|\vec{b}\|\right)$, then $\left(\vec{a} + \vec{b}\right)$ is always orthogonal to $\left(\vec{a} - \vec{b}\right)$. Thus, according to the first line of the matrix K(.,.), it comes out that the vector component of the solution **q** must be orthogonal to both $\vec{a}$ and $\vec{b}$. It can be verified that this information is already encoded in the bottom three-line bloc of this matrix. Indeed, it states that $\left(\vec{a} - \vec{b}\right)$ and $\left(\left(\vec{a} + \vec{b}\right) \times \vec{q}\right)$ must be collinear, which is possible iff $\rightarrow q$ is orthogonal to both vectors $\left(\vec{a} - \vec{b}\right)$ and $\left(\vec{a} + \vec{b}\right)$. These statements allow to discard redundant terms in matrix K(.,.) that can be rewritten as

$$\tilde{\mathrm{K}}(\mathbf{a},\mathbf{b}) = \left(\vec{a} - \vec{b}\ \ \left[\vec{a} + \vec{b}\right]_{\times}\right) \tag{12}$$

where $\tilde{\mathrm{K}}(\mathbf{a},\mathbf{b})$ is a 3×4 matrix defined for each pair of quaternions (**a**,**b**).

In practice, the rotation R_{A} and R_{B} are measured, and owing to the noise, they do not have the same trace. Hence, the full formulation of equation (11) is important to balance the scalar condition in presence of noise. The solution **q** for the rotation is determined in the least square sense by the Singular Value Decomposition (SVD) of

$$\mathrm{L} = \begin{pmatrix} \mathrm{K}(\mathbf{a}_1,\mathbf{b}_1) \\ \vdots \\ \mathrm{K}(\mathbf{a}_N,\mathbf{b}_N) \end{pmatrix} \tag{13}$$

with $\mathrm{K}(\mathbf{a}_i,\mathbf{b}_i)$, $i = 1, 2 \ldots, N$, being the matrices of the N considered motions $\{(\mathrm{A}_1,\mathrm{B}_1),\ldots,(\mathrm{A}_N,\mathrm{B}_N)\}$. Theoretically, only one motion is enough to solve the rotation part, which is not the case for the translation.

DOI: 10.1002/rcs

Solving for translation
The translation is determined after obtaining the rotation. Considering the translation part of equation (6), it comes that

$$(\mathrm{R_A} - \mathrm{I_3})\vec{t}_\mathrm{X} = \mathrm{R}_X\vec{t}_\mathrm{B} - \vec{t}_\mathrm{A} \tag{14}$$

The matrix $(\mathrm{R_A} - \mathrm{I_3})$ is of rank 2 because $((\mathrm{R}_A - \mathrm{I_3})\vec{l}_\mathrm{A} = 0)$, with $\vec{l}_A$ being the rotation vector associated to $\mathrm{R_A}$. Henceforth, at least $N = 2$ motions are needed to solve for the 3D translation vector. For $N \geq 2$ motions, the solution for the translation $\vec{t}_\mathrm{X}$ is formulated as a non-constrained least square problem:

$$\min \left\| \underbrace{\begin{pmatrix} \mathrm{R}_{\mathrm{A}_1} - \mathbf{I}_3 \\ \vdots \\ \mathrm{R}_{\mathrm{A}_N} - \mathbf{I}_3 \end{pmatrix}}_{D} \vec{t}_x - \begin{pmatrix} \mathrm{R_X}\vec{t}_{\mathrm{B}_1} - \vec{t}_{\mathrm{A}_1} \\ \vdots \\ \mathrm{R_X}\vec{t}_{\mathrm{B}_N} - \vec{t}_{\mathrm{A}_N} \end{pmatrix} \right\|_2 \tag{15}$$

with $\|.\|_2$ stating for the vector norm 2 in $\mathbb{R}^3$.

The motions need to have at least two non-parallel rotation vectors otherway the matrix D will remain of rank 2. Indeed if $\rightarrow l_A$ is that common rotation vector, then $(D \rightarrow l_A = 0)$. Remark that the hand–eye representation in SE(3) does not give any intrinsic constraint for the 3D translation like it gives for the rotation with the trace condition. According to the experiments, the rotation R_X tends to be well estimated, whereas $\rightarrow t_X$ is very noise sensitive for a small number of motions with little amplitudes.

Simultaneous solving

General description of the dual quaternion formulation
The key idea of this approach is to simultaneously estimate the rotation and the translation. Let us consider equation (4) in SE(3). Let $\hat{\mathbf{a}}$, $\hat{\mathbf{b}}$ and $\hat{\mathbf{q}}$ be the unit dual quaternions associated with A, B and X. Equation (4) can be rewritten as

$$\hat{\mathbf{a}}.\hat{\mathbf{q}} = \hat{\mathbf{q}}.\hat{\mathbf{b}} \tag{16}$$

Taking into account the multiplication of dual quaternion equation (A.11) (cf. Appendix A) and splitting the equation in its real and dual parts, it follows that the real part is

$$\mathbf{a}.\mathbf{q} = \mathbf{q}.\mathbf{b} \tag{17}$$

and the dual part is

$$\mathbf{a}'.\mathbf{q} + \mathbf{a}.\mathbf{q}' = \mathbf{q}.\mathbf{b}' + \mathbf{q}'.\mathbf{b} \tag{18}$$

Remark that the real part equation is similar to equation (9).

Specificity of the classic approach
In (16), it is stated that the scalar part of $\hat{\mathbf{a}}$ is equal to the scalar part of $\hat{\mathbf{b}}$. According to the screw representation, this means that the motions A and B have the same rotation angle (equality of the real scalar parts) and the same amplitude of pitch (equality of the dual scalar parts). This property is taken into account in the classic solving using dual quaternion to discard redundant equations. Considering the quaternion multiplication of equation (A.3) and the simplification for the redundant parts, the final matrix equation for the classic dual quaternion formulation is obtained as follows:

$$\underbrace{\begin{pmatrix} \tilde{\mathrm{K}}(\mathbf{a},\mathbf{b}) & 0_{3\times4} \\ \tilde{\mathrm{K}}(\mathbf{a}',\mathbf{b}') & \tilde{\mathrm{K}}(\mathbf{a},\mathbf{b}) \end{pmatrix}}_{M} \begin{pmatrix} \mathbf{q} \\ \mathbf{q}' \end{pmatrix} = 0 \tag{19}$$

with $K(.,.)$ being defined previously for equation (11). The matrix M is 6×8 of rank 5 (cf. (16)). If we assume N motions, then the global linear matrix is constructed as

$$\mathrm{P} = \begin{pmatrix} \mathrm{M}_1 \\ \vdots \\ \mathrm{M}_N \end{pmatrix} \tag{20}$$

P is a $6n \times 8$ matrix that has in general rank 6 in a noise-free case. If all the rotation axes are mutually parallel, then the rank falls to 5. In the general case, the algorithm uses an SVD decomposition to deduce the two dual quaternion bases that span the right null space of P. The valid quaternion is computed as the intersection of this null space with the subspace of unit dual quaternions represented by equation (A.4) for the real part and equation (A.13) for the dual part (cf. Appendix A).

Solving method for simultaneous estimation
Let us denote the two dual quaternions that generate the right null space of P by $\hat{\mathbf{u}}$ and $\hat{\mathbf{v}}$. The set of dual quaternion solution of equation (20) is described as

$$\hat{\mathbf{q}} = \alpha_1\hat{\mathbf{u}} + \alpha_2\hat{\mathbf{v}}, \quad \alpha_1\alpha_2 \in \mathbb{R} \tag{21}$$

The two real parameters α_1 and α_2 are suboptimally determined regarding to the constraints of equations (A.4) and (A.13). Two second-order polynomials $\Gamma_{\hat{\mathbf{u}},\hat{\mathbf{v}}}(\lambda)$ and $\Delta_{\hat{\mathbf{u}},\hat{\mathbf{v}}}(\lambda)$ are, respectively, obtained from the unit quaternion condition of equation (A.4) and the orthogonality condition of equation (A.13).

Because α_1 and α_2 never vanish together, $\lambda = \frac{\alpha_1}{\alpha_2}$ can be set. The algorithm solves the second-order equation $(\Delta_{\hat{\mathbf{u}},\hat{\mathbf{v}}}(\lambda) = 0)$ and from the two obtained solutions picks up the one that maximizes the polynomial function $\Gamma_{\hat{\mathbf{u}},\hat{\mathbf{v}}}(\lambda)$. Let us consider this solution as being λ_0; then, α_1 and α_2 are computed as

$$\alpha_2 = \frac{1}{\Gamma_{\hat{\mathbf{u}},\hat{\mathbf{v}}}(\lambda_0)} \text{and } \alpha_1 = \lambda_0\alpha_2 \tag{22}$$

Discussion
By construction and according to the experiments, this algorithm is not very stable to the noise perturbation. Mainly for three reasons,

DOI: 10.1002/rcs

- It does not use the scalar part in the estimation, and in presence of noise, the scalar equality is no longer verified.
- Even if the translation is robustly represented through the screw formalism, coupling the estimation makes the rotation part suffer from the noise in the translation estimation. Indeed, on the one hand, the rotation is fully characterized by equation (17). On the other hand, in the screw representation, the rotation is independent of the translation, whereas the translation depends on the rotation.
- The third reason concerns the constraints for a dual quaternion being unitary. The estimation process suggests to first estimate a 2D vector subspace in the dual quaternion space and then compute the intersection of this subspace with the subspace that fulfils the constraint of one-to-one correspondence with SE(*3*). In a noise-free case or even under small disturbance, this intersection is always one point that corresponds to the solution we are looking for. However, under average noise disturbance, there is no guarantee that the intersection is one point. Moreover, there is no guarantee that there is an intersection.

Our linear algorithm: improved dual quaternion

Specificity of the proposed approach

Given the analysis we drew in the previous sections, the strongest points of our algorithm are the following:

- It uses the dual quaternion formalization of the hand–eye problem that gives a robust representation of the rotation and the translation.
- Because the rotation is independent from the translation in the screw representation and to avoid the influence of the noise of the translation on the rotation, the rotation is estimated separately from the translation.
- The calibration uses the scalar parts to balance the estimate.
- The one-to-one conditions of correspondence with SE(*3*), encoded in equations (A.4) and (A.13), are optimally taken into account in the estimation.

Solving method for the proposed approach

The real part of the dual quaternion formulation of equation (17) follows exactly the formulation of equation (11), where the scalar parts are included. The dual part of the dual quaternion formulation of equation (18) can be written as

$$\mathrm{K}(\mathbf{a}', \mathbf{b}')\mathbf{q} + \mathrm{K}(\mathbf{a}, \mathbf{b})\mathbf{q}' = 0 \tag{23}$$

The expression of equation (23) contains only the dual part $\mathbf{q}'$ as unknown. Assuming the same set of N motions used to estimate $\mathbf{q}$, the following linear system is obtained:

$$\mathrm{L}\mathbf{q}' = -\mathrm{L}'\mathbf{q} \tag{24}$$

where L is defined in equation (13) and L' is defined as

$$\mathrm{L}' = \begin{pmatrix} \mathrm{K}(\mathbf{a}'_1, \mathbf{b}'_1) \\ \vdots \\ \mathrm{K}(\mathbf{a}'_N, \mathbf{b}'_N) \end{pmatrix} \tag{25}$$

Moreover, to represent a valid translation in SE(3), $\mathbf{q}'$ has to obey the constraint of equation (A.13). This can be easily formulated as a least squares problem subject to a linear constraint:

$$\min_{\text{Eq. (A.13)}} \left\| \mathrm{L}_{\mathbf{q}'} + \mathrm{L}'_{\mathbf{q}} \right\|_2 \tag{26}$$

with $\|.\|_2$ being the vector norm 2 in $\mathbb{R}^4$. Such a system can be easily solved using classic methods (23). Table 1 summarizes the steps of the improved dual quaternion approach *I.D.Q* for hand–eye calibration:

Algorithm 1: *"I.D.Q"* algorithm for hand-eye calibration

Data: N motions $\{(\mathbf{a}_i, \mathbf{a}'_i); (\mathbf{b}_i, \mathbf{b}'_i)\}$, with at least 2 non-parallel rotation axis.

Result: the hand-eye transform.

1 construct the matrix L of Eq.(13), using Eq.(11);
2 compute SVD decomposition of L, and deduce the rotation part $\mathbf{q}$;
3 construct L′ using Eq.(25);
4 compute the dual part $\mathbf{q}$ by solving Eq.(26);
5 compute $\vec{t_X}$ using Eq.(A.15). R_X is directly computed from $\mathbf{q}$.

Overview of classic non-linear solving

In the previous formulation, the camera is assumed to be calibrated beforehand at each instant t. The camera extrinsics are the matrices ${}^{e}\mathrm{T}_{g}$, and the camera intrinsics are in the case of a pinhole model; the ratio a, the skew s, the focal length f and the principal point $(\mathrm{pp}_x, \mathrm{pp}_y)$ that describe the affine transformation between the camera frame and the image frame:

Table 1. Real data: reprojection error analysis with three calibration images

	$\xi = 10\%$	$\xi = 20\%$	$\xi = 30\%$
X linear			
Mean	47.86	18.84	31.30
Std	13.38	12.70	16.41
(*a*,*s*,*f*,ξ,*X*,*Z*) refined			
Mean	0.83	0.74	1.66
Std	0.32	0.41	0.77
Ground truth calibration			
Mean	0.86	0.37	0.96
Std	0.43	0.10	0.70

DOI: 10.1002/rcs

$$\Gamma = \begin{pmatrix} af & sf & \mathrm{pp}_x \\ 0 & a^{-1}f & \mathrm{pp}_y \\ 0 & 0 & 1 \end{pmatrix} \tag{27}$$

A new formulation of equation (4) that uses the perspective projection from the grid frame to the image frame has been proposed by Horaud and Dornaika (7,19) as

$$\mathrm{P}(t)\mathrm{Y} = \mathrm{P}(t_0)\mathrm{YB} \tag{28}$$

With the perspective projection matrix,

$$P(.) = \left(\Gamma\,\vec{0}\right)^{\mathrm{e}}\mathrm{T}_{\mathrm{g}}(.) = \left(N(.) \quad n(.)\right) \tag{29}$$

where $\rightarrow 0$ is a 3×1 vector of zeros and $\mathrm{Y} = {}^{\mathrm{g}}\mathrm{T}_{\mathrm{e}}(t_0)\mathrm{X}$ is the transformation matrix from the hand frame to the grid frame. Owing to this formulation, all the movements have to be computed from a reference pose t_0 to keep Y constant. It has been proven in (7,19) that the two formulations are equivalent as it can be seen from the decomposition of equation (28) in the same fashion as equation (7):

$$\begin{aligned} \vec{n}_{\mathrm{N}} &= \mathrm{R}_{\mathrm{Y}}\vec{n}_{\mathrm{B}} \\ (\mathrm{N} - \mathrm{I}_3)\vec{t}_{\mathrm{Y}} &= \mathrm{R}_{\mathrm{Y}}\vec{t}_{\mathrm{B}} - \vec{t}_{\mathrm{N}} \end{aligned} \tag{30}$$

To estimate R_{Y} and $\vec{t}_{\mathrm{Y}}$ with n motions of the camera body, the problem is cast into a minimization of two positive weighed error functions:

$$\mathcal{J}_1 = \lambda_1 f_1(\mathrm{R}_{\mathrm{Y}}) + \lambda_2 f_2\left(\mathrm{R}_{\mathrm{Y}}, \vec{t}_{\mathrm{N}}\right) \tag{31}$$

λ_1 and λ_2 are two real positive values,

$$f_1(\mathrm{R}_Y) = \sum_{i=1}^{n} \left\| \vec{n}_{\mathrm{N}_i} - \mathrm{R}_{\mathrm{Y}}\vec{n}_{B_i} \right\|^2 \tag{32}$$

and

$$f_2\left(\mathrm{R}_{\mathrm{Y}}, \vec{t}_{\mathrm{N}}\right) = \sum_{i=1}^{n} \left\| (\mathrm{N} - \mathrm{I}_3)\vec{t}_{\mathrm{Y}} - \mathrm{R}_{\mathrm{Y}}\vec{t}_{\mathrm{B}} - \vec{t}_{\mathrm{N}} \right\|^2 \tag{33}$$

This minimization criterion has been implemented using a quaternion representation of the rotation matrix R_Y by adding a weighed unitary quaternion constraint. The weight values have been set experimentally to one for λ_1 and λ_2 and to 10^6 for the rotation quaternion constraint. The minimization is assessed over the six parameters of the Y transform.

This non-linear formulation offers the possibility to enable a refinement of the perspective camera parameters together with the hand–eye transform. However, it exhibits some gaps that might be fulfilled:

- Minimizing a sum of norms can be problematic when they do not represent the same measure. Indeed, a distance over unit vectors is represented by f_1, and a norm over metric distances is represented by f_2.
- Initially, it does not refine neither the intrinsics nor the extrinsics, and if it does so, there will be too much unconstrained parameters to refine: $6 * (n + 1)$ for the different poses and 5 for intrinsics in addition of six of the required matrix Y.
- Matrix Y is computed within a reference to one camera pose that can be the most noisy estimated matrix stemmed from the camera calibration step.

Our proposed non-linear unified approach

Camera with radial distortion

Let $(x,y,z,1)^{\mathrm{T}}$ be the homogenous coordinates of a given 3D point expressed in the grid/object frame. Let again $(d_x,d_y)^{\mathrm{T}}$ be the distorted pixel in the image plan defined as

$$\left(d_x, d_y\right)^{\mathrm{T}} = f_{\mathrm{c}}({}^{e}\mathrm{T}_g(x, y, z, 1)^{\mathrm{T}}) \tag{34}$$

Several choices of f_{c} have been proposed, and each one of them can be used in our approach. In our implementation, the used model is the one that allows us to obtain intrinsic and distortion calibration with minimum number of solution (1).

For a set of finite pixels $\left\{\left(d_x^i, d_y^i\right)\right\}_{i=1,\ldots,N}$ corresponding to a set of finite 3D points $\{(x^i,y^i,z^i)\}_{i=1,\ldots,N}$, the camera reprojection error is given by

$$\mathcal{J} = \sum_{i=1}^{N} \left\| \left(d_x^i, d_y^i\right)^{\mathrm{T}} - f_{\mathrm{c}}\left({}^{\mathrm{e}}\mathrm{T}_{\mathrm{g}}\left(x^i, y^i, z^i, 1\right)^{\mathrm{T}}\right) \right\|^2 \tag{35}$$

In a first glance, an intuitive way to improve the reprojection error is to integrate it in the minimization criterion proposed by (7,19). Unfortunately, adding another norm that is not in the same scale as the two other is not recommended and will not balance the minimization among the three norms as it should be. Now, if we express the camera extrinsic transform ${}^{e}T_g$ with respect to the hand–eye and the grid–base transforms using loop constraint equation (2),

$${}^{\mathrm{e}}\mathrm{T}_{\mathrm{g}}(t) = \mathrm{X}^{\mathrm{h}}\mathrm{T}_{\mathrm{w}}(t)\mathrm{Z} \tag{36}$$

this expression can then be used in the minimization criterion to replace the extrinsics in the reprojection error. This allows us to considerably reduce the number of unknowns from $6 \times (n + 1)$ for the extrinsic parameters down to 12 for both X and Z. In addition, the five affine intrinsic parameters $(a,s,f,\mathrm{pp}_x,\mathrm{pp}_y)$ and the distortion parameter ξ that can be either a single real value as in (24) or a vector as in (25) make the number of unknowns constant and independent of the number of motions. Thus, for $n + 1$ camera poses (which gives n motions) where for each we have N_j, $j = 1, \ldots, n + 1$, calibration points, the camera reprojection error through all the views is written as

DOI: 10.1002/rcs

$$\mathcal{J} = \sum_{j=1}^{n+1} \sum_{i=1}^{N_j} \left\| \left(d_x^i, d_y^i\right)_j^{\mathrm{T}} - f_c\left(\mathrm{X}^{\mathrm{h}}\mathrm{T}_{\mathrm{w}}(j) Z \left(x^i, y^i, z^i, 1\right)_j^{\mathrm{T}}\right) \right\|^2 \tag{37}$$

with ${}^{\mathrm{h}}T_{\mathrm{w}}(j)$ being the hand-to-base transform from the tracker sensor for each view $j = 1, \ldots, n+1$. The 3D grid points $\left(x^i, y^i, z^i 1\right)_j^{\mathrm{T}}$ are mapped in the pixels $(d_x^i, d_y^i)_j^{\mathrm{T}}$, $i = 1,$ $, N_j,$ by projection and distortion. This formulation implicitly introduces the conjugation constraint of equations (4) and (5) to ensure that both X and Z are consistent with the current system and the loop constraint of equation (2). Unlike the classic non-linear methods (7,19,21), the following advantages are presented by the proposed approach:

1. It relies on a robust initial values of X and Z estimated by a specifically designed Algorithm 1.
2. One homogeneous measure norm is used rather than mixing different measures for rotation, translation and image error as in (7,19,21).
3. The minimization is processed with respect to the intrinsics $(a,s,f,\mathrm{pp}_x,\mathrm{pp}_y,\xi)$, X and Z, which considerably reduces the number of parameters to refine when compared with classic non-linear method as in (7).
4. It does not depend on one single camera pose as in (7,19) and use an averaging estimation by introducing Z in the formulation.
5. It allows us to refine distortion parameters in the same process that was not used in any state-of-the-art hand–eye estimation methods.
6. The minimization term of equation 37 can be introduced directly in a bundle adjustment framework that has been proved very efficient for projective and Euclidean 3D reconstruction (26,27).

Choice of camera model

Several approaches have been proposed to model the imaging function with distortion effect f_c from the 3D scene points to the distorted imaged pixels (1,28,25). Even if our formulation is independent of this model and works with any of them, the model proposed in (1) is chosen to be used for our implementation with bundle adjustment because it allows one to obtain calibration parameters $(a,s,f,\mathrm{pp}_x,\mathrm{pp}_y,\xi)$ with one single image. It can be written as

$$\left(d_x, d_y\right)^{\mathrm{T}} = \mathrm{K} f_{\mathrm{d}}({}^{\mathrm{e}}\mathrm{T}_{\mathrm{g}}(x, y, z, 1)^{\mathrm{T}}) \tag{38}$$

where f_{d} is the first-order division model of distortion (24), which can be expressed as

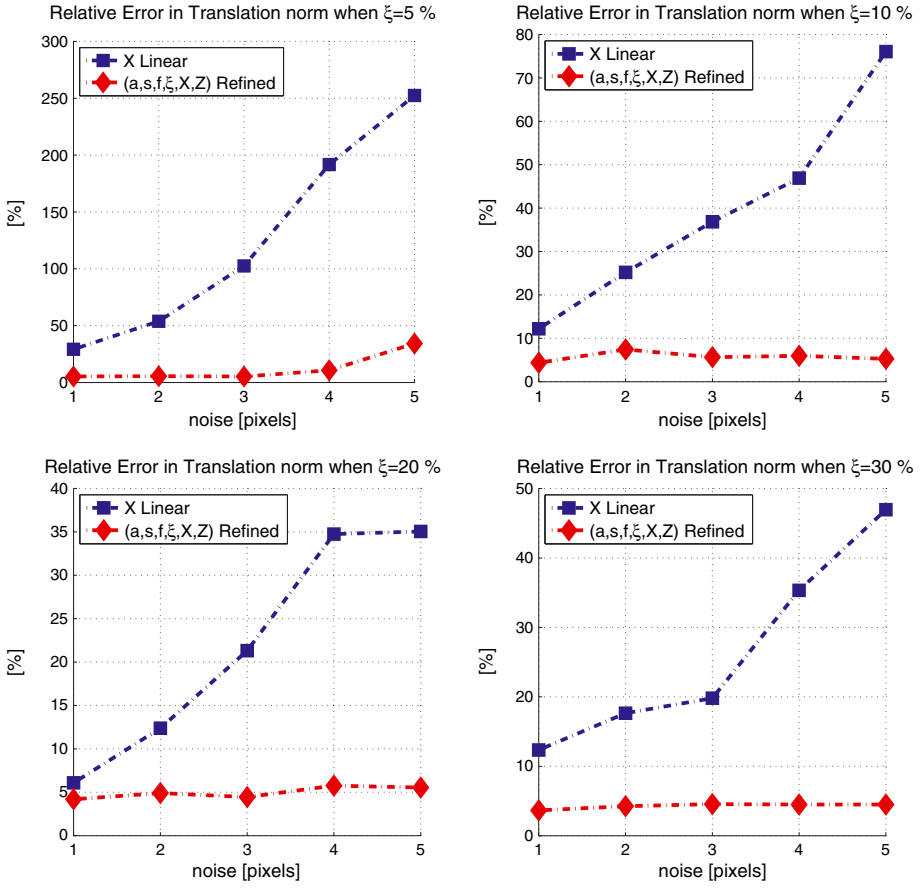

Figure 3. Synthetic data. The relative error in translation. The repeatability of the linear hand–eye estimation is improved

DOI: 10.1002/rcs

$$f_d((x,y,z,1)^T) = \begin{pmatrix} \dfrac{2x}{z+\sqrt{z^2-4\xi(x^2+y^2)}} \\ \dfrac{2y}{z+\sqrt{z^2-4\xi(x^2+y^2)}} \\ 1 \end{pmatrix} \quad (39)$$

with ξ a negative value representing the distortion factor. It represents how far the perspective projected points are pushed toward the centre of the image. To have meaningful representation, it will be referred as being the percentage ratio of the shifted amount distorted–undistorted above the undistorted radius for the corner point of the image.

Evaluation of the proposed non-linear method on synthetic data

Setup of the synthetic experiment

Simulations were conducted to show how far the method is robust to noise and how does it behave for a couple of distortion values. The simulation is conducted as follows: Let us assume the following ground truth hand-to-eye and grid-to-world transforms:

$$^{h}T_e = \begin{pmatrix} 0.433 & -0.866 & -0.250 & 25\ \text{mm} \\ 0.750 & 0.480 & -0.433 & 25\ \text{mm} \\ 0.480 & 0 & 0.866 & 90\ \text{mm} \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (40)$$

$$^{w}T_g = \begin{pmatrix} 0 & 0 & 1 & -100\ \text{mm} \\ 1 & 0 & 0 & 1800\ \text{mm} \\ 0 & 1 & 0 & 2000\ \text{mm} \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (41)$$

The simulated camera is assumed to have $f = 500$ mm of focal length, a null skew and a unitary aspect ratio. The resolution of the image is assumed to be of 640×480 pixels, and the principal point is assumed to be in the middle of the image. The used distortion factors are, respectively, $\xi\% = 5, 10, 20, 30$. The calibration grid is assumed to be planar with square patterns of 40 mm per side. $n = 5$ images are generated and perturbed with noise of zero mean Gaussian distribution with std = 1, 2, 3, 4, 5 pixels. They are calibrated with the method proposed in (1) with an average of $N = 50$ calibration grid points for each image. The camera motions are randomly generated with uniform distribution with a mean of 240 mm and a standard deviation of 100 mm in translation and a mean of 70 deg and a standard deviation of 30 deg in rotation. The hand pose measurements

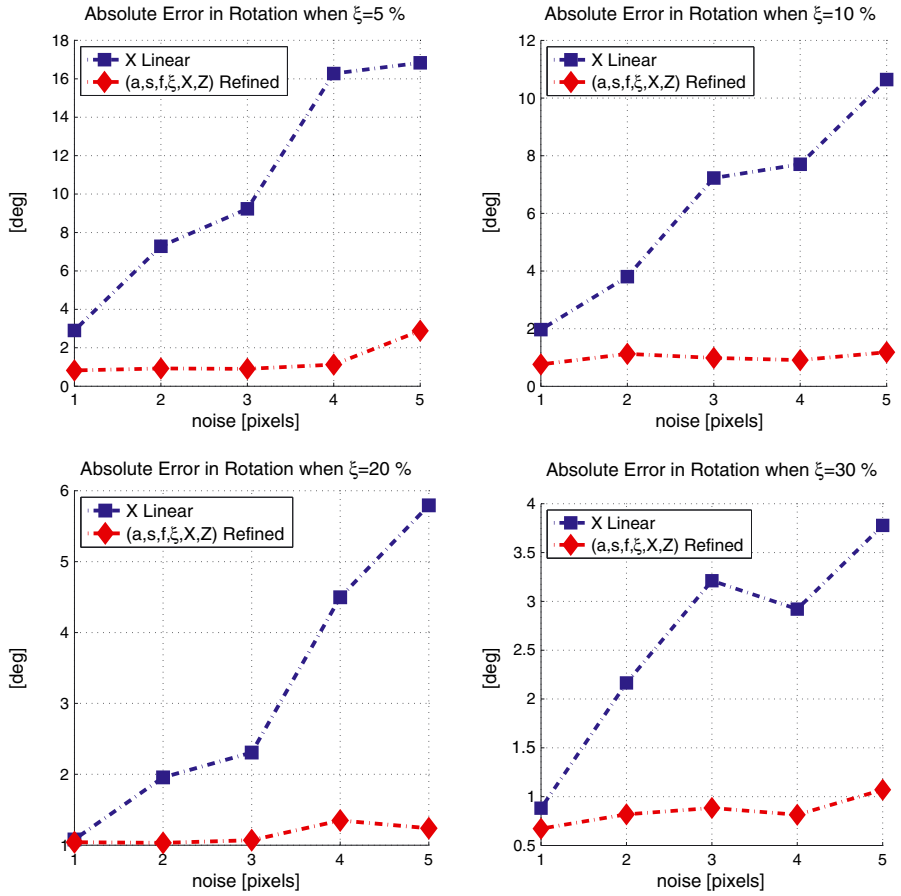

Figure 4. Synthetic data. The absolute error in rotation. The repeatability of the linear hand–eye estimation is improved

DOI: 10.1002/rcs

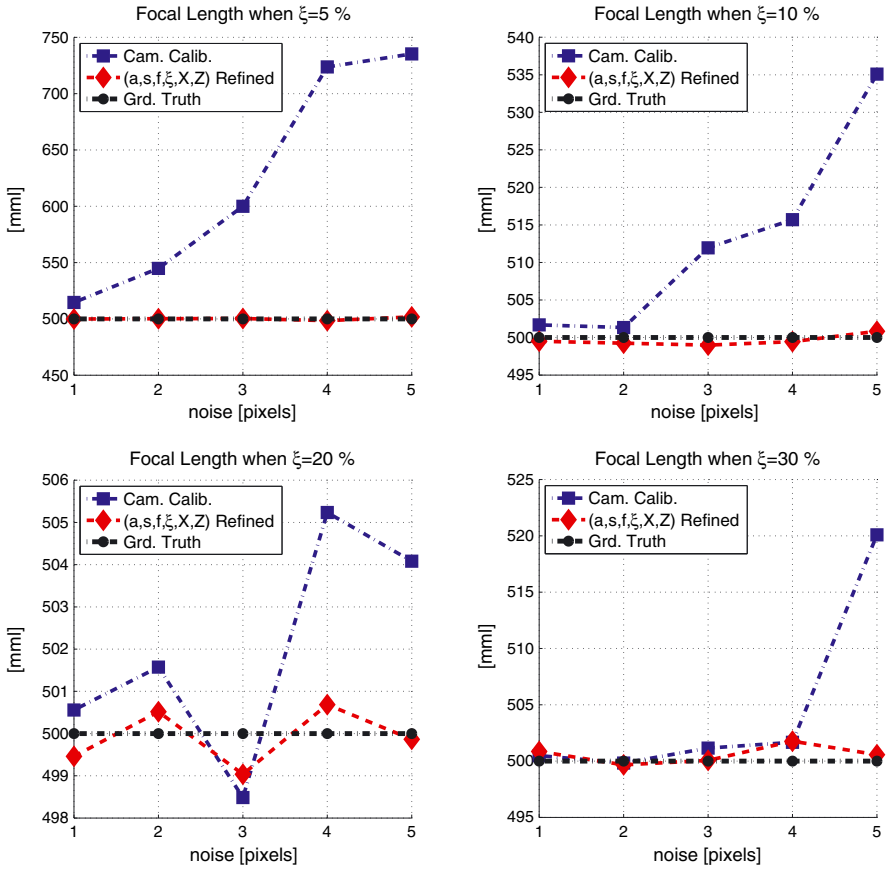

Figure 5. Synthetic data. Focal length estimation. The repeatability of the intrinsic calibration is improved

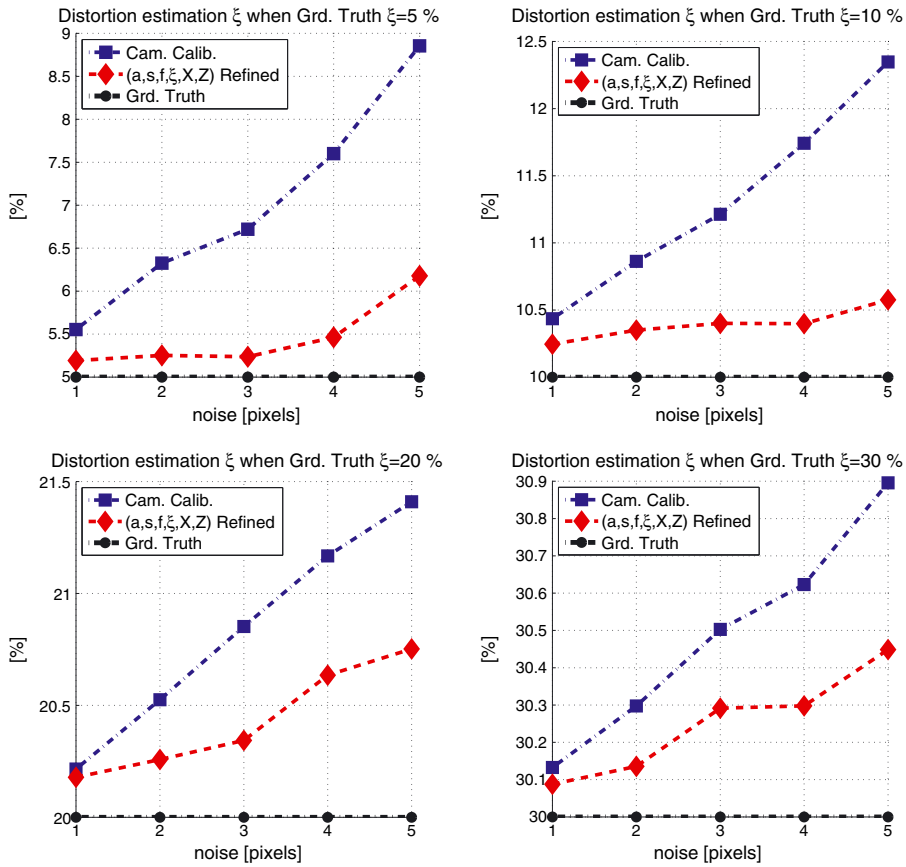

Figure 6. Synthetic data. Distortion estimation. The repeatability of the calibrated distortion factor is improved

DOI: 10.1002/rcs

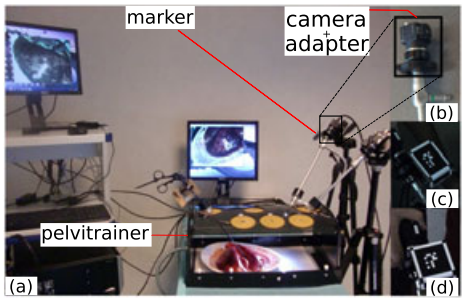

Figure 7. Experimental setup to acquire datasets with ground truth. (a) Two Point Grey cameras are synchronized to obtain reference ground truth data using stereo views. The endoscope with a planar marker is then moved around the object to reconstruct. The marker is tracked with two Point Grey cameras, and its pose is computed with triangulation. All the cameras of the setup are synchronized with 15 fps and run at a resolution of 640×480 pixels. (b) The Point Grey camera is mounted with a c-mount adapter to the laparoscope lens. (c) Left view of the planar marker. (d) Right view of the planar marker

are perturbed with a zero mean Gaussian distribution of 0.1 mm in translation and 0.2 deg in rotation. Each experiment is typically repeated 20 times to obtain statistically meaningful results. The reported graphics exhibit the Root Mean Square (*RMS*) errors through those runs for different noise and distortion factors.

Analysis of the synthetic results

The minimization of the reprojection error of equation (37) is processed using bundle adjustment with *levmar* toolbox (29). The camera intrinsics are initialized with the method described in (1). The hand-to-eye and grid-to-base transforms are initialized by linear solving as described in the fifth section. It is shown in the results that a simultaneous refinement of the camera and the hand–eye parameters outperforms an independent camera calibration from a hand–eye calibration. We can see in Figures 3–6 that the hand–eye, focal length and distortion are estimated with better accuracy. It can also be observed that our method improves the repeatability when the noise in the detection of the grid corners increases.

Real experimental results

The experimental setup consists of a *Point Grey* camera with a resolution of 960×1280 pixels to which is mounted a laparoscope lens with a c-mount adapter. A ring of Light Emitting Diode (LED) markers is fixed around the body of the camera. The poses of the markers (the hand pose) are given by the Optotrak Certus manufactured by the *NDI Motion Capture*. The ground truth intrinsics and extrinsics

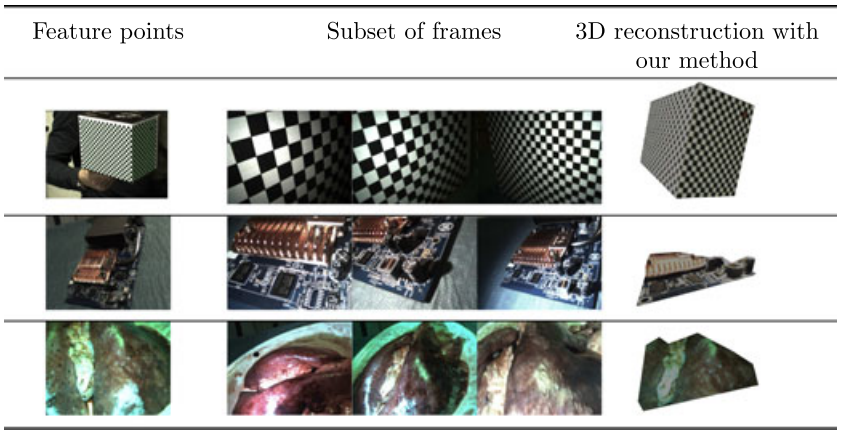

Figure 8. Real data. Qualitative results of 3D reconstructions using our proposed non-linear method. Top row: the checkerboard cube with bounding box of $180 \times 100 \times 140\,\text{mm}^3$. Middle row: the electronic boardcard with bounding box of $80 \times 60 \times 30\,\text{mm}^3$. Bottom row: the lung's lamb with bounding box of $110 \times 100 \times 50\,\text{mm}^3$

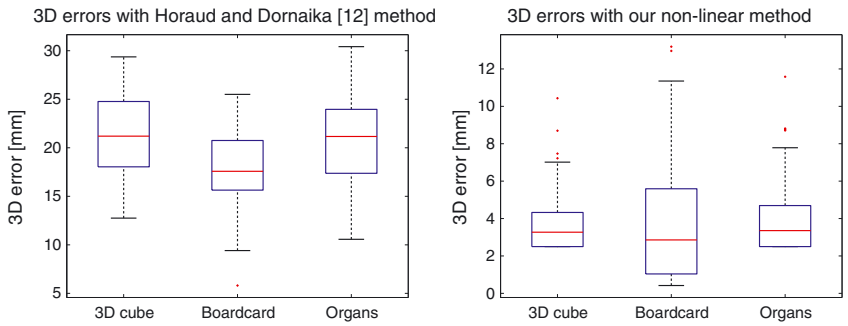

Figure 9. Real data. 3D errors of reconstructions: comparison between Horaud Dornaika method (7) and our method. Our method is substantially more accurate than Horaud and Dornaika's method with an average of 3.5 mm

DOI: 10.1002/rcs

of the camera are computed using *Bouguet* toolbox (30). A checker board of 5 mm per square side is used. An average set of 40 corner points per image is used in the calibration. Three different lenses with low (10%), medium (20%) and high (30%) levels of distortions are used (these are usually the distortion ranges in endoscopes). A set of 15 images is taken with each lens. According to the opto-sensor measurements, the hand was moved with a motion of about 200 mm in average in translation and an average of 20 deg in rotation. The hand-to-eye and grid-to-base transforms are initialized as described in the fifth section by using three images. In Table 1 is summarized a statistical analysis of the reprojection errors with the different ground truth distortion amounts. The initialization of the reprojection error is computed with the initial values of X and Z. We can observe that our unified optimization method improves the mean and the standard deviation (std) of the reprojection error.

3D reconstruction

To acquire real datasets with ground truth, an appropriate framework (see Figure 7) is set up, where we use two laparoscopes to acquire stereo views and two cameras to track a planar marker mounted on one of the two laparoscopes. The two laparoscopes are mounted to two Point Grey cameras with two c-mounts. The plane pose of the marker is computed with triangulation after detection (31). The four Point Grey Flea2 colour cameras are synchronized at 15 fps with a resolution of 640×480 pixels. Thanks to this setup, accurate ground truth 3D points of three different targets can be built: (i) a checkerboard cube, (ii) an electronic boardcard and (iii) an ex vivo organ (a lamb's lung). Scale Invariant Feature Transform (SIFT) (32) is used to compute feature points and correspondences. Three calibration images are used to compute the hand–eye transform and the camera intrinsics (a,s,f,ξ) with our unified method. Bouguet toolbox (30) is used to compute the hand–eye transform with Horaud and Dornaika method (7) by using the same three calibration images. The ground truth distortion value is of 20% as was given by Bouguet toolbox. A set of 50 image frames is taken for each target by moving the laparoscope with marker around the target. Outliers were removed using the hand–eye estimates, and the marker (the hand) poses to compute the epipolar constraint. In Figure 8, the obtained 3D reconstructions are shown using our method. In the case of the cube dataset, 882 feature points were gathered for the 3D reconstruction. In the case of the electronic boardcard and the ex vivo organ, the detection of feature points was hard because of the repetitive textures, and then, many features were rejected as outliers. Only 99 and 49 points were, respectively, used for the 3D reconstruction of the electronic board and the ex vivo organ. The 3D points were computed by triangulation (33) using the hand–eye transform and the marker poses. The reconstruction errors are computed as the difference between the stereo 3D points and the reconstructed 3D points of each target. Our method is substantially more accurate than Horaud and Dornaika's method with an average of 3.5 mm (see Figure 9). The electronic boardcard and the ex vivo organ have bigger standard deviation errors because of the slight variation in the feature position among frames. This is mainly due to the repetitive texture and a non-negligible specular reflection component of these surfaces.

Conclusion

The main contribution of this paper is a robust and stable estimation of both the hand–eye and camera intrinsics with radial distortion. Our approach relies on three steps: (i) linear initial estimates of hand–eye and radial distortion with minimum number of frames: one single image to estimate the radial distortion and three frames to estimate the initial hand–eye transform, (ii) we propose to express the camera extrinsic with respect to hand–eye and world–grid transforms and (iii) we run bundle adjustment on the reprojection error with respect to the distortion parameters, the camera intrinsics and the hand–eye transform. Our method is quantitatively compared with state-of-the-art linear and non-linear methods. We experimentally proved that embedding the distortion factor and the hand–eye transform in the minimization of a unified reprojection error improves substantially the estimation and the usage of a hand–eye-based 3D reconstruction.

Appendix A. Background in quaternion and dual quaternion

Appendix A.1. Representation of rotations using quaternions
A quaternion $\mathbf{q} \in \mathbb{Q}$ is a quadruplet of real numbers (34) that can be split in a scalar q_0 and a 3D vector component $\rightarrow q$:

$$\mathbf{q} = \begin{pmatrix} q_0 \\ \vec{q} \end{pmatrix} \tag{A.1}$$

The conjugate quaternion $\mathbf{q}^*$ of $\mathbf{q}$ is defined as

$$\mathbf{q}^* = \begin{pmatrix} q_0 \\ -\vec{q} \end{pmatrix} \tag{A.2}$$

The product of two quaternions a and b is defined as follows:

$$\mathbf{a}.\mathbf{b} = \begin{pmatrix} a_0 b_0 - \vec{a}^T.\vec{b} \\ a_0 \vec{b} + b_0 \vec{a} + \vec{a} \times \vec{b} \end{pmatrix} \tag{A.3}$$

A quaternion q represents uniquely a rotation matrix R SO(3) if and only if it is a unit quaternion:

DOI: 10.1002/rcs

$$\mathbf{q}^*.\mathbf{q} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{A.4}$$

This condition encodes the orthogonal constraint fulfilled by the SO(3) matrices. Because a rotation matrix has three DoF, this constraint defines a 3D submanifold in the 4D quaternion space. The one-to-one mapping from SO(3) to this submanifold is given by

$$\mathbf{q} = \begin{pmatrix} \sin\left(\frac{\theta}{2}\right) \\ \cos\left(\frac{\theta}{2}\right)\vec{l} \end{pmatrix} \tag{A.5}$$

where $\vec{l} \in \mathbb{R}^3$ (with $\|\vec{l}\| = 1$) is the rotation axis and $\theta \in [-\pi, \pi]$ is the rotation angle. The composition product of rotations is represented as a quaternion product, and the inverse rotation is represented as the conjugate quaternion.

Appendix A.2 Representation of rigid motions using dual quaternions

Consider a rigid motion in SE(3) represented by a 4×4 matrix T with six DoF (three DoF for the rotation R and three DoF for the translation $\vec{t}$):

$$\mathrm{T} = \begin{pmatrix} \mathrm{R} & \vec{t} \\ 0 & 1 \end{pmatrix} \tag{A.6}$$

Any rigid motion T can be carried by assuming a rotation of an angle θ around a 3D line $\mathbf{s}$ and a translation d along this same line (cf. *Chasles theorem* (35)). This leads to the screw representation for rigid transformations consisting in a line in 3D space (represented by a 6×1 vector $\mathbf{s}$), a rotation angle θ and a pitch value d. For a particular $T \in \mathrm{SE}(3)$, the screw axis is

$$\mathbf{s} = \begin{pmatrix} \vec{l} \\ \vec{m} \end{pmatrix} = \begin{pmatrix} \vec{l} \\ \frac{1}{2}[\vec{t}]_\times \vec{l} + \frac{1}{2} cot\left(\frac{\theta}{2}\right)[\vec{l}]_\times [\vec{t}]_\times \vec{l} \end{pmatrix} \tag{A.7}$$

with $\vec{t}$ being the translation component and $\vec{l}$ and θ being, respectively, the axis and the angle of the rotation R. Remark that $\vec{l}$ and the momentum vector $\vec{m}$ are always orthogonal to each other. The pitch d is given by

$$d = \left\langle \vec{t}, \vec{l} \right\rangle \tag{A.8}$$

In the same manner that a rotation R can be represented by a quaternion $\mathbf{q}$, a rigid displacement T can be described using a dual quaternion $\hat{\mathbf{q}}$. A dual quaternion has the following form:

$$\hat{\mathbf{q}} = \mathbf{q} + \varepsilon \mathbf{q}' = \begin{pmatrix} q_0 \\ \vec{q} \end{pmatrix} + \varepsilon \begin{pmatrix} q_0 \\ \vec{q}' \end{pmatrix} \tag{A.9}$$

with $\mathbf{q}$ and $\mathbf{q}'$ being quaternions and ε being a scalar constant such that $\varepsilon^2 = 0$. $\mathbf{q}$ and $\mathbf{q}'$ are usually referred as the real and the dual components. The conjugate of a dual quaternion $\hat{\mathbf{q}}$ is defined as

$$\hat{\mathbf{q}}* = \mathbf{q}* + \varepsilon \mathbf{q}'* \tag{A.10}$$

The product of two dual quaternions $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$ is carried as follows:

$$\hat{\mathbf{a}}.\hat{\mathbf{b}} = (\mathbf{a}.\mathbf{b}) + \varepsilon(\mathbf{a}.\mathbf{b}' + \mathbf{a}'.\mathbf{b}) \tag{A.11}$$

To represent an SE(3) element, $\hat{\mathbf{q}}$ has to be a unit dual quaternion:

$$\hat{\mathbf{q}}* .\hat{\mathbf{q}} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \tag{A.12}$$

In other terms, $\mathbf{q}$ has to be a unit quaternion and has to verify an orthogonality condition with $\mathbf{q}'$:

$$q_0 q'_0 + \left\langle \vec{q}, \vec{q}' \right\rangle = 0, \tag{A.13}$$

Consider the motion T and its screw representation discussed earlier. T can be represented by a dual quaternion $\hat{\mathbf{q}}$, where the real component $\mathbf{q}$ is the quaternion corresponding to the rotation R and where the dual part $\mathbf{q}'$ is

$$\mathbf{q}' = \begin{pmatrix} -\frac{d}{2}\sin\left(\frac{\theta}{2}\right) \\ \sin\left(\frac{\theta}{2}\right)\vec{m} + \frac{d}{2}\cos\left(\frac{\theta}{2}\right)\vec{l} \end{pmatrix} \tag{A.14}$$

Let A and B be two rigid transformations. The dual quaternion representation of T = AB is $\hat{\mathbf{q}} = \hat{\mathbf{a}}.\hat{\mathbf{b}}$.

Let $\hat{\mathbf{q}}$ be a dual quaternion representing a rigid motion. Because the real part $\mathbf{q}$ is the quaternion encoding the rotation, recovering matrix R is trivial. To determine the translational component of the motion, the following relation involving the conjugate of $\mathbf{q}$ can be used:

$$\begin{pmatrix} 0 \\ \vec{t} \end{pmatrix} = 2\mathbf{q}'.\mathbf{q}^* \tag{A.15}$$

Conflict of Interest

The authors have stated explicitly that there are no conflicts of interest in connection with this article.

Funding

No specific funding.

References

1. Barreto JP, Roquette J, Sturm P, Fonseca F. Automatic calibration of medical endoscopes using a single image of a planar grid. In BMVC, 2009.

DOI: 10.1002/rcs

2. Wijesoma S, Wolfe D, Richards R. Eye-to-hand coordination for vision-guided robot control applications. *International Journal of Robotics Research* 1993; **12**: 65–78.
3. Wong A, Mayorga R, Rong A, Liang X. A vision based online motion planning of robot manipulators. In IROS, 1996; 940–948.
4. Nicolau S, Pennec X, Soler L, Ayache N. A complete augmented reality guidance system for liver punctures: first clinical evaluation. In MICCAI, 2005; 539–547.
5. Wengert C, Cattin P, Duff J, Szekely G. Markerless endoscopic registration and referencing. In Medical Image Computing and Computer-Assisted Intervention. In MICCAI, 2006; 816–823.
6. Tsai R, Lenz R. Real time versatile robotics hand/eye calibration using 3D machine vision. In ICRA, 1988; 554–561.
7. Horaud R, Dornaika F. Hand-eye calibration. *International Journal of Robotics Research* 1995; **14**(3): 195–210.
8. Triggs B, McLauchlan P, Hartley R, Fitzgibbon A. Bundle adjustment — a modern synthesis. In Proceedings of the International Workshop on Vision Algorithms: Theory and Practice, 2000.
9. Malti A, Barreto JP. Robust hand-eye calibration for computer aided medical endoscopy. In ICRA, 2010; 5543–5549.
10. Shiu YC, Ahmad S. Calibration of wrist-mounted robotic sensors by solving homogeneous transform equations of the form ax = xb. *IEEE Transactions on Robotics and Automation* 1989; **5**(1): 16–29.
11. Chou JCK, Kamel M. Quaternions approach to solve the kinematic equation of rotation of a sensor-mounted robotic manipulator. In ICRA, 1988; 656–662.
12. Chou JCK, Kamel M. Finding the position and orientation of a sensor on a robot manipulator using quaternions. *International Journal of Robotics Research* 1991; **10**(2): 240–254.
13. Fassi I, Legnani G. Hand to sensor calibration: a geometrical interpretation of the matrix equation ax = xb. *Journal of Robotic Systems* 2005; **22**: 497–506.
14. Park F, Martin B. Robot sensor calibration: solving ax = xb on the Euclidean group. *IEEE Transactions on Robotics and Automation* 1994; **10**: 771–721.
15. Zhuang H, Wang K, Roth ZS. Simultaneous robot/world and tool/flange calibration by solving homogeneous transformation of the form ax = yb. *IEEE Transactions on Robotics and Automation* 1994; **10**(4): 549–554.
16. Daniilidis K. Hand-eye calibration using dual quaternions. *International Journal of Robotics Research* 1998; **18**: 286–298.
17. Rivera-Rovelo J, Herold-Garcia S, Bayro-Corrochano E. Geometric hand-eye calibration for an endoscopic neurosurgery system. In ICRA, 2008; 1418–1423.
18. Schmidt J, Niemann H. Data selection for hand-eye calibration: a vector quantization approach. *International Journal of Robotics Research* 2008; **27**: 1027–1053.
19. Dornaika F, Horaud R. Simultaneous robot-world and hand-eye calibration. *IEEE Transactions on Robotics and Automation* 1998; **14**(4): 617–622.
20. Strobl KH, Hirzinger G. Optimal hand-eye calibration. In IROS, 2006; 4647–4653.
21. Strobl KH, Sepp W, Hirzinger G. On the issue of camera calibration with narrow angular field of view. In IROS, 2009; 309–315.
22. Shahand MI, Eastman RD, Hong TH. An overview of robot-sensor calibration methods for evaluation of perception systems. In Performance Metrics for Intelligent Systems Workshop, 2012.
23. Golub GH, van Loan CF. *Matrix Computations* (3rd edn). Johns Hopkins University Press: Baltimore, USA, 1996.
24. Fitzgibbon AW. Simultaneous linear estimation of multiple view geometry and lens distortion. In CVPR, 2001; 125–132.
25. Heikkila J, Silven O. A four-step camera calibration procedure with implicit image correction. In CVPR, 1997; 1106–1112.
26. Triggs B, Mclauchlan P, Hartley R, Fitzgibbon A. Bundle adjustment – a modern synthesis. In Vision Algorithms: Theory and Practice, 2000.
27. Mouragnon E, Lhuillier M, Dhome M, Dekeyser F, Sayd P. 3D reconstruction of complex structures with bundle adjustment: an incremental approach. In ICRA, 2006; 3055–3061.
28. Hartley R, Kang S. Parameter-free radial distortion correction with center of distortion estimation. *IEEE Trans Pattern Anal Mach Intell* 2007; **29**(8): 1309–1321.
29. Lourakis M, Argyros A. The design and implementation of a generic sparse bundle adjustment software package based on the Levenberg-Marquardt algorithm. Tech. Rep. 340, Institute of Computer Science - FORTH, Heraklion, Crete, Greece. 2004. Available from http://www.ics.forth.gr/~lourakis/sba
30. Bouguet JY. Camera Calibration Toolbox for Matlab. 2003.
31. Kato H, Billinghurst M. Marker tracking and hmd calibration for a video-based augmented reality conferencing system. In Proceedings of the 2nd IEEE and ACM International Workshop on Augmented Reality, IWAR '99, 1999; 85–90.
32. Lowe DG. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* 2004; **60**(2): 91–110.
33. Hartley R, Sturm P. Triangulation. *Computer Vision and Image Understanding* 1997; **68**(2): 146–157.
34. Kuipers JB. *Quaternions and Rotation Sequences: A Primer with Applications to Orbits, Aerospace and Virtual Reality*. Princeton University Press: Princeton, USA, 2002.
35. Siciliano B, Khatib O. *Handbook of Robotics*. Springer-Verlag: Berlin Heidelberg, 2008.

DOI: 10.1002/rcs