Machine Vision and Applications
DOI 10.1007/s00138-007-0068-0

ORIGINAL PAPER

Calibration of 3D kinematic systems using orthogonality constraints

Tomislav Pribanić · Peter Sturm · Mario Cifrek

Received: 27 July 2006 / Accepted: 9 December 2006

Abstract Processing images acquired by multi-camera systems is nowadays an effective and convenient way of performing 3D reconstruction. The basic output, i.e. the 3D location of points, can easily be further processed to also acquire information about additional kinematic data: velocity and acceleration. Hence, many such reconstruction systems are referred to as 3D kinematic systems and are very broadly used, among other tasks, for human motion analysis. A prerequisite for the actual reconstruction of the unknown points is the calibration of the multi-camera system. At present, many popular 3D kinematic systems offer so-called wand calibration, using a rigid bar with attached markers, which is from the end user's point of view preferred over many traditional methods. During this work a brief criticism on different calibration strategies is given and typical calibration approaches for 3D kinematic systems are explained. In addition, alternative ways of calibration are proposed, especially for the initialization stage. More specifically, the proposed methods rely not only on the enforcement of known distances between markers, but also on the orthogonality of two or three rigidly linked wands. Besides, the proposed ideas utilize common present calibration tools and shorten the typical calibration procedure. The obtained reconstruction accuracy is quite comparable with that obtained by commercial 3D kinematic systems.

Keywords Calibration · 3D kinematic system

T. Pribanić (✉) · M. Cifrek
Faculty of Electrical Engineering and Computing,
University of Zagreb, Unska 3, Zagreb 10000, Croatia
e-mail: tomislav.pribanic@fer.hr

M. Cifrek
e-mail: mario.cifrek@fer.hr

P. Sturm
INRIA Rhône-Alpes, 655 Avenue de l'Europe,
38330 Montbonnot St Martin, France
e-mail: peter.sturm@inria.fr

1 Introduction

Three-dimensional reconstruction of points is a basic task in a variety of areas and applications: entertainment, animation, industrial design, sports/medicine, etc. Different applications contributed to parallel developments of a variety of principles (and instrumentations) to obtain 3D information, including for instance the use of electromagnetic or acoustic sensors, accelerometers, photogrammetric principles [7,9,29], etc. Perhaps one of the major advantages of photogrammetric methods is the fact that they are practically 100% non-invasive, being based on processing images acquired by cameras. That feature is almost an imperative in many applications such as human motion analysis. The image created by a camera represents a 2D projection of a 3D object. Two such images are sufficient to yield 3D coordinates by the means of photogrammetric reconstruction [10]. However, there are cases where under certain assumptions a single camera is sufficient [19]. One of those special cases includes the use of structured light where the role of one camera is substituted by a light source, for instance a video-projector [15,18,36,37,39,41]. In this paper, we only consider multi-camera systems, without structured light or other additional devices.

Prior to the reconstruction of the unknown object points, the multi-camera system has to be calibrated. The projection of a 3D point to a 2D camera image plane is described by a camera projection model and parameters

thereof. Calibration is the process during which these parameters are determined [49]. Over the course of the years different methods/tools have been developed in order to make calibration as simple as possible and at the same time satisfying a high degree of reconstruction accuracy. Typically, this is based on acquiring images of a scene with known geometry. An easy way to do this is to build a special calibration object, such as the traditional 3D calibration cages. Such an approach, apart from the apparent advantages of being rather accurate and reliable, has many disadvantages that nowadays are less and less acceptable. For instance, accurate fabrication, manipulation and storage of 3D cages very often ask for considerable amounts of money, patience (sometimes even with no guarantee that calibration will be successful after all) and space, respectively. All those conditions are somewhat relaxed if we use planar calibration objects [51,44]. However, the accurate calibration of a multi-camera system requires that the planar object be well visible in several cameras at each instant; this puts too many requirements on the calibration procedure, especially for non-expert users, to be acceptable for use in commercial systems.

Taking one more step in simplifying things is to use 1D calibration objects, or wands [12]; this further alleviates the above problems and solves the mentioned visibility issue. There is an ongoing tendency to make calibration methods as simple as possible and use as few data as possible. These approaches heavily rely on (hopefully) readily available scene constraints and/or assumptions about camera parameters (spatial configurations) [25,14,32]. Finally, the most flexible calibration approach is autocalibration, which is based on acquiring images of an object without known geometry besides the assumption of being rigid [34,26]. One potential problem is image matching that is required for all calibration approaches; this may be solved as proposed in [46], where calibration is done using images of a single LED, moved throughout the working volume by the user. Nevertheless, although autocalibration is theoretically feasible and works well for moving cameras, it requires expert knowledge in order to apply it for a static system of multiple cameras, especially with respect to camera placement, due to the existence of degenerate conditions, which commonly occur in practice [45]. Overall, it is difficult to provide performance guarantees (accuracy and required time) for autocalibration, which makes it difficult to use in commercial systems.

In our work, we thus adhere to the use of wands, i.e. 1D calibration objects with attached markers, located at known distances from one another. In this paper, we investigate the use of two or three rigidly linked wands. The aims are to obtain both, a more accurate calibration and a shorter calibration procedure, due to requiring fewer images. We study how to use the constraint of orthogonality of wands, besides known distances between markers.

2 Typical 3D kinematic system calibration

The commercial value of nowadays' popular kinematic systems is highly dictated by how user-friendly they are, which is in turn largely influenced by the particular calibration method involved [8,3,6,1,4,2]. Presently, many commercially available systems are offering camera calibration using simple wand(s) of known length, for example [3]. A typical calibration procedure consists of two steps. The first step requires imaging two or three orthogonal wands, placed on the ground (Fig. 1). The position and orientation of these wands actually determine the spatial coordinate system in accordance with user requirements. The second purpose of this first calibration phase is the initialization of camera parameters. Each wand has a certain number of markers, whose relative positions are accurately known. Thus correspondences between 3D space and camera image planes can be readily established and used to initialize camera parameters (internal parameters as well as external ones, i.e. position and orientation). Due to the existence of various sources of errors or noise, primarily those caused by lens imperfections and non-perspective distortions, these initial estimates are only rough approximations [27].

The task of the second calibration step is to refine the initial estimates by waving with a wand of known length (so-called wand dance) throughout the desired calibration volume. The goal is to acquire images of the wand for as many locations and orientations as possible. This refinement leads usually to non-linear optimization algorithms [16]. A detailled analysis of issues concerned with non-linear optimization algorithms is not within a scope of this paper. However, a few general remarks can be given. Problems concerning local/global minima, convergence and speed of convergence, etc. are very much influenced by the quality of the initial estimates of parameters, especially by how close they are to the optimal values [38]. It is not rare in practice, particularly with older versions of 3D kinematic systems, that the user is explicitly asked to repeat the entire calibration or strongly suggested to do so, based on various quantitative measures of calibration quality. Common causes are a bad positioning of the orthogonal triad with respect to the cameras (leading to bad initial parameter estimates) and/or badly performed wand dance such that the wand does not cover a sufficient amount of the cameras' field

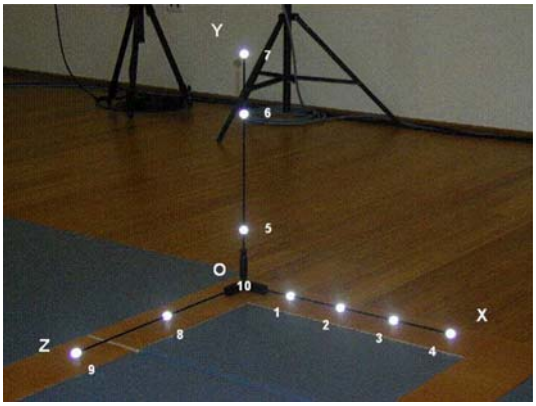

Fig. 1 Image of the orthogonal triad with attached markers. Marker distances with respect to the triad origin, i.e. 'marker' 10 are as follows. *X*-axis: *1* (15 cm), *2* (30 cm), *3* (45 cm), *4* (60 cm); *Y*-axis: *5* (15 cm), *6* (45 cm), *7* (60 cm); *Z*-axis: *8* (30 cm), *9* (60 cm)

Fig. 2 Single calibration wand used for 'wand dance'

of view (leading to an insufficient geometric redundancy and thus prohibiting the estimation of accurate and precise parameter estimates even during non-linear optimization, especially if the initial estimates are poor). It is thus crucial to obtain good initial estimates for camera parameters. This is the aim of the methods proposed and evaluated in this paper.

3 Background

In this section, we very briefly summarize some notions of projective and Euclidean geometry that are used in our calibration methods. A comprehensive treatment can be found in [42] as well as in [21,26] for more computer vision related descriptions.

Perspective projection is modeled using so-called intrinsic and extrinsic parameters, encapsulated in a 3×4 projection matrix:

$$\mathsf{P} = \mathsf{K}\mathsf{R}\left(\mathtt{I}_{3\times3} \; -\mathbf{t}\right)$$

The upper triangular 3×3 matrix K contains the intrinsic parameters, such as focal length, and the vector $\mathbf{t}$ and rotation matrix R constitute the camera's extrinsic parameters. The usual definition of calibration is the determination of K. In this paper, the goal is to compute the intrinsic parameters of all cameras in the multi-camera system, as well as their (relative) extrinsic parameters. This is done using images of wands; the available pieces of information are: matches between images (of markers on the wand), known distances between markers, orthogonality of wands (we use two or three rigidly attached, mutually orthogonal, wands). It is well known that from image matches alone, a projective reconstruction of the scene and the cameras is already possible [20,22,24]. The additional information mentioned above, as well as assumptions on the intrinsic parameters (e.g. zero skew) then allow to obtain a Euclidean reconstruction, i.e. to compute intrinsic and extrinsic camera parameters. Note that not all proposed methods follow this scheme, i.e. start with a projective reconstruction; this discussion is simply to show the feasibility of the problem.

In the following, we summarize a few ways of using the available information to obtain a Euclidean reconstruction. The 'absolute dual quadric' [47] is the only quadric invariant to Euclidean transformations; its determination in a projective reconstruction allows to upgrade the latter to a Euclidean reconstruction. One may use constraints on the scene, the camera intrinsic or extrinsic parameters to compute the absolute dual quadric. In this paper, we use constraints given by mutually perpendicular scene planes: two such planes Π_1 and Π_2 give the following linear equation on the matrix representing the dual absolute quadric Ω:

$$\Pi_1^\mathsf{T}\,\Omega\,\Pi_2 = 0$$

This constraint is used in method 2a below.

The upgrade from a projective to a Euclidean reconstruction can also be done via an intermediate affine reconstruction. An affine reconstruction is possible if the plane at infinity can be identified in a projective reconstruction. To do so, it is sufficient to determine three or more points at infinity. When using wands with three or more markers with known distances, their point at infinity can be computed using the cross-product. This way, a projective reconstruction of wands in three or more positions can be upgraded to affine. As for the additional upgrade to a Euclidean reconstruction, one may determine the 'absolute conic' on the plane at infinity. Like for the dual absolute quadric, constraints on the scene or the camera intrinsic or extrinsic parameters may be used. In method 2b below, we use constraints given by mutually orthogonal wands. Let $\mathbf{V}_1$ and $\mathbf{V}_2$ be the 2D coordinates of the points at infinity of two orthogonal wands. We then have a linear equation on the matrix representing the absolute conic:

$$\mathbf{V}_1^\mathsf{T}\,\Omega_\infty\,\mathbf{V}_2 = 0$$

The information about orthogonal wands can also be used differently. Instead of determining the absolute conic in a projective reconstruction, one may aim at computing its image, the so-called IAC (image of the absolute conic). Contrary to the actual absolute conic, this encodes only intrinsic parameters. The IAC is given by $\omega = \mathsf{K}^{-\mathsf{T}}\mathsf{K}^{-1}$. Let $\mathbf{v}_1$ and $\mathbf{v}_2$ be the vanishing points of two orthogonal wands, i.e. the projections of their points at infinity. We then have the following linear equation on ω:

$$\mathbf{v}_1^{\mathsf{T}}\omega\mathbf{v}_2 = 0$$

This is used in method 1a below.

4 Proposed methods

4.1 Motivation

One of our main goals is to simplify the current typical calibration procedure for 3D kinematic systems (cf. previous section). We also aim at proposing alternative methods for camera parameter initialization and optimization, that should be at least as accurate as current ones. It is imperative to use the commonly used calibration tools (Figs. 1, 2), in order to ensure a relatively easy incorporation into 3D kinematic systems and to confront the user with as small changes as possible in the calibration procedure.

In order to shorten and simplify the calibration procedure it is desirable to transform the two typical calibration phases (initialization and wand dance) into a single one, i.e. to start right away with the wand dance. Theoretically this is possible, using autocalibration (see above). Autocalibration algorithms typically proceed by first estimating a projective 3D reconstruction. This is then upgraded to a Euclidean reconstruction by estimating an appropriate 3D homography H. Unfortunately, practical implementations of several autocalibration algorithms did not give satisfactory results in the presence of real, noisy data. In the following paragraphs we will describe a few experiences regarding this.

We for example experimented with the Kruppa equations approach, using the assumptions of square pixels, zero skew and known principal point [25]. Hence, only the focal lengths were searched for. A few other similar approaches that also only compute focal lengths, were implemented too, e.g. [12]. All tested algorithms have proven to be extremely sensitive to nonlinear lens distortions, which is, generally speaking, true for most autocalibration algorithms. When preprocessing the images for distortion removal, the algorithms gave much better results. However, in practice this would mean that an additional calibration phase is required, for the compensation of lens distortion.[1]

A typical approach for the off-line calibration of distortion parameters relies on imaging straight line patterns (plumbline calibration). In distortion-free images, the imaged patterns should be straight, whereas they appear curved in distorted images. It is possible to estimate distortion parameters such that an associated distortion correction brings the curved patterns back to straight ones [13,17]. When working with fixed camera lenses and/or when not using the cameras' zoom, such an off-line calibration may have to be done only once per camera. This might thus be an acceptable solution in practice. However, in our case the idea of the proposed methods was to make them robust and flexible, applicable when the user wants to change lenses or to use the cameras' zoom. In that case, the potential nonlinear distortion calibration has to be done transparently for the user.

There exist methods for compensating for nonlinear distortion, solely based on image point correspondences between two cameras [23,50]. One of their (current) disadvantages is that the two cameras are supposed to have the same distortion parameters, which limits their applicability for calibrating cameras of a 3D kinematic system. Further, these methods are currently limited to two cameras at a time and they only give good results if the distortions are significant. With moderate distortions though, results are weak. All this limits the applicability of these methods in our case.

Finally, after some other negative experiences, we ruled out a pure autocalibration concept for our system. We also renounced on traditional calibration approaches using 3D calibration cages. This was mainly due to the requirement that the proposed method should be at least equally as practical and user-friendly as with existing systems. We thus opted for the third approach mentioned in the introduction, which in terms of (dis)advantages lies somewhere in between the first two mentioned (autocalibration and traditional). It uses certain properties of the scene such as orthogonality or parallelism of lines, known distance ratios, etc. [14,25,32]. It does not require accurate but cumbersome 3D structure, but can actually work with calibration tools such as those shown in Figs. 1 and 2. The equipment is lighter than for traditional calibration approaches and offers a much better visibility: markers are visible from all sides, which is essential for a stable estimation of the camera poses. Further, due to using geometrical information on the scene, i.e. the calibration tool, typical disadvantages of autocalibration

[1] Some earlier versions of [2] relied on such a solution.

methods are avoided, such as critical relative poses and instability in their vicinity.

4.2 Definition of proposed methods

The idea of the proposed methods is to sweep the calibration volume, during the so-called wand dance, with two (or three) rigidly attached orthogonal wands (wand pair or triad) instead of a single one as usual. In fact, such an approach effectively eliminates the need to also undertake the usual calibration phase of placing an orthogonal wand triad on the floor and imaging it (Fig. 1), prior to the wand dance itself. The proposed approach will allow us to set up constraints on some of the mentioned geometric entities, compute them and ultimately obtain the camera parameters from them. Within a same calibration phase both, data for initial parameter computation and data for refinement are provided, and there is no need for an extra calibration phase and/or assumptions on initial estimates (except maybe of the skew factor being zero, depending on the quality of the cameras). In the following, four different calibration methods are defined and investigated, that exploit the described input in different ways.

Method 1a

1. Perform the wand dance using two orthogonal wands, each having at least two markers whose relative distances are known. Note that besides the actual markers, we also extract the origin of the wands (marker 10 in Fig. 1) by intersecting lines fitted to the actual markers. Hence, for each wand, we have three or more markers, with known relative distances.
2. For every frame of every camera, find the vanishing points of the wands' directions from the known distance ratios between the markers (using the cross-product, as mentioned above).
3. Use the constraint that the wands are perpendicular to one another to form a linear equation system on the image of the absolute conic (IAC) ω, built from equations of the following type ($\mathbf{v}_1$ and $\mathbf{v}_2$ are vanishing points associated with two perpendicular wands):

$$\mathbf{v}_1^{\mathsf{T}} \omega \mathbf{v}_2 = 0 \tag{1}$$

4. Compute a least-squares solution for the above overdetermined equation system and apply the Cholesky decomposition on the computed ω to retrieve the camera's internal parameters.
5. Repeat steps 2 to 4 for each camera.
6. Choose one camera as reference, form pairs with it and all other cameras. Carry out steps 6 to 9 for each such pair.
7. Compute the fundamental matrix using all marker positions that were extracted in synchronous frames.
8. Knowing the internal camera parameters and fundamental matrix, compute the essential matrix and decompose it to obtain the second camera's pose relative to the reference camera (external parameters). At this stage, the translational part of the pose is only known up to a scale.
9. Perform a metric 3D reconstruction of marker positions using the known internal and external camera parameters. Compute the scaling for upgrading the metric to a Euclidean reconstruction and pose, as the ratio of the known wand lengths and the average of those computed from the 3D reconstruction. Then upgrade the 3D reconstruction and camera pose.
10. Refine the above calculated initial camera parameters enforcing the known wands lengths and/or orthogonality of wands (see Sect. 4.3).

Method 1b

1. Perform the wand dance using two orthogonal wands, each with at least two markers. The relative positions of the markers of both wands are assumed known.
2. For every frame of every camera, compute the homography H between the plane in space (formed by the two wands) and the image plane.
3. Compute the images of the two so-called circular points: $\mathbf{h}_1 \pm I\mathbf{h}_2$ where $\mathbf{h}_1$ and $\mathbf{h}_2$ are the first two columns of H. Both of these points lie on the IAC ω, i.e.:

$$(\mathbf{h}_1 \pm I\mathbf{h}_2)^{\mathsf{T}} \omega \, (\mathbf{h}_1 \pm I\mathbf{h}_2) = 0 \tag{2}$$

The real and imaginary parts of the two equations (one for '+' and one for '−') are identical (up to sign for the imaginary parts) and correspond to the following two linear equations on ω:

$$\mathbf{h}_1^{\mathsf{T}} \omega \mathbf{h}_1 - \mathbf{h}_2^{\mathsf{T}} \omega \mathbf{h}_2 = 0 \tag{3}$$

$$\mathbf{h}_1^{\mathsf{T}} \omega \mathbf{h}_2 = 0 \tag{4}$$

4. Solve the overdetermined linear equation system composed of equations of type (3) and (4) and apply the Cholesky decomposition on the computed ω in order to retrieve the camera's internal parameters.

This approach is effectively using the plane-based calibration method developed in [44,51].
5. Repeat steps 2 to 4 for every camera.
6. Follow steps 6 to 10 of Method 1a.

Method 2a

1. Perform the wand dance using the three orthogonal wands, each having at least two markers with known relative distances.
2. Choose one camera as reference, form pairs with it and all other cameras. Perform steps 3 to 6 for all such pairs.
3. Compute the fundamental matrix F and perform a projective 3D reconstruction. To do so, compute the pair of canonical projection matrices P_1 and P_2 from F:

$$\mathsf{P}_1 = (\mathtt{I}|\mathbf{0}) \qquad \mathsf{P}_2 = \left([\mathbf{e}_2]_\times + \mathsf{F} + \mathbf{e}_2\mathbf{v}^\mathsf{T}|\lambda\mathbf{e}_2\right) \tag{5}$$

Here, $\mathbf{e}_2$ is the epipole of the second camera and $\mathbf{v}$ and λ are an arbitrary vector and scalar, respectively. Using these projection matrices, compute a projective 3D reconstruction of the wand dance. This differs from the true one by an unknown projective transformation $\mathsf{H}_{4\times4}$.
4. For each time instant of the wand dance, compute the coordinates of the three planes formed by pairs of wands of the orthogonal triad. The perpendicularity of two planes Π_1 and Π_2 gives linear constraints on the absolute dual quadric Ω:

$$\Pi_1^\mathsf{T}\,\Omega\,\Pi_2 = 0 \tag{6}$$

5. Solve the system of above equations, i.e. compute the absolute dual quadric Ω. Find a projective transformation $\mathsf{H}_{4\times4}$ that will put the computed Ω, back into its canonical position, i.e. where it has the matrix representation[2]

$$\begin{pmatrix} \mathtt{I}_{3\times3} & \mathbf{0} \\ \mathbf{0}^\mathsf{T} & 0 \end{pmatrix}$$

Apply the same $\mathsf{H}_{4\times4}$ on the 3D projective reconstruction of the wand dance to obtain a metric reconstruction of it. Similarly, apply the transformation $\mathsf{H}_{4\times4}$ on the initially computed canonical projection matrices to obtain projection matrices coherent with the metric 3D reconstruction.
6. Use the information about true wand lengths to compute the scale factor needed to transform the metric reconstruction to Euclidean (step 9 of method 1a).
7. Follow step 10 described in method 1a.

Method 2b

1. Perform the wand dance using two orthogonal wands, each having at least two markers with known relative distances.
2. For every frame of one camera find the vanishing points of the wand directions using the known distance ratios of markers.
3. Choose one camera as reference, form pairs with it and all other cameras. Perform steps 4 to 8 for all such pairs.
4. Compute the fundamental matrix F and perform a projective 3D reconstruction, as in step 3 of method 2a. The 3D reconstruction includes the points at infinity associated with the wands' directions, reconstructed based on the wands' vanishing points extracted in the images.
5. Compute the plane at infinity Π_∞ by fitting it to the reconstructed points at infinity and find the homography $\mathsf{H}_{4\times4}$ that maps it to its canonical position, i.e. to coordinates $(0,0,0,1)^\mathsf{T}$. Apply H to the reconstructed marker positions, to obtain an affine reconstruction of them.
6. Consider the fact that points at infinity are effectively representing directions of lines, in our case orthogonal lines (the two wands at each instant of the wand dance). The perpendicularity of wands allows to form constraints on the absolute conic Ω_∞:

$$\mathbf{V}_1^\mathsf{T}\,\Omega_\infty\,\mathbf{V}_2 = 0 \tag{7}$$

Here, $\mathbf{V}_1$ and $\mathbf{V}_2$ are 2D coordinates of points at infinity and Ω_∞ is a 2D conic, all lying on the plane at infinity.
7. Solve the overdetermined system of above equations, i.e. compute the absolute conic Ω_∞. Find an affine transformation $\mathsf{H}_{4\times4}$ that will put the computed Ω_∞, back into its canonical position, i.e. where it is represented by the identity matrix. Apply the same H on the 3D affine reconstruction of the wand dance which is in turn equivalent to obtaining a metric reconstruction of it. Similarly, apply both transformation matrices, from projective to affine and from affine to metric, on the initially computed canonical projective matrices to obtain projection matrices that correspond to a metric 3D reconstruction.

[2] The transformation $\mathsf{H}_{4\times4}$ is not unique of course, since concatenating it with any metric transformation will still give a valid solution.

8. Use the information about true wand lengths to compute the scale factor needed to transform the metric reconstruction to Euclidean (step 9 of method 1a).
9. Follow step 10 described in method 1a.

One of the common details of all four proposed methods is the last step, i.e. the refinement of parameters. That fine tuning was performed using the bundle adjustment (BA) approach, as described in the following section.

4.3 Bundle adjustment

Bundle adjustment essentially means the simultaneous optimization of the unknowns in a 3D reconstruction or any other structure-from-motion problem [48]. This comes usually down to a non-linear optimization, with a cost function based on the so-called reprojection error (geometric distance between measured image points and image points generated by projecting current estimates of 3D points via current estimates of projection matrices). Concretely, if measurement errors on image point coordinates are identically and independently distributed according to a normal, or Gaussian, distribution, then minimizing the sum of squared reprojection errors gives the maximum likelihood solution. Note that this optimization problem may have many local minima, so recently researchers have started to develop methods for global solutions, using modified cost functions, for certain classes of structure-from-motion problems [30,31].

Bundle adjustment is usually formulated for reconstructing individual points, but if geometric constraints on them are available, they should be used. There exist approaches for incorporating various types of geometric constraints, such as known distances between points (often together with an estimate of the uncertainty of that knowledge), collinearity of points, perpendicularity of lines, etc. [33,35,43]. These usually proceed by adding terms to the cost function, that express the amount of deviation of the estimates from the given constraints. Another approach is to use the given constraints to reduce the number of unknowns to optimize. The main drawback of such an approach is that constraints are supposed to hold exactly, whereas in the first approach, an uncertainty on them can be incorporated. On the other hand, the first approach requires tuning parameters, regulating the relative influence of the different types of expressions in the cost function (reprojection error, deviation from a given distance or a given angle, etc.). Further, the second approach, due the reduction of the number of unknowns and the smaller number of terms in the cost function, allows for more time and memory efficient computations.

In our application, we have rather precise geometric constraints, since purpose-made wands are employed; constraints are the known distances between markers and the orthogonality of wands. Instead of using these constraints individually, we encode them simultaneously: if besides orthogonality of wands we also know their positional offsets, then we can encode all marker positions in a single 3D coordinate system. By definition, the encoded points will satisfy all constraints. Instead of optimizing the coordinates of all markers, it is now sufficient to optimize euclidean transformations: for each instant of the wand dance, one euclidean transformation represents the positions of all markers on the two or three wands used.

Bundle adjustment thus comes down to optimizing: internal camera parameters (including radial distortion), relative pose of all cameras, one euclidean transformation per time instant that represents the pose of the set of wands. Although smaller than at the outset, the number of unknowns is still rather huge, being larger than six times the number of frames (time instants), i.e. of the order of 10,000 for a 1-min wand dance and a fram rate of 25 frames per second. It would be possible to reduce this by sub-sampling the number of frames or by partitioning data into several sets and merging those later, alternating minimizing reprojection error by varying the cameras with minimizing reprojection error by varying the points [48]. However, these mentioned methods have the common weakness of not actually using all available data simultaneously.

We thus use a typical remedy used in bundle adjustment, namely to benefit from the sparse structure of the cost function's Jacobian matrix [26,43]. In our case, the euclidean transformations associated with different time instants, are not directly coupled in the cost function and the associated entries in the Jacobian matrix are zero. We straightforwardly follow the recipes given in [26,43] to implement a sparse version of Levenberg-Marquardt (LM) optimization [38]. In our case, each iteration of LM requires, among others, the inversion of n symmetric 6×6 matrices (n being the number of frames) and that of one symmetric $m \times m$ matrix (m being the number of optimized camera parameters: 6 extrinsic and 1–4 intrinsic parameters per camera). The most costly computation is the latter inversion; importantly, its complexity is independent of the number of frames used. The complexity of the other computations is linear in the number of frames.

Besides the actual optimization, we have to discuss the initialization of the unknowns. Part of this is described in the above methods: they all compute

Euclidean camera projection matrices in some common coordinate system, from which initial intrinsic and extrinsic parameters can be extracted. As for the remaining unknowns, the Euclidean transformations representing the pose of wands pairs or triads, they can be computed easily from the reconstructed 3D marker positions. One can for example use the absolute orientation procedure of [28], to compute an optimal estimate of the translation and rotation between the reconstructed markers and the ones in the model of the wand pair (triad). More details on the actual procedure we use are given in [40].

5 Equipment and methods

In order to test the proposed methods, a popular commercially available system was used, Smart from the eMotion company [3]. The system version used (version 1.10, Build 2.39) consists of nine cameras (50 Hz). It is a so-called opto-electronic system which actually reconstructs positions of passive retro-reflective markers, attached to the subject's points of interest. Markers are illuminated by stroboscopic IR sources of light attached to the cameras. The Smart system used is installed in the Biomechanic laboratory of Peharec Polyclinic in Pula, Croatia [5].

The first experiment consisted of a typical system calibration as proposed by the system manufacturer. An orthogonal triad of wands (60 cm long each) was positioned on the floor. Each wand defined one of the world coordinate axes and had a certain number of retro-reflective markers on it. The vertical axis has three markers and the two horizontal axes 4 and 2, respectively (cf. Fig. 1). The relative positions of markers are accurately known. A visual check was performed that each camera 'sees' all triad wands, i.e. the markers on it and image acquisition was undertaken for a few seconds. Afterwards, as apart of the second step, the orthogonal triad was removed and a wand dance with a single wand was performed for another couple of minutes. The entire procedure was carried out by trained polyclinic staff to ensure calibration results would not be impaired by the user's inexperience. Finally, Smart's software routines were run to compute camera parameters based on the acquired images from these two calibration steps.

The second experiment used our approaches where the wand dance started right away with the orthogonal pair or triad of axes. It lasted roughly 60 s (half of the time proposed by Smart for its wand dance with a single wand). At the end the wand pair/triad was simply put on the floor to set the origin of the coordinate system of the working volume, according to user preferences, if needed (sometimes the user does not care where the spatial coordinate system origin is). The working volume was approximately 3.2 m × 2.2 m × 2.0 m, in both experiments.

Among a rather comprehensive analyzing software Smart also has the capability to export/import various data into/from Matlab: 2D image data of marker centroids from the acquired sequences, 3D reconstructed marker positions, camera projection matrices, etc. The following data were exported for further analysis: 2D image data of markers on the wands when putting them on the floor which otherwise serves for camera parameter initialization in case of the Smart calibration procedure. Then, camera projection matrices calculated by the Smart calibration procedure and finally the 2D image sequence of the wand dance with the two (three) orthogonal axes, used for the proposed calibration methods.

6 Evaluation of the proposed calibration methods

There are various ways of evaluating the accuracy of a calibration result. We mainly use a rather common principle in the evaluation of 3D kinematic systems which is easy to understand and very informative. We have chosen the so-called static test where a wand of known length is imaged and reconstructed for a large number of positions/orientations in space. Essentially, we process a separate data set, obtained for another wand dance than the one used for calibration. That wand dance also covers the entire calibration volume on contains over 3,000 positions/orientations. The absolute mean error between the true wand length (30 cm) and the reconstructed ones is used as accuracy measure.

We also considered some more complex evaluation methods such as dynamic testing where for instance some 'unknown' velocity and/or acceleration is computed and tested against expected values. However, those are frequently combined with static testing, as the one just explained, and serve usually as confirmation of conclusions reached otherwise based on static evaluation tests. Another common way of evaluating calibration (and implicitly then also reconstruction) quality is computing the image residual between measured and predicted image point coordinates. Nevertheless, we have refrained from such an evaluation. We have experienced that this approach is not necessarily consistent with ground truth values (not only in case of our proposed methods). In other words, calibrations giving smaller image coordinate residuals do not always yield smaller 3D reconstruction errors.

The results shown below are samples from an extensive number of experiments with the Smart and

proposed calibration methods. We thus consider it reasonable to assume that the shown results are representative enough so that we can draw general conclusions from them. Nevertheless, one should always carry in mind that practically all calibration methods are subject to various individual factors, for instance (in)experience, time and dedication devoted to the calibration procedure and so forth. As mentioned, we have tried to eliminate the impact of such and similar influences so that we can justifiably regard the presented accuracy figures as representative.

6.1 Results and discussion: initialization part

Let us first consider results on the initial values of internal camera parameters, before bundle adjustment (Tables 1, 2, 3, 4, 5, 6). Table 2 shows internal parameters after refinement (bundle adjustment), as output by the SMART system. These are taken as reference for this evaluation. Although they are not perfectly accurate, they can be considered in the present context as close to the true ones, i.e. the values the optimization should converge to. Hence, the initialization method which produces parameters closest to those of Table 2 can be, generally speaking, considered as the method of choice: the closer the initial values to the final optimal ones, the higher the chance of avoiding convergence problems in the bundle adjustment. A close comparison clearly indicates that method 1a (Table 3) provides the best results. It also seems that the other three proposed methods (Tables 4, 5, 6) provide about equally good initial values for the intrinsic parameters. Furthermore, all proposed methods give better results than SMART's initialization method (Table 1).

Table 1 Internal camera parameters

Camera	f (pixels)		Skew	pp (pixels)	
	x	y		x	y
1	751.9	387.0	12.7	400.0	175.3
2	755.6	381.0	18.3	330.3	233.5
3	676.6	343.3	7.4	272.3	200.5
4	704.1	367.4	6.4	320.7	171.3
5	765.2	399.0	5.0	335.9	170.2
6	746.8	386.4	19.0	376.9	215.3
7	691.0	354.6	9.2	285.3	164.5
8	672.7	343.3	14.4	297.2	142.7
9	672.8	348.2	10.3	292.1	156.2

Initial values given by the SMART system

Table 2 Internal camera parameters

Camera	f (pixels)		Skew	pp (pixels)	
	x	y		x	y
1	727.7	375.9	0.0	349.4	153.7
2	723.8	374.6	0.0	304.4	145.4
3	723.9	375.2	0.0	290.3	138.4
4	724.3	374.7	0.0	325.4	140.0
5	724.2	375.1	0.0	347.7	137.0
6	724.8	375.0	0.0	349.9	143.4
7	719.2	371.8	0.0	328.9	134.7
8	730.1	377.2	0.0	350.7	133.4
9	715.9	370.5	0.0	345.0	138.6

Final optimized values given by the SMART system

Table 3 Internal camera parameters

Camera	f (pixels)		Skew	pp (pixels)	
	x	y		x	y
1	724.82	376.17	0.02	376.90	127.97
2	749.74	390.04	0.67	321.94	135.95
3	743.38	390.71	−0.49	267.50	132.06
4	728.30	376.46	−0.25	346.85	133.46
5	715.60	373.75	−1.46	332.82	145.18
6	734.55	380.96	0.17	381.45	136.57
7	715.34	372.93	−2.26	274.41	126.25
8	704.35	370.23	0.87	307.96	122.16
9	723.27	373.40	0.48	341.93	136.22

Initial values given by method 1a

Table 4 Internal camera parameters

Camera	f (pixels)		Skew	pp (pixels)	
	x	y		x	y
1	683.91	354.36	0.0	379.76	128.79
2	689.64	357.33	0.0	322.63	135.05
3	669.16	346.71	0.0	260.05	138.78
4	699.19	362.27	0.0	343.82	138.23
5	695.17	360.19	0.0	368.39	141.00
6	709.06	367.39	0.0	391.62	140.19
7	640.20	331.71	0.0	312.66	172.37
8	688.78	356.88	0.0	362.46	131.42
9	676.49	350.52	0.0	357.15	143.89

Initial values given by method 1b

Table 5 Internal camera parameters

Camera	f (pixels)		Skew	pp (pixels)	
	x	y		x	y
1	702.17	360.00	10.12	354.64	118.81
2	769.14	400.99	3.24	303.50	137.59
3	725.01	373.57	5.33	249.41	165.19
4	729.18	376.74	1.24	359.99	132.02
5	792.97	406.88	20.98	331.10	123.50
6	709.51	371.12	16.55	382.62	139.25
7	717.49	371.11	0.81	229.15	122.72
8	747.16	386.62	5.24	327.04	128.71
9	805.36	415.67	14.58	309.63	136.89

Initial values given by method 2a

Table 6 Internal camera parameters

Camera	f (pixels)		Skew	pp (pixels)	
	x	y		x	y
1	681.09	351.90	1.36	360.18	138.47
2	626.55	323.14	2.45	376.80	151.26
3	728.38	383.59	7.58	316.53	140.14
4	784.64	408.48	8.18	245.79	154.05
5	704.46	364.10	4.46	371.56	132.72
6	695.91	363.57	1.83	360.04	125.76
7	721.62	381.06	16.88	317.97	80.85
8	692.93	362.10	8.85	347.91	106.75
9	780.96	407.29	5.95	226.52	136.23

Initial values given by method 2b

Table 7 Mean error (mm) between reconstructed and true wand lengths

Pair	SI	SF	1a	1b	2a	2b
12	29.26	9.54	14.27	11.99	6.68	32.43
13	15.97	7.74	10.36	7.33	13.05	5.16
14	14.59	6.54	4.53	7.84	10.37	21.68
15	14.01	6.17	5.23	8.03	25.84	4.71
16	19.83	8.81	6.81	6.74	9.43	6.87
17	12.17	10.74	4.20	13.04	17.02	11.17
18	14.03	6.61	5.43	7.57	*795.66*	6.65
19	13.53	9.45	5.31	10.02	36.37	*670.53*
23	20.79	6.20	7.73	9.06	15.92	15.85
24	17.36	4.48	4.79	6.72	4.51	15.03
25	18.26	4.38	4.50	8.37	23.46	24.19
26	25.11	6.14	8.75	7.18	5.99	12.08
27	15.45	8.62	4.58	16.80	23.55	15.71
28	17.12	4.47	6.44	7.75	15.20	*135.56*
29	35.36	6.55	5.37	8.06	*634.62*	16.68
34	16.20	4.96	3.61	7.20	4.26	9.95
35	38.67	13.40	5.73	11.71	*210.52*	9.99
36	23.71	6.45	7.72	6.17	12.41	6.54
37	15.29	11.20	11.11	52.41	7.46	6.63
38	15.77	5.43	6.22	9.82	13.50	5.15
39	13.69	5.60	3.65	8.50	33.26	11.22
45	15.62	4.29	5.10	5.94	7.98	13.92
46	24.66	5.86	5.59	8.21	10.66	15.64
47	14.28	7.17	3.57	15.72	4.27	29.03
48	17.74	5.85	6.02	6.67	4.12	19.29
49	12.42	4.68	3.47	7.46	8.26	19.98
56	23.89	5.44	5.15	8.26	10.17	4.32
57	19.00	11.97	5.38	14.68	*101.77*	17.02
58	14.70	4.45	6.10	7.04	13.18	12.00
59	9.56	4.34	6.67	10.08	9.04	16.56
67	20.03	8.63	4.83	9.16	10.40	17.98
68	23.76	6.93	5.76	7.28	12.03	5.23
69	20.07	5.51	4.21	9.17	31.55	19.32
78	14.42	7.85	5.95	11.12	11.29	10.51
79	11.23	7.86	3.68	9.24	12.22	13.07
89	15.20	4.54	11.80	7.35	6.45	17.16
Avg.	18.41	6.91	6.10	10.27	60.35 (19.41)	35.41 (13.79)

SI using Smart's initial data; *SF* using Smart's optimized data
Values in the last row show the average computed only over values not given in italics

The superiority of method 1a is further demonstrated through the 3D reconstruction accuracy, cf. Table 7. Here, the accuracy is evaluated for all 36 possible pairs among the nine cameras used. Table 7 shows that the reconstruction accuracy for method 1a is roughly identical to that obtained using Smart's refined projection matrices. Remind that the latter were obtained after bundle adjustment, whereas the results for the other methods are given before bundle adjustment.

The reconstruction results of method 1b are not as good as those of method 1a, however they are still better than the results provided using Smart's initial parameter values. Methods 2a and 2b perform the worst among the proposed methods, in two respects. First, for some camera pairs we end up with an unacceptable reconstruction accuracy (values in *italics* in Table 7). Second, even when considering the average results after leaving out these 'outliers', they are worse than those for methods 1a and 1b.

In the following, we discuss on potential reasons for the different performances of the proposed methods and also on some heuristic measures we took to improve performance.

Method 1b Method 1b requires the computation of homographies between planes in space, defined by a pair of wands, and the image plane. This needs a minimum of four pairs of point matches, with no three being collinear. In general, the larger the number of point matches and the more they are distributed, the more stable the computation of homographies. Unfortunately, in our case neither a large number of point pairs nor a favorable distribution of them is available. Consequently, the steps following the computation of the homography may be negatively impacted. These issues come particularly into effect in those frames where the angle between wands, measured in the image plane, is small. We have found empirically that it proves to be beneficial to exclude all those frames where that angle is smaller than 80°. This rather strict threshold causes that we end up with few usable calibration data (redundancy) to form the final system of equations (2) for IAC computation.

Another issue we studied is the influence of assuming standard values for the skew factor (being zero) and the aspect ratio (being one). These assumptions are rather justified given the quality of the cameras used. At first we have used so-called soft constraints, by adding two more equations, measuring the difference between estimated and assumed values, to the otherwise largely overdetermined system of equations. As intuitively expected, soft constraining did not bring any significant improvement,

since adding two more equations to the overdetermined system of several thousands of equations (solved by the least squares method) can hardly do much difference. Alternatively, one could assign different weights to the added equations, but that has not been tried out yet. The other option is to use hard constraints, i.e. to exactly enforce the assumed values. This brought an improvement in accuracy; the results for method 1b shown in Table 4 were obtained this way (which is why the skew factors are exactly equal to zero).

Method 1a For this method, a good distribution of points in the image plane is considerably easier achievable than with method 1b, since vanishing points are computed for individual wands. Nevertheless, even here we face at least two potential problems: the possibility of the wand being (almost) perpendicular or parallel to the image plane. The first occurrence is relatively easy to handle by simply discarding all frames where the distance between the wand's end markers is less than some threshold. Similarly, the second problem could be taken care of by neglecting all frames where the distance of at least one vanishing point from the image center is larger than some value. However, we have undertaken another strategy: computations start by first using vanishing points close to the image center and gradually including further and further vanishing points. For each set of vanishing point data we obtain one set of internal camera parameters, and we have to decide which one to choose. Let us recall the fact that working with good quality cameras it is reasonable to assume a zero value for the skew parameter. Therefore we choose the set of internal parameters that has the smallest computed skew parameter.

Methods 2a and 2b Both methods are based on a projective reconstruction. The approaches for upgrading this to a metric reconstruction are unhappily not invariant to the choice of the initial projective reconstruction, i.e. to the values for $\mathbf{v}$ and λ (cf. Eq. 5). In practice we found that different values indeed gave final Euclidean reconstructions with rather different accuracies. Besides this, we also face the issue of critical motions, i.e. relative camera poses that lead to unstable self-calibration [45]. In order to alleviate these issues, we proceed the following way. Projective reconstruction and self-calibration are performed for every possible camera pair. For each camera, we select the pair involving it, that gives a Euclidean reconstruction with reconstructed wand lengths closest in average to the true ones. The projection matrix associated with that pair is used to compute the internal parameters for the considered camera. For instance, for the first camera and method 2b, camera pair 15 was considered (last column, fifth row in Table 7).

In the following, we summarize some more experiences/discussions. In previous paragraphs, it has been demonstrated that the application of certain thresholds or strategies of exploiting the available data has a positive effect for methods 1a and 1b. Similar approaches have also been tried for methods 2a and 2b. For example, in the case of method 2a it might be beneficial to apply thresholds on the angles between lines and/or planes in space before the actual formation of equation systems to compute the absolute dual quadric. However since we work with projective reconstructions and angles are metric characteristics, such an approach is not recommended. As for method 2b, we apply the same scheme as for method 1a, concerning the gradually increasing amount of vanishing points used for the calibration.

In our system, non-perspective distortions are estimated during bundle adjustment but not by the initialization methods 1a to 2b. Since radial distortion is most prominent towards the image borders, one might only use markers extracted within a certain distance from the image center. However, the number of wand positions where this holds for several cameras at the same time, is limited, i.e. the amount of the available data that will actually be used might be too small for an accurate calibration.

Methods 2a and 2b have in common that they calibrate two cameras simultaneously, indirectly via establishing a 3D Euclidean reconstruction. The quality of the Euclidean reconstruction depends on the cameras' spatial configuration (angle between optical axes, ratio of baseline and point depths, etc.), but also, as explained above, on the coordinate frame of the initial, projective, reconstruction. This may explain the bad results for some camera pairs, cf. Table 7.

For completeness, let us say that all proposed methods work equally well with noise-free synthetic data, outputting perfectly accurate results in terms of camera parameters and reconstructions. This simply proves the correctness of their theoretical foundations and of their implementation.

6.2 Results and discussion: optimization part

Table 8 shows mean errors for the proposed methods, after optimization with bundle adjustment. Additionally, the last column shows the results of our bundle adjustment procedure, when being initialized with the Smart's initial set of parameters (as opposed to results of Smart's own optimization procedure). Given the available number of markers on two or three wands (depending on the method), the natural question arises whether to use all available data during the optimization or only a certain part of it. We thus have run bundle adjustments

Table 8 Results of the bundle adjustment as described in section 4.3, i.e. enforcing distances between markers as well as orthogonality between wands: mean error (mm) between true and reconstructed wand lengths. In case of using all markers, the considered true wand length was 30 cm, in the other cases 60 cm (distance between end markers of wands)

Wands	Markers used	1a	1b	2a	2b	SI
XYZ	All (9)	0.56	0.60	0.59	0.67	0.68
XY	4, 7, 10	0.48	0.49	0.50	0.59	0.59
	All (7)	0.59	0.59	0.68	0.91	0.78
XZ	4, 9, 10	0.42	0.42	0.47	0.51	0.61
	All (6)	0.74	0.69	0.74	0.92	0.84
YZ	7, 9, 10	0.67	0.69	0.70	0.66	0.84
	All (5)	1.31	1.26	1.29	1.09	1.32

Table 9 Results when only enforcing distances between markers in bundle adjustment: mean error (mm) between true and reconstructed wand lengths. In case of using all markers, the considered true wand length was 30 cm, in the other cases 60 cm (distance between end markers of wands)

Wands	Markers used	1a	1b	2a	2b	SI
X	4, 10	0.57	0.50	0.51	0.59	0.67
Y	7, 10	0.82	0.63	0.75	0.84	0.69
Z	9, 10	0.52	0.46	0.58	0.54	0.65
XYZ	4, 7, 9, 10	0.58	0.42	0.48	0.46	0.58
XYZ	All (9)	0.71	0.44	0.70	1.27	1.08

with different numbers of markers. The second column of Table 8 gives the indices of the markers used (cf. Fig. 1). For example in the case of calibrating with the pair of wands *X* and *Y*, we either use all seven markers (1 through 7, cf. Fig. 1), or only a subset thereof, in particular the end markers of both wands and the origin (4, 7, 10). Note that this information only concerns the bundle adjustment; initial values were computed using all markers.

Another interesting issue is whether to enforce all available geometric constraints during bundle adjustment as described in Sect. 4.3, or only the known distances between markers, as in previous approaches. Table 9 shows results of the latter approach. Similarly as for Table 8, this was tested for different sets of markers.

Considering the effect of using the maximum or minimum number of markers for certain calibration methods, it seems that it is better to perform bundle adjustment using the minimal number of markers, independently of the initialization method and the type of constraints used in bundle adjustment. As a reminder, the minimum number of markers is three in the case of bundle adjustment as described in Sect. 4.3, or two if only enforcing the known wand length. At a first glance that may sound contradictory with the general principle that using more calibration data (higher redundancy) gives better results. However, at the same time, what also matters for all proposed methods is the distance between markers in the image plane which is also the function of marker distance in 3D space. Namely, for a calibration volume of a given size and a given image resolution, the calibration tool has to have some appropriate minimal size. For too small calibration wands its marker projections in the image would be too close to one another and in the presence of noise the relative error of their image distance (i.e. position) estimation with BA would be greater than when markers are quite apart. Thus, we believe that calibration wands of size 60 cm are optimal for the current spatial camera setup. In other words, calibrating with smaller wands is not recommendable at the present. Of course making calibration wands too big would make their manipulation harder and also would decrease the number of locations where the tool is imaged in as many cameras as possible. For completeness, let us mention that throughout the development of wand-based calibration approaches, newer and newer versions of 3D kinematic systems use smaller and smaller wands for calibration.

The importance of a good initialization of the parameters before bundle adjustment, has already been emphasized. On the other hand, almost equally important, are the constraints that can be enforced (like in our case, angles and distances), which help to converge to correct values even starting from mediocre initial values. Judging simply based on closeness of initial parameters to refined ones, the highest hopes were put in method 1a, then in method 1b and so forth. And indeed, a slight advantage goes in favor of methods 1a and 1b in the sense of the reconstruction accuracy obtained after optimization (Tables 8, 9). Moreover, all four proposed methods appear to have better outcomes than one where Smart's set of initial parameters value was used (last column in Tables 8, 9).

It is expected that using all constraints is better than only using, e.g. known wand lengths during bundle adjustment. This is confirmed by comparing Tables 8 and 9, particularly if we consider the results of Table 8 corresponding to minimal numbers of markers.

Eventually, we have reconstructed the 'unknown' wand lengths and computed their mean error from the true values using the 3D marker positions output by Smart. The mean error, for a wand of 30 cm length is 0.58 mm. 8). However, during the experiments we have deliberately made the wand dance for our methods twice shorter than for Smart calibration. Specifically, for our calibration volume Smart recommends to perform a wand dance of approximately 120 s, whereas we have reduced this to 60 s for our methods.

7 Conclusion

We have presented the problem of typical 3D kinematic system calibration, in the procedural sense (what the user is supposed to do) and also regarding many aspects from the computational point of view. We have summarized issues we encountered when trying to rely solely on self-calibration. This lead us to another solution, taking advantage of scene constraints, which are easy to provide in calibration tools for 3D kinematic systems. We have proposed four such methods for the initialization and subsequent refinement of camera parameters. The proposed methods use rather straightforward computational procedures to initialize camera parameters, using linear equation systems. For certain applications the accuracy obtained using these approaches may be satisfactory. Otherwise, bundle adjustment is recommended. Bundle adjustment should include all available constraints (here, orthogonality and known distances) and if necessary, should include non-linear lens distortion parameters.

After analyzing the typical calibration procedure for 3D kinematic systems we have found a way of simplifying and shortening it. Apart from being user-friendly and relying on already existing calibration tools (which warrants easy and fast implementation in present systems) our methods completely discard the necessity for two separate calibration steps: typically one for initialization and the other for parameter refinement—wand dance. Our proposed methods successfully acquire data for both in a single calibration step: wand dance with two or three orthogonal wands. In addition, the time required for wand dance is approximately twice shorter for our methods compared to what is recommended for the Smart system. One may argue that handling two or three wands is not as convenient as a single one. However, we believe it is a reasonable compromise given all mentioned advantages. Particularly, the fact that we completely discard the additional step of positioning an orthogonal triad on the floor which serves normally for parameter initialization and is frequently a cause for calibration to be completely repeated if a bad initialization set is acquired. Further, the gain of time and potential gain in accuracy, due to using richer geometric constraints.

We have shown that our methods provide a reconstruction accuracy at least as good as the commercial 3D kinematic system Smart. One could criticize that we have based our comparisons on a single commercial 3D kinematic system. We believe that due to its popularity, features and above all mostly positive critics, we can reasonably assume that similar results would have been obtained with any other high quality 3D kinematic system.

Acknowledgements Parts of this work were carried out during an internship of T. Pribanić at INRIA, which was partially funded by a student scholarship of the French Embassy in Croatia.

References

1. Ariel Dynamics Worldwide. http://www.arielnet.com/
2. BTS/Elite. http://www.bts.it/en/index.html
3. eMotion/Smart http://www.emotion3d.com, http://www.bts.it/eng/proser/elisma.htm
4. MotionAnalysis. http://www.motionanalysis.com/
5. http://www.peharec.com/index.html
6. Simi. http://www.simi.com/en/
7. Simple3D. http://www.simple3d.com/
8. Vicon Motion Systems. http://www.metrics.co.uk/
9. Allard, P., Stokes, I., Blanchi, J.-P.: Three Dimensional Analysis of Human Movement. Human Kinetics Champaign (1995)
10. Atkinson, K.B.: Close Range Photogrammetry and Machine Vision. Whittles Publishing (1996)
11. Baruh, H.: Analytical Dynamics. WCB/McGraw-Hill, New York (1999)
12. Borghese, N.A., Cerveri, P.: Calibrating a video camera pair with a rigid bar. Pattern Recogn. **33**, 81–95 (2000)
13. Brown, D.C.: Close-Range Camera Calibration. Photogramm. Eng. **37**(8), 855–866 (1971)
14. Caprile, B., Torre, V.: Using Vanishing Points for Camera Calibration. Int. J. Comput. Vis. **4**, 127–140 (1990)
15. Carrihill, B., Hummel, R.: Experiments with the intensity ratio depth sensor. Comput. Vis. Graph. Image Process. **32**, 337–358 (1985)
16. Chatterjee, C., Roychowdhury, V.P.: Camera calibration for image processing. In Webster, J.G. (ed.): Wiley Encyclopedia of Electrical and Electronics Engineering, Vol. 2, pp. 743–758. Wiley, New York (1999)
17. Devernay, F., Faugeras, O.: Straight lines have to be straight: automatic calibration and removal of distortion from scenes of structured environments. Mach. Vis. Appl. **13**, 14–24 (2001)
18. Douglas Hung, D.C.: 3D scene modelling by sinusoid encoded illumination. Image Vis. Comput. **11**(5), 251–256 (1993)
19. Eian, J., Poppele, R.E.: A single-camera method for three-dimensional video imaging. J. Neurosc. Methods **120**, 65–83 (2002)
20. Faugeras, O.: What can be seen in three dimensions with an uncalibrated stereo rig? In: Proceedings of the European Conference on Computer Vision, pp. 563–578. Santa Margherita Ligure, Italy (1992)
21. Faugeras, O.: Three-dimensional Computer Vision. MIT, Cambridge (1993)
22. Finsterwalder, S.: Die geometrischen Grundlagen der Photogrammetrie. Jahresbericht Deutscher Mathematik **6**, 1–44 (1899)
23. Fitzgibbon, A.W.: Simultaneous linear estimation of multiple view geometry and lens distortion. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 125–132. Kauai (2001)
24. Hartley, R.I., Gupta, R., Chang, T.: Stereo from uncalibrated cameras. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 761–764. Urbana-Champaign (1992)

25. Hartley, R.: Kruppa's equations derived from the fundamental matrix. IEEE Trans. Pattern Anal. Mach. Intell. **19**(2), 133–135 (1997)
26. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision, 2nd edn. Cambridge University Press, Cambridge (2004)
27. Hatze, H.: High-precision three-dimensional photogrammetric calibration and object space reconstruction using a modified DLT approach. J. Biomech. **21**, 533–538 (1988)
28. Horn, B.K.P., Hilden, H.M., Negahdaripour, S.: Closed-form solution of absolute orientation using orthonormal matrices. J. Opt. Soc. Am. A **5**(7), 1127–1135 (1988)
29. Jarvis, R.A.: A perspective on range-finding techniques for computer vision. IEEE Trans. Pattern Anal. Mach. Intell. **5**, 122–139 (1983)
30. Kahl, F., Henrion, D.: Globally optimal estimates for geometric reconstruction problems. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 978–985. Beijing (2005)
31. Ke, Q., Kanade, T.: Quasiconvex optimization for robust geometric reconstruction. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 986–993. Beijing (2005)
32. Liebowitz, D., Zisserman, A.: Metric rectification for perspective images of planes. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 482–488. Santa Barbara (1998)
33. Maas, H.-G.: Image sequence based automatic multi-camera system calibration techniques. Int. Arch. Photogramm. Remote Sens. **32**(B5), 763–768 (1998)
34. Maybank, S.J., Faugeras, O.D.: A theory of self calibration of a moving camera. Int. J. Comput. Vis. **8**(2), 123–151 (1992)
35. McGlone, C.: Bundle adjustment with object space geometric constraints for site modeling. In: Proceedings of the SPIE Conference on Integrating Photogrammetric Techniques with Scene Analysis and Machine Vision II, pp. 25–36. Orlando (1995)
36. Minou, M., Kanade, T., Sakai, T.: A method of time-coded parallel planes of light for depth measurement. Trans. IECE Jpn. **64**, 521–528 (1981)
37. Pages, J., Salvi, J., Garcia, R., Matabosch, C.: Overview of coded light projection techniques for automatic 3D profiling. In: Proceedings of the IEEE International Conference on Robotics and Automation, pp. 133–138 (2003)
38. Press, W.H., Teukolsky, S.A., Vetterling, W.T., Flannery, B.P.: Numerical Recipes in C. Cambridge University Press, Cambridge (1997)
39. Pribanić, T., Cifrek, M., Peharec, S.: 3D shape recovery with no explicit video projector calibration. In: Proceedings of the 12th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision, pp. 137–140. Plzen (2004)
40. Pribanić, T.: A method for modeling and analysis of human movement using optoelectronic system. Doctoral Thesis, Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia (2005)
41. Sansoni, G., Corini, S., Lazzari, S., Rodella, R., Docchio, F.: Three-dimensional imaging based on gray-code projection: characterization of the measuring algorithm and development of a measuring system for industrial applications. Appl. Opt. **36**(19), 4463–4472 (1997)
42. Semple, J.G., Kneebone, G.T.: Algebraic projective geometry. Oxford University Press, London (1979)
43. Slama, C.C. (editor): Manual of Photogrammetry, 4th edn. American Society of Photogrammetry and Remote Sensing, Falls Church (1980)
44. Sturm, P., Maybank, S.: On plane-based camera calibration: a general algorithm, singularities, applications. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 432–437. Fort Collins (1999)
45. Sturm, P.: Critical motion sequences for the self calibration of cameras and stereo systems with variable focal length. Image Vis. Comput. **20**(5–6), 415–426 (2002)
46. Svoboda, T., Martinec, D., Pajdla, T.: A convenient multi-camera self-calibration for virtual environments. PRESENCE: Teleoperators Virtual Environ. **14**(4), 407–422 (2005)
47. Triggs, B.: Autocalibration and the absolute quadric. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 609–614. Puerto Rico (1997)
48. Triggs, W., McLauchlan, P.F., Hartley, R.I., Fitzgibbon, A.: Bundle adjustment—a modern synthesis. In: Proceedings of the Workshop on Vision Algorithms, pp. 298–372. Springer, Berlin (1999)
49. Wei, G.Q., Ma, S.D.: Implicit and explicit camera calibration: theory and experiment. IEEE Trans. Pattern Anal. Mach. Intell. **16**(5), 469–480 (1994)
50. Zhang, Z.: On the epipolar geometry between two images with lens distortion. In: Proceedings of the 13th International Conference on Pattern Recognition, pp. 407–411. Vienna, Austria (1996)
51. Zhang, Z.: A flexible new technique for camera calibration. IEEE Trans. Pattern Anal. Mach. Intell. **22**(11), 1330–1334 (2000)

Author Biographies

Tomislav Pribanić received the Dipl.-Eng., M.Sc., and Ph.D. degrees in electrical engineering from the Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia in 1996, 2001, and 2005, respectively. His postgraduate thesis was awarded with Silver medal "Josip Lončar", award given for an outstanding thesis. In 2004 he spent three months as a visiting researcher at INRIA Rhône-Alpes, France and also, in 2006, 6 months at Fraunhofer Institute for Computer Graphics (IGD) in Darmstadt, Germany. He collaborates actively with Biomechanical laboratory of Peharec Polyclinic in Pula, Croatia, and has been taking part in development of number of SW packages/protocols for human motion analyses, many of which are implemented in commercials systems. Tomislav Pribanić is employed as an Assistant on the courses Digital Systems, Transducers in measurement systems, Biomedical informatics and Biomedical electronics for students of electrical engineering at the Faculty of Electrical Engineering and Computing, University of Zagreb. Also, he is a research collaborator under scientific project granted by Ministry of Science and Technology, Republic of Croatia and some other R&D projects. His research interests are human motion analysis, computer vision, electronic instrumentation and measurements, and biomedical engineering. He has been area chair for VIIP 2005 and ISPA 2005. In 2006 he has founded 'Biomechanical seed group' for Croatia, under the supervision of International society of biomechanics (ISB). Tomislav Pribanić is also member of IEEE (Institute of Electrical and Electronics

Engineers), EMBS (Engineering in Medicine and Biology Society), IFMBE (International Federation for Medical and Biological Engineering), and Croatian Medical and Biological Engineering Society (CROMBES). He has over 20 peer-reviewed scientific publications. He is happy father of one daughter (4) and son (1).

Peter Sturm holds a position of Research Director at INRIA. He graduated from the University of Karlsruhe (Germany) in 1994 and obtained a DEA (MSc) at INPG (Institut National Polytechnique de Grenoble) in the same year. During his undergraduate studies, Peter ran a 1-person software company, writing and selling software for sports events (used in 3 World Championships in Judo and Sumo and numerous international and national tournaments). In 1997, he obtained his Ph.D. from the INPG. His Ph.D. thesis was awarded with the SPECIF award, given to one French thesis per year in the Computer Science area. After a 2-year postdoc at Reading University (UK) Peter joined INRIA and the MOVI research team in 1999. He is a member of the editorial board of the *Image and Vision Computing* journal. He will co-organize ECCV 2008, has been Area Chair for ECCV 2006 and a member of programme committees of over 40 conferences or workshops. His research interests cover 3D reconstruction and reflectance modelling from images, object tracking and geometrical computer vision. Peter Sturm has over 80 peer-reviewed scientific publications.

Mario Cifrek received the Dipl.-Eng., M.Sc., and Ph.D. degrees in electrical engineering from the Faculty of Electrical Engineering and Computing, University of Zagreb, Croatia in 1987, 1992, and 1997, respectively. He is currently an Associate Professor of electronic measurements, instrumentation, and biomedical engineering with the Department of Electronic Systems and Information Processing, Faculty of Electrical Engineering and Computing, University of Zagreb. Also since 2006, he is assistant dean of the same faculty. His research interests are focused on design of biomedical instrumentation and signal analysis for research and clinical applications. He is a member of Institute of Electrical and Electronics Engineers (IEEE), Engineering in Medicine and Biology Society (EMBS), International Federation for Medical and Biological Engineering (IFMBE), International Federation for Automatic Control (IFAC) — TC Human-Machine Systems, Croatian Medical and Biological Engineering Society (CROMBES), KoREMA (Croatian Society for Communications, Computing, Electronics, Measurement and Control), and HMD (Croatian Metrology Society). Since 2005 he is a collaborating member of the Croatian Academy of Engineering.