ELSEVIER

Signal Processing 77 (1999) 309–334

SIGNAL PROCESSING

www.elsevier.nl/locate/sigpro

Accurate and simple geometric calibration of multi-camera systems☆

F. Pedersini, A. Sarti*, S. Tubaro

Dipartimento di Elettronica e Informazione, Politecnico di Milano, Piazza L. da Vinci 32, 20133, Milano, Italy

Received 5 March 1998; received in revised form 15 February 1999

Abstract

In this paper we present a low-cost, accurate and flexible approach to the calibration of multi-camera acquisition systems for 3D scene modeling. The adopted calibration target-set is just a marked planar surface, which is imaged in several positions in order to emulate a larger 3D target-frame. In order to obtain a better camera parameter estimation, the proposed approach is able to refine the a priori knowledge on the target-set through a process of self-calibration. This allows us to start with rough measurements of the coordinates of the calibration targets. We formalize our parameter estimation problem as a particular case of the more general class of inverse problem. In particular, we derive an analytic prediction of the calibration performance, based on error propagation analysis, whose correctness is demonstrated through simulation experiments. Finally, the results of a series of calibration experiments on real data is presented, which confirm the effectiveness of approach in a variety of experimental conditions.

Zusammenfassung

Wir präsentieren in diesem Artikel einen kostengünstigen, genauen und flexiblen Weg, um die Multi-Kamera-Erfassung für eine 3D-Szenariummodellierung zu kalibrieren. Die angenommene Zielmenge zur Kalibrierung ist lediglich eine markierte ebene Oberfläche, die in verschiedenen Positionen ins Bild gebracht wird, um einen größeren 3D-Zielrahmen zu emulieren. Um eine bessere Schätzung der Kameraparameter zu erhalten, ist die vorgeschlagene Methode in der Lage, die a priori-Kenntnis über die Zielmenge durch einen Selbstkalibrierungsprozeß auszunutzen. Wir formulieren unser Parameterschätzproblem als einen Spezialfall der allgemeineren Klasse der Invertierungsprobleme. Speziell leiten wir eine analytische Vorhersage für die Kalibrierungsleistung her, die auf der Fehlerfortplanzungsanalyse basiert, und deren Korrektheit anhand simulierter Experimente gezeigt wird. Abschließend werden die Ergebnisse einer Reihe von Kalibrierungsexperimente mit echten Daten vorgestellt, die die Wirksamkeit der Methode in einer Vielzahl experimenteller Bedingungen bestätigen.

Résumé

Dans cet article, nous présentons une approche de faible coût, précise et flexible pour calibrer des systèmes d'acquisition à caméras multiples pour la modélisation de scènes 3D. L'ensemble cible de calibration que nous avons

☆Work supported in part by the ACTS Project "PANORAMA", Project No. AC092.

* Corresponding author. Tel.: 39-2-2399-3647; fax: 39-2-2399-3413.

E-mail address: sarti@elet.polimi.it (A. Sarti)

0165-1684/99/$ - see front matter
PII: S0165-1684(99)00042-0

adopté est simplement une surface plane marquée, imagée dans plusieurs positions afin d'émuler une trame cible 3D plus large. Afin d'obtenir une meilleure estimation des paramètres des caméras, l'approche proposée permet de raffiner les connaissances a priori sur l'ensemble cible au moyen d'un processus d'auto-calibration. Ceci nous permet de commencer avec des mesures grossières des coordonnées des cibles de calibration. Nous formalisons notre problème d'estimation de paramètres comme un cas particulier d'une classe de problèmes plus générale. En particulier, nous dérivons une prédiction analytique de la performance de calibration, reposant sur une analyse de la propagation de l'erreur, dont l'exactitude est démontrée par des expériences de simulations. Finalement, nous présentons les résultats d'une série d'expériences de calibration sur des données réelles, qui confirment l'efficacité de l'approche dans une large variété de conditions expérimentales.

Keywords: Camera calibration; Parameter estimation; Inverse problems; 3D reconstruction

1. Introduction

Multi-camera acquisition systems are today often employed for 3D scene reconstruction in a variety of applications ranging from industrial quality control to content for virtual reality applications. In the past decades, in fact, a variety of methods have been developed for estimating the 3D structure of a scene through the joint analysis of a set of its views. Most of these methods rely on the a priori knowledge of a set of parameters that specifies the geometrical model of the acquisition system. The estimation process of such parameters is generally called *camera calibration* and represents a crucial step in the global reconstruction chain. As a matter of fact, the quality of the reconstruction is crucially dependent on the accuracy of the calibration process and a 3D reconstruction of "metric" quality often requires a long and cumbersome calibration procedure. It is the aim of this article to approach the calibration problem in a general fashion with the goal of keeping the complexity and the setup of the calibration procedure as simple as possible without giving up accuracy in the estimation results.

It is well-known that the 2D coordinates of some image features, as acquired with two or more cameras, can be used for recovering the 3D position of the scene details that originated them, through a process of "geometric triangulation". In order to do so, we need to know the physical (optical and electrical) and geometrical (positional) characteristics of the cameras and we need to make sure that the correspondences between image features are correctly determined. The "matching" of image features is usually a critical problem as the search space for stereo-correspondences is two-dimensional (the image plane). However, the knowledge of the model parameters of the acquisition system can be used for making it a one-dimensional search by exploiting the *epipolar geometry* of the camera setup. In fact, given a point on the first image, the stereo-corresponding point on the second one is bound to lie on the *epipolar line*, which is the projection of the first optical ray onto the second image plane [1]. This epipolar constraint, however, is generally not enough to guarantee the correctness of a stereo correspondence. A search for feature correspondences along the epipolar line is, in fact, often performed by comparing the luminance profiles in the neighborhood of the candidate matches on the two views, under some constraints on the relative ordering between them [13]. The risk of matching ambiguities can be reduced through the adoption of global consistency constraints, implemented through dynamic programming [13]. In alternative, we can *geometrically* remove the ambiguity with the introduction of a third camera. In this case, in fact, each point of a matched triplet is bound to lie on the intersection of the epipolar lines corresponding to the other two points.

The use of more than three cameras could be justified by the need of making the 3D reconstruction strategy more robust or by the need of expanding the class of 3D information that we can safely extract from a joint analysis of the available views. For example, *horizon* contours (i.e. extremal boundaries generated by smooth self-occlusions) [14,18] are known to provide valuable information on the

local 3D structure of the surface (position, tangent plane and curvature) near the visible rims of the objects, provided that at least three cameras be available. However, in order to make this reconstruction approach robust, the adoption of at least four cameras with known geometry would help.

It is important to mention that a number of single-camera methods are also available in the literature [19,23]. Such methods, often based on the analysis of a video sequence acquired with a single camera, are usually non-calibrated, in the sense that the parameters of a camera model are "implicitly" estimated together with the 3D structure of the imaged scene. However, if the goal is that of a high-accuracy "metric" 3D reconstruction, then a preliminary partial calibration (estimation of the intrinsic camera parameters) of the camera system becomes important, especially when the camera resolution is modest [17]. In conclusion, the acquisition systems that we are interested in are multi-camera systems and, although the approach we will illustrate can be applied to a broad range of cameras, we will focus on cameras of modest resolution (standard TV resolution) in order to verify what performance can be achieved with such low-cost devices.

One obvious way to obtain information on the camera parameters is, in fact, to decide them beforehand. This can be done through a mechanical adjustment (through high-precision mechanical supports) of the position and the orientation of each camera and on the use of "metric" lenses and sensors (optics with a priori known characteristics). This solution, however, is normally not applicable because of its complexity and its high cost. A more flexible approach is to estimate the parameters of the acquisition systems through a photogrammetric analysis of matched image features [2,4,17,22]. In general the estimation procedure consists of a joint analysis of one or more views of a number of points (*targets*), which could be *fiducial* marks placed in the scene volume or even some natural point-like features that belong to the scene to be reconstructed. This procedure can be implemented in a variety of ways, depending on the structure and on the available a priori information on the calibration targets. One common approach to camera parameter estimation is to make use of an artificial *target-set*, whose *targets* are attached to a rigid frame that occupies part of the 3D viewing space, with a priori known geometrical characteristics. As the exact 3D coordinates of the targets are assumed available (for example because they have been previously measured through some high-precision procedure), they can be used together with the image coordinates of their views in order to estimate the parameters of the acquisition system. This approach is commonly referred to as (simple) calibration, and is characterized by a complete knowledge of the calibration target-set. The opposite situation is produced by a set of targets that are scattered in the scene volume in locations that are completely unknown. This extreme situation occurs when, instead of using a pre-measured calibration target-frame, we use a set of targets that have been artificially added to the scene or natural point-like features that are already present in the scene to be reconstructed. This type of *blind calibration* problem is usually referred to as self-calibration and, due to the much larger number of unknowns, in total absence of a priori information on the targets, it is an undetermined problem [7], in the sense that it does not allow us to recover the whole geometry of the camera system. In between the two extremes of simple and blind calibration there is a whole range of situations in which only some information on the targets or on the cameras is available in a variety of forms, for example statistical information (nominal target coordinates and a measure of their uncertainty), rigidity constraints, etc. We will see that this partial information can be successfully exploited for making the self-calibration problem solvable.

It is important to emphasize that the estimated parameters of the acquisition system are expected to hold accurate only for measurements within the 3D volume "spanned" by the specific calibration target-set [8]. In fact, roughly speaking, the target-set plays the role of a *training set* for the simple calibration procedure; therefore it should be chosen in such a way to be "statistically representative" of the scene to be reconstructed. As a consequence, in order to achieve high accuracy in the calibration and in the 3D reconstruction, it is important for the targets to properly "fill up" the entire volume that will be later occupied by the object to be measured.

This implies that the size of an adequate calibration target-frame should be comparable with that of the scene to be reconstructed, with obvious difficulties in the calibration procedure. In order to overcome this difficulty, we virtually "enlarge" a target-set of modest size through the acquisition of a number of its views in different positions. The positions of the target-frame are chosen in such a way that the union of all targets will fill up volume of interest in such a way to be more representative of the scene to be reconstructed. Of course, every time we move the target-frame we introduce six new positional unknowns, unless we are able to force the frame into pre-determined positions through some high-precision positioning device. Due to the cost of high-precision mechanical positioners, the only feasible alternative is to move the pattern freely between acquisitions and to proceed with an a posteriori determination of this motion parameters by embedding their estimation into the calibration process itself. Notice that this way of proceeding corresponds to performing a partial self-calibration as some information on the global 3D set of targets (position of the target-frame) is not available and, therefore, must be estimated. However, we will keep referring to this method as a *simple calibration* technique, meaning that the available 3D coordinates of the targets within the target-frame will be taken as granted and trusted upon.

As we can easily expect, the quality of the simple calibration process is strongly influenced by the accuracy of the camera model. Because of that, the ideal projective camera [1], also known as "pinhole" camera, is usually not accurate enough as to guarantee high accuracy in the 3D measurements. In particular, accounting for the non-ideal behavior of the camera lenses can become a crucial aspect in applications of 3D reconstruction. However, although the accuracy of the camera model can be improved through the introduction of an adequate number of parameters [21], there is no point in using too sophisticated a model as one of the main sources of inaccuracy in the calibration methods is, in fact, the accuracy with which the 3D coordinates of the targets are known. Due to the high cost of accurate measurement procedures, the only option we have for improving the performance of the parameter estimation process is to improve the calibration performance through a sort of self-calibration approach [6], which allows us to go beyond the accuracy of the available target measurements.

As already said above, a *blind* calibration strategy (self-calibration in total absence of information on the targets) is an extremely ill-conditioned problem. However, some approximate a priori information on the target-set is usually available, or it can be easily obtained through rough measurements. If such measurements can be assumed to be fairly unbiased, even if they are rough, we can devise a self-calibration strategy that is able to refine the rough measurements of the target's coordinates while estimating the parameters of the acquisition system. In general, however, as we need to maximize the accuracy of the target's coordinates, we need the noise that the data is affected by (additive noise and the consequent error in the localization of the image coordinates of the targets) [6] to be of modest magnitude. In the following, we will refer to this approach as a self-calibration method, in the sense that the accuracy of the target's coordinates is improved by the estimation process, with a consequent improvement of the calibration's accuracy.

In this article we propose a simple and effective technique for calibrating CCD-based multi-camera acquisition systems, which is capable of highly accurate results even when using a low-cost planar calibration target-set of modest size, and low-cost imaging devices, such as standard TV-resolution cameras connected to commercial frame-grabbers. The key features of the method are the above-described "*multi-view*, *multi-camera*" (MVMC) approach,[1] based on the analysis of a number of views of a calibration target-set placed in different positions, combined with a self-calibration approach, which makes it able to refine (when necessary) rough information on the target's coordinates. Our goal is to show that accurate calibration can be a task of fairly modest difficulty and cost.

In order to devise and develop the method proposed in this manuscript, we formalized the simple calibration and the self-calibration methods as two

[1] In the following, the term *multi-view* will specifically refer to the availability of several multi-camera acquisitions of the same target-set in different positions.

particular instances of the more general class of inverse problems [16], which only differ in the input data. In fact, we derived an analytic prediction of the calibration performance, based on error propagation analysis, whose correctness is demonstrated in the manuscript through simulation experiments. Furthermore, a series of calibration experiments on real data has been carried out in order to evaluate the accuracy and the robustness of the proposed algorithm in a variety of experimental conditions. In particular, we conducted a series of experiments for comparing the performance of the self-calibration approach with that achievable through simple calibration.

The article is organized as follows: in Section 2 we summarize some basic concepts and define the notation that we will need to approach the considered problem. In particular, the camera model adopted for this work is illustrated in detail. Furthermore, we present the simple calibration and the self-calibration problems (Section 2.2) as particular cases of inverse problems. In particular, in Section 2.3 we discuss an approach to inverse problems that can be used for analytically predicting the performance of the simple calibration and the self-calibration methods. In Sections 3 and 4 we illustrate our approach to simple-calibration/self-calibration, based on multiple acquisitions of a planar target-set. This approach allows a great flexibility in the exploitation of the a priori knowledge of the acquisition system and on the target-set. Furthermore it allows us to obtain a level of accuracy which is at the same level as with 3D target-sets. Sections 5 and 6 are devoted to the presentation of the results of some simulation experiments on synthetic data and on some calibration experiments on real data. Such experiments confirm the validity of the analytical prediction of the performance of our method, and prove the effectiveness and the flexibility of the approach in a variety of experimental conditions. Section 7 concludes the manuscript with final remarks and suggestions for future improvements.

2. Preliminaries

The goal of this section is to present and formalize the camera simple-calibration and self-calibration problems as instances of the general theory of inverse problems. In order to do so, we will first provide a description of the adopted camera model and of the parameter estimation approach, in order to be able to discuss this approach as a particular case of inverse problem.

2.1. The camera model

A camera model is defined as the mathematical relationship between the 3D coordinates of a point in the scene space and its corresponding coordinates on the image plane. Even though this relationship can be defined in a variety of ways, a rough classification can be made between those models that are based on an operator (e.g. projection matrix) that maps object coordinates onto image coordinates through a homogeneous representation [1,20,22], and those that define a model by directly using all the optical and geometric parameters of the camera [17]. The camera model we adopted, shown in Fig. 1, belongs to this latter category, as we are interested in attributing to each parameter a precise physical meaning. This choice provides us with a certain flexibility in using all the a priori information on the camera setup. For example, if we know that the optical lens of the adopted acquisition system has a nominal focal length of 16 mm, then this information can be readily used for improving the reliability and the accuracy of the calibration process. Moreover, this type of camera model provides us with more physical

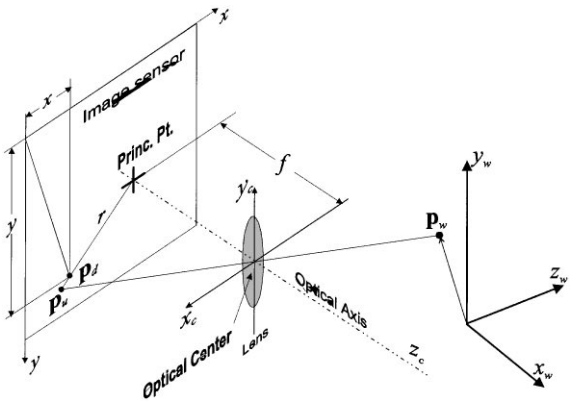

Fig. 1. Adopted camera model.

intuition and, as a consequence, allows us to readily judge the outcome of the calibration through a direct comparison between the estimated parameters and the "physical characteristics" of the camera (position, orientation, focal length etc.). Another important advantage of such camera models is that it is characterized by a non-redundant set of parameters.

Three different reference frames are defined and used [17] for the camera model of Fig. 1:

- *World reference frame* – rigidly attached to the scene; used for specifying the *world* coordinates of any point $\boldsymbol{p}_w = [x_w\ y_w\ z_w]^{\mathrm{T}}$ of the 3D scene;
- *Camera reference frame* – rigidly attached to the camera; z_c is the optical axis, while x_c and y_c are parallel to the horizontal and vertical axes of the image plane (which is assumed to be orthogonal to the optical axis), respectively. The origin is the *optical center* of the lens. The camera coordinates of a 3D point are specified as $\boldsymbol{p}_c = [x_c\ y_c\ z_c]^{\mathrm{T}}$. The intersection between optical axis and image plane is called *principal point*;
- *Image reference frame* – defined on the image plane; the origin is the center of the pixel at the bottom-left corner of the image, x_f and y_f denote rows and columns, respectively. The image coordinates $\boldsymbol{p} = [x\ y]^{\mathrm{T}}$ are expressed in pixels (see upper-left frame in Fig. 1).

The relationships between the world coordinates of a point $\boldsymbol{p}_w$ and the image coordinates $\boldsymbol{p}$ of its projection onto the image plane are given in the following:

1. *Conversion from world-coordinates* ($\boldsymbol{p}_w$) *to camera-coordinates* ($\boldsymbol{p}_c$)

$$\boldsymbol{p}_c = \boldsymbol{R}\cdot\boldsymbol{p}_w + \boldsymbol{t}, \tag{1}$$

where $\boldsymbol{R}$ is a rotation matrix and $\boldsymbol{t}$ a translation vector which specify the rigid displacement between world reference frame and camera frame.

2. *Perspective projection of a* 3*D point P onto the image plane*

$$\boldsymbol{p}_u = -\frac{f}{z_c}\boldsymbol{p}_c, \tag{2}$$

which results in $\boldsymbol{p}_u = [x_u\ y_u\ f]^{\mathrm{T}}$, f being the focal length of the optical lens.

3. *Lens distortion* – modeled as a shift of the image points from their ideal perspective projection, lens distortion can be thought of as a nonlinear stretching of the image plane. In order to accurately model lens distortion [11] both its *radial* and *tangential* components should be considered. With radial distortion, image coordinates are radially shifted from the principal point, while tangential distortion accounts for the component that is perpendicular to the radial direction. In this manuscript we only consider radial distortion as the tangential component is often negligible with respect to the radial one [17]. The radial distortion is usually modeled by the power series that expresses the undistorted image coordinates $\boldsymbol{p}_u = [x_u\ y_u]^{\mathrm{T}}$ as a function of the distorted ones $\boldsymbol{p}_d = [x_d\ y_d]^{\mathrm{T}}$:

$$\begin{aligned} x_u &= x_d\cdot(1 + k_3 r_d^2 + k_5 r_d^4 + \cdots)\\ y_u &= y_d\cdot(1 + k_3 r_d^2 + k_5 r_d^4 + \cdots)\\ r_d^2 &= x_d^2 + y_d^2, \end{aligned} \tag{3}$$

where r_d is the distance between the distorted image point and the principal point. The first two terms of the series (k_3, k_5) are usually sufficient for an accurate parameterization of the radial distortion [21].

4. *Conversion from camera coordinates to image coordinates*

$$x = c_x + \frac{x_d}{d_x}, \qquad y = c_y + \frac{y_d}{d_y}, \tag{4}$$

where $\boldsymbol{c} = [c_x\ c_y]^{\mathrm{T}}$ are the image coordinates of the principal point (expressed in pixels), while d_x and d_x are the horizontal and vertical size of the pixel, respectively [17].

The above set of equations allows us to directly compute the image coordinates of a point, given its position in the scene and the parameters of the camera model.

In conclusion, our camera model is completely and uniquely specified by the parameters involved in the above equations. In particular, $\boldsymbol{R}$ and $\boldsymbol{t}$ are called *extrinsic* parameters, as they define the geometric relationship between cameras and 3D scene, while the others are called *intrinsic*, as they only depend on the physical characteristics of the

cameras. Throughout this manuscript, the parameters of the jth camera will be specified by the following vector of eleven elements:

$$\boldsymbol{c}_j = [\varphi^{(j)}\ \theta^{(j)}\ \psi^{(j)}\ t_x^{(j)}\ t_y^{(j)}\ t_z^{(j)}\ f^{(j)}\ k_3^{(j)}\ k_5^{(j)}\ c_x^{(j)}\ c_y^{(j)}]^{\mathrm{T}}, \quad (5)$$

whose first six elements are the Euler angles and the translation components that characterize the rigid displacement $(\boldsymbol{R},\boldsymbol{t})$ from the world-frame to the camera-frame. It is important to emphasize that the size of the pixel (d_x, d_y) is assumed a priori known and is not included in the set of camera parameters that are to be estimated through camera calibration. This assumption is reasonable in the following two cases [17]:

- digital camera, provided that the actual pixel size is known;
- camera with an analog output connected to a frame grabber (image digitizer), provided that pixel size is known and that the ratio between the pixel-clock frequency of the camera and the sampling frequency of the frame grabber is known (our calibration experiments with analog cameras were conducted by synchronizing the frame grabber's clock with the camera pixel clock [4]).

2.2. Estimation of the acquisition system's parameters

The estimation of the camera parameters is carried out through the analysis of the views of a test object (calibration target-set). The target-set usually consists of a set of *fiducial marks*, also called *targets*, positioned within the 3D volume that is being imaged by the camera system (see Fig. 2).

As already said in the Introduction, we adopt a simple calibration approach when the knowledge on the 3D coordinates of the targets is complete and accurate, therefore it can be trusted upon as is. When, on the contrary, no information is available at all, then the estimation problem (*blind calibration*) is generally undetermined. When, finally, the 3D positions of the fiducial points are only partially known, then they need be estimated as well through some *self-calibration* process. We will see in the next Section under which conditions this approach is applicable as far as the number of cameras and the number of views of the calibration target-set are concerned [9].

Let $\boldsymbol{p}_w(i)$, $i = 1, \ldots, N$, be the world-coordinates of the ith target and let $\boldsymbol{c}_j$, $j = 1, \ldots, M$ be the parameter vectors of the M cameras. The image coordinates $\boldsymbol{p}^{(j)}(i) = [x^{(j)}(i)\ y^{(j)}(i)]^{\mathrm{T}}$ of the ith target, as seen from the jth camera, can be the written as a function of both camera system parameters and target's coordinates

$$\boldsymbol{p}^{(j)}(i) = g(\boldsymbol{m}_{i,j}), \quad (6)$$

where $\boldsymbol{m}_{i,j} = [\boldsymbol{p}_w^{\mathrm{T}}(i)\ \boldsymbol{c}_j^{\mathrm{T}}]^{\mathrm{T}}$. This global equation can be thought of as a *direct* formulation of the camera modeling problem. Roughly speaking, a self-calibration problem can be seen as a method for *inverting* the formulation (6) with respect to $\boldsymbol{m}_{i,j}$.

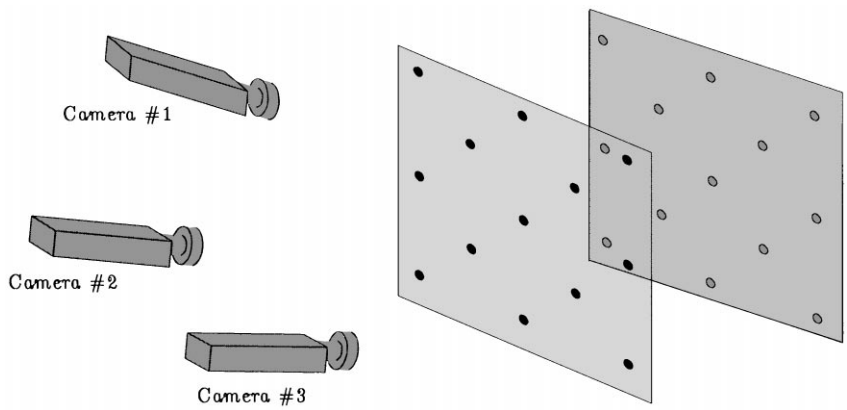

Fig. 2. Calibration setup.

When, on the contrary, the knowledge of the 3D coordinates of the targets is complete and accurate, then it can be considered as embedded in the direct model (*simple calibration*), which now becomes

$$\boldsymbol{p}^{(j)}(i) = g(\boldsymbol{p}_w(i), \boldsymbol{m}_j) = g^{(i)}(\boldsymbol{m}_j), \quad \boldsymbol{m}_j = \boldsymbol{c}_j. \tag{7}$$

In general, given a set of data and a parametric model $g(\cdot)$ of the system, estimating the parameters that characterize the system through the selected model represents an *inverse problem* [16]. In our case, the inverse problem consists of using the available data $\boldsymbol{p}^{(j)}(i)$ in order to determine the correct *parameters* that make the model $g(\cdot)$ correctly describe the mapping of the 3D points onto the image planes. Such parameters are, quite clearly, the intrinsic and the extrinsic ones that characterize the camera system.

In order to follow the terminology that is normally used when dealing with inverse problems [16], we will collect into a single vector $\boldsymbol{p}$ all the available image coordinates of all the targets $\boldsymbol{p}^{(j)}(i)$, $i = 1, \ldots, N$, $j = 1, \ldots, M$, while the 2D vector space $\mathscr{P}$ that can be spanned by $\boldsymbol{p}$ will be called *observation space*. Similarly, we will define a global parameter vector $\boldsymbol{m}$, which contains all the model vectors ($\boldsymbol{m}_{i,j}$, $i = 1, \ldots, N$, $j = 1, \ldots, M$, in the self-calibration case, or $\boldsymbol{m}_j$, $j = 1, \ldots, M$, in the simple calibration case) and spans the so-called *model space* $\mathscr{M}$. In accordance to this terminology, $g(\cdot)$ will be referred to as *direct model*.

From a practical standpoint, the simple-calibration/self-calibration process can be seen as a way of exploiting a large number of constraints that cumulate in a space made of a large number of coordinates. The constraint equations are those that force the projection of a target onto an image plane, computed through Eq. (6), to correspond to its actual image coordinates. In fact, the projection of a 3D point onto an image plane give rise to a pair of equations (one per image coordinate). It is customary (and advisable) to use a redundant number of fiducial points with respect to the number of unknows, so that the model space will result as *overconstrained* [17]. As a consequence, the determination of the model will have to be performed through a process of minimization of a measure of the error between the observed data $\boldsymbol{p}$ and the data computed through the model parameter vector $\boldsymbol{m}$ [4,6,21]. For example, adopting the MSE as a measure of this error, we will have to compute

$$\hat{\boldsymbol{m}} = \arg\min_{\boldsymbol{m}} \{\|\boldsymbol{p} - g(\boldsymbol{m})\|^2\}. \tag{8}$$

This minimization process is clearly nonlinear and a variety of methods can be used for determining the solution $\hat{\boldsymbol{m}}$. The procedures that are commonly adopted for solving this type of nonlinear problems are all iterative [21]; therefore an accurate initialization of the minimization process could become crucial for preventing the algorithm from being trapped into some local minima [17]. In order to take all the available information into account, each term of the cost function to be minimized in Eq. (8) can be weighted by a factor that takes into account the accuracy with which the 2D coordinates of the image point have been detected and the accuracy with which the coordinates of the corresponding 3D point are known [6,24].

2.3. Some remarks on inverse problems

As already said above, the camera simple-calibration/self-calibration process consists of the estimation of the model parameter vector $\boldsymbol{m}$, through the knowledge of the observed data $\boldsymbol{p}$ and the direct model $g(\cdot)$. This operation corresponds to inverting the model function

$$\boldsymbol{m} = g^{-1}(\boldsymbol{p}).$$

In inverse problems, such as calibration and self-calibration problems, the data vector $\boldsymbol{p}$ is usually the result of *physical* measurements. As a consequence, due to the unavoidable noise that affects the measuring process, the observed data vector $\tilde{\boldsymbol{p}}$ will generally differ from the data vector $\boldsymbol{p}$ that we would predict if the CCD sensor were noiseless and had infinite resolution and if our camera model were infinitely accurate.

The vector $\boldsymbol{p}$ contains the image coordinates of all the targets as viewed by all the cameras. The measuring process that provides the observed data vector $\tilde{\boldsymbol{p}}$ consists of the analysis of the luminance profiles of the acquired views and, as such, it is affected by errors [3] that are mostly due to the limited image resolution [5]. In order to be able to account for measurement's uncertainty,

a conditional probability density function (p.d.f.) of the form $f_{\tilde{P}|P}(\tilde{\boldsymbol{p}}|\boldsymbol{p})$ can be defined, where an uppercase letter denotes a random vector and its lowercase version denotes an instance of this vector.

It is important to keep in mind that any direct model $g(\cdot)$ used in practical applications can only be approximate. In fact, our camera model only involves elementary principles of geometrical optics, while a much more complex formulation would be required for a more correct and complete description of the optics and of the image sensor [11]. In order to take the model's uncertainty into account, a conditional p.d.f. of the form $f_{P|M}(\boldsymbol{p}|\boldsymbol{m}) = S(\boldsymbol{p} - g(\boldsymbol{m}))$ can be defined. Quite clearly, the "spread-function" S of a perfect model will be an ideal impulse $\delta(\cdot)$.

As already said in the Introduction, some a priori information is usually available (or easy to obtain) on the model's parameters, and this should not be ignored. In particular, the world-coordinates of the targets could be known with a certain (even very limited) accuracy. Of course, depending on how the available partial information is specified for the calibration problem, a variety of different problems can be formulated and approached. For example, rather than knowing the 3D coordinates of the targets with limited accuracy, our knowledge could be limited to the fact that the targets are scattered over a nearly-regular grid, or that the target-set undergoes a rigid motion during the acquisition of a series of multi-views. Other than on the target-set, some information could also be available on the camera parameters. For example it might be known that the focal length of a camera lies within a specific range. All the above a priori information can be incorporated [16] in the calibration/self-calibration process through the definition of some proper probability density functions.

A statistical description of the acquisition system is provided by the p.d.f. $f_{P,M}(\boldsymbol{p}, \boldsymbol{m})$. In general, the solution of an inverse problem and, in particular, of our calibration problem, is the value of $\boldsymbol{m}$ that maximizes the a posteriori information on the model's parameters $f_{M|P}(\boldsymbol{m}\,|\boldsymbol{p})$, which can be derived from $f_{P,M}(\boldsymbol{p}, \boldsymbol{m})$. By doing so, we perform a maximum likelihood estimation of the form

$$\boldsymbol{m}_{\mathrm{ML}} = \max_{\boldsymbol{m}} [f_{M|P}(\boldsymbol{m}|\boldsymbol{p})]. \tag{9}$$

Furthermore, when the sources of uncertainty that affect our inverse problem can be modeled by a zero-mean Gaussian p.d.f., it is also possible to predict the accuracy of the solution of the inverse problem in quite a general fashion (see Appendix A).

As shown in Appendix A for the general case of the nonlinear direct model $g(\cdot)$ (calibration problems are nonlinear) it is possible to estimate the a posteriori covariance using a relationship of the form

$$\boldsymbol{C}_{M|P} = (\boldsymbol{G}^{\mathrm{T}}\boldsymbol{C}_P^{-1}\boldsymbol{G} + \boldsymbol{C}_M^{-1})^{-1}, \tag{10}$$

where

$$\boldsymbol{G} = \left(\frac{\partial g}{\partial \boldsymbol{m}}\right)_{\boldsymbol{m}=\boldsymbol{m}_{\mathrm{ML}}}$$

is the Jacobian of the forward model, which represents a linearization of $g(\boldsymbol{m})$ about $\boldsymbol{m}_{\mathrm{ML}}$, $\boldsymbol{C}_M$ is the a priori covariance matrix of the model's parameter vector and $\boldsymbol{C}_P$ is the covariance matrix associated to both the "forward modeling uncertainty" and the "experimental uncertainty" (i.e. the statistical relationship between $\tilde{\boldsymbol{p}}$ and $\boldsymbol{p}$).

Notice that the a posteriori information on the model parameters ($\boldsymbol{C}_{M|P}$) is obtained as a combination of a priori information ($\boldsymbol{C}_M$) and information on the dispersion of the available data ($\boldsymbol{C}_P$). The diagonal elements of $\boldsymbol{C}_{M|P}$ represent the variance associated to the estimate of each model parameter $\boldsymbol{m}_{i,j}$. The other elements of $\boldsymbol{C}_M$ can be used to estimate the correlation between the various parameters and to have an idea on how "separable" such parameters are [16]. In conclusion, an inverse problem can always be seen as a way of "translating" information from the data space $\mathscr{P}$ into the model space $\mathscr{M}$; therefore the solution of a "well-posed" inverse problem should give an a posteriori uncertainty on the model parameters that is smaller than the a priori uncertainty [16].

3. Multi-view multi-camera approach

As already said in the previous Section, the aim of simple camera calibration and self-calibration is to estimate the model parameters $\boldsymbol{m}$ from the

knowledge about the observed data $\boldsymbol{p}$, i.e. to solve the inverse problem $\boldsymbol{m} = g^{-1}(\boldsymbol{p})$. The observed data $\boldsymbol{p}$ is a set of all image coordinates of the calibration targets, as seen on all available views, i.e. $\boldsymbol{p} = \{\boldsymbol{p}^{(j)}(i); 1 \leqslant i \leqslant N,\ 1 \leqslant j \leqslant M\}$ where N and M are the total number of targets and cameras, respectively.

3.1. The simple calibration approach

In the case of simple calibration, the 3D coordinates of the fiducial points are known. The (small) uncertainty on their positions can be included in the model through the computation of $f_{P|M}(\boldsymbol{p}|\boldsymbol{m})$. The dimensionality of the model space, in this case, is LM where L is the number of parameters of the camera model (from Eq. (5) we have $L = 11$) and M is the number of cameras to be calibrated.

In order for the calibration problem to be solvable, it is necessary to have at least as many constraints (equations) as unknowns. As each fiducial point gives rise to a pair of equations per camera, a minimum of six independent (non-collinear) points is required for determining the eleven parameters of the camera model.[2] However, since the problem is nonlinear and strongly ill-conditioned [20], a larger number of points should be considered.

Since the pixel size is a known parameter (see Section 2.1), the simple calibration problem can also be solved adopting a simple target-set whose fiducial points are all coplanar [17]. A planar target-set is much simpler to build than a 3D target-frame. In fact, it can be easily constructed applying a sufficient number of properly shaped stickers (targets) to a rigid planar surface. The coordinates of the targets (and their uncertainty), relative to a frame attached to the surface, can be quite easily determined while applying the stickers to the surface. Conversely, a 3D calibration target-frame always requires an accurate 3D measurement of the coordinates of the targets, which is generally performed with some photogrammetric technique [10]. The main drawback of 2D target-sets, however, is that of providing data that are rather correlated to each other. Furthermore, they occupy a rather limited volume of the scene, and this is true of all target-sets (also the 3D ones) that are small enough to be considered as "portable". It is well-known, in fact, that a reliable camera calibration can only be performed if the targets are not only numerous enough, but also well-distributed in the 3D space that will later be occupied by the object to be measured [8]. In order to overcome such limitations, we virtually enlarge the planar target-set through the acquisition of several of its views, as shown in Fig. 3. The positions of the target-frame are chosen in such a way that the union of all targets will occupy the entire volume of interest in a fairly uniform fashion.

The above strategy requires a modification of a standard simple calibration procedure as, even when the coordinates of the targets are known with respect to the frame attached 2D surface, the relative motion that the target undergoes between acquisitions is not known and needs to be determined. In order to do so, we could proceed by forcing the frame into pre-determined positions by means of high-precision positioning devices. This choice, however, would end up being more complex and expensive than the construction of a 3D target-frame. The only feasible alternative is thus to freely move the pattern between acquisitions and to proceed with an a posteriori determination of this motion by embedding its estimation into the calibration process itself. In order to do so, the six parameters that describe both position and orientation[3] of the target-set (relative to the world reference frame) will be added to the model parameters that need to be estimated for each position of the target-frame. By doing so, we modify the simple calibration method in the direction of self-calibration, even if we keep referring to it as a simple-calibration approach.

If we are considering V different positions of the targets-frame, then $6(V-1)$ new unknowns must

[2] It can be shown [17] that, when dealing with a single-camera system, all points should not be co-planar unless the pixel size is known beforehand.

[3] In the following we will use the term "position" to mean both position and orientation.

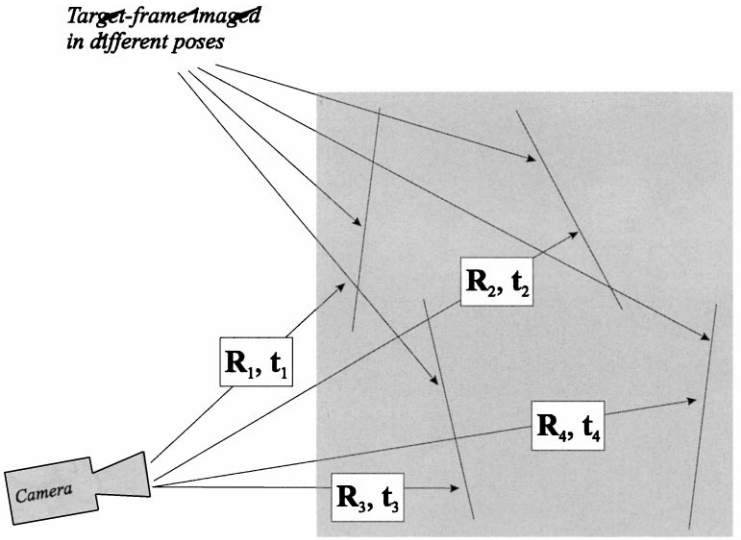

Fig. 3. Schematic description of the calibration setup in which a simple 2D target-set is being imaged by a camera, in a variety of poses. The 3D position and orientation of the calibration frame in each one of the considered poses is included in the set of unknowns to be estimated through calibration.

be added to the number of unknowns that we would have if only one position were considered. On the other hand, a total of $F = \sum_{k=1}^{V} 2H_{jk}$ equations can be written, H_{jk} being the number of targets that are imaged in the kth view taken from the jth camera ($0 \leqslant H_{jk} \leqslant H$ where H is the total number of targets of the calibration frame).

In the previous section, we assumed that the acquisition system to be calibrated is a multi-camera rig. This assumption, however, is not restrictive. In principle, in fact, we could individually calibrate the cameras by following the above procedure. We should keep in mind, however, that a joint calibration of all cameras is generally more efficient and introduces a larger number of constraints in the parameter estimation process, with the result of reducing the risk of an erroneous estimation. As a matter of fact, if we consider that the motion of the target-frame from view to view is the same for all cameras, then each camera gives its contribution to the estimation of this motion. For this reason, the simultaneous calibration of all cameras of the acquisition system increases the well-posedness of the calibration problem, with the result of making the estimation easier and more reliable. With respect to the case in which one camera is calibrated with V views of the target-set, each additional camera adds L unknowns (in our case $L = 11$) and approximately $2HV$ equations (assuming that all targets are imaged in the various acquisitions). The final approach is therefore a multi-view, multi-camera calibration (see Fig. 4), in which all parameters are estimated through the same global error minimization process.

3.2. The self-calibration approach

In all simple calibration procedures, including the case of multi-view multi-camera (MVMC) simple calibration, the quality of the parameter estimation strictly depends on the accuracy with which the world-coordinates of the targets are known. More precisely, Eq. (A.5) of the Appendix gives us the a posteriori covariance matrix $\boldsymbol{C}_{\boldsymbol{M}|\boldsymbol{P}}$ as a function of the uncertainty C_P on the world-coordinates of the targets and a priori uncertainty $\boldsymbol{C}_{\boldsymbol{M}}$ on the model's parameter vector. From Eq. (A.5) we deduce that, in order to obtain high accuracy in the estimation of the acquisition

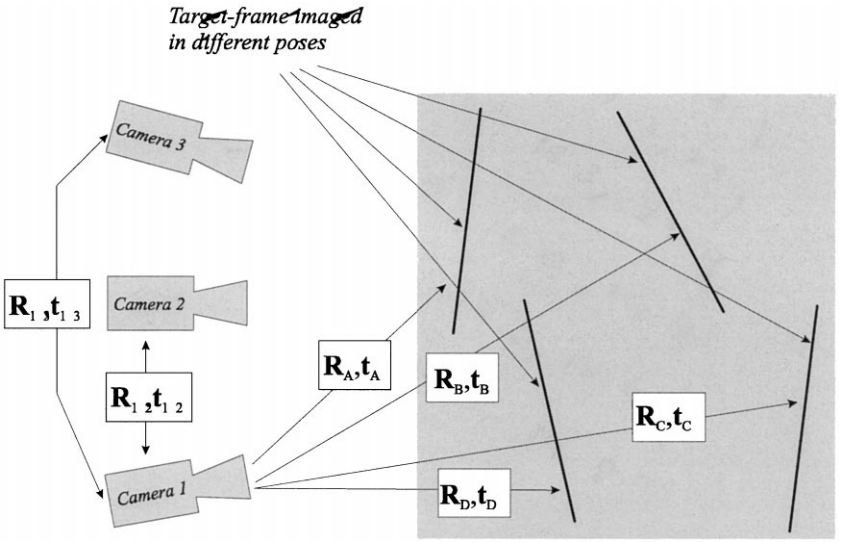

Fig. 4. Schematic description of the calibration setup in which a simple 2D target is used for calibrating an M-camera acquisition system with (in our case $M = 3$). As the positions and the orientations of the calibration frame are a priori unknown, they need to be estimated through calibration.

system's parameters, we need to use either a simple calibration strategy with very accurate information on the target's coordinates or a self-calibration approach for refining our knowledge of the target's coordinates while estimating the camera parameters. Notice that, in the simple calibration case, if the 3D target's coordinates were not accurate enough, the linearization (15) would no longer be a correct approximation of the (nonlinear) model. Notice also that the a priori information on the 3D coordinates of the targets is important for the convergence of the self-calibration process.

As in the simple calibration case, our self-calibration strategy is based on a multi-view multi-camera (MVMC) approach, and the reasons behind this choice are exactly the same: to virtually expand the target-frame and fill up the object space with targets; and to provide the estimator with more "independent" data.

If an M-camera acquisition system is used for acquiring a set of V views per camera of a target-frame that contains H targets, then the following inequality must hold

$$2MVH > 3H + ML + 6(V - 1). \tag{11}$$

On the left hand side of this inequality is the number of constraints (two equations per viewed target per camera), under the simplifying assumption that the image coordinates of all targets can actually be determined. On the right hand side of Eq. (11) is the number of unknowns to be estimated: in fact there are $3H$ coordinates of the targets (such points are usually not exactly coplanar); ML camera parameters (with $L = 11$); and $6(V - 1)$ parameters that characterize the motion of the target-frame.

As we can see from Eq. (11), a single-camera acquisition system obviously requires at least $V = 2$ views for the self-calibration problem to be solvable. It is important to remember, however, that the self-calibration problem is, in general, an undetermined one, and is made solvable by the assumption that the errors on the 3D coordinates of the targets be limited (although not necessarily small) and have zero mean. In general, given the number of views, there is a minimum number H of targets below which the problem is undetermined. In practice, however, due to the ill-conditioning of the problem, it is customary to use a number of targets that will make the problem substantially overdetermined.

4. Implementation

In both simple-calibration and self-calibration the observed data vector $\boldsymbol{p}$ contains the image coordinates of all the targets in all the available images. In order to guarantee an accurate localization (with sub-pixel accuracy) of the fiducial points, the image coordinates of the center (or some other relevant point such as a corner or an edge crossing) of the target must be detected with an appropriate sub-pixel technique. For example, the points to be detected could be the centers (or the vertices) of a grid of black square stickers on a white background. In our case, the centers of circular stickers (contrasting with the background) are considered, and the image coordinates are estimated through template matching [5]. In order to do so, the image coordinates of the fiducial points are first roughly estimated through luminance thresholding. Such coordinates are then refined by comparing the luminance profile, in a neighborhood of the rough estimate, with a synthetic luminance template through a mean-squares optimization process. By doing so we estimated the parameters of the (elliptical) template and, in particular, the refined coordinates of the fiducial point (center of the ellipse). The accuracy of the estimated coordinates of the fiducial points mainly depends on the size of the adopted template [5] and is described by the $2N \times 2N$ covariance matrix $\boldsymbol{C}_{\text{loc}}$, where $N = HVM$ is the number of data points. The localization error in template matching applications is mainly to be attributed to the fact that a comparison is performed between an ideal template and a sampled luminance profile. The pixels that contribute to this error can be numerous and their contributions can be assumed as independent. Therefore it is reasonable to treat the statistical distribution of the localization error as Gaussian and with zero mean. In conclusion, in normal conditions, the covariance matrix can be assumed as being diagonal, i.e;

$$\boldsymbol{C}_{\text{loc}} = \sigma_{\text{loc}}^2 \boldsymbol{I}_{2N}. \tag{12}$$

This is so as

- the localization error of all fiducial points on all available images can be considered to have the same Gaussian distribution, with standard deviation σ_{loc} and zero mean;
- the localization errors of two different points are uncorrelated.

Let $\boldsymbol{c} = [\boldsymbol{c}_1^{\text{T}} \ \cdots \ \boldsymbol{c}_M^{\text{T}}]^{\text{T}}$ be the vector of the (intrinsic and extrinsic) camera parameters of all the M cameras. The data vector $\boldsymbol{p} = [x_1 \ y_1 \ x_2 \ y_2 \ \cdots \ x_N \ y_N]^{\text{T}}$ contains the image coordinates of the visible targets, as extracted from all the V views of all the M cameras. In general we have $N \leqslant HVM$, H being the number of targets on the frame, which becomes an equality if all targets are visible. The model parameters vector $\boldsymbol{m}$ will thus contain:

- the world-coordinates $\boldsymbol{p}_w$ of the H targets (in self-calibration case); the world reference frame could be, for example, attached to the target-frame in its first position;
- the vector $\boldsymbol{c}$ of all camera parameters;
- the vector that describes the rigid motion undergone by the target-frame between two different positions; this motion can be specified with respect to any of the target-frame's position, for example the first one:

$$\boldsymbol{v} = [\boldsymbol{v}_{21}^{\text{T}} \quad \boldsymbol{v}_{31}^{\text{T}} \quad \cdots \quad \boldsymbol{v}_{V1}^{\text{T}}]^{\text{T}},$$

where $\boldsymbol{v}_{k1} \in \Re^6$, $k = 2, \ldots, V$, contains the translational coordinates and the Euler angles that describe the position and the orientation of the target-frame in its kth position, relative to its first position.

The complete parameter vector is thus given by $\boldsymbol{m} = [\boldsymbol{p}_w^{\text{T}} \ \boldsymbol{c}^{\text{T}} \ \boldsymbol{v}^{\text{T}}]^{\text{T}}$. Given an estimate $\hat{\boldsymbol{m}}$ of this vector, we can predict the corresponding data vector $\hat{\boldsymbol{p}}$ through the direct model: $\hat{\boldsymbol{p}} = g(\hat{\boldsymbol{m}})$.

The estimation algorithm is based on an iterative procedure whose aim is to determine the maximum likelihood estimation $\boldsymbol{m}_{\text{ML}}$ of $\boldsymbol{m}$ corresponding to the prediction $\boldsymbol{p}_{\text{ML}} = g(\boldsymbol{m}_{\text{ML}})$ that differs the least (in the MSE sense) from the observed data $\tilde{\boldsymbol{p}}$:

$$\boldsymbol{m}_{\text{ML}} = \min_{\boldsymbol{m}} \{|g(\boldsymbol{m}) - \tilde{\boldsymbol{p}}|^2\}. \tag{13}$$

Due to the large number of unknowns and to the ill-conditioning of the problem, the search for the global minimum of the cost function $C(\boldsymbol{m}) = |g(\boldsymbol{m}) - \tilde{\boldsymbol{p}}|^2$ can be very difficult and could easily return some local minimum instead. Quite clearly, we could avoid this problem by having the minimization process start from a point which is close

enough to the global minimum, but this would require a rather accurate approximation of the final solution to begin with. The multi-view multi-camera approach to simple calibration (and self-calibration), in fact, offers an elegant way to solve this problem through *fractionating* the estimation process in a *multi-resolution* fashion.

In order to better explain the approach, let us first assume that the world-coordinates of the targets are known and consider the case of simple calibration. As a first step, we proceed with the individual calibration of each camera through a separate analysis of the V views of the target-frame. By averaging the V sets of camera parameters obtained from the individual calibrations we obtain a good starting point for the next calibration step. This second step still concerns the individual cameras but it uses all the available views simultaneously. At the end of this process we obtain a refined version of the camera parameter vector $\boldsymbol{c}$ and M estimates of the vector $\boldsymbol{v}$ that describes the target positions. As a last step, we can proceed with a global (all cameras) calibration process based on the simultaneous analysis of all the views. This last minimization step will perform a refinement of the previous estimate of the parameters, which consists of the parameter vector $\boldsymbol{c}$ and an average between all the target-frame's motion vectors. More specifically, the simple-calibration process can be organized as follows:

Step 1: single-view single-camera (SVSC) simple calibration. Each camera is individually calibrated V times, one for each view of the calibration target-frame. Calibration is performed through a non-linear optimization process. In particular, some parameters are first roughly estimated using Tsai's calibration algorithm [17], and then refined while estimating the other parameters through nonlinear optimization.

Step 2: multi-view single camera (MVSC) simple calibration. The calibration of each individual camera is refined through a joint analysis of all the V views of the target-frame. Both the camera parameters and the vector that specifies the positions of the target-frame are estimated through nonlinear optimization. The process starts from the camera parameters estimated through SVSC calibration. As far as the intrinsic parameters, however, an average between V solutions is used.

Step 3: multi-view multi-camera (MVMC) simple calibration. All the M cameras are calibrated simultaneously through a joint analysis of all the views of the target-frame. Camera parameter vectors and target-frame's positions are refined through nonlinear optimization. The starting point for the minimization process is the vector of parameters $\boldsymbol{c}$ obtained through MVSC calibration. As far as the vector $\boldsymbol{v}$ that describes the positions of the target-frame is concerned, its initial vector is given by the average of the estimates obtained through MVSC calibration.

Although the nonlinear minimization process has been implemented in a "multi-resolution" fashion, the intrinsic ill-conditioning of the problem suggests the adoption of a search strategy that exhibits a certain robustness against local minima. Preference should be given to algorithms that are able to explore the space of the unknowns in a more exhaustive fashion than standard *gradient*-based optimization methods. Our best results have been achieved with the Nelder–Mead algorithm [15], which is a modified version of the simplex method.[4]

Often the world-coordinates of the targets are only approximately known. Their "nominal" position, in fact, can be simply determined only if we can accept a significant uncertainty, which can usually be determined as well. In this case, a self-calibration process must be adopted for reducing this uncertainty. The optimal approach to self-calibration is the one that simultaneously estimates camera parameters and targets' world-coordinates through global nonlinear optimization. In order to avoid local minima, the camera parameters can first be roughly estimated through MVMC calibration by assuming that our a priori knowledge of the targets' world-coordinates is not affected by uncertainty. Then the whole vector $\boldsymbol{m}$ containing both targets' world-coordinates and camera parameters can be refined through the minimization of the cost function (13).

If the a priori information on the world-coordinates of the targets is sufficiently accurate, then

[4] An even better robustness against local minima would have been achieved by using techniques of simulated annealing [15], but in this case the computation time would be prohibitive in most applications.

a suboptimal approach to self-calibration can be adopted. In this case, in fact, the camera parameters can first be roughly estimated through MVMC calibration by assuming that the targets' world-coordinates are correct. Then the 3D coordinates of the targets can be refined by taking the previously estimated camera model as given. The process can be iterated until a stable global solution is reached.

Notice that, even when the uncertainty of the world-coordinates of the fiducial points is very limited, we can always estimate the uncertainty of the model's parameters through a linearization of the direct model. More precisely, the uncertainty of the targets' world-coordinates can be treated in the same way as the data uncertainty σ_{loc} associated with the limited precision in the localization of the fiducial points on the image plane (see Eq. (A.5)). For a first and rough estimate of this uncertainty, it is reasonable to assume that the image coordinates of targets that are Gaussian-distributed over a planar frame around their nominal location, are still Gaussian-distributed, and their standard deviation on the image plane is

$$\sigma_{\boldsymbol{P}} = \sigma_{\boldsymbol{P}_w} \frac{f}{z_c},$$

where $\sigma_{\boldsymbol{P}_w}$ is the standard deviation of the uncertainty of the targets' world-coordinates, f is the focal length of the camera and z_c is the distance between camera and targets. As the 3D measurements and the image feature localization error can be considered as statistically independent, a measure of the uncertainty of the global data can be computed. The uncertainty can still be considered as Gaussian, with variance

$$\sigma_{\tilde{p}}^2 = \sigma_{\mathrm{loc}}^2 + \sigma_{\tilde{\boldsymbol{P}}}^2. \tag{14}$$

The global covariance matrix $\boldsymbol{C}_{\tilde{p}} = \sigma_{\tilde{p}}^2 \boldsymbol{I}_{2N}$ is then used to predict the uncertainty $\boldsymbol{C}_{\boldsymbol{M}}$ on the estimated model by using Eq. (A.5).

The knowledge of the uncertainty on the world-coordinates of the targets plays a crucial role in the estimation process as it is used for reducing the search space of the minimization process. Roughly speaking, the smaller the uncertainty on targets' world-coordinates, the stronger the constraints on their refinement. This approach can thus be thought of as a *soft transition* between calibration and self-calibration, depending on accuracy of the a priori knowledge on the targets' world-coordinates. This is a key feature of the proposed technique, as it allows us to fully exploit the available a priori information.

5. Simulation results

In this section we present the results of a series of simulations on synthetic data, which have been carried out in order to verify the correctness of Eq. (A.5), for the performance evaluation of the proposed calibration (self-calibration) algorithm.

If we have very limited a priori information on the model parameters, or even no information at all, the a posteriori covariance matrix $\boldsymbol{C}_{\boldsymbol{M}|\boldsymbol{P}}$ becomes

$$\boldsymbol{C}_{\boldsymbol{M}|\boldsymbol{P}} = (\boldsymbol{G}^{\mathrm{T}} \boldsymbol{C}_{\boldsymbol{P}}^{-1} \boldsymbol{G})^{-1}, \qquad \boldsymbol{G} = \left(\frac{\partial g}{\partial \boldsymbol{m}} \right)_{\boldsymbol{m} = \boldsymbol{m}_{\mathrm{ML}}}. \tag{15}$$

The diagonal elements of $\boldsymbol{C}_{\boldsymbol{M}|\boldsymbol{P}}$ quantify the dispersion of the model parameters, while the extradiagonal elements describe the correlations between parameters.

5.1. The simple calibration problem

In order to verify the accuracy of the results predicted by Eq. (15), we first simulated the simple calibration of a single camera using a single view of the target-frame (SVSC calibration). The considered camera was characterized by a standard TV resolution (720×576 pixel) and a pixel size of $11 \times 11\,\mu\mathrm{m}$. The focal length was 16 mm, the principal point was chosen in the image center and the radial distortion was limited to the first term of Eq. (3) ($k_3 = 10^{-3}\,\mathrm{mm}^{-2}$). The simulated target-frame was a square-shaped planar grid of 256 points with a step-size of 50 mm, placed at a distance of 1100 mm from the camera and tilted 15° with respect both the x_c and y_c axes of the camera reference frame. Its position was chosen in such a way that its center would be seen exactly in the center of the image. We estimated the dispersion of the camera parameters (with respect to the correct values) through the simulations of a series of 400

Table 1
Dispersion of the camera parameters in SVSC calibration (the dispersion σ_i was analytically estimated while $\hat{\sigma}_i$ was estimated through simulation) for different RMS magnitudes of the 2D positional noise. The target-set was a rigid set of 256 point positioned on a square grid with a step-size of 50 mm. The target was placed at a distance of 1100 mm from the camera, tilted about 15 degrees with respect to the x_c and y_c axes of the camera reference frame, and positioned in such a way that its center would be projected onto the center of the image

	$\sigma_D = 0.1$ pixel		$\sigma_D = 0.2$ pixel		$\sigma_D = 0.5$ pixel		$\sigma_D = 1$ pixel	
Parameter	$\hat{\sigma}_i$	σ_i	$\hat{\sigma}_i$	σ_i	$\hat{\sigma}_i$	σ_i	$\hat{\sigma}_i$	σ_i
ϕ (degree)	0.048	0.051	0.096	0.100	0.240	0.259	0.481	0.544
ψ (degree)	0.013	0.012	0.026	0.023	0.065	0.055	0.129	0.119
θ (degree)	0.045	0.040	0.070	0.079	0.227	0.187	0.454	0.416
t_x (mm)	0.927	0.820	1.853	1.641	4.635	3.858	9.267	8.494
t_y (mm)	0.912	0.983	1.823	1.923	3.647	5.055	9.117	10.539
t_z (mm)	1.391	1.312	2.782	2.609	6.956	6.268	13.911	13.624
f (mm)	0.020	0.019	0.040	0.038	0.101	0.082	0.202	0.199
c_x (pixel)	1.219	1.079	2.439	2.158	5.647	5.074	12.193	11.183
c_y (pixel)	1.200	1.294	2.400	2.534	5.999	6.662	11.997	13.887

calibrations in which the available data (image coordinates of the fiducial points) are affected by a Gaussian noise with a given variance. In particular, we considered positional RMS errors of 0.1, 0.5 and 1 pixel in both horizontal and vertical directions. Such errors account for both localization errors and uncertainty of the world-coordinates of the targets. The results are collected in Table 1.

We also carried out some simulations for studying the influence of the target-frame's orientation on the quality of the estimates. Also in this case the simulations concern the SVSC approach, but the results are significant for the MVMC case as well. We used Eq. (15) to estimate the camera parameter's dispersions in two different calibration setups. In the former the above-described calibration pattern (planar surface with 256 targets on a square grid with 50 mm step-size) was tilted 5°, 0° and 0° with respect the x_c, z_c and y_c axes, respectively. In the second setup tilt angles were 15°, 0° and 15°, respectively. In both cases the positional error of each image coordinate of the targets was $\sigma_P = 0.15$ pixel. The results of these simulations are collected in Tables 2–4. As we can see, the quality of the estimates is strongly influenced by the orientation of the target (especially that of the parameters t_z and f). Furthermore, the orientation of the target-set also significantly influences the extra-diagonal elements of the a posteriori correlation matrix $C_{M|P}$, which characterize the correlation between parameters. This reveals a strong correlation between c_y and t_x and between c_x and t_y.

From the above simulation results and further experiments we observed that, in general, choosing an orientation of the target-frame which is far from parallel to the image plane of the camera improves the estimate of f and t_z. However, tilting a planar calibration frame too much would cause the targets to occupy only a portion of the image, with the result of losing the benefits gained through tilting. This is especially true for smaller target-frames. With our choice of calibration frame, however, we found that a good compromise between such two contrasting needs was to use tilt angles of approximately 30°. In particular, tilting the target-frame about 30° around the x_c axis particularly improves the estimates of c_x and t_x, while tilting the frame of about 30° around the y_c axis particularly improved the estimates of c_y and t_y.

We carried out some further investigation for determining the relationship between the number of targets and the dispersion of the estimates (see Fig. 5). The size of the targets was the same for all acquisitions, while their total number was modified by changing the step-size of the square grid where they lied. Also in this case we performed SVSC

Table 2
Accuracy of the camera parameter's estimation method. Two different calibration setups are considered: $\boldsymbol{m}$ is the actual parameter vector, $\hat{\boldsymbol{m}}$ is its estimate obtained through calibration and $\hat{\sigma}_m$ are the computed values of the dispersion of the parameters. In both the considered calibration setups a 2D positional error with an RMS magnitude of 0.15 pixel (for each coordinate) is considered

	First case: $\sigma_D = 0.15$ pixel			Second case: $\sigma_D = 0.15$ pixel		
Parameter	m_0	$\hat{m}$	$\hat{\sigma}_m$	m_0	$\hat{m}$	$\hat{\sigma}_m$
ϕ (degree)	5.0	5.005	0.066	15.0	14.988	0.048
ψ (degree)	0.0	−0.002	0.004	0.0	0.000	0.013
θ (degree)	0.0	−0.015	0.034	15.0	14.993	0.045
t_x (mm)	0.0	0.401	0.944	0.0	−0.176	0.927
t_y (mm)	0.0	0.632	1.123	0.0	−0.255	0.912
t_z (mm)	1100.0	1096.248	8.969	1100.0	1098.457	1.391
f (mm)	16.0	15.944	0.131	16.0	15.978	0.020
c_x (pixel)	360.0	359.468	1.241	360.0	360.236	1.219
c_y (pixel)	288.0	287.161	1.476	288.0	288.338	1.200
k_3($\mathrm{mm}^{-2}10^{-3}$)	1.0	0.990	0.006	1.0	1.000	0.006

Table 3
Correlation between parameters: first calibration setup

	ϕ	ψ	θ	t_x	t_y	t_z	c_x	c_y	f	k_3
ϕ	1.0000	0.0000	0.0000	0.0000	0.8712	0.8567	0.0000	−0.8711	0.8561	−0.0145
ψ	0.0000	1.0000	0.9077	−0.9195	0.0000	0.0000	0.9200	0.0000	0.0000	0.0000
θ	0.0000	0.9077	1.0000	−0.9399	0.0000	0.0000	0.9411	0.0000	0.0000	0.0000
t_x	0.0000	−0.9195	−0.9399	1.0000	0.0000	0.0000	−1.0000	0.0000	0.0000	0.0000
t_y	0.8712	0.0000	0.0000	0.0000	1.0000	0.5399	0.0000	−1.0000	0.5385	−0.0182
t_z	0.8567	0.0000	0.0000	0.0000	0.5399	1.0000	0.0000	−0.5389	1.0000	−0.0313
c_x	0.0000	0.9200	0.9411	−1.0000	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000
c_y	−0.8711	0.0000	0.0000	0.0000	−1.0000	−0.5389	0.0000	1.0000	−0.5376	0.0181
f	0.8561	0.0000	0.0000	0.0000	0.5385	1.0000	0.0000	−0.5376	1.0000	−0.0238
k_3	−0.0145	0.0000	0.0000	0.0000	−0.0182	−0.0313	0.0000	0.0181	−0.0238	1.0000

Table 4
Correlation between parameters: second calibration setup

	ϕ	ψ	θ	t_x	t_y	t_z	c_x	c_y	f	k_3
ϕ	1.0000	0.6061	0.6274	−0.5947	0.9828	0.7831	0.5930	−0.9827	0.7836	−0.0506
ψ	0.6061	1.0000	0.9868	−0.9870	0.5143	0.7570	0.9870	−0.5125	0.7575	−0.0498
θ	0.6274	0.9868	1.0000	−0.9865	0.5489	0.7735	0.9866	−0.5470	0.7740	−0.0507
t_x	−0.5947	−0.9870	−0.9865	1.0000	−0.5159	−0.7405	−1.0000	0.5141	−0.7382	0.0627
t_y	0.9828	0.5143	0.5489	−0.5159	1.0000	0.7273	0.5141	−1.0000	0.7250	−0.0604
t_z	0.7831	0.7570	0.7735	−0.7405	0.7273	1.0000	0.7384	−0.7251	0.9984	−0.1716
c_x	0.5930	0.9870	0.9866	−1.0000	0.5141	0.7384	1.0000	−0.5123	0.7361	−0.0623
c_y	−0.9827	−0.5125	−0.5470	0.5141	−1.0000	−0.7251	−0.5123	1.0000	−0.7229	0.0600
f	0.7836	0.7575	0.7740	−0.7382	0.7250	0.9984	0.7361	−0.7229	1.0000	−0.1216
k_3	−0.0506	−0.0498	−0.0507	0.0627	−0.0604	−0.1716	−0.0623	0.0600	−0.1216	1.0000

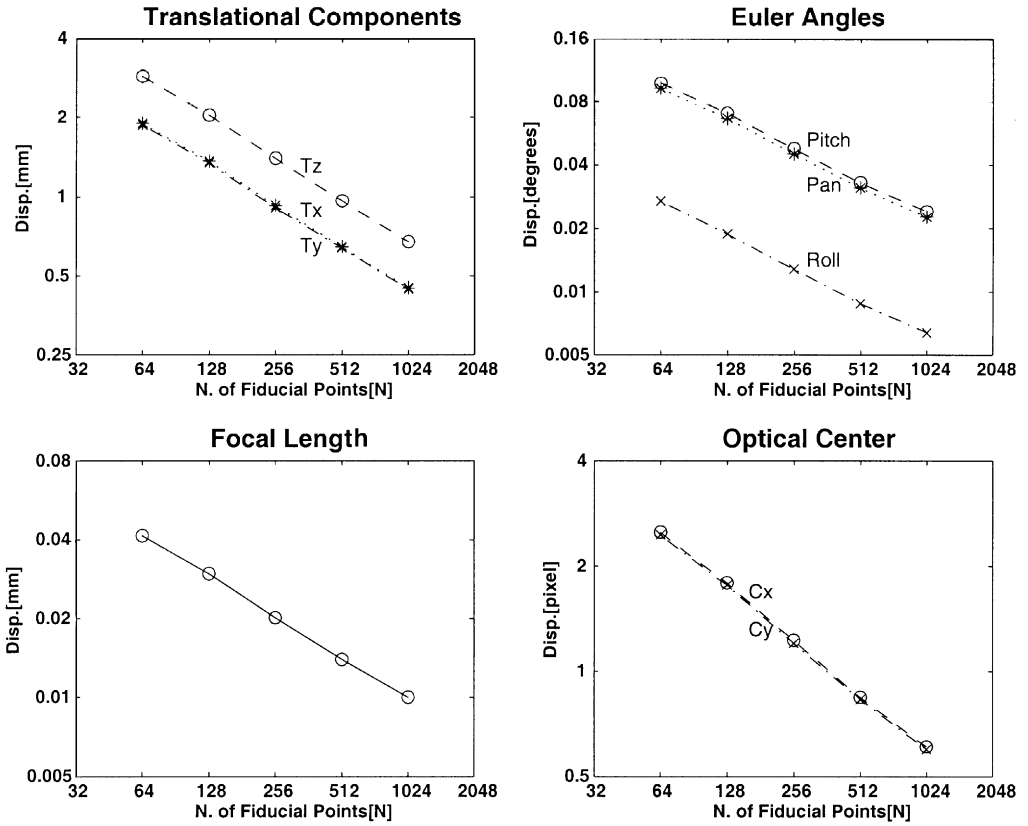

Fig. 5. Dispersion of the camera parameters versus number of fiducial points (log scale) in a single-view single-camera calibration. Notice that the parameter's dispersion is always inversely proportional to the square root of the number of targets, as expected.

calibration, but the results can be extended to the MVMC case. As we can see from Fig. 5, the parameter's dispersion is always inversely proportional to the square root of the number of targets. This suggests that all of the fiducial points contribute with "independent" information to the estimation process. Quite obviously, this fact holds true as long as the image localization error of the fiducial points can be assumed as independent from the density of the targets. In fact, when the localized points are the centers of black circles on a white background, too dense a target would force the radius of such circles to be too small to allow good localization accuracy.

We also observed that, instead of increasing the number of targets, it is possible to consider more views of the target-frame with changes in the orientation and in the camera-frame distance, with very little loss of accuracy (see Table 5). In general, however, the best calibration results are obtained by selecting the orientation according to the above-listed indications.

It is important to notice that the calibration residuals, i.e. the differences between the observed image coordinates and those predicted through the estimated camera model, are very important for judging the quality of the calibration results. Fig. 6 shows the vector field of such residuals. Its vectors, placed in correspondence to the image locations of the targets, represent the (magnified) differences between observed and estimated image coordinates. It is reasonable to expect that the residual vectors of a good calibration will have a random and uniform distribution of orientations,

Table 5
Comparison between SVSC calibration (1024 fiducial points) and a MVSC calibration (256 fiducial points): $\boldsymbol{m}$ is the vector of the actual parameters; $\hat{\boldsymbol{m}}$ is the vector of parameters estimated with a typical calibration; and $\hat{\sigma}_m$ is the computed dispersion of the parameters

	First case: 1 view, 1024 points			Second case: 4 views, 256 points		
Parameter	m_0	$\hat{m}$	$\hat{\sigma}_m$	m_0	$\hat{m}$	$\hat{\sigma}_m$
f (mm)	16.0	15.962	0.064	16.0	15.978	0.068
c_x (pixel)	360.0	359.508	0.745	360.0	360.236	0.825
c_y (pixel)	288.0	287.461	0.833	288.0	288.338	0.744
k_3 (mm^{-2} 10^{-3})	1.0	1.0	0.006	1.0	0.992	0.006

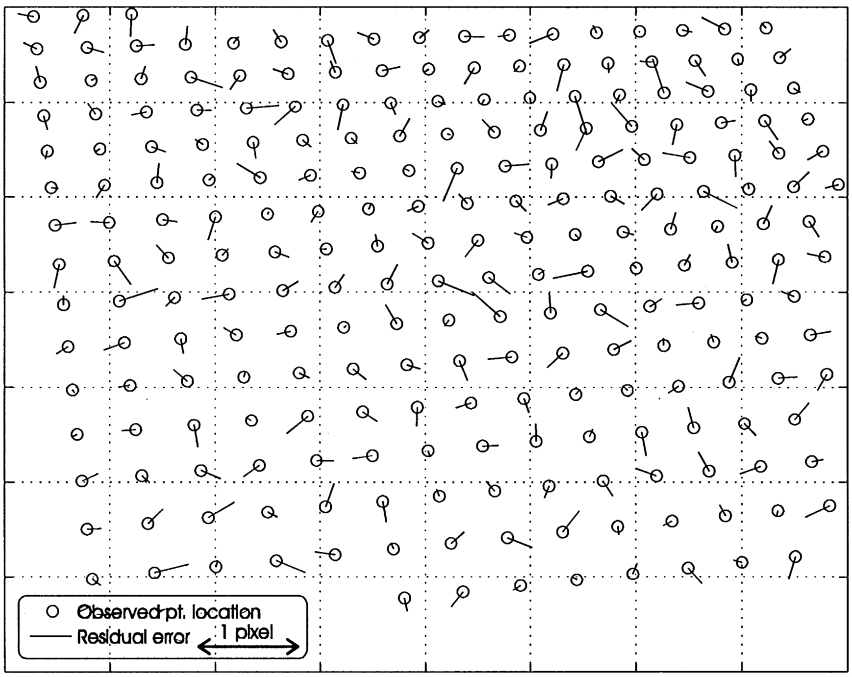

Fig. 6. Vector field of typical calibration residuals (differences between the observed image coordinates and those predicted through the estimated camera model). It is reasonable to expect that the residual vectors of a good calibration will have uniformly distributed orientations, and will be characterized by a magnitude that is comparable with the uncertainty of the observed data.

and will be characterized by a magnitude that is comparable with the uncertainty of the observed data. In fact, if the residuals were larger than the data uncertainty, then we could conclude that the acquisition system is poorly modeled. In the opposite case, the model would be exceedingly accurate for calibration purposes, and could be used for noise characterization.

As a concluding remark, it is important to mention the fact that, when the radial distortion is very modest (high-quality optical lenses), a n accurate estimation of the principal point becomes more difficult, as already shown in the literature [20]. In this case it is necessary to estimate such parameters with some other methods, as shown in [12].

5.2. Self-calibration

In order to quantitatively evaluate the performance of the proposed self-calibration we performed some simulation experiments based on Eq. (15). The simulated setup was based on a trinocular TV-resolution acquisition system whose cameras were positioned at the vertices of a triangle with 600 mm sides. The focal length of the lenses was approximately 12 mm. The target-frame was a nearly-planar surface with 600 targets, approximately arranged at the crosspoints of a square grid of 50 mm step-size, with a zero-mean positional error, so that the reference grid could be used as a "nominal" rough measurement of the target's coordinates.

As a first experiment, we performed self-calibration using three different positions of the target-frame, all modestly tilted with respect to the image planes and placed at a distance of about 1.3 m from them. The feature localization error was about 0.07 pixel. The estimated dispersion σ of the parameters resulted as in Table 6.

As we can see, the estimated residual uncertainty of the targets' world-coordinates were found to be very close to the targets' positional error over the nominal grid. In order to compute this uncertainty, we followed a geometrical approach that was similar to that shown in Fig. 7 for the simplified case of a binocular acquisition system. As we can see, the regions of uncertainty on the image planes are projected onto the 3D space, giving rise to generalized cones. The intersection of such cones represents the corresponding dispersion of the targets' location in the object space.

Table 6
Dispersion of the parameters with a self-calibration setup based on three cameras. The target-frame is acquired in three different positions. In all views the target-set is approximately parallel to the image planes

Parameter	σ
Euler angles	0.002 degree
Translational components	0.0015 mm
Focal length	0.0009 mm
Radial distortion coefficient	$10^{-6}\,\mathrm{mm}^{-2}$
Optical center's coordinates	0.018 pixel
x and y coordinates on target-plane	0.06 mm
Elevation from target-plane	0.11 mm

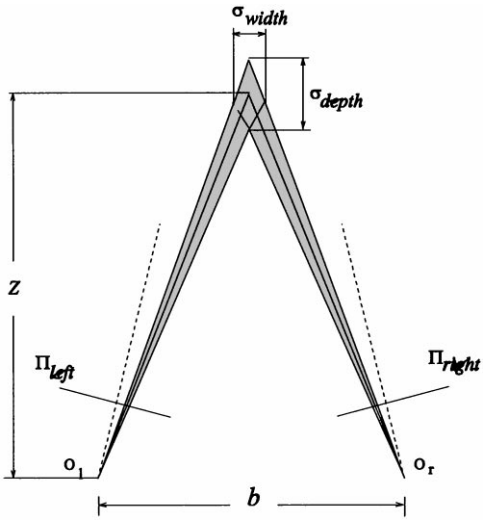

Fig. 7. Estimation of the uncertainty of the coordinates of a 3D point, starting from the region of uncertainty associated with its projection on the image planes.

Table 7
Parameter dispersion for a trinocular camera system, with five views of the target-frame. In two of the views, the target is significantly tilted with respect to the image planes

Parameter	σ_m
x and y coordinates on the target plane	0.045 mm
Elevation from the target plane	0.08 mm

Quite clearly, the results improve significantly if we add two more views in which the target-frame is more tilted with respect to the image planes. In this case the world-coordinates were improved by the self-calibration approach, as shown in Table 7.

Such results confirm the importance, for calibration as well as self-calibration, of using several views of the target-frame in a variety of positions.

6. Experimental results

In order to test the reliability of our (simple/self) calibration methods, we performed a series of tests

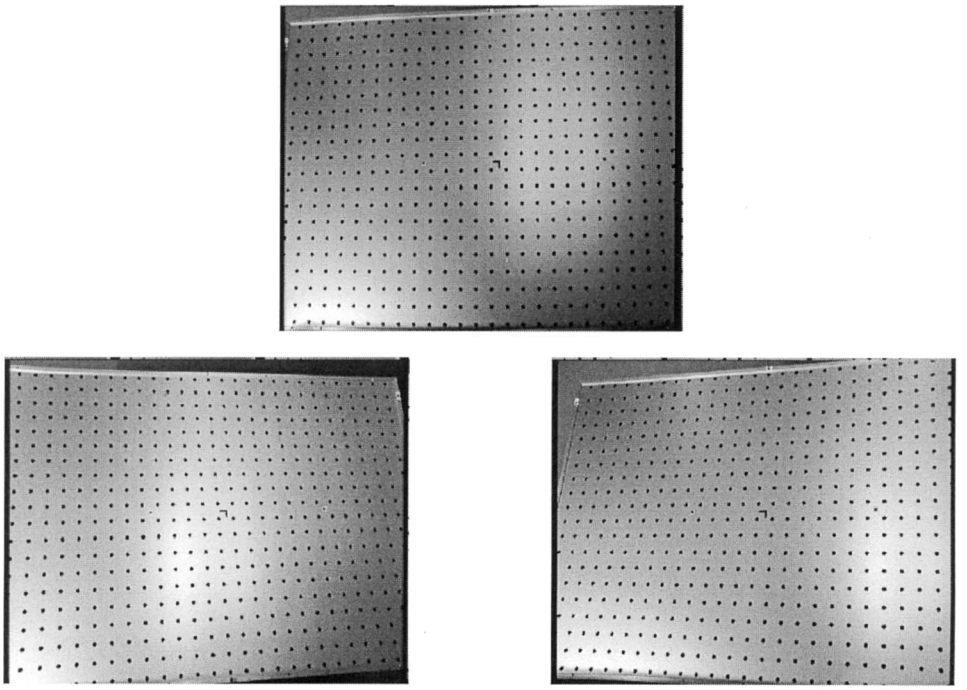

Fig. 8. A trinocular view of the *high-quality* target-frame.

in a variety of experimental conditions. In this section, we present the results of two series of tests: the former is relative to a high-quality target-frame whose targets are not just known in their nominal 3D coordinates but have been accurately measured through a photogrammetric procedure in order to quantify the positional displacement relative to the nominal coordinates. For the second experiment we used a less expensive target frame, of which we only knew the nominal 3D coordinates of the targets. For all our experiments we adopted an acquisition system made of three standard TV-resolution CCD cameras.

The first test, was conducted with three Sony DCX950 color cameras, each with a 3CCD sensor of 2/3″ (diagonal size). The cameras, whose nominal focal length was 12 mm, were mounted on a rigid frame at the vertices of a triangle with a baseline of 800 mm and the other two sides of about 500 mm. The volume to be calibrated was about 2 m wide, 1.5 m deep and 1.5 m tall, placed at an average distance of about 2 m from the camera set.

The adopted high-quality target-set was made of a grid of 29×20 fiducial points (centers of black circular stickers with a radius of 12.5 mm), placed on the surface of an aluminium "wafer" with honeycomb structure for improved rigidity and light weight (see Fig. 8). The grid's nominal step-size was 50 mm, while the exact world-coordinates of the targets had been measured through classical photogrammetric methods (whose accuracy was better that 0.1 mm). With this target-set we performed a series of simple calibration experiments, in which the target-frame was placed in five different positions, two of which were significantly tilted with respect to the image planes.

As a first step, we determined the image coordinates of the targets with sub-pixel accuracy through a procedure based on template matching. The estimate of the coordinates of the center of a matched elliptic template is affected by uncertainty due to a number of reasons, including the finite resolution of the digital sensor and the template's modeling uncertainty. As the causes of the localization

uncertainty are numerous and nearly independent from each other, it is reasonable to treat the statistical distribution of the localization error as Gaussian. With elliptic spots having an area of approximately 20–30 pixels (such as in our case), the localization error is found to be approximately distributed as a Gaussian with a standard deviation of about 0.1–0.2 pixels, as suggested in [5] for a situation that was very similar to ours. In order to analytically verify our experimental results, we chose a std. dev. of 0.15 pixels. With this uncertainty on the image localization of the fiducial points, the estimated dispersion of the camera parameters was found to be in complete agreement with that computed through Eq. (15). We also carried out a series of self-calibration experiments using five different positions of the same target-frame. By using self-calibration for refining the nominal (low-accuracy) world-coordinates of the targets, we obtained the results listed in Table 8. As we can see, there is a good agreement between the values obtained experimentally and those predicted analytically through Eq. (15) applied to the self-calibration experiment. In these measurements the z_c axis is chosen to correspond to the average of the optical axes of the three cameras. Due to the limited baseline of the triangular camera frame, along this direction the dispersion of the measurements is larger than in the other two directions. The results of Table 8 confirm the accuracy of the self-calibration approach.

When dealing with a smaller calibration volume, it is much easier to construct low-cost target-frames. In order to show this fact, we carried out a series of experiments on a scene volume of about $60 \times 60 \times 60$ cm, placed at an average distance of about 80 cm from the camera set. For the second set of experiments, we adopted three B/W SONY XC77CE cameras with 2/3″ (diagonal size) CCD sensors, with a nominal focal length of 16 mm, placed on a rigid frame at the vertices of a triangle that was approximately 40 cm tall and had a baseline of about 60 cm (see Fig. 9). We used an inexpensive target-set made with an A4-size sheet of laser-printed paper glued on a flat surface for testing our self-calibration MCMV approach. The target-set was made of 10×14 circular dots with a radius of 5 mm, laser-printed at a resolution of 600 dpi. The targets were nominally positioned at the crosspoints of a square grid with a step-size of 20 mm (see Fig. 10), and the extracted image-coordinates were those corresponding to the centers of the circular dots.

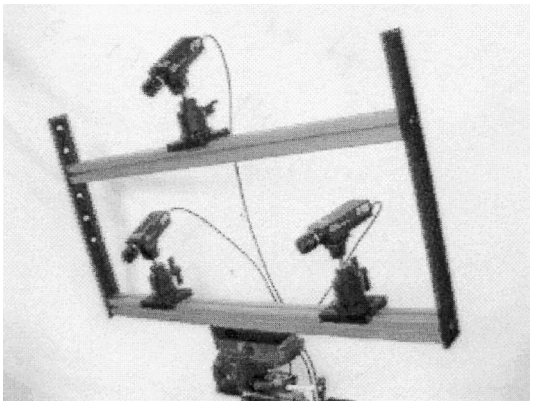

Fig. 9. Trinocular system used for the second set of calibration (and self-calibration) experiments.

In Fig. 11 the a priori locations (defined directly on the drawing to be printed) of the fiducial points and the corresponding positional error (computed using the a posteriori coordinates estimated through self-calibration) are visualized. As we can see, the action of the paper drive system of the laser printer introduces a positional error, which is also confirmed through visual inspection by means of a high-precision ruler. In conclusion, typical laser printers are able to guarantee high resolution but poor positional accuracy; therefore this type of

Table 8
Self-calibration results using a high-quality target-frame. Dispersion of the estimated world-coordinates of the targets with respect to the actual (measured) ones: $\hat{\sigma}_m$ is the dispersion estimated through several self-calibrations relative to different image acquisitions. The dispersion σ_m is computed analytically and is given as a range of values because it depends on which acquisition is being considered (we only have limited a priori information on the views)

Parameter	$\hat{\sigma}_m$	σ_m
x coordinate (mm)	0.064	0.03–0.07
y coordinate (mm)	0.078	0.03–0.07
z coordinate (mm)	0.1108	0.05–0.12

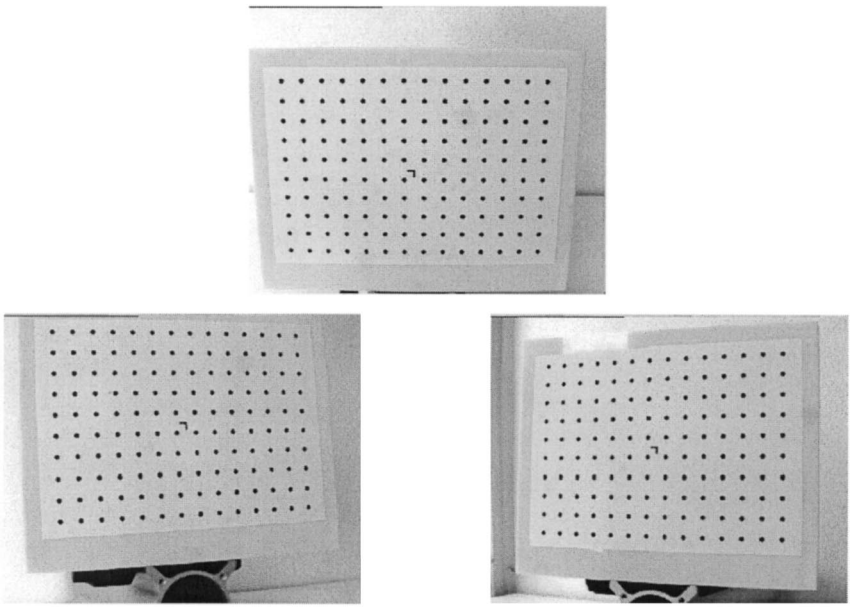

Fig. 10. A trinocular view of the low-cost target-frame (laser-printed sheet of paper glued to a planar glass surface).

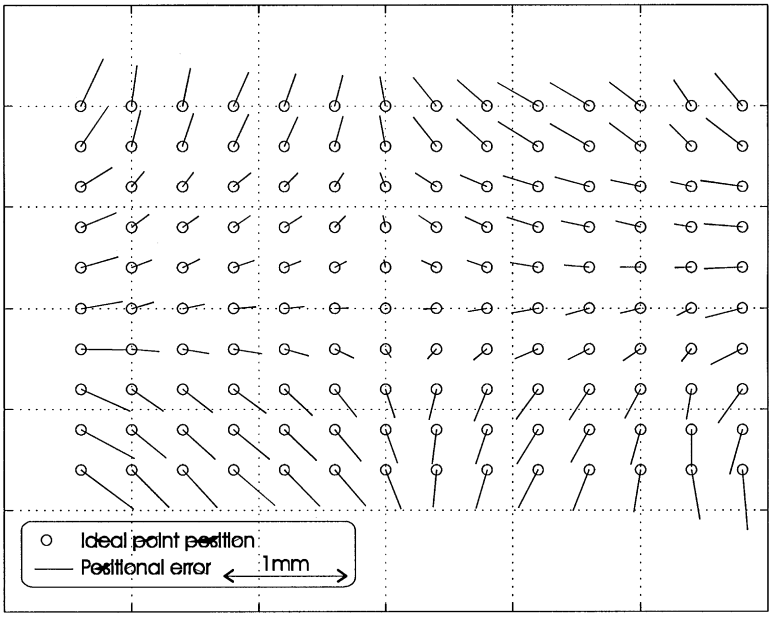

Fig. 11. A priori locations of the fiducial points and corresponding positional errors computed from the coordinates estimated through self-calibration. Notice that the error pattern corresponds to the deformation of the sheet of paper due to the mechanism for paper traction in the laser printer.

target-set is suitable for self-calibration but not so much for simple calibration. However, we must keep in mind that the systematic error caused by the dragging action of the printer's mechanics is not guaranteed to be unbiased. As a consequence, as it is reasonable to expect, better results are achieved when using targets that are printed on a large-format sheet of paper through a professional ink-jet plotter. This also was confirmed experimentally.

Anyway, in our experiments we found a good agreement between the a priori and the a posteriori world-coordinates of the fiducial points; therefore these types of target-sets can be effectively adopted in simple calibration applications.

We finally carried out some experiments for evaluating the maximum accuracy that can be reached by a 3D reconstruction procedure based on stereo-correspondences, when using the above-described trinocular camera system, calibrated with the proposed MCMV method. In order to do so, we considered a set of views of the target-frame that had not already been used for calibration. We estimated the distance between fiducial points through back-projection of their image-coordinates. The obtained accuracy was better than 0.2 mm, with an average distance of 2000 mm between cameras and object and a maximum object size of approximately 1500 mm (corresponding to a relative accuracy of about 130 ppm). Similar results were found with some other 3D reconstruction experiments, performed on a variety of test objects.

7. Conclusions

In this paper we presented a simple and effective technique for calibrating CCD-based multi-camera acquisition systems. The proposed method was proven to be capable of highly accurate results even when using very simple calibration target-sets and low-cost imaging devices, such as standard TV-resolution cameras connected to commercial frame-grabbers. In fact, the performance of our calibration approach is found to be about the same as that of other traditional calibration methods based on 3D target sets [17,20], but our planar target is much easier to construct, carry and handle.

The proposed calibration strategy is based on a "multi-view, multi-camera" approach, whose aim is to calibrate the multi-camera system through the analysis of a number of views of a simple calibration target-set, placed in different (unknown) positions. Furthermore, the method is based on a self-calibration approach, which is able to refine the a priori knowledge of the world-coordinates of the targets (even when such information is very poor) while estimating the parameters of the camera model.

The proposed method was proven to be flexible enough to allow the user to incorporate the a priori knowledge on the targets' locations in a variety of ways.

The accuracy and the robustness of the proposed calibration strategy was confirmed by a series of experiments, carried out with a variety of calibration setups. The accuracy of the analytical prediction of the uncertainty of the calibration results was also proven through simulation experiments.

Further research is currently being carried out in order to minimize the complexity of the calibration (self-calibration) process, by simplifying as much as possible the structure of the target-set while improving the management of the a priori information on the calibration setup.

Appendix A. Additional remarks on inverse problems

In this appendix, we provide some additional information on inverse problems in order to explain how to derive Eqs. (9) and (10). For a definition of the adopted notation, see Section 2.3.

Let us first define $f_{P,M}(\boldsymbol{p},\boldsymbol{m})$ as the p.d.f. that statistically describes the whole acquisition system. From this function we can derive all marginal and conditional p.d.f.'s of interest. For example, we have

$$f_M(\boldsymbol{m}) = \int_{\mathscr{P}} f_{P,M}(\boldsymbol{p},\boldsymbol{m})\, d\boldsymbol{p},$$

which incorporates our a priori information on the model's parameters. Notice that the term a priori specifies information that is not based on the data

$\boldsymbol{p}$. Conversely, the a posteriori information on the model's parameters is a p.d.f. of the form

$$f_{M|P}(\boldsymbol{m}|\boldsymbol{p}) = \frac{f_{P,M}(\boldsymbol{p},\boldsymbol{m})}{\int_{\mathcal{M}} f_{P,M}(\boldsymbol{p},\boldsymbol{m})\,d\boldsymbol{m}},$$

conditioned by the knowledge of the data vector $\boldsymbol{p}$. In turn, the data vector $\boldsymbol{p}$ is known in terms of the observations $\tilde{\boldsymbol{p}}$, as the model is affected by some degree of uncertainty. As the model uncertainty is specified by $f_{\tilde{P}|P}(\tilde{\boldsymbol{p}}|\boldsymbol{p})$, we can write:

$$f_{\tilde{P},M}(\tilde{\boldsymbol{p}},\boldsymbol{m}) = f_M(\boldsymbol{m}) \int_{\mathcal{P}} f_{\tilde{P}|P}(\tilde{\boldsymbol{p}}|\boldsymbol{p}) f_{P|M}(\boldsymbol{p}|\boldsymbol{m})\,d\boldsymbol{p}$$

and

$$f_{M|P}(\boldsymbol{m}|\boldsymbol{p}) = \frac{f_M(\boldsymbol{m})\int_{\mathcal{P}} f_{\tilde{P}|P}(\tilde{\boldsymbol{p}}|\boldsymbol{p}) f_{P|M}(\boldsymbol{p}|\boldsymbol{m})\,d\boldsymbol{p}}{\int_{\mathcal{M}} f_M(\boldsymbol{m})\int_{\mathcal{P}} f_{\tilde{P}|P}(\tilde{\boldsymbol{p}}|\boldsymbol{p}) f_{P|M}(\boldsymbol{p}|\boldsymbol{m})\,d\boldsymbol{p}\,d\boldsymbol{m}},$$

whose denominator plays the role of a normalization factor; therefore we can also write

$$f_{M|P}(\boldsymbol{m}|\boldsymbol{p}) = \alpha \cdot f_M(\boldsymbol{m}) \int_{\mathcal{P}} f_{\tilde{P}|P}(\tilde{\boldsymbol{p}}|\boldsymbol{p}) f_{P|M}(\boldsymbol{p}|\boldsymbol{m})\,d\boldsymbol{p},$$

which represents the solution of the inverse problem in its general formulation. As a matter of fact, from $f_{M|P}(\boldsymbol{m}|\boldsymbol{p})$ it is possible to extract any type of information we need on the model parameters (e.g. mean values, median values, maximum likelihood values, errors etc.).

When all sources of uncertainty that affect our inverse problem can be modeled by a zero-mean Gaussian p.d.f., it is possible to predict the accuracy of the solution of the inverse problem in quite a general fashion. In fact we have:

$$\begin{aligned} f_{M|P}(\boldsymbol{m}|\boldsymbol{p}) = K \cdot \exp\Bigg\{ & -\frac{1}{2}(g(\boldsymbol{m}) - \tilde{\boldsymbol{p}})^{\mathrm{T}} \boldsymbol{C}_P^{-1} (g(\boldsymbol{m}) - \tilde{\boldsymbol{p}}) \\ & -\frac{1}{2}(\boldsymbol{m} - \bar{\boldsymbol{m}})^{\mathrm{T}} \boldsymbol{C}_M^{-1} (\boldsymbol{m} - \bar{\boldsymbol{m}}) \Bigg\} \\ = K \cdot \exp[& -h(\boldsymbol{m})] \end{aligned} \tag{A.1}$$

where $\bar{\boldsymbol{m}} = E[\boldsymbol{m}]$ is the a priori estimate of the model's parameter vector, which can be derived from $f_M(\boldsymbol{m})$. Similarly, $\boldsymbol{C}_M$ is the a-priori covariance matrix of the model's parameter vector. Similarly, $\boldsymbol{C}_P$ is the covariance matrix associated to both the "forward modeling uncertainty" and the "experimental uncertainty" (i.e. the statistical relationship between $\tilde{\boldsymbol{p}}$ and $\boldsymbol{p}$). Finally K is, as usual, a normalization factor.

The solution of a general inverse problem and, in particular, of our calibration problem, is the value of $\boldsymbol{m}$ that maximizes $f_{M|P}(\boldsymbol{m}|\boldsymbol{p})$, i.e. the following maximum likelihood estimation

$$\boldsymbol{m}_{\mathrm{ML}} = \max_{\boldsymbol{m}} [f_{M|P}(\boldsymbol{m}|\boldsymbol{p})] = \min_{\boldsymbol{m}} [h(\boldsymbol{m})].$$

If the direct model $g(\cdot)$ is linear, then instead of writing $\boldsymbol{p} = g(\boldsymbol{m})$ we can write $\boldsymbol{p} = \boldsymbol{G}\boldsymbol{m}$, therefore Eq. (A.1) can be rewritten as:

$$f_{M|P}(\boldsymbol{m}|\boldsymbol{p}) = K \cdot \exp\left[-\frac{1}{2}(\boldsymbol{m} - \hat{\boldsymbol{m}})^{\mathrm{T}} C_M^{-1} (\boldsymbol{m} - \hat{\boldsymbol{m}}) \right], \tag{A.2}$$

where

$$\boldsymbol{m}_{\mathrm{ML}} = (\boldsymbol{G}^{\mathrm{T}} \boldsymbol{C}_P^{-1} \boldsymbol{G} + \boldsymbol{C}_M^{-1})^{-1} (\boldsymbol{G}^{\mathrm{T}} \boldsymbol{C}_P^{-1} \tilde{\boldsymbol{d}} + \boldsymbol{C}_M^{-1} \bar{\boldsymbol{m}}), \tag{A.3}$$

$$\boldsymbol{C}_{M|P} = (\boldsymbol{G}^{\mathrm{T}} \boldsymbol{C}_P^{-1} \boldsymbol{G} + \boldsymbol{C}_M^{-1})^{-1}. \tag{A.4}$$

Eq. (A.2) shows that, when the forward problem is linear and the uncertainty can be modeled as Gaussian, the a posteriori p.d.f. in the model space is Gaussian.[5]

In the case in which the direct model $g(\cdot)$ is nonlinear, as happens with camera calibration problems (see Eqs. (1)–(4)), an approximate result can still be achieved through an iterative computation of the maximum likelihood estimation $\boldsymbol{m}_{\mathrm{ML}}$, which is based on a linearization of $g(\boldsymbol{m})$ about $\boldsymbol{m}_{\mathrm{ML}}$ for estimating the a posteriori covariance

$$\boldsymbol{C}_{M|P} = (\boldsymbol{G}^{\mathrm{T}} \boldsymbol{C}_P^{-1} \boldsymbol{G} + \boldsymbol{C}_M^{-1})^{-1}, \tag{A.5}$$

where

$$\boldsymbol{G} = \left(\frac{\partial g}{\partial \boldsymbol{m}} \right)_{\boldsymbol{m} = \boldsymbol{m}_{\mathrm{ML}}} \tag{A.6}$$

[5] The term *least-square estimation* is justified by the fact that $\tilde{\boldsymbol{m}}$ maximizes $f_{M|P}(\boldsymbol{m}|\boldsymbol{p})$; therefore it also minimizes the quadratic expression $(\boldsymbol{G}\boldsymbol{m} - \tilde{\boldsymbol{d}})^{\mathrm{T}} \boldsymbol{C}_P^{-1} (\boldsymbol{G}\boldsymbol{m} - \tilde{\boldsymbol{d}}) + (\boldsymbol{m} - \bar{\boldsymbol{m}})^{\mathrm{T}} \boldsymbol{C}_M^{-1} (\boldsymbol{m} - \bar{\boldsymbol{m}})$.

is the Jacobian of the forward model. As we can see, the reliability of such results depends on how well $g(\cdot)$ can be linearized about $\boldsymbol{m}_{\mathrm{ML}}$.

References

[1] N. Ayache, Artificial Vision for Mobile Robots, MIT Press, 1991.

[2] Y.I. Aziz, H.M. Karara, Direct linear transformation into object space coordinates in close-range photogrammetry, in: Proc. Symp. Close-Range Photogrammetry, University of Illinois at Urbana-Champaign, Urbana, 1971, pp. 1–18.

[3] D. Barbe, Imaging devices using the charge-coupled concept, in: Proc. IEEE, Vol. 63, No. 1, January 1975.

[4] H.A. Beyer, Some aspects of the geometric calibration of CCD-cameras, ISPRS Intercomm. Conference on Fast Processing of Photogrammetric Data, Interlaken, 1987.

[5] H.A. Beyer, Geometric and radiometric analysis of a CCD-camera based photogrammetric close-range system, Ph.D. Thesis, No. 51, Institut für Geodäsie und Photogrammetrie, ETH, Zürich, May 1992.

[6] W. Faig, Manual of Photogrammetry, 4th edition, American Society of Photogrammetry, 1990.

[7] O. Faugeras, Stratification of three-dimensional vision: Projective, affine, and metric representations, Journal of the Optical Society of America (Optics, Image Science and Vision) 12 (3) (March 1995) 465–84.

[8] G. Ferrigno, N.A. Borghese, A. Pedotti, Pattern recognition in 3D automatic human motion analysis, ISPRS Journal of Photogrammetry and Remote Sensing 45 (1990) 227–246.

[9] A. Gruen, H. Beyer, System calibration through self-calibration, Invited paper, Workshop on Camera Calibration and Orientataion in Computer Vision, XVII ISPRS Congress, Washington, D.C., August 1992.

[10] C.F. Laizet, Determination of video cameras parameters in stereoscopic mode, Fourth European Workshop on Three-Dimensional Television, Rome, 20–21 October 1993.

[11] R. Lenz, U. Lenz, New developments in high resolution image acquisition with CCD area sensors, Optical 3-D Measurement Techniques II, Gruen/Kahmen Editors, Wichmann, 1993.

[12] R. Lenz, U. Lenz, New developments in high resolution image acquisition with CCD area sensors, Optical 3-D Measurement Techniques II, Gruen/Kahmen (Eds.), Wichmann, 1993.

[13] Y. Otha, T. Kanade, Stereo by intra- and inter-scanline search using dynamic programming, IEEE Trans. PAMI 7 (2) (1985) 139–154.

[14] F. Pedersini, A. Sarti, S. Tubaro, 3D surface reconstruction from horizons, International Workshop on Synthetic-Natural Hybrid Coding and Three-Dimensional (3D) Imaging (IWSNHC3DI'97), Rhodes, Greece, 5–9 September 1997.

[15] W. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery, Numerical Recipes – The Art of Scientific Computing, Cambridge University Press, 1986.

[16] A. Tarantola, Inverse Problem Theory, Elsevier, 1987.

[17] R.Y. Tsai, A versatile camera calibration technique for high-accuracy 3D machine vision metrology using off-the-shelf TV cameras and lenses, IEEE Journal on Robotics and Automation RA-3 (4) (August 1987) 323–344.

[18] R. Vaillant, O.D. Faugeras, Using extremal boundaries for 3D object modeling, IEEE Trans. Pattern Analysis and Machine Intelligence 14 (2) (February 1986) 157–173.

[19] L. Van Gool, A. Zisserman, Automatic 3D model building from video sequences, European Transactions on Telecommunications 8 (4) (July–August 1997) 369–378.

[20] G.Q. Wei, S. De Ma, Implicit and explicit camera calibration: Theory and experiments, IEEE Trans. PAMI 16 (5) (May 1994) 469–480.

[21] J. Weng, P. Cohen, M. Herniou, Camera calibration with distortion model and accuracy evaluation, IEEE Trans. on PAMI 14 (10) (October 1992) 965–980.

[22] Y. Yakimowsky, R. Cunningham, A system for extracting three-dimensional measurements from a stereo pair of TV cameras, Computer Graphics and Image Processing 7 (1978) 195–210.

[23] Z. Zhang, O. Faugeras, 3D Dynamic Scene Analysis, Springer-Verlag, 1992.

[24] D. Zhang, Y. Nomura, S. Fujii, Error analysis and optimization of camera calibration, in: Proc. of IEEE/RSJ Internat. Workshop on Intelligent Robots and Systems IROS-91, Osaka, Japan, 3–5 November 1991, pp. 292–296.