Brief Papers

Combined Evolution Strategies for Dynamic Calibration of Video-Based Measurement Systems

P. Cerveri, A. Pedotti, and N. A. Borghese

***Abstract*—Calibration is a crucial step to obtaining three-dimensional (3-D) measurement using video camera-based stereo systems. Approaches based on epipolar geometry are particularly appealing as there is no need to know the 3-D position of the control points *a priori* and because the solution is found by solving a set of linear equations through matrix manipulation. Indeed, all the parameters can be determined except for the pair of principal points, which poses a considerable drawback. Whereas in low-accuracy systems (two-dimensional measurement error $<$ 0.2 pixels) such points can be assumed to lie at the image center without degrading the overall 3-D accuracy, in high-accuracy systems their true position must be computed accurately. In this case, all the calibration parameters (including the principal points) can still be estimated through epipolar geometry, but it is necessary to minimize a highly nonlinear cost function. It is shown here that by combining two evolutionary optimization strategies this minimization can be carried out, both efficiently (in quasi-real time) and reliably (avoiding local minima). The resulting strategy, which we call enhanced evolutionary search (EES), allows the full calibration of a stereo system using only a rigid bar; this simplicity is a definite step forward in stereo-camera calibration. Moreover, EES can be applied to a wide range of applications where the cost function contains complex nonlinear relationships among the optimization variables.**

***Index Terms*—Covariance matrix, epipolar geometry, evolution strategies, optimization, stereo camera calibration.**

I. Introduction

Video camera-based stereo systems are used widely in many different fields, including close-range photogrammetry, robotic vision, computer-aided design, biomechanics, and virtual reality. Two major applications are the measurement of the three-dimensional (3-D) shape of objects (3-D scanners [1]) and the 3-D reconstruction of motion (trackers or motion capture [2]). The input for most systems is a set of matched features from each of the two camera image streams. Tailored low-level hardware [3] or software processing [4] detects these features and returns their accurate 3-D measurement. To transform a pair of two-dimensional (2-D) features into a feature positioned within the 3-D space correctly, the video camera-based stereo system must be calibrated, i.e., the parameters that determine its geometry must be determined. Such parameters are the orientation and location of one camera with respect to the other (external parameters), and the focal length and the principal point of each camera (internal parameters) [see Fig. 1(a)]. The principal points, defined as the intersection of the optical axis with the image plane, are of particular importance as they are the parameters involved in the optimization.

In the classical calibration approach, the parameters are computed from a set of 3-D control points positioned precisely in an external 3-D reference frame [5], [6]. Reliable parameters are obtained by distributing the points throughout the entire working volume, which requires large support structures. These structures must be constructed accurately to achieve high accuracy in estimating the parameters; thus, this approach is very expensive. Moreover, these large structures must be moved and positioned accurately within the working volume, making calibration both laborious and complex. To avoid this, self-calibration algorithms have been developed, mainly in photogrammetry with bundle adjustment [7], [8] and in computer vision with the structure-from-motion (SfM) approach [9]–[11]. The most general approach is that of bundle adjustment, both the 3-D position of the control points and the calibration parameters being estimated. This formulation leads to a nonlinear least squares estimate, which is susceptible to local minima unless initialized carefully. A more appealing solution is that offered by SfM[1] ; in this case, calibration is reduced to the estimation of the nine entries of a 3×3 matrix, called the fundamental matrix [13], (3). The result is a linear homogeneous system that does not need any initialization and can be solved by matrix manipulation. However, SfM is limited in that it cannot estimate more than eight calibration parameters [14], (1), which are the relative location and orientation of one camera with respect to the other and the two focal lengths [11], [15], [16]. The principal points must be determined apart. As a first approximation, they can be assumed to lie at the image center; only when the measurement error is high is this approximation considered adequate [17]. In reality, the principal points can be offset with respect to the image center by several pixels due to imperfect assembling of the optical system (Table IV). This leads to relevant error in

Manuscript received February 28, 2000; revised November 9, 2000.

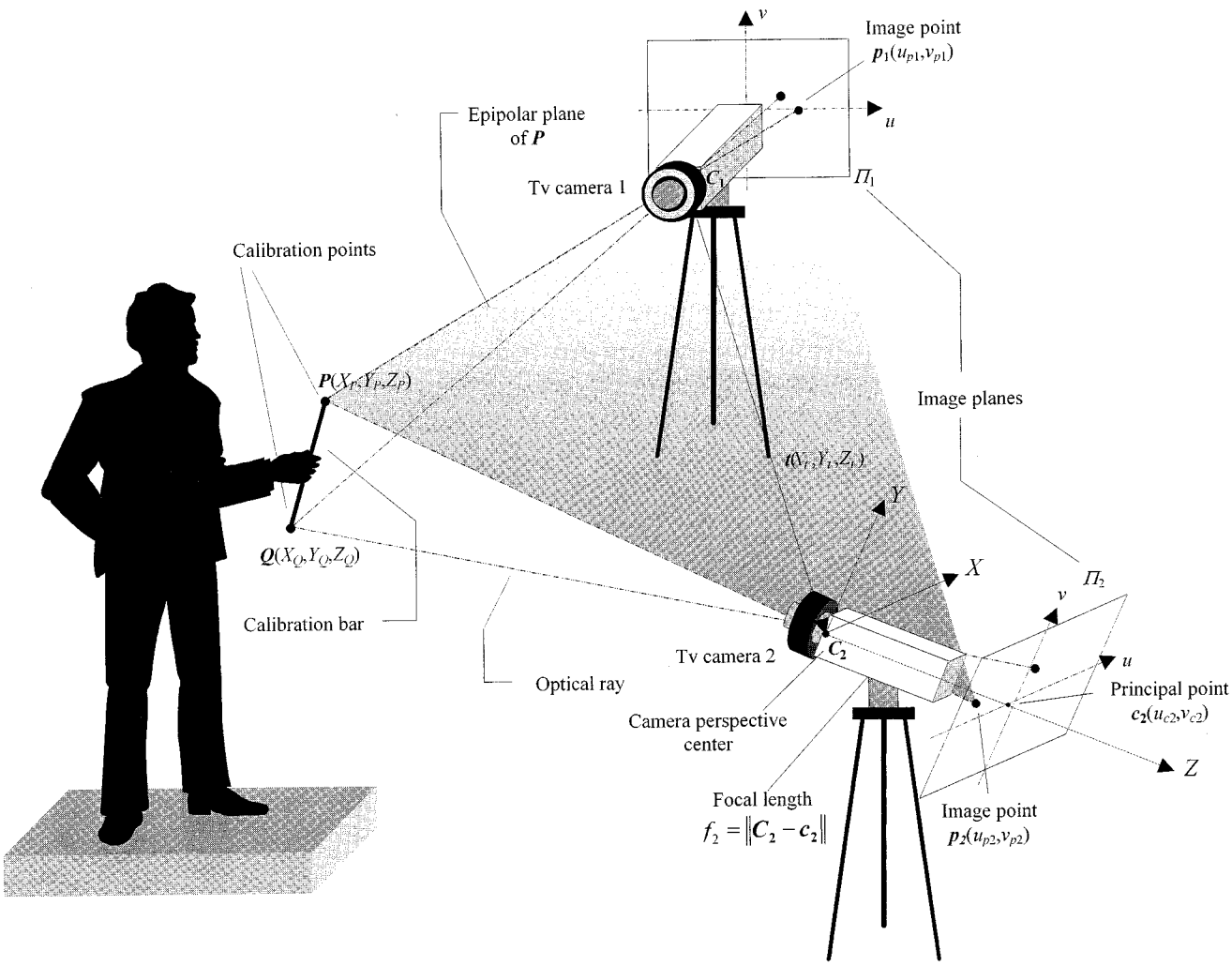
P. Cerveri is with the Bioengineering Department, Politecnico di Milano and Centro di Bioingegneria, Fondazione ProJuventute, 20148 Milano, Italy (e-mail: cerveri@biomed.polimi.it) and also with the Laboratory of Human Motion Analysis and Virtual Reality, (MAVR), Istituto Neuroscienze e Bioimmagini CNR, 20090 Segrate, Milano, Italy.

A. Pedotti is with the Bioengineering Department, Politecnico di Milano and Centro di Bioingegneria, Fondazione ProJuventute, 20148 Milano, Italy.

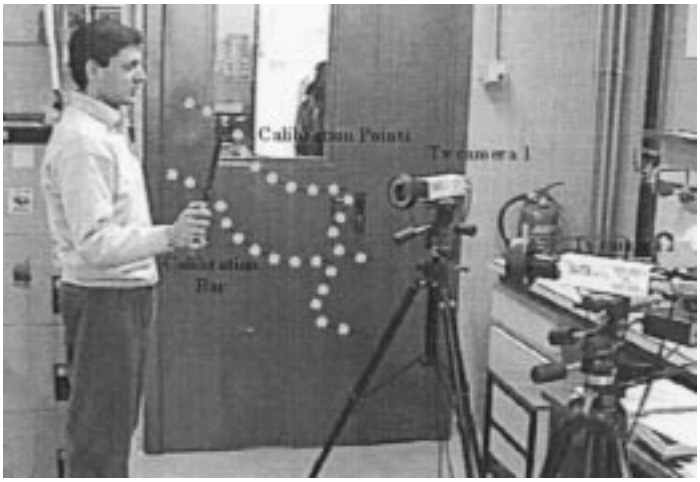
N. A. Borghese is with the Laboratory of Human Motion Analysis and Virtual Reality (MAVR), Istituto Neuroscienze e Bioimmagini CNR, 20090 Segrate, Milano, Italy (e-mail: borghese@inb.mi.cnr.it).

Publisher Item Identifier S 1089-778X(01)02397-9.

[1]SfM deals with the estimate of the internal parameters of one camera, its motion parameters, and the 3-D position of a set of control points. This has been linked to calibration of a stereo system in [12], where it is shown that a scene surveyed by a single camera in two positions is equivalent to surveying it with two cameras.



(a)



(b)

Fig. 1. (a) Geometrical setup and the calibration parameters: relative location $\boldsymbol{T}(X_t, Y_t, Z_t)$, relative orientation $\boldsymbol{R}(\omega, \phi, \kappa)$, focal length f_1 and f_2, and principal points $\boldsymbol{c}_1(u_{\boldsymbol{c}_1}, v_{\boldsymbol{c}_1})$ and $\boldsymbol{c}_2(u_{\boldsymbol{c}_2}, v_{\boldsymbol{c}_2})$. $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$ are the intersections of the optical axis with the image plane; r_1 and r_2 are the optical rays for the point $\boldsymbol{Q}$. (b) Dots represent the bar extremes in successive frames of the calibration sequence.

estimating the other calibration parameters and, in turn, to poor 3-D reconstruction accuracy when 2-D measurement error < 0.2 pixels root-mean-square error (rms). In this case, the position of the principal points must be determined precisely. To do this, we reframed calibration into an optimization problem, where an adequate cost function [see (4)] is minimized with respect to all the calibration parameters [15]. However, gradient-based algorithms fail to compute the global minimum unless initialized

carefully (Section V-A and Table III) and global search algorithms (that have to work in a 12-fold parameter space) are too slow for such an application.

This prompted us to derive a cost function where only the two principal points are the optimization parameters, all the other calibration parameters being determined by SfM [see (5)]. This reduces the optimization space to four dimensions. The price to be paid is that the resulting cost function is hard to optimize through gradient-descent algorithms, as SfM involves matrix operations like singular value decomposition and sign checks. We show here that combining two different evolutionary optimization algorithms generates a reliable and global optimization procedure that efficiently solves the calibration problem. The procedure requires only pairs of matched points and one 3-D metric information. Thus, the calibration tool we are now proposing is a simple rigid bar carrying markers at its ends. Moving the bar within the working volume [see Fig. 1(b)] collects the required calibration points and 3-D metric information easily [14].

The paper is arranged as follows: Section II briefly describes the stereo-camera calibration problem, Section III introduces the cost function and the optimization problem, Section IV describes the enhanced evolutionary search (EES) optimization, and Section V reports and discusses the results on simulated and real data in terms of computational load, reliability, and 3-D accuracy. The conclusions are drawn in Section VI and a full taxonomy is summarized in Appendix A.

II. Calibration of a Stereo–Camera System

The image of a 3-D point $\boldsymbol{P} = [X_P, Y_P, Z_P, 1]^T$ on the target of a video-camera $\boldsymbol{p} = [u_p, v_p, 1]^T$ can be described as a perspective projection of $\boldsymbol{P}$ on $\boldsymbol{p}$ [see Fig. 1(a)] through the center $\boldsymbol{C}$, expressed in homogenous notation in a single matrix equation [10]

$$\begin{aligned}\boldsymbol{p} &= \begin{bmatrix} u_p \\ v_p \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} -f & 0 & u_c \\ 0 & -f & v_c \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} \boldsymbol{R} & -\boldsymbol{RT} \\ \boldsymbol{0} & 1 \end{bmatrix} \begin{bmatrix} X_p \\ Y_p \\ Z_p \\ 1 \end{bmatrix} \\ &= \boldsymbol{KMDP} = \boldsymbol{AP}. \end{aligned} \tag{1}$$

The matrix $\mathbf{A}$ defines a linear projective mapping model that is adequate if distortions are sufficiently weak or corrected in advance [5], [19]. Equation (1) contains the following calibration parameters: $\boldsymbol{T} = [X_T, Y_T, Z_T]^T$ is the location of the camera, $\boldsymbol{R} = \boldsymbol{R}(\omega, \phi, k)$ is the orientation (function of the three independent orientation angles ω, ϕ, k), f is the focal length, and $\boldsymbol{c} = [u_c, v_c]^T$ is the principal point of the camera.

When the 3-D coordinate system is located in the perspective center of one camera with two axes parallel to those of the image plane, $\boldsymbol{D} = \boldsymbol{I}$ for that camera, (1) can be rewritten as

$$\boldsymbol{p}_1 = \boldsymbol{K}_1\boldsymbol{MP} \quad \text{for the first camera} \tag{2a}$$

$$\boldsymbol{p}_2 = \boldsymbol{K}_2\boldsymbol{MDP} \quad \text{for the second camera.} \tag{2b}$$

Let us now consider the three vectors $\overline{\boldsymbol{PC}_2}$, $\overline{\boldsymbol{C}_1\boldsymbol{C}_2}$, $\overline{\boldsymbol{PC}_1}$ (Fig. 1). These lie in a single plane; taking into account Eq. (2a) and (2b), Eq. (1) can be written as the following homogeneous linear system [11]

$$\begin{aligned}&\overline{\boldsymbol{PC}_2} \cdot \overline{\boldsymbol{C}_1\boldsymbol{C}_2} \wedge \overline{\boldsymbol{PC}_1} \\ &\quad = \boldsymbol{p}_2^{\boldsymbol{T}}\boldsymbol{K}_2^{-\boldsymbol{T}}\boldsymbol{R}(\boldsymbol{T}\wedge)\boldsymbol{K}_1^{-\boldsymbol{1}}\boldsymbol{p}_1 = 0 \Rightarrow \boldsymbol{p}_2^T\boldsymbol{F}\boldsymbol{p}_1 = 0 \end{aligned} \tag{3}$$

where $\boldsymbol{F}$ is the fundamental matrix and contains all the calibration parameters. The nine entries of $\boldsymbol{F}$ can be obtained through linear least squares estimation when at least eight 2-D matched points are available from the two cameras. From $\boldsymbol{F}$, we can compute the focal lengths, the relative orientation, and the relative location vector (with unit norm $\boldsymbol{T}_u, ||\boldsymbol{T}_u|| = 1$) through matrix manipulation that includes singular value decomposition and sign checking [11], [14]. The true relative location vector ($\boldsymbol{T} = k\boldsymbol{T}_u$) can be recovered, taking into account that the 3-D scale factor k is the ratio of the true 3-D distance between two calibration points to the distance between the 3-D position of the same points reconstructed with the estimated parameters [14], [18]. When the rigid bar is used as the calibration object, a large number of distances (those between the bar ends) can be collected easily; this facilitates the reliable estimation of the 3-D scale factor by averaging its value over all the collected bars. Fig. 2(a) shows the flow chart of SfM calibration [10], [14]; the only parameters left at this stage are the principal points.

III. Calibration Through Optimization

The key observation is that any bias in the calibration parameters is reflected in the accuracy of the reconstruction of the 3-D position of the bar ends [$\boldsymbol{P}$, $\boldsymbol{Q}$ in Fig. 1(a)], and their mutual distance [d in (B7)]. Another geometrical quantity affected by the calibration error is the distance between the two optical rays belonging to the same 3-D point [e_{int} in (B6) and r_1,and r_2 in Fig. 1(a)]. These two geometrical quantities form our cost function J

$$J = \alpha\sqrt{\frac{1}{N}\sum_{j=1}^{N}(d_j - L)^2} + \beta\sqrt{\frac{1}{N}\sum_{j=1}^{N}\left(\frac{e_{\text{int}j\boldsymbol{P}}^2 + e_{\text{int}j\boldsymbol{Q}}^2}{2}\right)} \tag{4}$$

where

- N number of surveyed bars;
- L true distance between the bar ends $\boldsymbol{P}$ and $\boldsymbol{Q}$;
- d_j bar length reconstructed at time j.

The variables $e_{\text{int}_{jP}}$and $e_{\text{int}_{jQ}}$ are the intersection error for the two bar ends at time j (Appendix B) and α and β are weight factors. When $\alpha \gg \beta$, their values have been shown not to be critical for convergence [14]. In our experiments, α and β were set equal to 1 and 0.1, respectively.

A search for the calibration parameters that minimize (4) implies searching a 12-fold parameter space that is incompatible with the time requirements of real video systems. The solution was to partition the parameters into a first set S_1 (f_1, f_2, k, $\boldsymbol{R}$, and $\boldsymbol{T}$), determined in closed form through SfM from a pair of principal points and matched pairs of

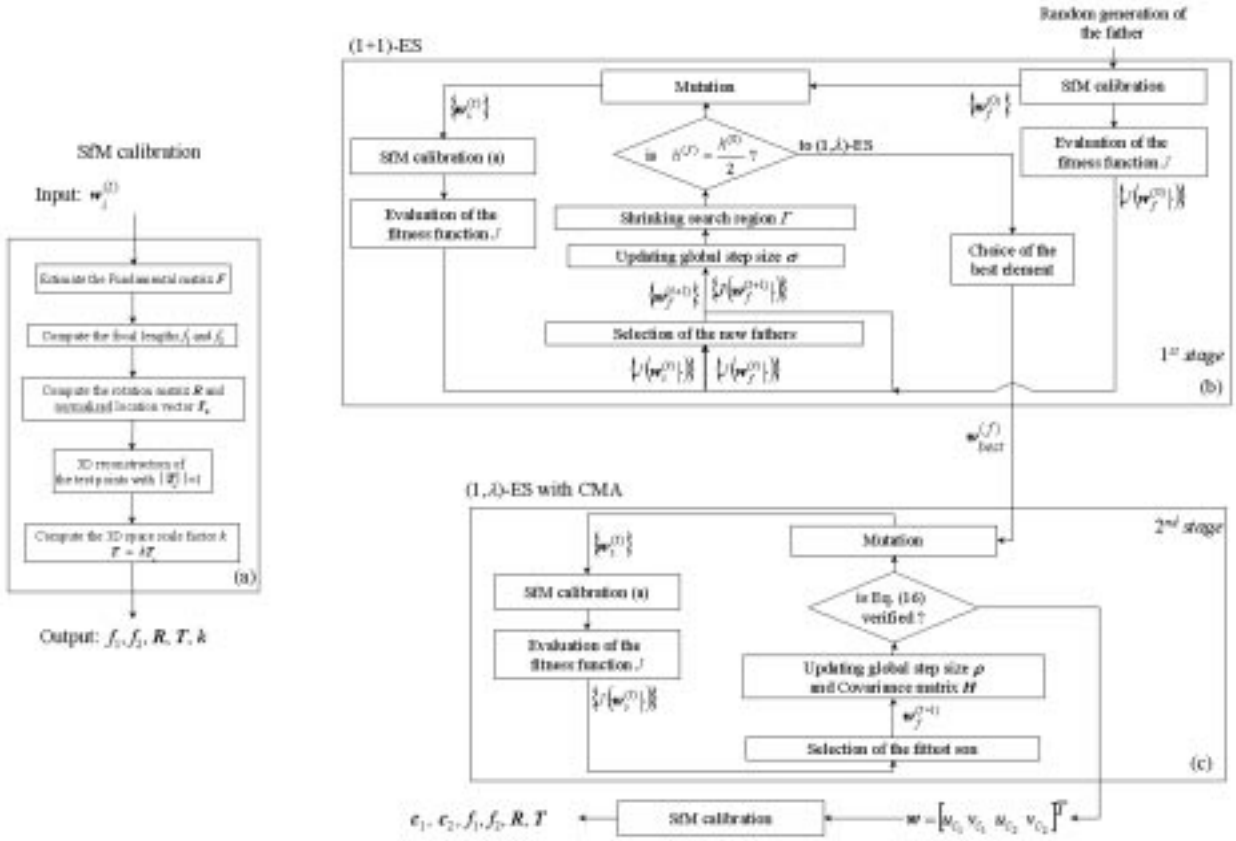


Fig. 2. Flowchart of the EES optimization algorithm. (a) Closed-form solution for the calibration parameters. (b) $(1+1)$-ES is adopted until the amplitude of the search region goes underthershold ($h^{(f)} = h^{(0)}/2$). (c) Solution is then refined with $(1, \lambda)$-ES with covariance matrix adaptation until (16) is satisfied. At this stage, the best offspring give the position of the principal points. Other parameters f_1, f_2, k, $\boldsymbol{R}$, $\boldsymbol{T}$ are those computed with these principal points through SfM solution.

2-D points $\mathrm{S}_1 = \mathrm{SfM}(\boldsymbol{c}_1, \boldsymbol{c}_2|\{\boldsymbol{p}_1\}, \{\boldsymbol{p}_2\})$ and a second set consisting only of the two principal points S_2 ($\boldsymbol{c}_1, \boldsymbol{c}_2$), which have to be determined through evolutionary optimization. In this framework, the cost function $J(.)$ can be rewritten as $J(\mathrm{SfM}(\mathrm{S}_2|\{\boldsymbol{p}_1\}, \{\boldsymbol{p}_2\}), \mathrm{S}_2|\{\boldsymbol{p}_1\}, \{\boldsymbol{p}_2\})$, which is a function of only the parameters in S_2. Such parameters can be determined by minimizing $J(.)$

$$\min_{\mathrm{S}_2=[\boldsymbol{c}_1, \boldsymbol{c}_2]} J(\mathrm{SfM}(\mathrm{S}_2|\{\boldsymbol{p}_1\}, \{\boldsymbol{p}_2\}), \mathrm{S}_2|\{\boldsymbol{p}_1\}, \{\boldsymbol{p}_2\}) = \min_{\mathrm{S}_2=[\boldsymbol{c}_1, \boldsymbol{c}_2]} J([\boldsymbol{c}_1, \boldsymbol{c}_2]|\{\boldsymbol{p}_1\}, \{\boldsymbol{p}_2\}). \quad (5)$$

Gradient-based techniques are not easily applied to find the minimum of (5), as the $\mathrm{SfM}(.)$ solution involves matrix manipulations (Section II); we introduced evolution optimization to accomplish the task.

IV. EES Optimization

In the evolutionary framework, the principal points constitute the elements of the population $\{\boldsymbol{w}[\boldsymbol{c}_1(u_{\boldsymbol{c}_1}, v_{\boldsymbol{c}_1}), \boldsymbol{c}_2(u_{\boldsymbol{c}_2}, v_{\boldsymbol{c}_2})]^T\}$ and the cost function $J(\boldsymbol{w}|.)$ the fitness [see (5)] of the element $\boldsymbol{w}$. The estimation of the principal points is reframed as a search in a four-dimensional domain ($D = 4$) for the best population element. EES, a novel algorithm in which the optimization is carried out in two stages, is proposed here. In the first stage, a $(1+1)$-ES strategy is used to locate the best search region. In the second stage, the optimal solution is found within this smaller search region through a multimembered evolutionary search, or (μ, λ)-ES, with covariance matrix adaptation.

A. Identification of the Search Region with $(1+1)$-ES.

$(1+1)$-ES [20] is used to explore the solution space to estimate a subregion $\Gamma^{(f)}$, where the optimum is located [see Fig. 2(b)]. The algorithm is initialized by randomly generating a set of N parents $\{\boldsymbol{w}_f^{(0)}\}$ inside the initial search region $\Gamma^{(0)}$. If no information is available, $\Gamma^{(0)}$ can be as large as the image plane; if information is available, a smaller region can be defined, as in Figs. 3(a) and 4(a). At each generation t, a set of N parents $\{\boldsymbol{w}_f^{(t)}\}$ is analyzed: SfM calibration is carried out through the procedure outlined in Section II for each parent and fitness is evaluated through (5). Then, through mutation, a set of N offspring $\{\boldsymbol{w}_s^{(t)}\}$, one for each parent, is generated inside $\Gamma^{(t)}$ as

$$\boldsymbol{w}_{s_i}^{(t)} = \boldsymbol{w}_{f_i}^{(t)} + \sigma^{(t)} \boldsymbol{z}_i^{(t)} \qquad \text{subjected to } \boldsymbol{w}_{s_i}^{(t)} \in \Gamma^{(t)} \quad (6)$$

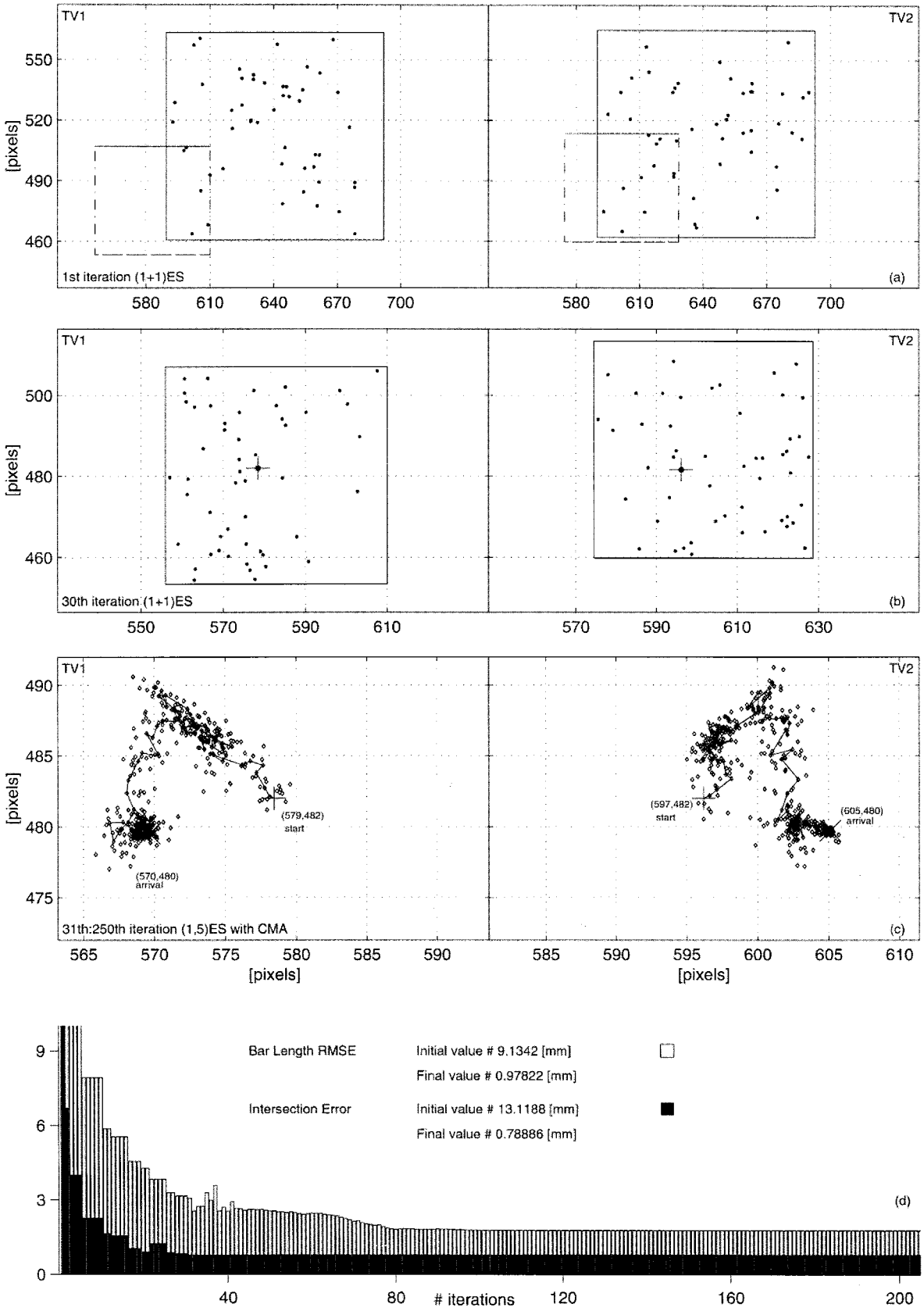


Fig. 3. Plots refer to the data of Table I. Search for the true principal points progresses in parallel on the two cameras. (a) It goes first through a $(1 + 1)$-ES optimization that starts inside the region Γ^0 (square with solid perimeter), 100 pixels wide ($h^{(0)} = 50$) centered in [512, 512, 512, 512] and ends in the region $h^{(f)} = h^{(0)}/2$ centered in the element [585, 480, 604, 486]. (b) At the last step of the first stage, the best population element $\boldsymbol{w}_{\text{best}} = [579, 482, 597, 482]$ (marked with a cross) is extracted inside this region (cf.Fig. 2); this constitutes the first parent of the $(1, \lambda)$-ES. (c) Evolution path. It is the line connecting the winning offspring of each step of $(1, \lambda)$-ES. Global minumum is reached at the individual $\boldsymbol{w}[596.73, 497.17, 605.27, 479.94]^T$. (d) Error on both bar length and ray intersection is reported as evolution progresses.

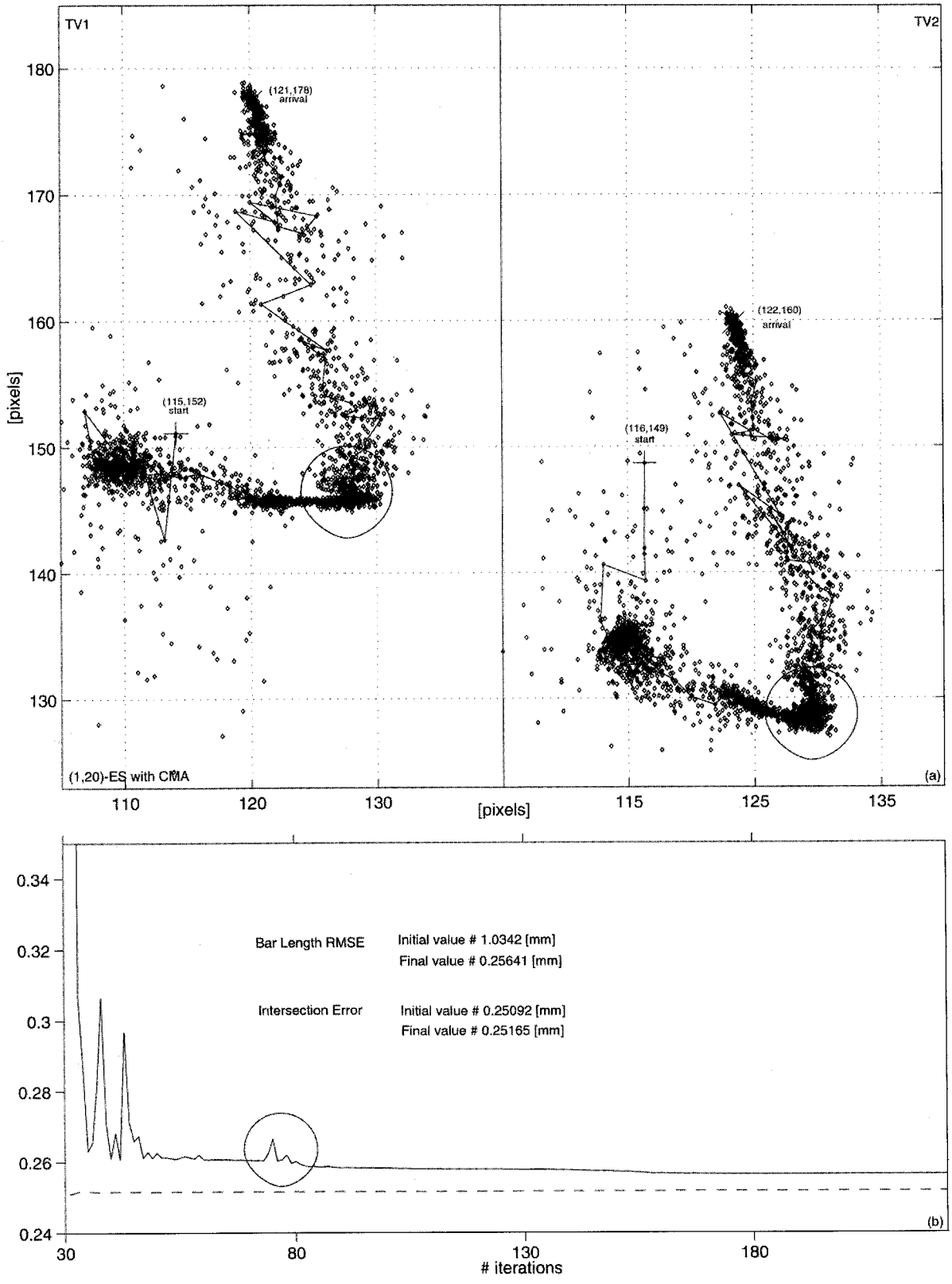


Fig. 4. (a) Evolution path pertaining to the second stage of the EES along with (b) its associated cost for real data. Region where the EES founded a local minimum is marked with a circle in panel (a). Note the corresponding transients in the cost function in panel (b).

where $z_i^{(t)}$ is obtained from the Gaussian distribution $z = N(0, 1)$. The fitness of the N offspring $\{\boldsymbol{w}_s^{(t)}\}$ is evaluated through (5) as for the parents by computing J for each of them. $\sigma^{(t)}$determines the actual amplitude of the mutation

[see (6)] and is updated according to the Rechenberg 1/5 rule [20] as

$$\text{if}\begin{cases} q < \frac{1}{5} \Rightarrow \sigma^{(t+1)} = \dfrac{\sigma^{(t)}}{m} \\ q = \frac{1}{5} \Rightarrow \sigma^{(t+1)} = \sigma^{(t)} \\ q > \frac{1}{5} \Rightarrow \sigma^{(t+1)} = m\sigma^{(t)} \end{cases} \tag{7}$$

where q is the ratio between the number of winning offspring and winning parents. Setting $m = 0.85^{1/D}$ gives a heuristic value resulting in linear order convergence rates [20]. The offspring are then placed in competition with the corresponding parent: the one that exhibits higher fitness (lower value of J) survives and becomes the parent in the next generation

$$\begin{aligned} J_{f_i}^{(t)} \le J_{s_i}^{(t)} &\Rightarrow \boldsymbol{w}_{f_i}^{(t+1)} = \boldsymbol{w}_{f_i}^{(t)} \\ J_{f_i}^{(t)} > J_{s_i}^{(t)} &\Rightarrow \boldsymbol{w}_{f_i}^{(t+1)} = \boldsymbol{w}_{s_i}^{(t)}. \end{aligned} \tag{8}$$

At each step, the resolution of the search is increased by reducing the search region amplitude according to the following schedule [see Figs. 3(a) and (b) and 4(a) and (b)]

$$h^{(t+1)} = \frac{h^{(0)}}{\sqrt{\log(t+2)}}. \tag{9}$$

The logarithmic function allows a faster reduction in the first generations. The reduction is complemented with a translation of the search region $\Gamma^{(t)}$, which at each generation, is centered at the fittest element of the previous generation. However, as no information on the topology of the solution space is considered, the convergence is slow. We arbitrarily stop this first stage when $t = 30$ ($h^{(t)} \approx h^{(0)}/2$). The best population element at the last step of this stage $\boldsymbol{w}_{\text{best}}^{(f)}$ becomes the parent for the (1, λ)-ES algorithm [see Figs. 2(b), 3(b), and 4(b)].

B. Local Search Through (1, λ)-ES

The key element here is the runtime adaptation of the best search path in the D-dimensional solution space; this is done through an analysis of the population elements and the fitness history. For this task, two strategic variables are defined: the covariance matrix $\boldsymbol{H}^{(t)}(D \times D)$ and the global step size $\rho^{(t)}$. $\boldsymbol{H}^{(t)}$ learns the local topology of the objective function and determines the actual shape of the search region. It is a D-dimensional ellipsoid oriented in the solution space through $\boldsymbol{B}^{(t)}$ that contains the eigenvectors of $\boldsymbol{H}^{(t)}$

$$\boldsymbol{H}^{(t)} = \boldsymbol{B}^{(t)}\boldsymbol{G}^{(t)^2}\boldsymbol{B}^{(t)^T}. \tag{10}$$

$\boldsymbol{G}^{(t)}$ contains the square root of the eigenvalues of $\boldsymbol{H}^{(t)}$ and sets the elongation of the ellipsoid $\boldsymbol{H}^{(t)}$ in the D principal directions determined by $\boldsymbol{B}^{(t)}$.

The value $\rho^{(t)}$ modulates the amplitude of the ellipsoid; its role is to widen the search region rapidly when better fitness is found repeatedly in a certain direction and to restrict it when fitness variability increases around a certain element, performing a local search at higher resolution. $\boldsymbol{H}^{(t)}$ and $\rho^{(t)}$ are used in the mutation process as in [21]

$$\boldsymbol{w}_{s_i}^{(t)} = \boldsymbol{w}_f^{(t)} + \rho^{(t)}\boldsymbol{B}^{(t)}\boldsymbol{G}^{(t)}\boldsymbol{z}_i^{(t)} \qquad i = 1, \ldots, \lambda \tag{11}$$

where λ is the number of offspring. As in (1 + 1)-ES, $\boldsymbol{z}_i^{(t)}$ is extracted from the Gaussian distribution $z = N(0, 1)$. After mutation, the fittest of the λ offspring is picked as the parent of the next generation and the strategic variables $\boldsymbol{H}^{(t)}$ and $\boldsymbol{\rho}^{(t)}$ are updated using the *evolutionary path* $\boldsymbol{s}^{(t)}$ [22], [23]. Such a path is achieved through a discounted sum of the displacements of the winning offspring (*derandomization* [21]) in the previous evolutionary steps

$$\boldsymbol{s}^{(t+1)} = (1 - a_s)\boldsymbol{s}^{(t)} + a_u\boldsymbol{B}^{(t)}\boldsymbol{G}^{(t)}\boldsymbol{z}_{i_{sel}}^{(t)}. \tag{12}$$

The weights a_s and a_u balance the effects of past history and innovation, smoothing out random deviations that could otherwise greatly disturb adaptation. The weights are chosen so that the variance of $\boldsymbol{s}^{(t)}$, $\boldsymbol{S}^{(t)} = \boldsymbol{s}^{(t)}(\boldsymbol{s}^{(t)})^T$, is normalized [see also (13)]: $a_u^2 + (1 - a_s)^2 = 1$. By choosing $a_s = 1/\sqrt{D}$ (see later), $a_u = \sqrt{(2\sqrt{D} - 1)/D}$. In the same way, $\boldsymbol{H}^{(t+1)}$ is computed as the discounted sum of the actual covariance matrix of the evolution path, $\boldsymbol{S}^{(t+1)}$ and the covariance matrix computed in the previous step $\boldsymbol{H}^{(t)}$

$$\boldsymbol{H}^{(t+1)} = (1 - a_H)\boldsymbol{H}^{(t)} + a_H\boldsymbol{S}^{(t+1)}. \tag{13}$$

Equations (12) and (13) are very similar to the discounted reward adopted by reinforcement learning paradigms in machine learning [24]. Such paradigms ensure that only mutation steps moving in the same direction and chosen repeatedly will be reinforced over time: the mutation distribution $\boldsymbol{H}^{(t)}$ and the evolutionary path $\boldsymbol{s}^{(t)}$ will be elongated in this direction.

The role of global variance $\rho^{(t)}$ is to detect discontinuity in the direction of the evolution path. When there are repeated changes in path direction, updating follows the principal: "reasonable adaptation has to reduce the difference between the distribution of the actual evolution path and an evolution path under random selection" [22]. At each step, ρ is updated as

$$\rho^{(t+1)} = \rho^{(t)}e^{\gamma(\|s_\rho^{(t+1)}\| - \overline{c})} \tag{14}$$

where $\boldsymbol{s}_\rho^{(t+1)}$ is the normalized path containing pure directional information [cf. (12) and (15)]

$$\boldsymbol{s}_\rho^{(t+1)} = (1 - a_s)\boldsymbol{s}_\rho^{(t)} + a_u\boldsymbol{B}^{(t)}\boldsymbol{z}_{i_{sel}}^{(t)} \tag{15}$$

and $\overline{c}_n = (1 - 1/4D + 1/21D^2)$ is the second-order approximation of the expectation value of the length distribution of vectors extracted from $N(0, \boldsymbol{I})$. From (14), it can be seen that $\rho^{(t)}$ is decreased when the evolution path direction changes often. When the same direction is chosen repeatedly, the mutation step is made larger.

At this stage, the parameters a_H, a_s, and γ have to be set properly. As the updating of $\rho^{(t)}$ is regulated by $\boldsymbol{s}_\rho^{(t+1)}$, which depends on D parameters, $a_s = 1/D$ could be a possible setting choice. However, as the role of $\rho^{(t)}$ is to detect when the evolutionary process finds regions where the best direction has changed or a minimum found, it is safer to give it an even shorter time span by choosing $a_s = 1/\sqrt{D}$. To avoid an uncontrolled increase in global step size, γ should be larger than zero and smaller than a_s. We chose $\gamma = 0.1$. As far as $\boldsymbol{H}^{(t)}$ is concerned, as it is defined by $(D^2/2 + D/2)$ free parameters, the time scale of adaptation is in the order $O(D^2)$ and a suitable choice of

TABLE I
CALIBRATION RESULTS WITH SIMULATED DATA (ZOOM LENSES)

Setup # 1 Simulated data Zoom lenses	Working volume [m^3]	2.5×2.5×2.5		
	Angle between the cameras	~44 degrees		
	Bar length [m]	0.5		
	Target Dimension [pixels]	1,280 × 1,024		
	Image measurement error [pixels]	0.1 (rms)		
	# Calibration points	400		
	# Test bars	200		
	Initial principal points [pixels]	[640, 512, 640, 512]		
	$h^{(0)}$ side [pixels]	~100		
	$h^{(f)}$ side [pixels]	~$h^{(0)}/2$		
	$\sigma^{(0)}$ (1+1)-ES [pixels]	$h^{(0)}/5$		
	#Fathers in (1+1)-ES	50		
	$\rho^{(0)}$ (1+λ)-ES [pixels]	$h^{(f)}/5$		
	#Sons in (1,λ)-ES	5		
	True parameters	Initialization	Evolutionary algorithm	Gradient-based algorithm
$T_X\ T_Y\ T_Z$ [mm]	-2,802.20 192.25 -1,072.70	-2,862.99 156.85 -1,075.16	-2,802.61 193.29 -1,071.33	-2,808.01 193.86 -1,065.36
$\omega\ \phi\ \kappa$ [deg]	1.28 -43.99 0.52	2.02 –43.19 2.42	1.24 -44.00 0.45	1.26 -44.00 0.47
$u_{c1}\ v_{c1}\ f_1$ [pixel]	570.00 480.00 1,000.00	640.0 512.0 1,009.08	569.73 479.17 1,000.15	569.83 479.23 999.35
$u_{c2}\ v_{c2}\ f_2$ [pixel]	605.00 480.00 1,000.00	640.0 512.0 964.61	605.27 478.94 1,000.73	605.11 479.47 1,001.68
RI error [mm]	0.34	0.37	0.36	0.37
BL error [mm]	0.00 ± 0.48	0.05 ± 10.26	0.00 ± 0.48	-0.06 ± 0.48
GT error [mm] X, Y, Z	0.00 ± 0.25 0.00 ± 0.20 0.00± 0.42	244 ± 52 -107 ± 22 176± 54	0.00 ± 0.42 0.00 ± 0.29 0.00± 0.50	0.00 ± 0.42 0.00 ± 0.32 0.00± 0.49
#Iterations	*	*	30 + 177	11
Time [s]	*	0.05	13	5

a_H in (13) is $2/D^2$. This guarantees that high-frequency variability in the mutation direction disturbs the adaptation of $\boldsymbol{H}^{(t)}$ minimally. As a result of these choices, a_H is smaller than a_s, reflecting the different time span on which the two strategic variables $\boldsymbol{H}^{(t)}$ and $\rho^{(t)}$ operate.

Fig. 2 summarizes the overall combined evolution strategy. The algorithm stops when the residual normalized increment in the parameters goes below the threshold

$$\sqrt{\frac{1}{M}\sum_{j=1}^{M}\left(\frac{p_j^{(t+1)}-p_j^{(t)}}{p_j^{(t+1)}}\right)^2} < r_{\text{out}} \tag{16}$$

where $p_j^{(t)}$ is the jth calibration parameter ($M = 12$); in our case, $r_{\text{out}} = 0.001$.

V. EXPERIMENTAL RESULTS

The algorithm has been tested on simulated and real data.

A. Simulations

Simulations were carried out to assess the accuracy and reliability of the method. Two different camera setups were adopted and algorithm performance is reported in Tables I and II. In the first setup (Table I, Fig. 3), zoom lenses surveyed a working volume of $2.5 \times 2.5 \times 2.5\ \text{m}^3$, in the second (Table II), wide-angle lenses covered a working volume of $4.5 \times 4.5 \times 4.5\ \text{m}^3$. The working volume is defined here as the largest parallelepiped volume within the common field of view of the two cameras. In both simulations, the position and orientation of a set of 200 bars (length $L = 0.5$ m for the first experiment, $L = 1$ m for the second) were generated randomly inside the working volume to emulate the sampling of the bar in motion. The ends of the bars were then projected onto the image planes of the two cameras using the calibration parameters reported in Table s I and II (true parameters). The 2-D measurement error was simulated by adding a quantity z, extracted from the Gaussian distribution $z = N(0, 0.1)$, to each coordinate of the projected points. Starting from the measured control points $\{\boldsymbol{p}_m\}$ and the bar length L, the calibration parameters are estimated through EES. Different experiments have shown that the initial value of the parents, chosen at the beginning of the first stage, does not affect the convergence. The algorithm was implemented in C and its running time was measured on a personal computer, MMX-200 MHz Pentium II central processing unit. The accuracy indices are considered to be the statistical distribution of the error in the reconstruction of the bar length [BL, see (B7)] and for each point of the mean distance between the optical rays [RI, (B6)]. These practical measurements are easily carried out in the field. To evaluate EES with respect to ground truth (GT), the displacement of the reconstructed bar ends with respect to their nominal 3-D position is calculated and reported as GT error (see Tables I–IV). The X and Y directions are parallel to the axes of the image plane of the first camera and the Z direction is perpendicular to both. For the sake of comparison, we have included the results of when the cost function $J(.)$ in (5) is minimized with respect to all the calibration parameters without resorting to SfM for determining some of them. Gradient descent is initialized by setting the principal points in the image center, determining the other initial parameters through SfM machinery.

TABLE II
CALIBRATION RESULTS WITH SIMULATED DATA (WIDE ANGLE LENSES)

Setup # 2 Simulated data Wide angle lenses	Working volume [m^3]		$4.5 \times 4.5 \times 4.5$	
	Angle between the cameras		~74 degrees	
	Bar length [m]		1.0	
	Target Dimension [pixels]		$1{,}280 \times 1{,}024$	
	Image measurement error [pixels]		0.1 (rms)	
	# Calibration points		400	
	# Test bars		200	
	Initial principal points [pixels]		[640, 512, 640, 512]	
	$h^{(0)}$ side [pixels]		~100	
	$h^{(f)}$ side [pixels]		~$h^{(0)}/2$	
	$\sigma^{(0)}$ (1+1)-ES [pixels]		$h^{(0)}/5$	
	#Fathers in (1+1)-ES		50	
	$\rho^{(0)}$ (1+λ)-ES [pixels]		$h^{(f)}/5$	
	#Sons in (1,λ)-ES		5	
	True parameters	Initialization	Evolutionary algorithm	Gradient-based algorithm
$T_X\ T_Y\ T_Z$ [mm]	-7,664.7 29.2 -7,030.0	-8192.99 -91.33 -7817.59	-7665.39 31.08 -7028.74	-7663.24 32.01 -7025.42
$\omega\ \phi\ \kappa$ [deg]	5.84 -72.9 3 4.56	8.67 -74.68 8.21	5.77 -72.95 4.47	5.76 -72.92 4.46
$u_{c1}\ v_{c1}\ f_1$ [pixel]	570.0 480.0 2,000.0	512.0 512.0 2025.91	569.87 479.20 1999.92	569.23 480.00 1999.94
$u_{c2}\ v_{c2}\ f_2$ [pixel]	605.0 480.0 2,000.0	512.0 512.0 2112.94	605.72 479.01 2000.3	604.98 479.22 1999.8
RI error [mm]	0.24	0.32	0.29	0.30
BL error [mm]	0.02 ± 0.31	-0.19 ± 16.98	0.00 ± 0.31	-0.04 ± 0.32
GT error [mm] X, Y, Z	0.00 ± 0.21 0.00 ± 0.15 0.00± 0.25	310 ± 19.5 -157 ± 17.31 -289±57.27	0.00 ± 0.25 0.00 ± 0.15 0.00± 0.21	0.00 ± 0.28 0.00 ± 0.18 0.00± 0.22
#Iterations	*	*	30 + 264	15
Time [s]	*	0.05	19	8

TABLE III
SAME EXPERIMENT AS TABLE II WITH A DIFFERENT PAIR OF PRINCIPAL POINTS

The same set-up reported in Table I is used here with the exception of the principal points whose true position is now: [600, 450, 635,510]			
	Initialisation	Evolutionary algorithm	Gradient-based algorithm
$T_X\ T_Y\ T_Z$ [mm]	-5,284.38 417.65 -2,779.09	-2,802.60 193.14 -1,071.20	*
$\omega\ \phi\ \kappa$ [deg]	2.21 -58.46 1.32	1.24 -44.00 0.46	*
$u_{c1}\ v_{c1}\ f_1$ [pixel]	600.0 450.0 1,392.09	569.33 479.33 1,000.54	*
$u_{c2}\ v_{c2}\ f_2$ [pixel]	635.0 510.0 1,377.37	604.76 479.09 1,000.82	*
RI error [mm]	0.80	0.36	*
BL error [mm]	0.66 ± 72.52	0.00 ± 0.48	*
GT error [mm] X, Y, Z	-654.3 ± 130.5	0.00 ± 0.25 0.00 ± 0.15 0.00± 0.21	
#Iterations	*	28 + 302	*
Time [s]	0.5	21	*

As seen from Tables I–IV, EES converges to parameter values and accuracy indices equal to or even better than the gradient descent approach. The GT data show that there is no bias in the reconstruction of the bar ends and that the standard deviation is very close to that obtained when reconstructing the bar ends with the true parameters. Moreover, although gradient descent is a little faster, the main drawback is that there is no convergence for some geometrical camera arrangements. For example, if the principal points in the first experiment (Table I) are moved to the location [600, 450, 635, 510], which is even closer to the image center, gradient descent cannot converge while EES optimization does.

Fig. 3 shows a typical search for the pair of optimal principal points. The data correspond to the experiment with setup #1 (Table I). Comparing Fig. 3(a) and (b), it can be seen that the reduced search region in the first stage leads to a denser sampling of the population space. At the end of the first stage, EES switches to (1, λ)-ES. This switch is highlighted by a transient increase in the cost function $J(.)$ due to the time required to learn the local topology of $J(.)$ [see Fig. 3(d)]. It should be noted that EES "evolves" within particular regions before heading toward the optimum position. Such regions possibly have shallow cost functions, local minima, or abrupt changes in the decreasing direction of the cost function. This is evident

TABLE IV
CALIBRATION RESULTS WITH REAL DATA (ZOOM LENSES)

Setup # 3 Real data Zoom lenses	Working volume [m^3]	$\sim 0.4 \times 0.4 \times 1$	
	Angle between the cameras	~23 degrees	
	Bar length [mm]	86.0	
	Target Dimension [pixels]	256×256	
	Shrinkage factor	0.6766	
	# Calibration points	400	
	# Test bars	400	
	Initial principal points [pixels]	[128, 128, 128, 128]	
	$h^{(0)}$ side [pixels]	~25	
	$h^{(f)}$ side [pixels]	$\approx h^{(0)}/2$	
	$\sigma^{(0)}$ (1+1)-ES [pixels]	$h^{(0)}/5$	
	#Parents in (1+1)-ES	50	
	$\rho^{(0)}$ (1+λ)-ES [pixels]	$h^{(f)}/5$	
	#Offspring in (1,λ)-ES	20	
	Initialization	Evolutionary algorithm	Gradient-based algorithm
$T_X\ T_Y\ T_Z$ [mm]	41.25 -865.94 -365.14	45.29 -1283.39 -680.335	46.32 -1281.39 -680.423
$\omega\ \phi\ \kappa$ [deg]	19.22 1.43 0.83	22.99 1.67 0.25	22.38 1.70 0.24
$u_{c1}\ v_{c1}\ f_1$	128.0 128.0 1793.69	162.56 150.34 2248.86	164.56 149.98 2247.08
$u_{c2}\ v_{c2}\ f_2$ [pixel]	128.0 128.0 1667.12	152.92 128.31 1998.9	153.52 129.23 1999.51
RI error [mm]	0.34	0.35	0.35
BL error [mm]	-0.23 ± 3.0	0.00 ± 0.37	0.05 ± 0.41
GT error [mm] (Ex, Ey, Ez)	-0.50 ± 0.87 -1.08 ± 0.94 40.79± 1.04	0.16 ± 0.89 0.17 ± 0.54 0.20± 0.81	0.20 ± 0.90 0.18 ± 0.56 0.19 ± 0.82
#Iterations	*	30 + 180	21
Time [s]	0.5	38	12

in the plot of the evolution path [see Fig. 3(c)] that is the line connecting the winning offspring at each evolution step.

B. Real Data Experiments

We applied EES optimization widely to calibrate motion capture systems [2] and 3-D scanners [1] equipped with a wide range of setups and lenses. The bar length used was approximately 1/20th to 1/10th of the working volume diagonal and the bar was moved to cover the entire working volume. During motion, the bar is oriented in the depth direction, although never perfectly horizontal so as to avoid problems in classifying the ends. This allows the internal parameters to be determined with better confidence [10], [19]. An automatic classification of the bar ends can be achieved through the fundamental matrix constraint [25], the same algorithm allowing the bar, during motion, to slip from the field of view of the cameras. Table IV and Fig. 4 show some typical results on real data. In these experiments, a bar that carries two markers 86 mm apart was moved within the field of view of two cameras for a working volume of about $0.4 \times 0.4 \times 1\ \mathrm{m}^3$ [see Fig. 1(b)]. The marker projection onto the image plane of the two cameras was measured automatically by the Elite system [3] with subpixel accuracy. A total of 1250 bars were collected in 12.5 s; from these, 400 were extracted randomly for calibration. In this case, unlike in simulations, the true 3-D position of the bar extremes is not available, so a different GT accuracy test was introduced. A rectangular grid carrying 30 equally spaced markers (five rows by six columns, with adjacent marker spacing of 50 mm) is surveyed in four parallel positions with 200-mm separation one from the other. The 3-D marker position is reconstructed with the calibration parameters and any error in distance between a marker and the grid center is assumed as the horizontal error (Ex) if taken horizontally, as vertical error (Ey) if taken vertically, and as depth error (Ez) if taken perpendicular to the planes. As can be seen from Table IV, the obtained accuracy is equivalent to gradient descent (and it is also comparable to that obtained when control points of known location are used, i.e., when calibration structures are adopted [5], [6], [17]). The error values confirm that the parameters estimated through EES have no bias and accuracy equal to or better than that achieved through gradient descent. It should be noted that for these cameras and setups, the principal points are, for the first camera, offset by more than 34 pixels horizontally and 22 vertically and, for the second, 24 pixels horizontally. A typical evolution path (second stage of optimization) for these data is reported in Fig. 4. Note how when good direction is found, the algorithm elongates the mutation function rapidly in that direction optimizing the speed without being trapped in local minima. Regions with local minima are small areas characterized by heavier sampling (denser population), where the algorithm spends some iterations before finding the way out. For example, a local minimum is found just before the 80th iteration in the circled area [see Fig. 4(a)], where there is a transient increase in the cost function [see Fig. 4(b)].

This indirect evolution path is typical and is obtained using a wide range of setups and lenses. The tortuous path is due to the high cost in the region between the image center and the true position of the principal points that discourages any EES there. A denser sampling of the image plane in the first stage would provide the second stage with a starting parent closer to the optimum. However, as the region around the optimum is a very narrow valley, the search for a population element lying at the bottom of the valley would take much longer. Thus, the high cost in terms of computational time has discouraged this approach.

VI. Summary

EES optimization was applied to calibrate motion capture systems [3] and 3-D scanners [1] that had a wide range of setups and lenses. Cumbersome calibration grids [6] were substituted by a simple bar that moves within the working volume; this new powerful calibration technique has the same accuracy and reliability as the grids, but expends much less time and effort. It can be applied in many fields, allowing motion capture systems and 3-D scanner instruments to be calibrated easily in the most variable working conditions. The combination of a stochastic search, such as the one implemented in the $(1+1)$-ES and the covariance matrix-based search in $(1, \lambda)$-ES, led to a solution in quasi-real time and, above all, to the avoidance of local minima, which can originate from poor initialization. The EES technique is suitable for all those optimization problems where the cost function contains complex nonlinear relationships between the optimization variables. Discovering when and where EES optimization is the most, or the least, suitable technique remains for future work.

Appendix A
Taxonomy

Scalar values are denoted in italics, vectors and matrices in bold italics. The parameters are subdivided into calibration and optimization and into parameters to be set by the user or automatically set internal parameters.

A. Parameters to be Set by the User

L	True distance between the bar extremes (bar length).
$q_u \times q_v$	Target dimension.
D	Dimension of the search domain ($D = 4$ for camera calibration).
$h^{(0)}$	Initial dimension of the search region for (1 + 1)-ES algorithm (1st stage). $h^{(0)} = \min(q_u/10, q_v/10)$ in our case.

B. Parameters Internal to EES

1) Calibration Parameters:

$\boldsymbol{P} = [X_P, Y_P, Z_P, 1]^T$	3-D point.
$\boldsymbol{C}_i = [X_{\boldsymbol{C}_i}, Y_{\boldsymbol{C}_i}, Z_{\boldsymbol{C}_i}]^T$	3-D coordinates of the point perspective center $\boldsymbol{C}_i$ of the ith camera.
$\boldsymbol{p}_i = [u_{\boldsymbol{p}_i}, v_{\boldsymbol{p}_i}]^T$	2-D point on the image plane of the ith camera.
$\boldsymbol{F}$	Fundamental matrix.
$\boldsymbol{K}_i$	Matrix containing the internal parameters of the ith camera.
$\boldsymbol{D}$	Matrix containing the external parameters.
$\boldsymbol{T}_u = \boldsymbol{T}/\|\boldsymbol{T}\|$	Relative location vector with unitary norm.
k	3-D space scale factor ($\boldsymbol{T} = k\boldsymbol{T}_u$).
d_j	Distance between the bar extremes (bar length) computed at frame j. through the estimated geometrical parameters.
$e_{j\boldsymbol{P}}$	Intersection error for point $\boldsymbol{P}$ computed at frame j.

2) Optimization Parameters:

t	Generation (iteration step).
J	Fitness

$$J = \alpha\sqrt{\frac{1}{N/2}\sum_{j=1}^{N/2}(d_j - L)^2} + \beta\sqrt{\frac{1}{N/2}\sum_{j=1}^{N/2}\left(\frac{e^2_{\text{int}j\boldsymbol{P}} + e^2_{\text{int}j\boldsymbol{Q}}}{2}\right)}.$$

$\boldsymbol{w}$	$\boldsymbol{w} = [u_{\boldsymbol{c}_1}\ v_{\boldsymbol{c}_1}\ u_{\boldsymbol{c}_2}\ v_{\boldsymbol{c}_2}]^T$. Population element: $[\boldsymbol{c}_1 \boldsymbol{c}_2]$.
$\boldsymbol{w}_f$, $\boldsymbol{w}_s$	Parents and offspring of the population.

$(1 + 1)$-ES Optimization (First Stage)

$h^{(t)}$	Amplitude of the region of search region: $h^{(t+1)} = h^{(0)}/\sqrt{\log(t+2)}$.
$h^{(f)}$	Amplitude of the search region when the first stage is completed: $h^{(f)} \approx h^{(0)}/2 \Rightarrow t = 30$.
$\Gamma^{(t)}$	Region of search.
$\sigma^{(t)}$	Standard deviation of the mutation function: $\sigma^{(0)} = h^{(0)}/5$.
m	Empirical value for updating σ [see (7)]: $m = 0.85^{1/D}$ [19].
q	Ratio between winning offspring and winning parents.

$(1, \lambda)$-ES Optimization (Second Stage)

$\rho^{(t)}$	Global step size: $\rho^{(0)} = h^{(f)}/5$.
$\boldsymbol{H}^{(t)}$	Covariance matrix of the mutation function: $\boldsymbol{H}^{(0)} = \boldsymbol{I}$.
$\boldsymbol{G}^{(t)}$	Matrix containing the square root of the eigenvalues of $\boldsymbol{H}^{(t)}$.
$\boldsymbol{B}^{(t)}$	Matrix containing the eigenvectors of $\boldsymbol{H}^{(t)}$.
$\boldsymbol{s}^{(t)}$	Evolution path at step t.
$\boldsymbol{s}_\rho^{(t)}$	Normalized evolution path at step t.
$\boldsymbol{S}^{(t)}$	Covariance matrix of the evolution path at iteration step t.
$a_H \in]0, 1]$	Determines the time of averaging the distribution: $\boldsymbol{s}^{(t)}\boldsymbol{s}^{(t)^T} a_H = 2/D^2$.
$a_s \in]0, 1]$	Determines the decay constant for the evolution path: $\boldsymbol{s}^{(t)} a_s = 1/\sqrt{D} > a_H$.
$a_u = \sqrt{a_s(2 - a_s)}$	Normalizes the variance of $\boldsymbol{s}^{(t)}$ by solving $1 = (1 - a_s)^2 + a_u^2$. For $a_s = 1/\sqrt{D}$, $a_u = \sqrt{(2\sqrt{D} - 1)/D}$.
γ	Damping of the step size variation between successive generations: $\gamma = 0.1$.
$\overline{e} = \sqrt{D}(1 - 1/4D + 1/21D^2)$	Approximates the expectation of the χ_D distribution, which is the distribution of the length of a $N(0, \boldsymbol{I})$ distributed random vector in $\Re^D$.

C. Output Parameters (Calibration Parameters)

$\boldsymbol{c}_i = [u_{\boldsymbol{c}_i}, v_{\boldsymbol{c}_i}, 1]^T$	Principal point of the image plane of the ith camera.
f_i	Focal length of the ith camera.

$\boldsymbol{R} = \boldsymbol{R}(\omega, \phi, \kappa)$ Relative orientation (a function of three independent rotations: ω, ϕ, κ).

$\boldsymbol{T} = (\boldsymbol{C}_2 - \boldsymbol{C}_1) = [X_T\ Y_T\ Z_T]^T$ Relative location (base line).

APPENDIX B
3-D RECONSTRUCTION

When the stereo-camera parameters have been estimated the 3-D position of a point $\boldsymbol{P}$ can be determined as the intersection point of the two optical rays. These are the straight lines [r_1 and r_2 in Fig. 1(a)] through the projection of $\boldsymbol{P}$, $\boldsymbol{p}_1$ and $\boldsymbol{p}_2$ onto the image plane of the two cameras and the perspective centers $\boldsymbol{C}_1$ and $\boldsymbol{C}_2$. Due to noise on $\boldsymbol{p}_j$ and $\boldsymbol{q}_j$, these two straight lines generally do not intersect and the 3-D reconstruction of $\boldsymbol{P}$ lies in the midpoint of the minimal distance segment [6]. The two straight lines have the following:

$$\boldsymbol{P}_j = \boldsymbol{C}_j + h\frac{(\boldsymbol{p}_j - \boldsymbol{C}_j)}{\|\boldsymbol{p}_j - \boldsymbol{C}_j\|} = \boldsymbol{C}_j + h_j\boldsymbol{l}_j \tag{B1}$$

where

$\boldsymbol{C}_1$ equals zero;
$\boldsymbol{C}_2$ equals $\boldsymbol{T}$;
$\boldsymbol{l}_j$ director cosines.

Minimizing the 3-D distance e_{int} between $\boldsymbol{P}_1$ and $\boldsymbol{P}_2$

$$e_{\text{int}} = \|\boldsymbol{P}_1 - \boldsymbol{P}_2\| \tag{B2}$$

a linear system is obtained, where

$$h_1 = \frac{(\boldsymbol{l}_1^{\boldsymbol{T}} + (\boldsymbol{l}_2^{\boldsymbol{T}}\boldsymbol{l}_1)\boldsymbol{l}_2^{\boldsymbol{T}})\boldsymbol{T}}{1 - \boldsymbol{l}_2^{\boldsymbol{T}}\boldsymbol{l}_1} \tag{B3a}$$

$$h_2 = \frac{(\boldsymbol{l}_2^{\boldsymbol{T}} + \boldsymbol{l}_2^{\boldsymbol{T}}\boldsymbol{l}_1)\boldsymbol{l}_1{}^{\boldsymbol{T}})\boldsymbol{T}}{1 - \boldsymbol{l}_2^{\boldsymbol{T}}\boldsymbol{l}_1}. \tag{B3b}$$

The 3-D reconstruction of $\boldsymbol{P}$ is obtained as

$$\boldsymbol{P} = \frac{h_1\boldsymbol{l}_1 + (\boldsymbol{T} + h_2\boldsymbol{l}_2)}{2} \tag{B4}$$

where the $\boldsymbol{l}_j$s are a function of the calibration parameters

$$\boldsymbol{l}_j = \frac{\boldsymbol{R}_j(\boldsymbol{p}_j - \boldsymbol{p}_{o_j})}{\|\boldsymbol{R}_j(\boldsymbol{p}_j - \boldsymbol{p}_{o_j})\|} \tag{B5}$$

where $\boldsymbol{R}_1 = \boldsymbol{I}$ for the first camera and $\boldsymbol{R}_2 = \boldsymbol{R}$ for the second camera. $\boldsymbol{p}_i = [u_{p_i}, v_{p_i}, -f_i]^T$ [10]. The minimum distance between the two intersecting rays can be obtained from (B4)

$$e_{\text{int}} = \|h_1\boldsymbol{l}_1 - (\boldsymbol{T} + h_2\boldsymbol{l}_2)\|. \tag{B6}$$

The distance between two 3-D points $\boldsymbol{P}$ and $\boldsymbol{Q}$ is computed as

$$d = \left\| \frac{h_1\boldsymbol{l}_1 + (\boldsymbol{T} + h_2\boldsymbol{l}_2)}{2}\bigg|_{\boldsymbol{P}} - \frac{h_1\boldsymbol{l}_1 + (\boldsymbol{T} + h_2\boldsymbol{l}_2)}{2}\bigg|_{\boldsymbol{Q}} \right\|. \tag{B7}$$

ACKNOWLEDGMENT

The authors would like to thank Dr. Hansen for his critical reading of the manuscript and an anonymous referee for the suggestions.

REFERENCES

[1] N. A. Borghese, G. Ferrigno, G. Baroni, R. Savarè, S. Ferrari, and A. Pedotti, "AUTOSCAN: A flexible and portable scanner of 3-D surfaces," *IEEE Comput. Graph. Appl.*, vol. 18, pp. 2–5, May/June 1998.
[2] N. A. Borghese, M. Di Rienzo, G. Ferrigno, and A. Pedotti, "Elite: A goal-oriented vision system for moving objects detection," *Robotica*, vol. 9, pp. 275–282, Sept. 1990.
[3] G. Ferrigno and A. Pedotti, "Modularly expansible system for real-time processing of a TV display, useful in particular for the acquisition of coordinates of known shapes objects," U.S. Patent 4 706 296, 1990.
[4] X. Hu and N. Ahuja, "Matching point features with ordered geometric, rigidity and disparity constraints," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. 16, pp. 1041–1045, Oct. 1994.
[5] J. Weng, P. Cohen, and M. Herniou, "Camera calibration with distortion models and accuracy evaluation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 14, pp. 965–979, Oct. 1992.
[6] N. A. Borghese and G. Ferrigno, "An algorithm for 3-D automatic movement detection by means of standard TV cameras," *IEEE Trans. Bio-Med. Eng.*, vol. 37, pp. 1221–1225, Dec. 1990.
[7] A. Gruen and H. A. Beyer, "System calibration through self-calibration," in *Proc. Workshop on Camera Calibration and Orientation in Computer Vision*, Washington, DC, Aug. 1992, pp. 218–233.
[8] H. G. Maas, "Dynamic photogrammetric calibration of industrial robots," in *Proc. Videometrics V*, vol. 3174, 1997, pp. 106–112.
[9] H. C. Longuet-Higgins, "A computer algorithm for reconstructing a scene from two projections," *Nature*, vol. 293, pp. 133–135, Sept. 1981.
[10] O. D. Faugeras, *Three-Dimensional Computer Vision*. Cambridge, MA: M.I.T. Press, 1992.
[11] R. I. Hartley, "In defence of 8-point algorithm," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, pp. 580–593, June 1997.
[12] N. A. Borghese and P. Perona, "Calibration of a stereo system with points of unknown location," in *Proc. 14th Int. Conf. Soc. Biomech. ISB*, Paris, France, June 1993, pp. 202–203.
[13] Q. T. Luong and O. D. Faugeras, "The fundamental matrix: Theory, algorithms, and stability analysis," *Int. J. Comp. Vision*, vol. 17, no. 1, pp. 43–76, 1996.
[14] N. A. Borghese and P. Cerveri, "Calibrating a video camera pair with a rigid bar," *Pattern Recognit.*, vol. 33, pp. 81–95, Jan. 2000.
[15] S. Bougnoux, "From projective to Euclidean space under any practical situation, a criticism of self-calibration," in *Proc. 1998 IEEE Int. Conf. Computer Vision*, Bombay, India, Sept. 1998, pp. 790–796.
[16] R. I. Hartley, "Kruppa equations derived from the fundamental matrix," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, pp. 133–135, Feb. 1997.
[17] R. K. Lenz and R. Y. Tsai, "Techniques for calibration of the scale factor and image center for high accuracy 3-D machine vision metrology," in *Proc. IEEE Int. Conf. Robotics Automation*, Raleigh, NC, July 1987, pp. 68–75.
[18] P. Cerveri, N. A. Borghese, and A. Pedotti, "Complete calibration of a stereo photogrammetric system through control points of unknown coordinates," *J. Biomech.*, vol. 31, no. 10, pp. 935–940, 1998.
[19] P. R. Wolf, *Elements of Photogrammetry*. New York: McGraw-Hill, 1983.
[20] T. Bäck, G. Rudolph, and H.-P. Schwefel, "Evolutionary programming and evolution strategies: similarities and differences," in *Proceedings of the Second Annual Conference on Evolutionary Programming*, D. B. Fogel and W. Atmar, Eds. La Jolla, CA: Evol. Programm. Soc., 1993, pp. 11–22.
[21] N. Hansen and A. Ostermeier, "Completely derandomised self- adaptation in evolution strategies," Evol. Comput., to be published.
[22] A. Ostermeier, A. Gawelczyk, and N. Hansen, "A derandomized approach to self-adaptation of evolution strategies," *Evol. Comput.*, vol. 2, no. 4, pp. 369–380, 1994.
[23] N. Hansen and A. Ostermeier, "Adapting arbitrary normal mutation distributions in evolution strategies: The covariance matrix adaptation," in *Proc. 1996 IEEE Int. Conf. Evolutionary Computation*, 1996, pp. 312–317.
[24] L. P. Kaelbing, M. L. Littman, and A. W. Moore, "Reinforcement learning: A survey," *J. Artif. Intell. Res.*, vol. 4, pp. 237–285, June 1996.
[25] Z. Zhang, R. Deriche, O. Faugeras, and Q.-T. Luong, "A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry," *Artif. Intel. J.*, vol. 78, pp. 87–119, Oct. 1995.