Nonlinear Estimation of the Fundamental Matrix with Minimal Parameters

Adrien Bartoli and
Peter Sturm, *Member*, *IEEE Computer Society*

Abstract—The purpose of this paper is to give a very simple method for nonlinearly estimating the fundamental matrix using the minimum number of seven parameters. Instead of minimally parameterizing it, we rather update what we call its orthonormal representation, which is based on its singular value decomposition. We show how this method can be used for efficient bundle adjustment of point features seen in two views. Experiments on simulated and real data show that this implementation performs better than others in terms of computational cost, i.e., convergence is faster, although methods based on minimal parameters are more likely to fall into local minima than methods based on redundant parameters.

Index Terms—Structure-from-motion, bundle adjustment, minimal parameterization, fundamental matrix.

1 Introduction

THE fundamental matrix has received a great interest in the computer vision community, see, e.g., [5], [6], [11], [12], [20], [23], [24]. It encapsulates the epipolar geometry or the projective motion between two uncalibrated perspective cameras and can be used for 3D reconstruction, motion segmentation, self-calibration, etc. Accurately estimating the fundamental matrix is therefore a major research issue. Most of the time, point correspondences between the two images are used. A linear solution is obtained using the 8-point algorithm [5], [11] optionally embedded in a robust estimation scheme [20], [23]. This estimate is then nonlinearly refined by minimizing a physically meaningful criterion that may involve reconstructed 3D point coordinates as well (in particular for bundle adjustment). However, nonlinearly estimating the fundamental matrix suffers from the lack of a simple technique to represent it efficiently. This paper, which is an extension of [2], provides such a technique in Section 3, based on the orthonormal representation of the fundamental matrix that we introduce. We show in Section 4 how this method can be used to refine the fundamental matrix by bundle adjustment of point features. We demonstrate experimentally in Sections 5.1 and 5.2 that the resulting algorithm performs better than existing ones in terms of computational cost.

2 Notations and Relation to Previous Work

The fundamental matrix denoted as F is a homogeneous (i.e., defined up to scale) (3×3) rank-2 matrix. It therefore has nine entries, but only 7 degrees of freedom.

There have been many attempts to minimally parameterize it, i.e., to represent it with seven parameters. Most of the previous works deal with directly parameterizing the epipolar geometry. The fundamental matrix F is decomposed into the epipoles $\mathbf{e}$ and $\mathbf{e}'$ and the epipolar transformation, which is a 1D projective transformation relating the epipolar pencils, represented by a homogeneous (2×2) matrix g [4], [12], [23].

Representing these entities with minimal parameters requires eliminating their arbitrary scale factors. This can be done by fixing, e.g., the 2-norm of homogeneous entities, but then the parameterization would not be minimal. Another solution is to freeze one entry of each homogeneous entity (in practice, the largest entry), which yields three possibilities for each epipole and four for the epipolar transformation, so $3 \cdot 3 \cdot 4 = 36$ possible parameterizations.

In [12], the authors propose to restrict the two-view configurations considered to the cases where both epipoles are finite and can therefore be expressed in affine coordinates. Consequently, this parameterization can be used only when both epipoles do not lie at infinity. Due to the homogeneity of the epipolar transformation, four distinct parameterizations are still necessary for g. A total of four parameterizations are then needed to represent this restricted set of fundamental matrices.

The method has been extended in [23] to the general case, i.e., when the epipoles can be either finite or infinite. In this case, it is shown that all 36 distinct parameterizations are necessary. This leads to a cumbersome and error-prone implementation of the optimization process.

Note that there are nine different possibilities to form the fundamental matrix—or any other 2D entity such as the extended epipolar transformation [4] or the canonic plane homography $\mathrm{H}^\star$ [13]—from $\mathbf{e}$, $\mathbf{e}'$, and g [23].

In [4], [24], the method has been revised so as to reduce the number of parameterizations using image transformations. In [4], the image transformations used are metric and the number of distinct parameterizations is restricted to three plus one bilinear constraint on the entries of g, while, in [24], the transformations used are projective, which allows one to reduce the number of parameterizations to one. The main drawback is that in the transformed image space, the original noise model on the image features is not preserved. A means to preserve it, up to first order approximation, has been proposed in [24] for the gradient-weighted criterion, which is not the one used for bundle adjustment.

Another solution is the point-based parameterization of [19]. The idea is to represent the fundamental matrix by a set of 7-point correspondences. Minimal optimization can then be conducted by varying one coordinate for each point correspondence. The fundamental matrix is obtained at each minimization step by computing the standard 7-point solution, which means that the null-space of a (7×9) matrix has to be computed and a cubic equation has to be solved. There may be up to three solutions. The one giving the lowest residual error is kept. The disadvantage of this parameterization is that it is costly to obtain the fundamental matrix given its parameters (i.e., the 7-point correspondences). Also, analytic differentiation is not possible.

3 Nonlinear Optimization with Seven Parameters

In contrast to the existing work, we do not try to represent the entire set of fundamental matrices using seven parameters. We rather locally update it with seven parameters. Before going further, we illustrate this idea by considering the case of the nonlinear estimation of 3D rotations, which is simpler and, as will be seen later, has similarities with the case of the fundamental matrix.

3.1 The Case of 3D Rotations

There exist many representations of 3D rotations, see, e.g., [18], including Euler angles, the Gibbs vector, Cayley-Klein parameters, Pauli spin matrices, axis-and-angle systems, $SO(3)$ matrices,[1] and

• *The authors are with INRIA Rhone-Alpes, 655 avenue de l'Europe, 38334 Saint Ismier cedex, France. E-mail: {Adrien.Bartoli, Peter.Sturm}@inria.fr.*

Manuscript received 30 Apr. 2002; revised 26 Jan. 2003; accepted 19 July 2003.
Recommended for acceptance by M. Irani.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 116437.

1. $SO(3)$ is the Lie group of (3×3) matrices R satisfying $\mathrm{R}^\top \mathrm{R} = \mathrm{I}$ and $\det \mathrm{R} = 1$.

unit quaternions. None of these representations is able to uniquely represent all 3D rotations with the minimum three parameters. For that reason, the following scheme is often used for their nonlinear estimation, see, e.g., [1], [7], [21]. The rotation is represented by an $SO(3)$ matrix R and is locally updated using three parameters by any well-behaved (locally nonsingular) representation, such as three Euler angles $\boldsymbol{\theta}^\top = (\theta_1\ \theta_2\ \theta_3)$ as:

$$\mathrm{R} \leftarrow \mathrm{R}\,\mathrm{R}(\boldsymbol{\theta}), \tag{1}$$

where $\mathrm{R}(\boldsymbol{\theta}) = \mathrm{R}_x(\theta_1)\,\mathrm{R}_y(\theta_2)\,\mathrm{R}_z(\theta_3)$ is the $SO(3)$ matrix representation of the 3D rotation corresponding to $\boldsymbol{\theta}$ with

$$\mathrm{R}_x(\theta_1) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_1 & -\sin\theta_1 \\ 0 & \sin\theta_1 & \cos\theta_1 \end{pmatrix},$$
$$\mathrm{R}_y(\theta_2) = \begin{pmatrix} \cos\theta_2 & 0 & \sin\theta_2 \\ 0 & 1 & 0 \\ -\sin\theta_2 & 0 & \cos\theta_2 \end{pmatrix},$$
$$\mathrm{R}_z(\theta_3) = \begin{pmatrix} \cos\theta_3 & -\sin\theta_3 & 0 \\ \sin\theta_3 & \cos\theta_3 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

At the end of each iteration, R is updated and $\boldsymbol{\theta}$ is reset to zero. Hence, at each iteration, the estimated Euler angles are small (initialized as zero), which makes this representation nonsingular.

3.2 Minimal Update

Following the example of 3D rotations, we propose the *orthonormal representation* of the fundamental matrix where more parameters than degrees of freedom are needed, but that can be easily updated using the minimum seven parameters.

Given an estimate of the fundamental matrix F obtained using, e.g., the 8-point algorithm, consider its singular value decomposition $\mathrm{F} \sim \mathrm{U}\Sigma\mathrm{V}^\top$, where U and V are $O(3)$ matrices[2] and Σ a diagonal one containing the singular values of F. Since F has rank 2, $\Sigma \sim \mathrm{diag}(\sigma_1, \sigma_2, 0)$, where $\sigma_1 \geq \sigma_2 > 0$ [22]. We can scale Σ such that $\mathrm{F} \sim \mathrm{U}\,\mathrm{diag}(1, \sigma, 0)\,\mathrm{V}^\top$, where $\sigma = \sigma_2/\sigma_1$ ($\sigma_1 \neq 0$ since $\mathrm{F} \neq 0$) and $1 \geq \sigma > 0$.

This decomposition shows that any fundamental matrix can be represented by $(\mathrm{U}, \mathrm{V}, \sigma)$, i.e., two $O(3)$ matrices and a scalar, which form what we call its orthonormal representation. Note that, in the case $\sigma = 1$, i.e., when the fundamental matrix is an essential matrix [8], the orthonormal representation is not unique (see below).

The orthonormal representation is consistent in that it yields $3 + 3 + 1 = 7$ degrees of freedom. The fundamental matrix can be recovered as:

$$\mathrm{F} \sim \mathbf{u}_1\mathbf{v}_1^\top + \sigma\mathbf{u}_2\mathbf{v}_2^\top, \tag{2}$$

where $\mathbf{u}_i$ and $\mathbf{v}_i$ are the columns of U and V, respectively.

This representation suggests the following update scheme. Each $O(3)$ matrix can be updated using an $SO(3)$ matrix, using (1) as in the case of 3D rotations, while σ can be included as such into the optimization:

$$\mathrm{U} \leftarrow \mathrm{U}\,\mathrm{R}(\mathbf{x}) \quad \mathrm{V} \leftarrow \mathrm{V}\,\mathrm{R}(\mathbf{y}) \quad \sigma \leftarrow \sigma + \delta_\sigma. \tag{3}$$

Here, $\mathbf{x}$ and $\mathbf{y}$ are 3-vectors of Euler angles. Intuitively, the orthonormal representation should be intrinsically well-conditioned since U and V are $O(3)$ matrices.

Completeness. A first question that immediately follows about the above-proposed method is whether all two-view configurations are covered. Clearly, any fundamental matrix can be decomposed into two $O(3)$ matrices and a scalar. The question arises from the fact that U and V are $O(3)$ matrices, which may have positive or negative determinants, and are updated using $SO(3)$ matrices, $\mathrm{R}(\mathbf{x})$ and $\mathrm{R}(\mathbf{y})$, respectively, which have positive determinants. Actually, this is not a problem since the signs of U and V can be freely switched, which accordingly switches the signs of their determinants, while leaving the corresponding F invariant: $\mathrm{F} \sim (\pm\mathrm{U})\,\Sigma(\pm\mathrm{V})^\top$.

Ensuring bounds on σ. A second remark is about the bounds on σ: $0 < \sigma \leq 1$. There are several possibilities to ensure them while leaving the corresponding F invariant. However, we have found during our experiments that, in practice, this does not affect the behavior of the underlying optimization process.

Essential matrices. As pointed out previously, in the case of $\sigma = 1$, where the fundamental matrix considered is an essential matrix, the proposed orthonormal representation is not unique: If U and V represent F, then also $\mathrm{U}\,\mathrm{R}_z(\alpha)$ and $\mathrm{V}\,\mathrm{R}_z(\alpha)$ for any α. This induces that the Jacobian matrix (6) has rank 6, as shown in Section 4.3. We propose two ways to deal with this singularity.

First, one can use a nonlinear optimization technique that handles singular parameterizations, e.g., damped Newton-type techniques. Using Levenberg-Marquardt, we found in our experiments that the singularity does not induce numerical instabilities.

Second, one can avoid singular configurations by properly normalizing the image points. Indeed, an essential matrix arises usually from a semicalibrated configuration where the origin of the coordinate frame in the image lies close to the principal point and where the image coordinates have been scaled by approximately the inverse focal length. In practice, the principal point position is unknown, but it is likely to be close to the image center. Hence, singular configurations can be avoided by translating the origin of the coordinate frame off the image center.

4 Bundle Adjustment

In this section, we show how the orthonormal representation can be used for bundle adjustment of point features $\mathbf{q}_i \leftrightarrow \mathbf{q}'_i$, $i \in 1 \ldots m$ seen in two views, through the minimization of the reprojection error. Similar results can be derived for other criteria, such as the minimization of the distances between points and epipolar lines or the gradient-weighted criterion [12], [23]. However, in order to obtain the maximum likelihood estimate of the fundamental matrix, one has also to estimate corrected point positions $\widehat{\mathbf{q}}_i \leftrightarrow \widehat{\mathbf{q}}'_i$, i.e., which satisfy exactly the epipolar geometry and, therefore, correspond to 3D points $\mathbf{Q}_i$.

Bundle adjustment consists in minimizing a cost function described in Section 4.1 over structure and motion parameters. In projective space, there are 15 inherent degrees of gauge freedom, due to the coordinate-frame ambiguity. In [9], a general framework consisting in incorporating gauge constraints up to first order in numerical estimation is introduced. The method of [15] falls in that category. Another technique is to let the gauge be free to drift, sometimes partially, while it is ensured that it does not move too far at each iteration. These methods are compared to ours in Section 5.

When the motion is represented by the fundamental matrix, the gauge is completely eliminated. We call any pair of camera matrices P and P′ a *realization*. In Section 4.2, we give analytical formulae to compute a realization from the orthonormal representation of F (as opposed to [12], [19], [23], [24]).

The algorithm is summarized in Table 1.

4.1 Cost Function

Bundle adjustment consists in solving the following optimization problem, see e.g., [15], [21], [23]: $\min_{\mathrm{a},\mathrm{b}} \sum_j r_j^2$, where $\mathbf{a}$ and $\mathbf{b}$ are respectively motion and structure parameters (or parameters used to update them), $\mathbf{r}$ is the $4m$-vector of residual errors defined by:

2. $O(3)$ is the Lie group of (3×3) matrices R satisfying $\mathrm{R}^\top\mathrm{R} = \mathrm{I}$.

TABLE 1
Implementing Our Minimal Estimator within the Bundle Adjustment Levenberg-Marquardt-Based Framework Given in [7, p. 574] (Algorithm A4.1)

Two-view projective bundle adjustment expressed within the framework of [7, p.574] (algorithm A4.1). The initial guess of the fundamental matrix is F_0.

Add the following steps:

(i') Initialize the orthonormal representation $(\mathrm{U}, \mathrm{V}, \sigma)$ by a scaled singular value decomposition of F_0: $\boxed{\mathsf{F}_0 \sim \mathsf{U}\ \mathrm{diag}(1, \sigma, 0)\ \mathsf{V}^\top}$.

(ii') Turn the full $(r \times 12)$ camera Jacobian matrix $\mathsf{A} = \bar{\mathsf{A}}$ into the minimal $(r \times 7)$ Jacobian matrix of the orthonormal representation: $\boxed{\mathsf{A} \leftarrow \mathsf{A}\ \mathsf{A}^\perp}$, where $\mathsf{A}^\perp$ is given by equations (6,7).

Change the parameter update step as:

(viii) Update the orthonormal representation as:

$$\boxed{\mathsf{U} \leftarrow \mathsf{U}\ \mathsf{R}(\mathbf{x})} \quad \boxed{\mathsf{V} \leftarrow \mathsf{V}\ \mathsf{R}(\mathbf{y})} \quad \boxed{\sigma \leftarrow \sigma + \delta_\sigma},$$

where $\boldsymbol{\delta}_a^\top = (\mathbf{x}^\top\ \mathbf{y}^\top\ \delta_\sigma)$ are the 7 motion update parameters, update the structure parameters by adding the incremental vector $\boldsymbol{\delta}_b$ and compute the new residual vector.

Note that r is the number of residuals and that the second projection matrix has to be extracted from the orthonormal representation using (5) (e.g., for computing the residual vector).

$$\mathbf{r}^\top_{(4m\times 1)} = (\ldots \quad \mathbf{q}_{i1} - \widehat{\mathbf{q}}_{i1} \quad \mathbf{q}_{i2} - \widehat{\mathbf{q}}_{i2} \quad \mathbf{q}'_{i1} - \widehat{\mathbf{q}}'_{i1} \quad \mathbf{q}'_{i2} - \widehat{\mathbf{q}}'_{i2} \quad \ldots),$$

where $\widehat{\mathbf{q}}_i \sim \mathrm{P}\mathbf{Q}_i$ and $\widehat{\mathbf{q}}'_i \sim \mathrm{P}'\mathbf{Q}_i$ are predicted image points.

4.2 Computing a Realization

Due to the projective frame ambiguity, there exists a 15-parameter family of realizations for a given fundamental matrix. A common choice is the canonic projection matrices given by [13]:

$$\mathrm{P} \sim (\mathrm{I}_{(3\times 3)}\ 0_{(3\times 1)}) \text{ and } \mathrm{P}' \sim (\mathrm{H}^\star\ \gamma \mathbf{e}'), \tag{4}$$

where $\mathbf{e}'$ is the second epipole, given by the left null-vector of F, $\mathrm{F}^\top \mathbf{e}' \sim 0_{(3\times 1)}$, and $\mathrm{H}^\star \sim [\mathbf{e}']_\times \mathrm{F}$ is the canonic plane homography [13]. The arbitrary scalar γ fixes the relative scale between $\mathrm{H}^\star$ and $\mathbf{e}'$. Without loss of generality, we assume that $\gamma = \|\mathbf{e}'\| = 1$. Any other realization can then be obtained by postmultiplying P and P′ by a nonsingular 3D homography.

Computing the canonic projection matrices (4) can be achieved directly from the orthonormal representation of F. The second epipole is the last column of U: $\mathbf{e}' \sim \mathbf{u}_3$ ($\|\mathbf{u}_3\| = 1$), so the canonic plane homography can be formulated as:

$$\mathrm{H}^\star \sim [\mathbf{e}']_\times \mathrm{F} \sim [\mathbf{u}_3]_\times \left(\mathbf{u}_1 \mathbf{v}_1^\top + \sigma \mathbf{u}_2 \mathbf{v}_2^\top\right).$$

Since U is an $O(3)$ matrix, $[\mathbf{u}_3]_\times \mathbf{u}_1 = \pm \mathbf{u}_2$ and $[\mathbf{u}_3{}_\times \mathbf{u}_2 = \mp \mathbf{u}_1$ which yields $\mathrm{H}^\star \sim \mathbf{u}_2 \mathbf{v}_1^\top - \sigma \mathbf{u}_1 \mathbf{v}_2^\top$ and, thus, the particularly simple and direct form of the second projection matrix:

$$\mathrm{P}' \sim \left(\mathbf{u}_2 \mathbf{v}_1^\top - \sigma \mathbf{u}_1 \mathbf{v}_2^\top \mid \mathbf{u}_3\right). \tag{5}$$

4.3 Analytical Differentiation

Many nonlinear optimization methods necessitate computing the Jacobian matrix $\mathrm{J} = (\mathrm{A} \mid \mathrm{B})$ of the residual vector r with respect to motion and structure parameters $\mathbf{a}$ and $\mathbf{b}$. While this can be achieved numerically using, e.g., finite differences [16], it may be better to use an analytical form for both computational efficiency and numerical accuracy. We focus on the computation of $\mathrm{A} = \frac{\partial \mathbf{r}}{\partial \mathbf{a}}$ since $\mathrm{B} = \frac{\partial \mathbf{r}}{\partial \mathbf{b}}$ only depends upon structure parameterization. Let $\mathbf{p}' = \mathrm{vect}(\mathrm{P}')$, where $\mathrm{vect}(.)$ is the row-wise vectorization. We decompose $\mathrm{A}_{(4m\times 7)} = \frac{\partial \mathbf{r}}{\partial \mathbf{p}'} \frac{\partial \mathbf{p}'}{\partial \mathbf{a}} = \bar{\mathrm{A}}_{(4m\times 12)}\ \mathrm{A}^\perp_{(12\times 7)}$. Only the 12 entries of P′ are considered since P is fixed in the canonic reconstruction basis (4). The matrix $\bar{\mathrm{A}} = \frac{\partial \mathbf{r}}{\partial \mathbf{p}'}$ depends on the chosen realization of the fundamental matrix, i.e., on the coordinate frame employed. We have chosen the canonic projection matrices (4). This Jacobian matrix is employed directly for the overparameterization proposed in [6]. Deriving its analytical form is straightforward. We therefore concentrate on deriving a closed-form expression for $\mathrm{A}^\perp$. One of the advantages of the update rule (3) is that there exists a simple closed-form expression for $\mathrm{A}^\perp$. Nonlinear least squares with analytical differentiation can be applied based on $\mathrm{A}^\perp$.

Let us consider the orthonormal representation $(\mathrm{U}, \mathrm{V}, \sigma)$. The motion update parameters are minimal and defined by $\mathbf{a}^\top = (x_1\ x_2\ x_3\ y_1\ y_2\ y_3\ \sigma)$, where $\mathbf{x}^\top = (x_1\ x_2\ x_3)$ and $\mathbf{y}^\top = (y_1\ y_2\ y_3)$ are used to update U and V, respectively. Since U and V are updated with respect to the current estimate, $\mathrm{A}^\perp$ is evaluated at $(\mathrm{U}, \mathrm{V}, \sigma)$, i.e., at $\mathbf{a}^\top = \mathbf{a}_0^\top = (0^\top_{(6\times 1)}\ \sigma)$. Equation (5) is used to derive a closed-form expression of the second canonic projection matrix after updating. By expanding, differentiating and evaluating this expression at a_0, we obtain:

$$\mathrm{A}^\perp = \frac{\partial \mathbf{p}'}{\partial \mathbf{a}} = \left(\left(\frac{\partial \mathbf{p}'}{\partial x_1}\right) \cdots \left(\frac{\partial \mathbf{p}'}{\partial y_3}\right)\ \left(\frac{\partial \mathbf{p}'}{\partial \sigma}\right)\right), \tag{6}$$

where:

$$\begin{aligned}
\partial \mathbf{p}'/\partial x_1 &= \mathrm{vect}(\mathbf{u}_3 \mathbf{v}_1^\top \mid -\mathbf{u}_2) \\
\partial \mathbf{p}'/\partial x_2 &= \mathrm{vect}(\sigma \mathbf{u}_3 \mathbf{v}_2^\top \mid \mathbf{u}_1) \\
\partial \mathbf{p}'/\partial x_3 &= \mathrm{vect}(-\mathbf{u}_1 \mathbf{v}_1^\top - \sigma \mathbf{u}_2 \mathbf{v}_2^\top \mid \mathbf{0}_{3\times 1}) \\
\partial \mathbf{p}'/\partial y_1 &= \mathrm{vect}(-\sigma \mathbf{u}_1 \mathbf{v}_3^\top \mid \mathbf{0}_{3\times 1}) \\
\partial \mathbf{p}'/\partial y_2 &= \mathrm{vect}(-\mathbf{u}_2 \mathbf{v}_3^\top \mid \mathbf{0}_{3\times 1}) \\
\partial \mathbf{p}'/\partial y_3 &= \mathrm{vect}(\mathbf{u}_2 \mathbf{v}_2^\top + \sigma \mathbf{u}_1 \mathbf{v}_1^\top \mid \mathbf{0}_{3\times 1}) \\
\partial \mathbf{p}'/\partial \sigma &= \mathrm{vect}(-\mathbf{u}_1 \mathbf{v}_2^\top \mid \mathbf{0}_{3\times 1}).
\end{aligned} \tag{7}$$

In the general case, $\mathrm{rank}(\mathrm{A}^\perp) = 7$, but when $\sigma = 1$, $\mathrm{rank}(\mathrm{A}^\perp) = 6$ since $\partial \mathbf{p}'/\partial x_3 + \partial \mathbf{p}'/\partial y_3 = 0$.

If the minimal method of, e.g., [23] were used, 36 different Jacobian matrices, one for each parameterization, would have to be derived.

4.4 Particular Configurations

The epipolar geometry can be decomposed as a pair of epipoles and the 3-degrees of freedom epipolar transformation [12], [23]. If one or two of these components are a priori known, it may be convenient to leave them invariant during optimization of the fundamental matrix. Such features are easily added to our estimation method, as follows.

Leaving an epipole invariant. Consider, e.g., the second epipole encapsulated in the orthonormal representation as the third column of U. The update $\mathrm{U} \leftarrow \mathrm{U}\ \mathrm{R}(\mathbf{x})$ does not affect $\mathbf{u}_3$ if $x_1 = x_2 = 0$. Therefore, freezing the left or the right epipole can be done by removing x_1, x_2 or y_1, y_2, respectively, from the estimation and updating as $\mathrm{U} \leftarrow \mathrm{U}\ \mathrm{R}_z(x_3)$ or $\mathrm{V} \leftarrow \mathrm{V}\ \mathrm{R}_z(y_3)$, respectively.

Leaving the epipolar transformation invariant. The epipoles are encapsulated by the x_1, x_2 and the y_1, y_2 update parameters. Hence, the 3 degrees of freedom of the epipolar transformation are contained in the remaining update parameters: x_3, y_3, and σ. Removing them from the optimization freezes the underlying epipolar transformation.

5 Experimental Results

We compare an algorithm based on the orthonormal representation to other algorithms. We use simulated and real data in Sections 5.1 and 5.2, respectively. Below, we give details about the compared methods, the measured quantities, the computation of an initial suboptimal solution for structure and motion, and the nonlinear optimization scheme we use.

Compared methods. We compare the following motion parameterizations:

- FREE directly optimizes the 24 entries of the camera matrices. The gauge is left free to drift. The $24 - 7 = 17$ extra parameters are the homogeneous factors of each camera matrix and the 15-dimensional projective basis.
- NORMALIZED [15] is similar to FREE, but the gauge is fixed since a normalized coordinate frame is used. This is done by renormalizing the reconstruction before each step of the nonlinear minimization and by including first-order gauge constraints into the minimization. The reconstruction basis, as well as the homogeneous scale of the camera matrices are constrained.
- PARFREE [6] partially fixes the gauge by optimizing only the entries of the second camera matrix, while keeping $\mathrm{P} \sim (\mathrm{I}\ 0)$. The $12 - 7 = 5$ extra parameters are the homogeneous scale of the second camera matrix, the global scene scale, and the position of the plane at infinity.
- MAPS [3], [23] is a minimal parameterization based on multiple maps.
- ORTHO uses the orthonormal representation proposed in this paper.

Measured quantities. We measure two quantities characterisic of a bundle adjustment process, computational cost, i.e., CPU time to convergence and the error at convergence.

Structure parameterization. We use the structure parameterization proposed in [7] which consists in scaling the reconstructed points such that their third element is unity. The three remaining free elements are then optimized. Note that this parameterization can be used only when a canonical basis enforcing $\mathrm{P} \sim (\mathrm{I}\ 0)$ is used. Therefore, methods FREE and NORMALIZED have their own structure parameterization: They optimize the four elements of each point.

Initialization. We compute an initial solution for the motion using the normalized 8-point algorithm [5]. Image point coordinates are standardized such that they lie in $[-1 \ldots 1]$. Each point is reconstructed by minimizing its reprojection error.

Nonlinear optimization. We use the Levenberg-Marquardt technique with analytic differentiation. This is a damped Gauss-Newton method. Let J be the Jacobian matrix and $\mathrm{H} = \mathrm{J}^\top \mathrm{J}$ the Gauss-Newton approximation of the Hessian matrix. The damping consists in augmenting the normal equations $\mathrm{H}\delta = -\mathrm{J}^\top \mathbf{r}$ to be solved at each iteration: $\mathrm{H} \leftarrow \mathrm{H} + \mathrm{W}(\lambda)$. The parameter $\lambda \in \mathbb{R}$ is tuned heuristically, as described in [7], [21]. We try two approaches for the step control strategy, i.e., the choice of matrix $\mathrm{W}(\lambda)$. First, in [21], the authors recommend $\mathrm{W}(\lambda) = \lambda \mathrm{I}$. This is the original idea of the Levenberg-Marquardt algorithm [10], [14]. This will be referred to as LM. Second, in [7], the authors recommend $\mathrm{W}(\lambda) = (1 + \lambda)\ \mathrm{diag}(\mathrm{H})$, i.e., multiply the diagonal entries of H by $1 + \lambda$. This strategy is recommended in [16] and is due to [17]. This will be referred to as SEBER.

Note that gauge freedoms cause $\mathrm{H} = \mathrm{J}^\top \mathrm{J}$ to be rank-deficient, but that the damped matrix is guaranteed to have full-rank. Hence, Levenberg-Marquardt iterations change both the actual estimated geometry as well as the gauge.

We take advantage of the sparse structure of H and J to efficiently solve the augmented normal equations, as described in [7], [21]. More precisely, the sparseness of the structure parameters is exploited, and the complexity of the computation is $\mathcal{O}(mp^3)$, where m is the number of points and p is the number of motion parameters. Hence, we can expect the computational cost for an iteration to be similar for all parameterizations when the number of points is very large, and to be very different when the number of points is low.

We stop the estimation when the difference between two consecutive residual errors is lower than a threshold ξ, chosen typically in the range $10^{-8} \leq \xi \leq 10^{-4}$.

5.1 Simulated Data

5.1.1 Experimental Setup

We simulate points lying in a cube with one meter side length, observed by two cameras looking at the center of the cube. The standard configuration is the following: The focal length of the cameras is 1,000 (expressed in number of pixels). They are situated 10 meters away from the center of the cube and the baseline between them is one meter. The number of simulated points is 50. We add a centered Gaussian noise on true point positions with a 2-pixel variance. The normal equations are augmented using method LM. Each parameter of the above-described setup is independently varied to compare the parameterizations in different situations. The results are averaged over 100 trials. Computing the median gives similar results.

5.1.2 Results

Fig. 1 shows the results. We observe that all methods have roughly the same accuracy, i.e., they give the same reprojection errors, up to small discrepancies. Further comments on these discrepancies are given in the next paragraph.

On the other hand, there are quite large discrepancies between the computational costs of each method. The methods that have the highest computational costs are NORM and FREE, followed by PARFREE. The minimal methods MAPS and ORTHO have the lowest computational cost, roughly the same. These discrepancies are explained by the fact that redundant methods have more unknowns to estimate than minimal ones. Solving the normal equations is therefore more expensive (see below). These observations are valid for other experiments (not shown here) where the focal length of the cameras is varied from 500 to 2,000 pixels and where the baseline is varied from one to three meters. We also conduct the same experiments while augmenting the normal equation using SEBER. The same observations as above are valid. The results for all methods, compared to the LM augmentation, are worse in terms of both computational cost and reprojection error, while the discrepancies between the different methods for the reprojection error are reduced.

We observe that, in our C implementation, the computational cost of each iteration is dominated by the resolution of the normal equations, whose size is directly linked to the number of parameters. We measure the computational cost of an iteration for the different parameterizations. As said above, the complexity is linear in the number of points and cubic in the number of motion parameters. For different numbers of points, we obtain the results shown in Table 2.

These results show that the differences in computational costs are largely dominated by the number of motion parameters. The discrepancies become smaller when the number of points increases beyond 10,000, which is very large in the case of structure from motion for two views.

5.1.3 Convergence

As said above, there are small discrepancies in the reprojection errors achieved by the different methods, see in particular Fig. 1a.

TABLE 2
Computation Time (Seconds) of an Iteration for Different Parameterizations

	number of points			
parameterization	10	100	1,000	10,000
ORTHO	0.0045	0.0251	0.2152	2.0658
MAPS	0.0046	0.0251	0.2151	2.0658
PARFREE	0.0056	0.0307	0.2591	2.0753
FREE	0.0120	0.0589	0.5729	2.3231
NORM	0.0130	0.0664	0.6791	2.4148

These small discrepancies are due to the fact that each parameterization may lead to a different local minimum of the cost function. To better characterize this phenomenon, we measure the rate of successful estimations for the different methods against the distance from the scene to the cameras. An estimation is successful if it is not improved by any of the other compared method. More precisely, let M and M' designate two methods and $\mathcal{E}_{\mathrm{M}}(\mathrm{M}')$ be the error achieved by method M initialized by the result of method M'. We define the success of an estimation made with method M as:

$$\mathrm{Success}(\mathrm{M}) \equiv \left[\forall \mathrm{M}' \neq \mathrm{M},\ |\mathcal{E}_{\mathrm{M}}(\mathrm{INIT}) - \mathcal{E}_{\mathrm{M}'}(\mathrm{M})| \leq \xi\right],$$

where ξ is the threshold used to stop the iterations. We obtain the results as shown in Table 3.

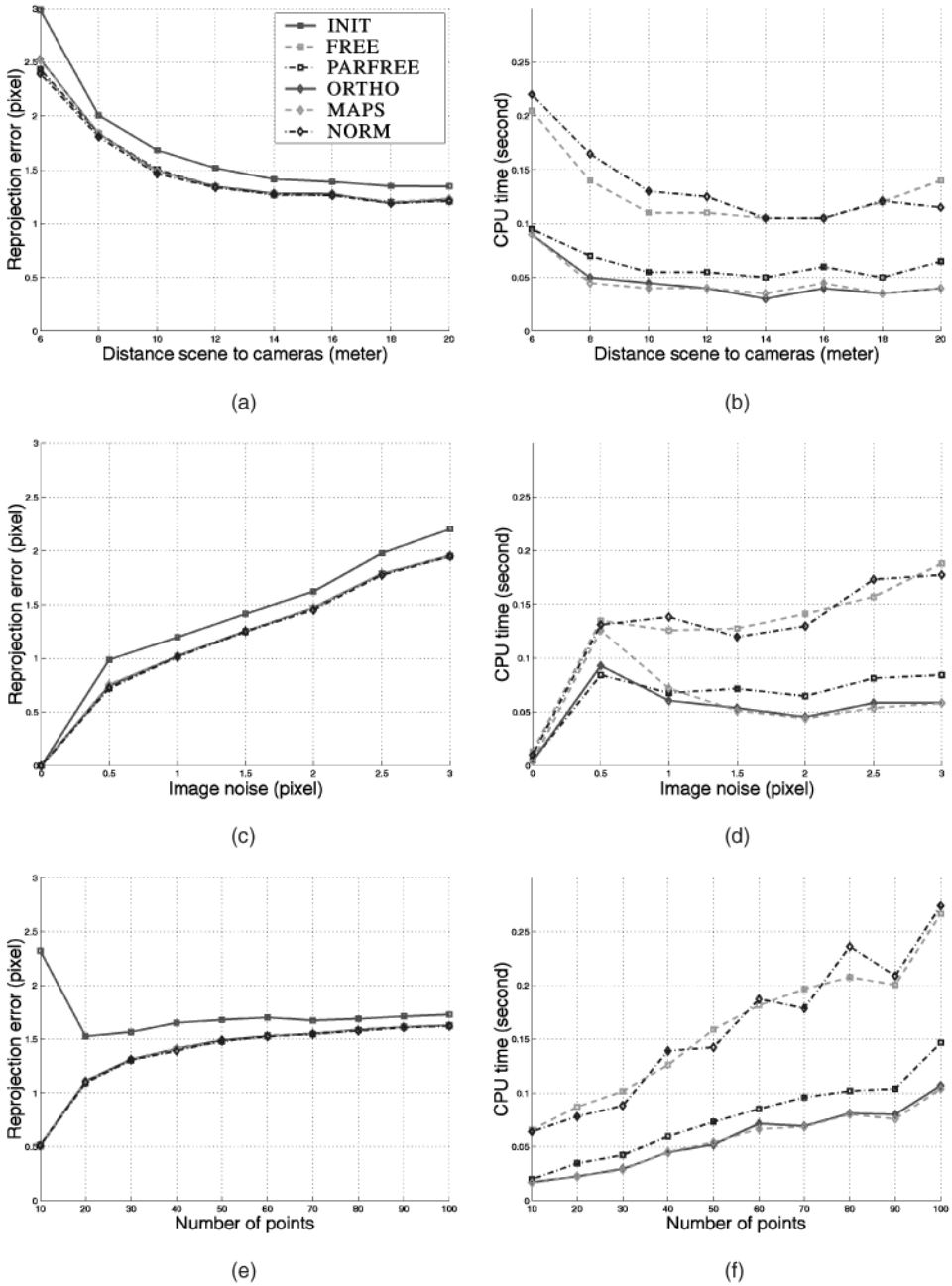


Fig. 1. Reprojection error (left column) and CPU time to convergence (right column) measured against different simulation parameters: distance scene to cameras (first row), image noise (second row), and number of points (third row). Concerning the reprojection error, the curves are almost always undistinguishable, apart from the initialization. For the CPU time, methods are divided into three groups: (from top to bottom) FREE and NORM, PARFREE, then MAPS and ORTHO.

TABLE 3
Convergence Results Shown as Success Rates in Percent

		distance scene to cameras (meter)							
	parameterization	6	8	10	12	14	16	18	20
minimal	ORTHO	83	97	97	100	100	98	100	99
	MAPS	88	95	96	100	100	99	98	99
redundent	PARFREE	100	100	94	100	100	100	100	100
	FREE & NORM	100	100	100	100	100	100	100	100

In the light of these results, we can say that methods using minimal parameters fall into local minima more often than methods based on redundant parameters. An explanation is that the minimal parameterizations are nonlinear, while the overparameterizations are linear, in the entries of the projection matrices. Hence, the local quadratic approximation of the cost fonction used in Levenberg-Marquardt is more accurate for overparameterizations.

5.1.4 Essential Matrix

As pointed out in Section 3.2, the orthonormal representation has a one-dimensional ambiguity when an essential matrix is considered. We want to check if, in the essential or near-essential cases, the orthonormal representation could induce numerical instabilities in the optimization process. For that purpose, we repeat the previous experiments, with the following two changes.

First, we map the fundamental matrix given by the 8-point algorithm to the closest essential matrix [8] and use this as an initial solution for the nonlinear optimization. Hence, the target epipolar geometry is a fundamental matrix, but the initial solution is an essential one.

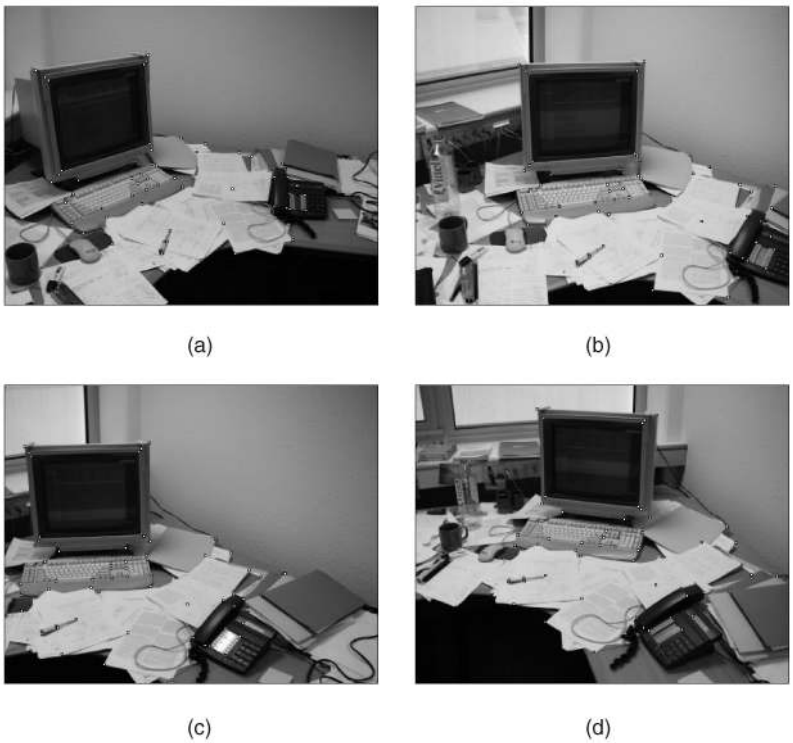
Second, instead of using the coordinates of the points in the images, we use the coordinates of the points on the retina. Hence, the underlying true epipolar geometry is represented by an essential matrix. We run the experiments based on varying the geometry of the problem for both SEBER and LM.

We obtained results very similar to the previous experiments. This means that the orthonormal representation can be used for both fundamental and essential matrices, without inducing numerical instabilities, when an appropriate nonlinear optimizer is employed.

5.2 Real Data

We use different pairs of the images shown in Table 4, in order to cover all possibilities for the epipoles to be close to the images or at

TABLE 4
Reprojection Error at Convergence, $\mathcal{E}$, and CPU Time to Convergence, $\mathcal{T}$, Obtained When Combining Pairs of Images to Obtain Epipoles Close to the Images or Toward Infinity



epipoles		views	INIT		FREE		PARFREE		ORTHO		MAPS		NORM	
$\mathbf{e}$	$\mathbf{e}'$		$\mathcal{E}$	$\mathcal{T}$	$\mathcal{E}$	$\mathcal{T}$	$\mathcal{E}$	$\mathcal{T}$	$\mathcal{E}$	$\mathcal{T}$	$\mathcal{E}$	$\mathcal{T}$	$\mathcal{E}$	$\mathcal{T}$
$-\infty$	∞	$\mathcal{A}, \mathcal{B}$	0.49	-	0.47	0.99	0.47	0.54	0.47	0.64	0.47	0.65	0.47	1.10
		$\mathcal{A}, \mathcal{C}$	0.68	-	0.65	0.67	0.65	0.39	0.65	0.29	0.65	0.34	0.65	0.70
$-\infty$	$-\infty$	$\mathcal{A}, \mathcal{D}$	0.84	-	0.67	0.70	0.67	0.38	0.67	0.33	0.67	0.33	0.67	0.74
∞	∞	$\mathcal{B}, \mathcal{C}$	0.57	-	0.53	0.27	0.53	0.14	0.53	0.15	0.53	0.14	0.53	0.23
∞	$-\infty$	$\mathcal{B}, \mathcal{D}$	0.79	-	0.55	0.25	0.55	0.23	0.55	0.18	0.55	0.19	0.55	0.25
		$\mathcal{C}, \mathcal{B}$	0.57	-	0.53	0.30	0.53	0.10	0.53	0.12	0.53	0.21	0.53	0.20
average $\bar{\mathcal{E}}$ and $\bar{\mathcal{T}}$			0.66	-	0.57	0.53	0.57	0.30	0.57	0.28	0.57	0.31	0.57	0.54

infinity, with 60 point correspondences. The results are shown in Table 4. For each combination of images and each algorithm, we estimate the computational cost and the reprojection error. The last row of the table shows mean values for each algorithm over the set of image pairs. Note that, for any image pair, the reprojection error is the same for all algorithms. Methods ORTHO, PARFREE, and MAPS give the lowest computational costs, roughly twice as low as those of methods FREE and NORM. We obtain similar results using SEBER.

6 Conclusions

We studied the problem of estimating the fundamental matrix over a minimal set of seven parameters. We proposed the orthonormal representation which enables to easily update an estimate of the fundamental matrix using seven parameters. The canonic projection matrices can be directly extracted from the orthonormal representation. The method can be plugged into most of the (possibly sparse) nonlinear optimizers such as Levenberg-Marquardt. We gave a closed-form expression for the Jacobian matrix of the residuals with respect to the motion parameters for bundle adjustment purposes, necessary for Newton-type optimization techniques.

We conducted experiments on simulated and real data. Our conclusions are that the methods based on minimal parameter sets have lower computational cost, but may be more frequently trapped in local minima.

Acknowledgments

The authors would like to thank Bill Triggs for discussions and one of the anonymous reviewers for very useful comments.

References

[1] K.B. Atkinson, ed., *Close Range Photogrammetry and Machine Vision.* Whittles Publishing, 1996.
[2] A. Bartoli, "On the Non-Linear Optimization of Projective Motion Using Minimal Parameters," *Proc. Seventh European Conf. Computer Vision,* vol. 2, pp. 340-354, May 2002.
[3] A. Bartoli and P. Sturm, "Three New Algorithms for Projective Bundle Adjustment with Minimum Parameters," Research Report 4236, INRIA, Grenoble, France, Aug. 2001.
[4] A. Bartoli, P. Sturm, and R. Horaud, "Projective Structure and Motion from Two Views of a Piecewise Planar Scene," *Proc. Eighth Int'l Conf. Computer Vision,* vol. 1, pp. 593-598, July 2001.
[5] R. Hartley, "In Defence of the 8-Point Algorithm," *Proc. Fifth Int'l Conf. Computer Vision,* pp. 1064-1070, June 1995.
[6] R.I. Hartley, "Projective Reconstruction and Invariants from Multiple Images," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 16, no. 10, pp. 1036-1041, Oct. 1994.
[7] R.I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision.* Cambridge Univ. Press, June 2000.
[8] T.S. Huang and O.D. Faugeras, "Some Properties of the E Matrix in Two-View Motion Estimation," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 11, no. 12, pp. 1310-1312, Dec. 1989.
[9] K. Kanatani and D.D. Morris, "Gauges and Gauge Transformations for Uncertainty Description of Geometric Structure with Indeterminacy," *IEEE Trans. Information Theory,* vol. 47, no. 5, July 2001.
[10] K. Levenberg, "A Method for the Solution of Certain Non-Linear Problems in Least Squares," *Quarterly of Applied Math.,* pp. 164-168, 1944.
[11] H.C. Longuet-Higgins, "A Computer Program for Reconstructing a Scene from Two Projections," *Nature,* vol. 293, pp. 133-135, Sept. 1981.
[12] Q.T. Luong and O. Faugeras, "The Fundamental Matrix: Theory, Algorithms and Stability Analysis," *Int'l J. Computer Vision,* vol. 17, no. 1, pp. 43-76, 1996.
[13] Q.T. Luong and T. Vieville, "Canonic Representations for the Geometries of Multiple Projective Views," *Computer Vision and Image Understanding,* vol. 64, no. 2, pp. 193-229, 1996.
[14] D.W. Marquardt, "An Algorithm for Least-Squares Estimation of Nonlinear Parameters," *J. Soc. for Industrial and Applied Math.,* vol. 11, no. 2, pp. 431-441, June 1963.
[15] P.F. McLauchlan, "Gauge Invariance in Projective 3D Reconstruction," *Proc. Multi-View Workshop,* 1999.
[16] W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery, *Numerical Recipes in C—The Art of Scientific Computing,* second ed. Cambridge Univ. Press, 1992.
[17] G.A.F. Seber and C.J. Wild, *Non-Linear Regression.* John Wiley & Sons, 1989.
[18] J. Stuelpnagel, "On the Parametrization of the Three-Dimensional Rotation Group," *SIAM Rev.,* vol. 6, no. 4, pp. 422-430, Oct. 1964.
[19] P. Torr and A. Zisserman, "MLESAC: A New Robust Estimator with Application to Estimating Image Geometry," *Computer Vision and Image Understanding,* vol. 78, no. 1, 2000.
[20] P.H.S. Torr, A. Zisserman, and S.J. Maybank, "Robust Detection of Degenerate Configurations While Estimating the Fundamental Matrix," *Computer Vision and Image Understanding,* vol. 71, no. 3, pp. 312-333, Sept. 1998.
[21] B. Triggs, P.F. McLauchlan, R.I. Hartley, and A. Fitzgibbon, "Bundle Ajustment—A Modern Synthesis," *Proc. Int'l Workshop Vision Algorithms: Theory and Practice,* B. Triggs, A. Zisserman, and R. Szeliski, eds., 2000.
[22] R.Y. Tsai and T.S. Huang, "Uniqueness and Estimation of Three-Dimensional Motion Parameters of Rigid Objects with Curved Surfaces," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 6, no. 1, pp. 13-27, Jan. 1984.
[23] Z. Zhang, "Determining the Epipolar Geometry and Its Uncertainty: A Review," *Int'l J. Computer Vision,* vol. 27, no. 2, pp. 161-195, Mar. 1998.
[24] Z. Zhang and C. Loop, "Estimating the Fundamental Matrix by Transforming Image Points in Projective Space," *Computer Vision and Image Understanding,* vol. 82, no. 2, pp. 174-180, May 2001.

▷ **For more information on this or any computing topic, please visit our Digital Library at** http://computer.org/publications/dlib.