On the Non-linear Optimization of Projective Motion Using Minimal Parameters

Adrien Bartoli

INRIA Rhône-Alpes
655, av. de l'Europe
38334 St. Ismier cedex, France.
Adrien.Bartoli@inria.fr
www.inrialpes.fr/movi/people/Bartoli

Abstract. I address the problem of optimizing projective motion over a minimal set of parameters. Most of the existing works overparameterize the problem. While this can simplify the estimation process and may ensure well-conditioning of the parameters, this also increases the computational cost since more unknowns than necessary are involved.
I propose a method whose key feature is that the number of parameters employed is minimal. The method requires singular value decomposition and minor algebraic manipulations and is therefore straightforward to implement. It can be plugged into most of the optimization algorithms such as Levenberg-Marquardt as well as the corresponding sparse versions. The method relies on the orthonormal camera motion representation that I introduce here. This representation can be locally updated using minimal parameters.
I give a detailled description for the implementation of the two-view case within a bundle adjustment framework, which corresponds to the maximum likelihood estimation of the fundamental matrix and scene structure. Extending the algorithm to the multiple-view case is straightforward. Experimental results using simulated and real data show that algorithms based on minimal parameters perform better than the others in terms of the computational cost, i.e. their convergence is faster, while achieving comparable results in terms of convergence to a local optimum. An implementation of the method will be made available.

1 Introduction

The problem of recovering structure and motion from images is one of the central challenges for computer vision. The use of image feature correspondences (e.g. points, lines) through different views and the study of geometrical aspects of the image formation process have led to numerous techniques acting in metric, affine or projective space, depending on whether camera calibration is fully, partially or not available.

Most of the time, a sub-optimal solution is obtained using linear techniques for motion then for structure recovery [3] or jointly [16] and subsequently refined. One of the most efficient techniques for such a structure and motion optimization is bundle adjustment. It involves minimizing a non-linear cost function based on the discrepancy between reprojected and original image features [15,17]. The behaviour of such techniques, in terms of convergence properties to a local optimum of the cost function and

A. Heyden et al. (Eds.): ECCV 2002, LNCS 2351, pp. 340–354, 2002.

computational cost, greatly depends on the algebraic representation of the problem, i.e. of structure and motion, and in particular, numerical conditioning and whether the number of parameters employed is minimal. For bundle adjustment, preserving the original noise model on image features is also crucial.

While bundle adjustment is theoretically well-defined (it corresponds to the maximum likelihood estimator, see e.g. [17]) there are still practical optimization problems since the employed cost functions have many local minima [13] where optimization processes may get trapped.

I address the problem of representing motion. The goal is to obtain minimal estimators, i.e. where the number of parameters considered for optimization is minimal. I focus on the projective case, i.e. when camera calibration is unknown. This topic is of primary importance since an accurate projective reconstruction is necessary to subsequently succeed in self-calibration.

Consider two perspective views of a rigid scene. The image signature of their relative position is the projective two-view motion or the epipolar geometry, usually described by the (3×3) rank-2 and homogeneous fundamental matrix [10,18]. A fundamental matrix has 7 degrees of freedom. Therefore, 7 parameters should be enough to optimize the projective two-view motion.

However, it has been seen that there does not exist a universal parameterization of the fundamental matrix using 7 parameters. This is due to the non-linear rank-2 constraint and the free-scale ambiguity. Existing works may fall into the following categories:

- overparameterization, e.g. the 12 entries of a perspective camera matrix [7]. More unknowns than necessary are involved to simplify the representation;
- multiple minimal parameterizations, 3 in [2] or 36 in [18];
- minimal parameterizations combined to image transformations [2,19] to reduce the number of parameterizations.

Other techniques optimize over the 9 parameters of the fundamental matrix while adding the non-linear rank-2 constraint and the normalization constraint as virtual measurements.

This paper makes the following contributions.

Firstly, I address the projective two-view motion case in §2. I introduce what I call the orthonormal representation of projective two-view motion. Based on this, I show how one can non-linearly estimate the projective two-view motion using a minimal number of 7 parameters. An important point is that this method does not depend upon the optimization technique considered.

Secondly, I illustrate the use of this method in a bundle adjustment framework based on [8] in §3. The result is a minimal maximum likelihood estimator for the fundamental matrix as well as scene structure. The reader who is interested into practical issues only should refer directly to this section.

Thirdly, I extend the framework to multiple views in §4 where I introduce the orthonormal representation of projective multiple-view motion. I derive, similarly to the two-view case, a means to perform optimization over a minimal set of parameters.

Finally, experimental results on simulated and real data are shown in §§5 and 6 respectively. They show that algorithms based on minimal motion parameters, and in

particular on the orthonormal representation, perform better than the others, in terms of computational cost, while achieving equivalent performances in terms of convergence properties. This is followed by my conclusions in §7.

2 The Projective Two-View Motion

2.1 Preliminaries

Let us consider two (3×4) uncalibrated perspective camera matrices. Due to homogeneity, each one has 11 degrees of freedom. Since there is a 15-degrees of freedom coordinate frame ambiguity on structure and motion, the projective two-view motion has $11 \cdot 2 - 15 = 7$ degrees of freedom.

The corresponding (3×3) fundamental matrix F has 9 entries but 7 degrees of freedom since it is homogeneous and has rank 2. It allows one to extract projection matrices for the two views while fixing the coordinate frame. These projection matrices constitute a realization of the fundamental matrix. Among the 15-parameter family of realizations, a common choice is the canonic projection matrices P and P' [11]:

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

where $\mathbf{e}'$ is the second epipole defined as the right kernel of F, $\mathsf{F}^\mathsf{T}\mathbf{e}' \sim \mathbf{0}_{(3\times 1)}$ and $\mathsf{H}^\star \sim [\mathbf{e}']_\times \mathsf{F}$ is the canonic plane homography. This defines the canonic coordinate frame which is unique, provided normalization constraints for $\mathsf{H}^\star$ and $\mathbf{e}'$. It will be seen in §2.3 that $||\mathbf{e}'||^2 = 1$ and $||\mathsf{H}^\star||^2 = \gamma$, where γ is an unknown constant scalar, is a convenient choice for my method. Note that $\sim$ means "equal up to scale" and $[.]_\times$ is the cross-product skew-symmetric (3×3)-matrix. All entities are represented in homogeneous coordinates, i.e. are defined up to scale.

2.2 Relation to Previous Work

Most of the previous work on minimally parameterizing projective two-view motion deals 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 an homogeneous (2×2)-matrix g.

Representing these entities with minimal parameters requires eliminating their homogeneity. This is usually done by normalizing each of them so that their largest entry is unity, which yields 3 possibilities for each epipole and 4 for the epipolar transformation, so $3 \cdot 3 \cdot 4 = 36$ possible parameterizations.

Obtaining the fundamental matrix |or any other 2D entity such as the extended epipolar transformation [2] or the canonic plane homography $\mathsf{H}^\star$| from $\mathbf{e}$, $\mathbf{e}'$ and g requires then the use of 9 distinct parameterizations to model all cases [18]. These cases coincide with 9 of the 36 previous ones.

In [10], the author proposes to restrict the two-view configurations considered to the cases where both epipoles are finite and can therefore be expressed in affine coordinates. Due to the homogeneity of the epipolar transformation, 4 distinct parameterizations are

still necessary for g. A unique parameterization can then be used to form the fundamental matrix.

The method has been extended in [18] to the general case, i.e. when the epipoles can be either finite or infinite. In this case, it is shown that the full 36 distinct parameterizations are necessary. This leads to a cumbersome and error-prone implementation of the optimization process.

In [2,19], the method has been revised so as to reduce the number of parameterizations using image transformations. In [2], the image transformations used are metric and the number of distinct parameterizations restricted to 3 plus one bilinear constraint on the entries of g, while in [19], the transformations used are projective, which allows one to reduce the number of parameterizations to 1. 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 [19] for the gradient-weighted criterion, which is not the one used for bundle adjustment.

2.3 The Orthonormal Representation

Derivation. I introduce what I call the orthonormal representation of projective two-view motion. I consider the fundamental matrix representation of the motion. Any (3×3) rank-2 matrix is a fundamental matrix, i.e. represents a motion. Conversely, a projective two-view motion is represented by a unique fundamental matrix (up to scale). Therefore, deriving a minimal representation of projective two-view motion from its fundamental matrix representation implies considering two constraints; the rank-2 constraint and a normalization constraint, which fixes the relative scale of the fundamental matrix. Previous work has shown that these constraints are quite tricky to enforce directly on the fundamental matrix [2,10,18,19].

To overcome this problem, instead of considering directly the fundamental matrix, I rather analyze its singular value decomposition $\mathsf{F} \sim \mathsf{U}\Sigma\mathsf{V}^\mathsf{T}$, where U and V are (3×3) orthonormal matrices and Σ a diagonal one containing the singular values of F. The orthonormal representation is then derived while enforcing the constraints:

- *rank-2:* since F is a rank-2 matrix, $\Sigma \sim \mathrm{diag}(\sigma_1, \sigma_2, 0)$ where σ_1 and σ_2 are strictly positive scalars and $\sigma_1 \geq \sigma_2 > 0$;
- *normalization:* since F is an homogeneous entity, I can scale its singular value decomposition such that $\mathsf{F} \sim \mathsf{U}\cdot\mathrm{diag}(1, \sigma, 0)\cdot\mathsf{V}^\mathsf{T}$ where $\sigma = \sigma_2/\sigma_1$ and $0 < \sigma \leq 1$ ($\sigma_1 \neq 0$ since F is rank-2).

This decomposition shows that any projective two-view motion can be represented by two (3×3) orthonormal matrices and a scalar.

This gives the orthonormal representation of projective two-view motion as:

$$(\mathsf{U}, \mathsf{V}, \sigma) \in \mathbb{F} \text{ where } \mathbb{F} \equiv O(3)^2 \times \{\sigma \mid 0 < \sigma \leq 1\}, \tag{2}$$

where $O(3)$ is the Lie group of (3×3) orthonormal matrices. This representation is minimal in that it has $3 + 3 + 1 = 7$ degrees of freedom. It can easily be computed from the singular value decomposition of the fundamental matrix. Note that $\sigma = 1$ may correspond to the case of an essential matrix, i.e. when cameras are calibrated.

Any element of $\mathbb{F}$ represents a unique two-view motion since it can be used to recompose a unique fundamental matrix (see next paragraph), i.e. a (3×3) matrix where both the rank-2 and a normalization constraints have been enforced. However, a fundamental matrix has more than one orthonormal representations. For instance, given an orthonormal representation $(\mathsf{U}, \mathsf{V}, \sigma) \in \mathbb{F}$, one can freely switch the signs of $\mathbf{u}_3$ or $\mathbf{v}_3$ while leaving the represented motion invariant. However, the space of fundamental matrices and the orthonormal representation are both 7-dimensional, which allows for minimal estimation.

Recovering 2D *entities.* The fundamental matrix corresponding to an orthonormal representation $(\mathsf{U}, \mathsf{V}, \sigma) \in \mathbb{F}$ can be recovered by simply recomposing the singular value decomposition:

$$\mathsf{F} \sim \mathbf{u}_1 \mathbf{v}_1^\mathsf{T} + \sigma \mathbf{u}_2 \mathbf{v}_2^\mathsf{T}, \tag{3}$$

where $\mathbf{u}_i$ and $\mathbf{v}_i$ are the columns of U and V respectively. Among all potential applications of the orthonormal representation, I will use it for bundle adjustment. Therefore, I will need to extract projection matrices from the fundamental matrix. This can be achieved directly from the orthonormal representation by recovering the second epipole and the canonic plane homography of [11], equation (1).

The second epipole is the last column of U: $\mathbf{e}' \sim \mathbf{u}_3$, while the canonic plane homography can be formulated as $\mathsf{H}^\star \sim [\mathbf{e}']_\times \mathsf{F} \sim [\mathbf{u}_3]_\times \left(\mathbf{u}_1 \mathbf{v}_1^\mathsf{T} + \sigma \mathbf{u}_2 \mathbf{v}_2^\mathsf{T} \right)$. 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 the canonic plane homography as $\mathsf{H}^\star \sim \mathbf{u}_2 \mathbf{v}_1^\mathsf{T} - \sigma \mathbf{u}_1 \mathbf{v}_2^\mathsf{T}$ and the second projection matrix as:

$$\mathsf{P}' \sim \left(\mathbf{u}_2 \mathbf{v}_1^\mathsf{T} - \sigma \mathbf{u}_1 \mathbf{v}_2^\mathsf{T} \mid \mathbf{u}_3 \right). \tag{4}$$

Normalization constraints discussed in §2.1 are clearly satisfied. This guarantees that the same canonic basis will be used through the optimization process.

2.4 Estimation Using Minimal Parameters

In this section, I use the previously described orthonormal representation of projective two-view motion to locally update a "base" estimate using the minimum 7 parameters.

Update using minimal parameters. Before going further, let us examine the case of 3D rotations. There does not seem to exist a minimal, complete and non-singular parameterization of the 3-dimensional set of rotations in 3D space. For example, consider the (3×3) rotation matrix $\mathsf{R} \in SO(3)$. Minimal representations, such as the 3 Euler angles $\boldsymbol{\theta}$ lead to singularities. However, one can find representations that are locally minimal and non-singular, e.g. in a neighbourhood of $\mathsf{R} = \mathrm{I}_{(3\times3)}$, i.e. $\boldsymbol{\theta} = \mathbf{0}_{(3\times1)}$. Therefore, most of the estimation processes of 3D rotations do not minimally parameterize the current estimate, but rather locally update its overparameterization. A typical example is to use a (3×3) rotation matrix representation that is locally updated as $\mathsf{R} \leftarrow \mathsf{R} \cdot \mathsf{R}(\boldsymbol{\theta})$ where $\mathsf{R}(\boldsymbol{\theta})$ is any minimal and locally non-singular parameterization of 3D rotations, e.g. Euler angles. This method is used in [6] for the non-linear recovery of metric structure and motion, where R is updated and $\boldsymbol{\theta}$ reset to zero after each iteration of the optimizer.

The same scenario arises for the projective two-view motion. There does not seem to exist a minimal, complete and non-singular parameterization of the corresponding 7-dimensional space. Consequently, I propose locally updating a given estimate using minimal parameters.

Let us consider the orthonormal representation $(\mathsf{U}, \mathsf{V}, \sigma) \in \mathbb{F}$ of a given projective two-view motion. Each matrix $(\mathsf{U}, \mathsf{V}) \in O(3)^2$ can be locally updated using 3 parameters by considering the method described above for 3D rotations. The scalar $\sigma \in \{\sigma \mid 0 < \sigma \leq 1\}$ is completely included into the parameterization. A means to ensure $0 < \sigma \leq 1$ is described below.

Completness. A first remark that immediately follows about the above-proposed method is whether all two-view configurations are covered. This 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, $\mathsf{R}(\mathbf{x})$ and $\mathsf{R}(\mathbf{y})$ respectively, which have only positive determinants. Therefore, the signs of $\det(\mathsf{U})$ and $\det(\mathsf{V})$ are not allowed to change during the optimization process. I claim that this is not a problem and that any fundamental matrix F can be reached from any initial guess F_0, even if $\text{sign}(\det(\mathsf{U}_0)) \neq \text{sign}(\det(\mathsf{U}))$, $\text{sign}(\det(\mathsf{V}_0)) \neq \text{sign}(\det(\mathsf{V}))$ or both. To prove this claim, I show that any fundamental matrix F represented by $(\mathsf{U}, \mathsf{V}, \sigma)$ has alternative representations where the signs of either $\det(\mathsf{U})$ or $\det(\mathsf{V})$, or both, have been switched. This is due to the non-uniqueness of the singular value decomposition, see e.g. [14]. Consider the recomposition equation (3) and observe that $\mathbf{u}_3$ and $\mathbf{v}_3$, the third columns of U and V respectively, do not affect the result. Therefore, they are only constrained by the orthonormality of U and V. Hence, their signs can be arbitrarily switched, which accordingly switches the sign of the determinant of the corresponding matrix. For example, $\mathbf{u}_3 \leftarrow -\mathbf{u}_3$ switches the sign of $\det(\mathsf{U})$ while leaving invariant the represented fundamental matrix. This concludes the proof.

Implementation details. Through the iterations, σ may go outside of its boundaries. This is not a problem since the corresponding motion is still valid.

There are several possibilities to ensure the boundaries of σ such as using linear constraints. I propose to enforce these boundaries at each iteration while leaving the current estimate invariant. However, I have found during my experiments of §§5 and 6 that in practice, this does not affect the behaviour of the underlying optimization process. A way to proceed is:

- if $\sigma < 0$ then $\sigma \leftarrow -\sigma$, $\mathbf{u}_2 \leftarrow -\mathbf{u}_2$, $\mathbf{u}_3 \leftarrow -\mathbf{u}_3$;
- if $\sigma > 1$ then $\sigma \leftarrow \frac{1}{\sigma}$, swap($\mathbf{u}_1$,$\mathbf{u}_2$), $\mathbf{u}_3 \leftarrow -\mathbf{u}_3$, swap($\mathbf{v}_1$,$\mathbf{v}_2$), $\mathbf{v}_3 \leftarrow -\mathbf{v}_3$.

One can easily check that these changes on the orthonormal representation leave the represented motion invariant.

3 Bundle Adjustment

In this section, I show how the orthonormal representation can be used for bundle adjustment of point features seen in two views. This is summarized in table 1. Similar results could be derived for other criteria, such as the minimization of the distances between points and epipolar lines or the gradient-weighted criterion [10,18].

Table 1. Implementing my minimal estimator within the bundle adjustment Levenberg-Marquardt-based framework given in [8], p.574 (algorithm A4.1). Note that r is the number of residuals and that the second projection matrix have to be extracted from the orthonormal representation using equation (4) (for e.g. computing the error vector).

Two-view projective bundle adjustment expressed within the framework of [8], p.574 (algorithm A4.1). The initial guess of the fundamental matrix is denoted by F_0.

Add the following steps:

(i') Initialize the orthonormal representation $(\mathsf{U}, \mathsf{V}, \sigma)$ by a scaled singular value decomposition of F_0:

$$\mathsf{F}_0 \sim \mathsf{U} \cdot \mathrm{diag}(1, \sigma, 0) \cdot \mathsf{V}^\mathsf{T}.$$

(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:

$$\mathsf{A} \leftarrow \mathsf{A} \cdot \mathsf{A}^{\mathrm{ortho}},$$

where $\mathsf{A}^{\mathrm{ortho}}$ is given by equations (5,6);

Change the parameter update step as:

(viii) Update the orthonormal representation as:

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

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

Add the last step:

(xi) Return the computed F using equation (3) as:

$$\mathsf{F} \sim \mathbf{u}_1 \mathbf{v}_1^\mathsf{T} + \sigma \mathbf{u}_2 \mathbf{v}_2^\mathsf{T}.$$

Cost function. Bundle adjustment consists in solving the following optimization problem, see e.g. [12,17,18]:

$$\min_{\mathbf{a},\mathbf{b}} \mathbf{r}^\mathsf{T} \mathbf{r},$$

where:

- $\mathbf{a}$ and $\mathbf{b}$ are respectively motion and structure parameters (or parameters used to update them);
- $\mathbf{r}$ is the vector of residual errors;
- $r_i^2 = d^2(\mathbf{q}_i, \mathsf{P}\mathbf{Q}_i) + d^2(\mathbf{q}'_i, \mathsf{P}'\mathbf{Q}_i)$ is the i-th point residual error (d is the 2D Euclidean distance) corresponding to its reprojection error;
- $\mathbf{q}_i$ and $\mathbf{q}'_i$ are corresponding image points for the first and second images;
- $\mathbf{Q}_i$ are 3D reconstructed points and depend upon $\mathbf{b}$;

– P and P' are projection matrices corresponding to the current motion estimate represented by $\mathbf{a}$. They must correspond to a realization of the fundamental matrix. I have shown in §2.3, equation (4), how the canonic realization can be directly obtained from the orthonormal representation.

Analytical differentiation. Newton-type optimization methods, such as the widely used Levenberg-Marquardt, necessitate computing the Jacobian matrix $\mathsf{J} = (\mathsf{A} \mid \mathsf{B})$ of the residual vector $\mathbf{r}$ with respect to motion and structure parameters $\mathbf{a}$ and $\mathbf{b}$. While this can be achieved numerically using e.g. finite differences [14], it may be better to use an analytical form for both computational efficiency and numerical accuracy. I focus on the computation of $\mathsf{A} = \frac{\partial \mathbf{r}}{\partial \mathbf{a}}$ since $\mathsf{B} = \frac{\partial \mathbf{r}}{\partial \mathbf{b}}$ only depends upon structure parameterization. I decompose it as $\mathsf{A}_{(r\times 7)} = \bar{\mathsf{A}}_{(r\times 12)} \cdot \mathsf{A}^{\text{ortho}}_{(12\times 7)}$ where:

– r is the number of residuals;
– only the 12 entries of P' are considered since P is fixed in the canonic reconstruction basis (1);
– $\bar{\mathsf{A}} = \frac{\partial \mathbf{r}}{\partial \mathbf{p}'}$ ($\mathbf{p}' = \text{vect}(\mathsf{P}')$ where vect(.) is the row-wise vectorization) depends on the chosen realization of the fundamental matrices, i.e. on the coordinate frame employed. I have chosen the canonic projection matrices (1). This Jacobian matrix is employed directly for the overparameterization of [8]. Deriving its analytical form is straightforward;
– $\mathsf{A}^{\text{ortho}} = \frac{\partial \mathbf{p}'}{\partial \mathbf{a}}$ is related to the orthonormal motion representation.

I therefore concentrate on deriving a closed-form expression for $\mathsf{A}^{\text{ortho}}$. If the minimal method of e.g. [18] were used, 36 different Jacobian matrices corresponding to each parameterization would have to be derived.

One of the advantages of my update scheme shown in table 1 and based on the orthonormal representation is that there exists a simple closed-form expression for $\mathsf{A}^{\text{ortho}}$.

Let us consider the orthonormal representation $(\mathsf{U}, \mathsf{V}, \sigma)$. In this case, the motion update parameters are minimal and defined by $\mathbf{a} = (x_1\ x_2\ x_3\ y_1\ y_2\ y_3\ \sigma)^\mathsf{T}$, where $\mathbf{x} = (x_1\ x_2\ x_3)^\mathsf{T}$ and $\mathbf{y} = (y_1\ y_2\ y_3)^\mathsf{T}$ are used to update U and V respectively. Since U and V are updated with respect to the current estimate, $\mathsf{A}^{\text{ortho}}$ is evaluated at $(\mathsf{U}, \mathsf{V}, \sigma)$, i.e. at $\mathbf{a} = \mathbf{a}_0 = (\mathbf{0}_{(6\times 1)}\ \sigma)^\mathsf{T}$. Let $\widetilde{\mathsf{U}} = \mathsf{U} \cdot \mathsf{R}(\mathbf{x})$ and $\widetilde{\mathsf{V}} = \mathsf{V} \cdot \mathsf{R}(\mathbf{y})$ be the updated U and V. Equation (4) is used to derive a closed-form expression of the second canonic projection matrix after updating, i.e. corresponding to the orthonormal representation $(\widetilde{\mathsf{U}}, \widetilde{\mathsf{V}}, \sigma)$. By expanding, differentiating and evaluating this expression at $\mathbf{a}_0$, I obtain:

$$\mathsf{A}^{\text{ortho}} = \frac{\partial \mathbf{p}'}{\partial \mathbf{a}} = \frac{\partial \mathbf{p}'}{\partial (x_1\ \ldots\ y_3\ \sigma)} = \left(\left(\frac{\partial \mathbf{p}'}{\partial x_1} \right) \cdots \left(\frac{\partial \mathbf{p}'}{\partial y_3} \right) \left(\frac{\partial \mathbf{p}'}{\partial \sigma} \right) \right), \tag{5}$$

where:

$$\partial \mathbf{p}' = \begin{cases} \text{vect}(\mathbf{u}_3 \mathbf{v}_1{}^\mathsf{T} \mid -\mathbf{u}_2) \cdot \partial x_1 \\ \text{vect}(-\sigma \mathbf{u}_1 \mathbf{v}_3{}^\mathsf{T} \mid \mathbf{0}) \cdot \partial x_2 \\ \text{vect}(\sigma \mathbf{u}_3 \mathbf{v}_2{}^\mathsf{T} \mid \mathbf{u}_1) \cdot \partial x_3 \\ \text{vect}(-\mathbf{u}_2 \mathbf{v}_3{}^\mathsf{T} \mid \mathbf{0}) \cdot \partial y_1 \\ \text{vect}(-\mathbf{u}_1 \mathbf{v}_1{}^\mathsf{T} - \sigma \mathbf{u}_2 \mathbf{v}_2{}^\mathsf{T} \mid \mathbf{0}) \cdot \partial y_2 \\ \text{vect}(\mathbf{u}_2 \mathbf{v}_2{}^\mathsf{T} + \sigma \mathbf{u}_1 \mathbf{v}_1{}^\mathsf{T} \mid \mathbf{0}) \cdot \partial y_3 \\ \text{vect}(-\mathbf{u}_1 \mathbf{v}_2{}^\mathsf{T} \mid \mathbf{0}) \cdot \partial \sigma. \end{cases} \tag{6}$$

4 The Multiple-View Case

In this section, I extend my projective two-view motion modelisation to multiple views. I analyse how to model additional views. I propose the orthonormal representation of projective multiple-view motion. As in the two-view case, this serves to devise elements for optimizing projective multiple-view motion over minimal parameters.

4.1 Modeling Additional Views

Once two views have been modeled, the coordinate frame is fixed. Therefore, an additional view does not have any free gauge and its complete projection matrix has to be modeled. Let P be such a (3×4) projection matrix. Since it is homogeneous, it has 11 degrees of freedom. This can be seen in several other ways. For example, one can consider the metric decomposition $\mathsf{P} \sim \mathsf{K}(\mathsf{R}\ \mathbf{t})$. However, this decomposition is not available here since I deal with uncalibrated cameras. One may also interpret $\mathsf{P} \sim (\mathsf{H}\ \alpha\mathbf{e})$, where H is a (3×3)-matrix, as a plane homography H with respect to the reference view, i.e. thus for which the projection matrix is $(\mathsf{I}_{(3\times3)}\ \mathbf{0}_{(3\times1)})$, $\mathbf{e}$ a 3-vector that represents an epipole with the same view and α a scalar that accounts for the relative scale between H and $\mathbf{e}$. Therefore, P has $8 + 2 + 1 = 11$ degrees of freedom. This interpretation is related to that of plane+parallax, see e.g. [9].

4.2 Relation to Previous Work

A common strategy for optimizing an homogeneous entity such as P is to overparameterize it by using all its entries. A normalization constraint is then softly imposed by using an hallucinated measurement, e.g. on the norm of P as $||\mathsf{P}||^2 - 1 = 0$. The drawback of this method is that more parameters than necessary are estimated, which increases the computational cost of the estimation process and may cause numerical instabilities. One could also renormalize P after each iteration as $\mathsf{P} \leftarrow \mathsf{P}/||\mathsf{P}||^2$. Alternatively, one could fix one entry of P to a given value, e.g. 1, but this representation would have singularities.

The main drawback of these techniques is that a unique minimal parameterization does not suffice to express all cases. This leads to the necessity for multiple expressions of e.g. the Jacobian matrix for Newton-type optimizers, which might complexify implementation issues.

4.3 The Orthonormal Representation

The orthonormal representation of $\mathsf{P} \sim (\mathsf{H}\ \alpha\mathbf{e})$ can be derived as follows. Let $\mathbf{s} = \alpha\mathbf{e}$. This inhomogeneous 3-vector is a scaled version of $\mathbf{e}$ which has 3 degrees of freedom since it also encapsulates the relative scale α between H and $\mathbf{e}$. Therefore, $\mathbf{s} \in \mathbb{R}^3$ and it can be directly parameterized by its 3 elements.

Consider now the homogeneous (3×3)-matrix H. As in the case of the fundamental matrix, see §2.3, I examine its singular value decomposition $\mathsf{H} \sim \mathsf{U}\Sigma\mathsf{V}^\mathsf{T}$ where U and V are (3×3) orthonormal matrices and Σ a diagonal one containing the singular values of H. Since H may be singular, see e.g. in §2.1 the canonic plane homography of [11], but

must not be rank-1 or null, $\Sigma \sim \mathrm{diag}(1, \sigma_1, \sigma_2)$, where $0 < \sigma_1 \leq 1$ and $0 \leq \sigma_2 \leq \sigma_1$. Therefore, the orthonormal representation of H writes as:

$$(\mathsf{U}, \mathsf{V}, \sigma_1, \sigma_2) \in \mathbb{H} \text{ where } \mathbb{H} \equiv \mathbb{F} \times \{\sigma_2 \mid 0 \leq \sigma_2 \leq \sigma_1\},$$

and $\mathsf{F} \equiv O(3)^2 \times \{\sigma_1 \mid 0 < \sigma_1 \leq 1\}$, see equation (2). As a byproduct, one can observe that I have derived the orthonormal representation $\mathbb{H}$ of 2D homographies which can be used as a starting point to devise minimal 8-parameter estimators for these transformations. Finally, I obtain the orthonormal representation of P, denoted by $\mathbb{P}$ as:

$$(\mathsf{U}, \mathsf{V}, \sigma_1, \sigma_2, \mathbf{s}) \in \mathbb{P} \text{ where } \mathbb{P} \equiv \mathbb{H} \times \mathbb{R}^3.$$

It is minimal in that it has $3 + 3 + 1 + 1 + 3 = 11$ degrees of freedom.

4.4 Optimization With Minimal Parameters

Mimicking the method of §2.4 for the projective two-view motion case, I obtain a way to minimally estimate projective multiple-view motion. Given a set of camera matrices, I represent two of them using the orthonormal projective two-view motion of §2.3. This fixes the coordinate frame. Each other view is then modeled by the orthonormal representation $(\mathsf{U}, \mathsf{V}, \sigma_1, \sigma_2, \mathbf{s}) \in \mathbb{P}$ described above. Each $O(3)$ matrix U, V can be updated using minimal parameters as e.g. $\mathsf{U} \leftarrow \mathsf{U} \cdot \mathsf{R}(\mathbf{x})$ whereas σ_1, σ_2 and $\mathbf{s}$ are directly optimized. As in the two-view case, one can derive algorithms to ensure the boundaries on σ_1 and σ_2. A closed-form solution for the Jacobian matrix of the residuals with respect to the motion parameters can be derived in a manner similar to the two-view case. Another advantage of this representation is that one can directly compute the inverse of H, the (3×3) leading part of P, from its orthonormal representation. This may be useful for e.g. projecting 3D lines or estimating 2D homographies using a symmetric cost function.

5 Experimental Results Using Simulated Data

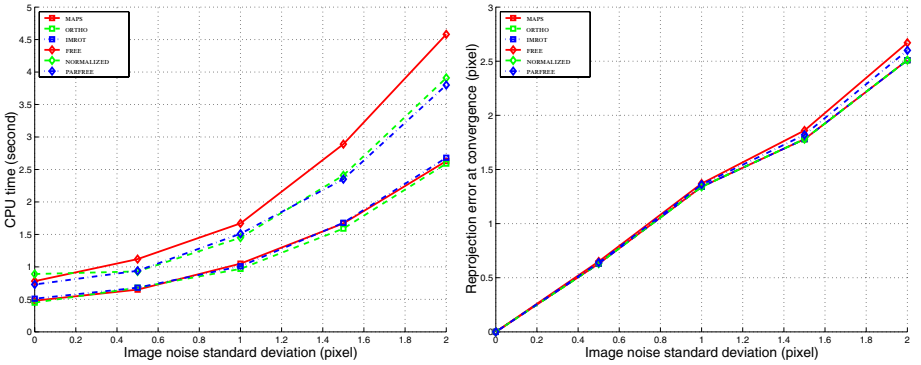
In this section, I compare the algorithm using the orthonormal motion representation (see table 1 and §§2.4 and 4.4) to existing ones using simulated data. I use the Levenberg-Marquardt method to perform optimization. Points are minimally parameterized as in [8], p.579. The test bench consists of 50 points lying inside a sphere with a radius of 1 meter observed by cameras with a focal length of 1000 (expressed in number of pixels). Each of these cameras is looking at the center of the sphere and is situated at a distance of 10 meters from it. The baseline between two consecutive cameras is 1 meter.

Points are generated in 3D space, projected onto the images and corrupted by an additive centered gaussian noise with varying standard deviation.

I measure the two quantities characteristic of a bundle adjustment process, computational cost, i.e. CPU time to convergence and the error at convergence, versus the standard deviation of added image noise. I also measure the error of the current estimate as a function of time through the optimization processes. The plots correspond to median values over 300 trials. The bundle adjustments are initialized by the values obtained using the

8 point algorithm [4] and the triangulation method described in [5] for two views. Each other view is then registered in turn by linearly computing its camera matrix. I compare the following algorithms:

- ORTHO: uses the minimal methods given in table 1 and §§2.4 and 4.4;
- MAPS and IMROT (Image Rotation): are other minimal methods given in [1], the associated research report. These methods are equivalent to those described in [18, 19] in the sense that the number of unknowns is minimal;
- FREE: uses an overparameterization with free gauges, namely all the entries of the camera matrices are optimized;
- NORMALIZED: uses an overparameterization plus hallucinated measurements to prevent the gauge to drift [12];
- PARFREE (Partially Free): uses a partially free gauge by completely parameterizing all camera matrices except the first one [7];



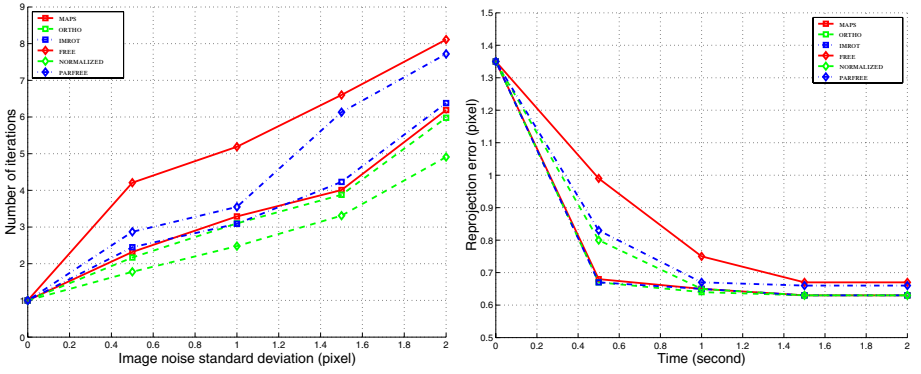
(a) (b)

Fig. 1. Comparison of (a): the CPU time to convergence and (b): value of the error at convergence versus varying image noise for different methods.

I conduct a serie of experiments using two cameras. One can observe on figure 1 (b) that, roughly speaking, all methods converge to the same minimum of the cost function. Methods that have a slightly less reliable convergence than the others are FREE and PARFREE.

Figure 1 (a) shows that, for roughly the same convergence properties, there are quite big discrepancies between the computational cost of each method. The method that has the highest computational cost is FREE, followed by NORMALIZED and PARFREE. This can be explained by the fact that these methods have more unknowns to estimate than the minimal ones. This requires more computational time for each iteration to be performed. Finally, methods using the minimal number of parameters, MAPS, ORTHO and IMROT have the lowest computational cost, roughly the same.

In the light of these results, it is clear that methods using minimal parameters should be preferred for both computational cost and convergence properties. The method ORTHO, relying on the orthonormal representation given in this paper has the advantage of simplicity. However, in order to understand and explain the behaviour of the different methods, I have measured the number of iterations and the computational cost of these iterations. These results are shown on figure 2.



(a) (b)

Fig. 2. Comparison of (a): the number of iterations to convergence versus varying image noise and (b): the evolution of reprojection errors.

Table 2. CPU time per iteration (second) for each method.

MAPS	ORTHO	IMROT	FREE	NORMALIZED	PARFREE
0.3841	0.3838	0.3864	0.4741	0.7043	0.4717

In more detail, I have found that methods FREE or PARFREE, leaving the gauge drift freely have very bad convergence properties, performing more iterations, roughly twice more, than the others, see figure 2 (a). Method NORMALIZED performs a number of iterations smaller than all the other methods but involves solving a much more costly linear system at each iteration, see table 2. Methods using the minimal number of parameters are trade-offs between the number of iterations and their computational cost: each iteration has a low computational cost and the number of iterations needed is in-between those of free gauge methods and NORMALIZED. This explains why these methods achieve the lowest total computational cost.

Figure 2 (b) shows the evolution of reprojection error for the different optimization processes. This experiment is useful in the sense that the time to convergence previously measured is highly dependent on how convergence is determined, e.g. by thresholding

two consecutive errors, and does not account for the ability of the algorithms to quickly, or not, reach almost the value of convergence. This experiment has been conducted using the same test bench as previously with a noise level on image point positions of 0.5 pixel. One can see on this figure that methods based on a minimal parameterization reach their value of convergence before the others. The NORMALIZED and PARFREE methods take roughly twice the same time, while the FREE method takes three times more.

Finally, I conduct experiments using 10 views. I observe that the differences between the algorithms observed in the two-view case are decreased while those requiring the lowest computation time are the same, i.e. MAPS, ORTHO and IMROT. Other experiments on the convergence properties of the algorithms in the multiple-view case yield conclusions similar to the two-view case.

6 Experiments on Real Images

In this section, I validate my algorithms using real images. I first consider the case of two images. In order to cover all possibilities for the epipoles to be close to the images or at infinity, I use pairs of the images shown on figure 3. Initial values for structure and motion are computed as in the case of simulated data.

Fig. 3. Real images used to validate the algorithms.

Results are shown in table 3. For each combination of image pair and each algorithm, I estimate the CPU time to convergence $\mathcal{T}$ and the error at convergence $\mathcal{E}$. The last row of the table show the mean values $\bar{\mathcal{T}}$ and $\bar{\mathcal{E}}$ of $\mathcal{T}$ and $\mathcal{E}$ for each algorithm over the set of image pairs. These results confirmed those obtained using simulated data.

I have also tested the algorithms on all four images of figure 3. Initial values have been obtained by registering each view to an initial guess of structure and motion obtained from the two first ones. The results are the followings: all algorithms converge with a final error of 0.73 pixels and their relative performances in terms of computation times to convergence were equivalent to those obtained in the case of two views.

7 Conclusions

I studied the problem of optimizing projective motion over minimal sets of parameters. I proposed the orthonormal representation of projective two-view motion. I showed

Table 3. Error at convergence $\mathcal{E}$ and time to convergence $\mathcal{T}$ obtained when combining pairs of images from figure 3.

epipoles		views	MAPS		ORTHO		IMROT		FREE		NORMALIZED		PARFREE	
$\mathbf{e}$	$\mathbf{e}'$		$\mathcal{E}$	$\mathcal{T}$	$\mathcal{E}$	$\mathcal{T}$	$\mathcal{E}$	$\mathcal{T}$	$\mathcal{E}$	$\mathcal{T}$	$\mathcal{E}$	$\mathcal{T}$	$\mathcal{E}$	$\mathcal{T}$
$\not\infty$	∞	A, B	0.63	2.45	0.63	2.39	0.63	2.47	0.68	3.98	0.63	2.99	0.68	3.02
		A, C	0.71	2.38	0.71	2.41	0.71	2.40	0.77	4.01	0.71	3.56	0.71	3.71
$\not\infty$	$\not\infty$	A, D	0.45	2.03	0.45	1.76	0.45	2.19	0.57	3.13	0.45	3.09	0.45	2.93
∞	∞	B, C	0.88	3.53	0.88	3.39	0.88	3.55	1.23	6.70	0.88	5.12	0.88	4.63
∞	$\not\infty$	B, D	0.59	2.33	0.59	2.10	0.59	2.81	0.59	3.99	0.59	3.41	0.59	3.56
		C, B	0.51	1.91	0.51	1.92	0.51	2.02	0.51	3.39	0.51	2.79	0.51	3.04
average $\bar{\mathcal{E}}$ and $\bar{\mathcal{T}}$			0.628	2.430	0.628	2.328	0.628	2.573	0.725	4.200	0.628	3.493	0.637	3.482

how this can be used to locally update projective two-view motion using a minimal set of 7 parameters. The canonic projection matrices can be directly extracted from the orthonormal representation. I extent this representation to projective multiple-view motion. As a byproduct, I derive the orthonormal representation of 2D homographies. The method can be plugged into most of the (possibly sparse) non-linear optimizers such as Levenberg-Marquardt. I gave a closed-form expression for the Jacobian matrix of the residuals with respect to the motion parameters, necessary for Newton-type optimization techniques.

The introduced orthonormal representation seems to be a powerful tool for minimal optimization of homogeneous entities in particular.

I conducted experiments on simulated and real data. My conclusions are that methods based on minimal parameter sets perform better than the others, in terms of computational cost while achieving equivalent results in terms of convergence properties. The most interesting results are obtained in the two-view case. Existing algorithms that do not constrain the gauge by any means perform worse than the others.

I will make an implementation of the method available on my home-page.

In future work, I plan to investigate the use of the orthonormal representation introduced in this paper to model other algebraic entities and devise minimal estimation techniques for them.

References

1. A. Bartoli and P. Sturm. Three new algorithms for projective bundle adjustment with minimum parameters. Research Report 4236, INRIA, Grenoble, France, August 2001.
2. A. Bartoli, P. Sturm, and R. Horaud. Projective structure and motion from two views of a piecewise planar scene. In *Proceedings of the 8th International Conference on Computer Vision, Vancouver, Canada*, volume 1, pages 593–598, July 2001.
3. P. Beardsley, P. Torr, and A. Zisserman. 3D model acquisition from extended image sequences. In B. Buxton and R. Cipolla, editors, *Proceedings of the 4th European Conference on Computer Vision, Cambridge, England*, volume 1065 of *Lecture Notes in Computer Science*, pages 683–695. Springer-Verlag, April 1996.

4. R. Hartley. In defence of the 8-point algorithm. In *Proceedings of the 5th International Conference on Computer Vision, Cambridge, Massachusetts, USA*, pages 1064–1070, June 1995.
5. R. Hartley and P. Sturm. Triangulation. *Computer Vision and Image Understanding*, 68(2):146–157, 1997.
6. R.I. Hartley. Euclidean reconstruction from uncalibrated views. In *Proceeding of the* DARPA–ESPRIT *workshop on Applications of Invariants in Computer Vision, Azores, Portugal*, pages 187–202, October 1993.
7. R.I. Hartley. Projective reconstruction and invariants from multiple images. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 16(10):1036–1041, October 1994.
8. R.I. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, June 2000.
9. M. Irani and P. Anadan. Parallax geometry of pairs of points for 3d scene analysis. In *Proceedings of the 4th European Conference on Computer Vision, Cambridge, England*, pages 17–30. Springer-Verlag, 1996.
10. Q.T. Luong and O. Faugeras. The fundamental matrix: Theory, algorithms and stability analysis. *International Journal of Computer Vision*, 17(1):43–76, 1996.
11. Q.T. Luong and T. Vieville. Canonic representations for the geometries of multiple projective views. *Computer Vision and Image Understanding*, 64(2):193–229, 1996.
12. P. F. McLauchlan. Gauge invariance in projective 3D reconstruction. In *Proceedings of the Multi-View Workshop, Fort Collins, Colorado, USA*, 1999.
13. J. Oliensis. The error surface for structure and motion. Technical report, NEC, 2001.
14. W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery. *Numerical Recipes in C - The Art of Scientific Computing*. Cambridge University Press, 2nd edition, 1992.
15. C.C. Slama, editor. *Manual of Photogrammetry, Fourth Edition*. American Society of Photogrammetry and Remote Sensing, Falls Church, Virginia, USA, 1980.
16. P. Sturm and B. Triggs. A factorization based algorithm for multi-image projective structure and motion. In B. Buxton and R. Cipolla, editors, *Proceedings of the 4th European Conference on Computer Vision, Cambridge, England*, volume 1065 of *Lecture Notes in Computer Science*, pages 709–720. Springer-Verlag, April 1996.
17. B. Triggs, P.F. McLauchlan, R.I. Hartley, and A. Fitzgibbon. Bundle ajustment — a modern synthesis. In B. Triggs, A. Zisserman, and R. Szeliski, editors, *Vision Algorithms: Theory and Practice*, volume 1883 of *Lecture Notes in Computer Science*, pages 298–372. Springer-Verlag, 2000.
18. Z. Zhang. Determining the epipolar geometry and its uncertainty: A review. *International Journal of Computer Vision*, 27(2):161–195, March 1998.
19. Z. Zhang and C. Loop. Estimating the fundamental matrix by transforming image points in projective space. *Computer Vision and Image Understanding*, 82(2):174–180, May 2001.