PETER STURM AND BILL TRIGGS : A FACTORIZATION BASED ALGORITHM FOR MULTI-IMAGE PROJECTIVE STRUCTURE AND MOTION.
*4*th *European Conference on Computer Vision, Cambridge, England, April 1996*, pp. 709-720 .
Version of 18/04/1996.

A Factorization Based Algorithm for Multi-Image Projective Structure and Motion

Peter Sturm and Bill Triggs

GRAVIR-IMAG & INRIA Rhône-Alpes*
655, Avenue de l'Europe, 38330 Montbonnot, France
Peter.Sturm@inrialpes.fr Bill.Triggs@inrialpes.fr

Abstract. We propose a method for the recovery of projective shape and motion from multiple images of a scene by the factorization of a matrix containing the images of all points in all views. This factorization is only possible when the image points are correctly scaled. The major technical contribution of this paper is a practical method for the recovery of these scalings, using only fundamental matrices and epipoles estimated from the image data. The resulting projective reconstruction algorithm runs quickly and provides accurate reconstructions. Results are presented for simulated and real images.

1 Introduction

In the last few years, the geometric and algebraic relations between uncalibrated views have found lively interest in the computer vision community. A first key result states that, from two uncalibrated views, one can recover the 3D structure of a scene up to an unknown projective transformation [Fau92, HGC92]. The information one needs to do so is entirely contained in the fundamental matrix, which represents the epipolar geometry of the 2 views.

Up to now, projective reconstruction has been investigated mainly for the case of 2 views. Faugeras [Fau92] studied projective reconstruction using 5 reference points. Hartley [HGC92] derives from the fundamental matrix 2 projection matrices, equal to the true ones up to an unknown projective transformation. These are then used to perform reconstruction by triangulation[HS94]. As for multiple images, most of the current methods [MVQ93, Har93, MM95] initially privilege a few views or points and thus do not treat all data uniformly.

Recently, multi-linear matching constraints have been discovered that extend the epipolar geometry of 2 views to 3 and 4 views. Shashua [Sha95] described the trilinear relationships between 3 views. Faugeras and Mourrain [FM95], and independently Triggs [Tri95b] have systematically studied the relationships between N images. Triggs introduced a new way of thinking about projective reconstruction. The image coordinates of the projections of a 3D point are combined into a single "joint image vector". Then, projective reconstruction consists essentially of rescaling the image coordinates in order to place the joint image vector in a certain 4-dimensional subspace of the joint image space called the *joint image*. This subspace is characterized by the multi-linear matching constraints between the views.

The projective reconstruction method we propose in this paper is based on the joint image formalism, but it is not necessary to understand this formalism to read the paper. We show that by rescaling the image coordinates we can obtain a *measurement matrix* (the combined image coordinates of all the points in all the images), which is of rank 4. Projective structure and motion can then be obtained by a singular value factorization of this matrix. So, in a sense this work can be considered as an extension of Tomasi-Kanade's and Poelman-Kanade's factorization methods [TK92, PK94] from affine to perspective projections.

The paper is organized as follows. (1) We motivate the idea of reconstruction through the rescaling of image coordinates. Throughout this paper we will restrict attention to the case of bilinear matching constraints (fundamental

* This work has been done in the context of the MOVI project which belongs to CNRS, INPG, INRIA and UJF.

matrix), although the full theory [Tri95a] also allows tri- and quadrilinear matching constraints to be used. (2) We discuss some numerical considerations and describe the proposed projective reconstruction algorithm. (3) We show results that we have obtained with real and simulated data. (4) We conclude and discuss several open issues, which will be part of our future work.

2 Projective Reconstruction from Multiple Views

2.1 The Projective Reconstruction Problem

Suppose we have a set of n 3D points visible in m perspective images. Our goal is to recover 3D structure (point locations) and motion (camera locations) from the image measurements. We will assume no camera calibration or additional 3D information, so we will only be able to reconstruct the scene up to an overall projective transformation of the 3D space [Fau92, HGC92].

We will work in homogeneous coordinates with respect to arbitrary projective coordinate frames. Let $\boldsymbol{Q}_p$ be the unknown homogeneous coordinate vectors of the 3D points, $\boldsymbol{P}_i$ the unknown 3×4 image projection matrices, and $\boldsymbol{q}_{ip}$ the measured homogeneous coordinate vectors of the image points, where $p = 1, \ldots, n$ labels points and $i = 1, \ldots, m$ labels images. Each object is defined only up to an arbitrary nonzero rescaling, *e.g.* $\boldsymbol{Q}_p \sim \mu_p \boldsymbol{Q}_p$. The basic image projection equations say that — up to a set of unknown scale factors — the $\boldsymbol{q}_{ip}$ are the projections of the $\boldsymbol{Q}_p$:

$$\lambda_{ip} \boldsymbol{q}_{ip} \;=\; \boldsymbol{P}_i \boldsymbol{Q}_p$$

We will call the unknown scale factors λ_{ip} **projective depths**[2]. If the $\boldsymbol{Q}_p$ and the $\boldsymbol{q}_{ip}$ are chosen to have affine normalization ('weight' components equal to 1) and the $\boldsymbol{P}_i$ are normalized so that the vectorial part of the 'weight' component row has norm 1, the projective depths become true optical depths, *i.e.* true orthogonal distances from the focal plane of the camera.

The complete set of image projections can be gathered into a single $3m \times n$ matrix equation:

$$\boldsymbol{W} \;\equiv\; \begin{pmatrix} \lambda_{11}\boldsymbol{q}_{11} & \lambda_{12}\boldsymbol{q}_{12} & \cdots & \lambda_{1n}\boldsymbol{q}_{1n} \\ \lambda_{21}\boldsymbol{q}_{21} & \lambda_{22}\boldsymbol{q}_{22} & \cdots & \lambda_{2n}\boldsymbol{q}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{m1}\boldsymbol{q}_{m1} & \lambda_{m2}\boldsymbol{q}_{m2} & \cdots & \lambda_{mn}\boldsymbol{q}_{mn} \end{pmatrix} \;=\; \begin{pmatrix} \boldsymbol{P}_1 \\ \boldsymbol{P}_2 \\ \vdots \\ \boldsymbol{P}_m \end{pmatrix} \begin{pmatrix} \boldsymbol{Q}_1 & \boldsymbol{Q}_2 & \cdots & \boldsymbol{Q}_n \end{pmatrix}$$

Notice that *with the correct projective depths* λ_{ip}, the $3m \times n$ **rescaled measurement matrix** $\boldsymbol{W}$ has rank at most 4. If we could recover the depths, we could apply an SVD based factorization technique similar to that used by Tomasi and Kanade [TK92] to $\boldsymbol{W}$, and thereby recover both 3D structure and camera motion for the scene. The main technical advance of this paper is a practical method for the recovery of the unknown projective depths, using fundamental matrices and epipoles estimated from the image data.

Taken individually, the projective depths are arbitrary because they depend on the arbitrary scale factors chosen for the $\boldsymbol{P}_i$, the $\boldsymbol{Q}_p$ and the $\boldsymbol{q}_{ip}$. However taken as a whole the rescaled measurements $\boldsymbol{W}$ have a strong internal coherence. The overall scale of each triple of rows and each column of $\boldsymbol{W}$ can be chosen arbitrarily (*c.f.* the arbitrary scales of the projections $\boldsymbol{P}_i$ and the 3D points $\boldsymbol{Q}_p$), but once these $m + n$ overall scales have been fixed there is no further freedom of choice for the remaining $mn - m - n$ scale factors in λ_{ip}. Hence, the projective depths really do contain useful information.

2.2 Recovery of Projective Depths

Now we will show how the projective depths can be recovered from fundamental matrices and epipoles, modulo overall row and column rescalings. The point projection equation $\lambda_{ip}\boldsymbol{q}_{ip} = \boldsymbol{P}_i\boldsymbol{Q}_p$ implies that the 6×5 matrix

$$\left(\begin{array}{c|c} \boldsymbol{P}_i & \lambda_{ip}\boldsymbol{q}_{ip} \\ \boldsymbol{P}_j & \lambda_{jp}\boldsymbol{q}_{jp} \end{array}\right) = \left(\begin{array}{c|c} \boldsymbol{P}_i & \boldsymbol{P}_i\boldsymbol{Q}_p \\ \boldsymbol{P}_j & \boldsymbol{P}_j\boldsymbol{Q}_p \end{array}\right) = \begin{pmatrix} \boldsymbol{P}_i \\ \boldsymbol{P}_j \end{pmatrix} \left(\boldsymbol{I}_{4\times 4} \middle| \boldsymbol{Q}_p\right)$$

has rank at most 4. Hence, all of its 5×5 minors vanish. We can expand these by cofactors in the last column to get homogeneous linear equations in the components of $\lambda_{ip}\boldsymbol{q}_{ip}$ and $\lambda_{jp}\boldsymbol{q}_{jp}$. The coefficients are 4×4 determinants

[2] This is not the same notion as the "projective depth" of Shashua, which is a cross ratio of distances along epipolar lines [Sha94]

of projection matrix rows. These turn out to be just fundamental matrix and epipole components [Tri95b, FM95]. In particular, if abc and $a'b'c'$ are even permutations of 123 and $\boldsymbol{P}_i^a$ denotes row a of $\boldsymbol{P}_i$, we have:

$$[\boldsymbol{F}_{ij}]_{a\,a'} = \begin{vmatrix} \boldsymbol{P}_i^b \\ \boldsymbol{P}_i^c \\ \boldsymbol{P}_j^{b'} \\ \boldsymbol{P}_j^{c'} \end{vmatrix} \qquad [\boldsymbol{e}_{ij}]^a = \begin{vmatrix} \boldsymbol{P}_i^a \\ \boldsymbol{P}_j^1 \\ \boldsymbol{P}_j^2 \\ \boldsymbol{P}_j^3 \end{vmatrix} \tag{1}$$

Applying these relations to the three 5×5 determinants built from two rows of image i and three rows of image j gives the following fundamental relation between epipolar lines:

$$(\boldsymbol{F}_{ij}\boldsymbol{q}_{jp})\,\lambda_{jp} = (\boldsymbol{e}_{ij} \wedge \boldsymbol{q}_{ip})\,\lambda_{ip} \tag{2}$$

This relation says two things:

- **Equality up to scale:** The epipolar line of $\boldsymbol{q}_{jp}$ in image i is the line through the corresponding point $\boldsymbol{q}_{ip}$ and the epipole $\boldsymbol{e}_{ij}$. This is just a direct re-statement of the standard epipolar constraint.
- **Equality of scale factors:** If the correct projective depths are used in (2), the two terms have *exactly the same size* — the equality is exact, not just up to scale. This is the new result that allows us to recover projective depths using fundamental matrices and epipoles. Analogous results based on higher order matching tensors can be found in [Tri95a], but in this paper we will use only equation (2).

Our strategy for the recovery of projective depths is quite straightforward. Equation (2) relates the projective depths of a single 3D point in two images. By estimating a sufficient number of fundamental matrices and epipoles, we can amass a system of homogeneous linear equations that allows the complete set of projective depths of a given point to be found, up to an arbitrary overall scale factor. At a minimum, this can be done with any set of $m-1$ fundamental matrices that link the m images into a single connected graph. If additional fundamental matrices are available, the equations become redundant and (hopefully) more robust. In the limit, all $m(m-1)/2$ fundamental matrices and all $m(m-1)$ equations could be used to find the m unknown depths for each point, but this would be computationally very expensive. We are currently investigating policies for choosing economical but robust sets of equations, but in this paper we will restrict ourselves to the simplest possible choice: the images are taken pairwise in sequence, $\boldsymbol{F}_{12}, \boldsymbol{F}_{23}, \ldots, \boldsymbol{F}_{m-1\,m}$.

This is almost certainly not the most robust choice, but it (or any other minimal selection) has the advantage that it makes the depth recovery equations trivial to solve. Solving the vector equation (2) in least squares for λ_{ip} in terms of λ_{jp} gives:

$$\lambda_{ip} = \frac{(\boldsymbol{e}_{ij} \wedge \boldsymbol{q}_{ip}) \cdot (\boldsymbol{F}_{ij}\boldsymbol{q}_{jp})}{\|\boldsymbol{e}_{ij} \wedge \boldsymbol{q}_{ip}\|^2}\,\lambda_{jp} \tag{3}$$

Such equations can be recursively chained together to give estimates for the complete set of depths for point p, starting from some arbitrary initial value such as $\lambda_{1p} = 1$.

However there is a flaw in the above argument: fundamental matrices and epipoles can only be recovered up to an unknown scale factor, so we do not actually know the scale factors in equations (1) or (2) after all! In fact this does not turn out to be a major problem. It is a non-issue if a minimal set of depth-recovery equations is used, because the arbitrary overall scale factor for each image can absorb the arbitrary relative scale of the $\boldsymbol{F}$ and $\boldsymbol{e}$ used to recover the projective depths for that image. However if redundant depth-recovery equations are used it is essential to choose a self-consistent scaling for the estimated fundamental matrices and epipoles. We will not describe this process here, except to mention that it is based on the quadratic identities between matching tensors described in [Tri95a].

Note that with unbalanced choices of scale for the fundamental matrices and epipoles, the average scale of the recovered depths might tend to increase or decrease exponentially during the recursive chaining process. Theoretically this is not a problem because the overall scales are arbitrary, but it could well make the factorization phase of the reconstruction algorithm numerically ill-conditioned. To counter this we re-balance the recovered matrix of projective depths after it has been built, by judicious overall row and column scalings.

2.3 Projective Shape and Motion by Factorization

Once we have obtained the projective depths, we can extract projective shape and motion from the rescaled measurement matrix $\boldsymbol{W}$.

Let

$$\boldsymbol{W} = \boldsymbol{U}\ \mathrm{diag}(\sigma_1, \sigma_2, \ldots, \sigma_s)\ \boldsymbol{V}$$

be a Singular Value Decomposition (SVD) of $\boldsymbol{W}$, with $s = \min\{3m, n\}$ and singular values $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_s \geq 0$. Since $\boldsymbol{W}$ is of rank 4, the σ_i for $i > 4$ vanish. Thus, only the first 4 columns (rows) of $\boldsymbol{U}$ ($\boldsymbol{V}$) contribute to this matrix product. Let $\boldsymbol{U}'$ ($\boldsymbol{V}'$) the matrix of the first 4 columns (rows) of $\boldsymbol{U}$ ($\boldsymbol{V}$). Then,

$$\boldsymbol{W} = \boldsymbol{U}'_{3m\times 4}\ \underbrace{\mathrm{diag}(\sigma_1, \sigma_2, \sigma_3, \sigma_4)}_{\Sigma}\ \boldsymbol{V}'_{4\times n} = \boldsymbol{U}'\ \boldsymbol{\Sigma}\ \boldsymbol{V}' \quad .$$

Any factorization of $\boldsymbol{\Sigma}$ into two 4×4 matrices $\boldsymbol{\Sigma}'$ and $\boldsymbol{\Sigma}''$, $\boldsymbol{\Sigma} = \boldsymbol{\Sigma}'\boldsymbol{\Sigma}''$, leads to

$$\boldsymbol{W} = \underbrace{\boldsymbol{U}'\boldsymbol{\Sigma}'}_{\mathbb{U}}\underbrace{\boldsymbol{\Sigma}''\boldsymbol{V}'}_{\mathbb{V}} = \hat{\boldsymbol{U}}_{3m\times 4}\hat{\boldsymbol{V}}_{4\times n} \quad .$$

We can interpret the matrix $\hat{\boldsymbol{U}}$ as a collection of m (3×4) projection matrices $\hat{\boldsymbol{P}}_i$ and $\hat{\boldsymbol{V}}$ as collection of n 4-vectors $\hat{\boldsymbol{Q}}_p$, representing 3D shape :

$$\boldsymbol{W} = \hat{\boldsymbol{U}}\hat{\boldsymbol{V}} = \begin{pmatrix} \hat{\boldsymbol{P}}_1 \\ \hat{\boldsymbol{P}}_2 \\ \vdots \\ \hat{\boldsymbol{P}}_m \end{pmatrix}_{3m\times 4} \begin{pmatrix} \hat{\boldsymbol{Q}}_1 & \hat{\boldsymbol{Q}}_2 & \cdots & \hat{\boldsymbol{Q}}_n \end{pmatrix}_{4\times n} \tag{4}$$

Equation (4) shows that the $\hat{\boldsymbol{P}}_i$ and $\hat{\boldsymbol{Q}}_p$ represent at least projective motion and shape, since

$$\hat{\boldsymbol{P}}_i\hat{\boldsymbol{Q}}_p = \lambda_{ip}\,\boldsymbol{q}_{ip} \sim \boldsymbol{q}_{ip} \quad .$$

Unlike the case of orthographic projections [TK92], there are no further constraints on the $\hat{\boldsymbol{P}}_i$ or $\hat{\boldsymbol{Q}}_p$: we can *only* recover projective shape and motion. For any non singular projective transformation $\boldsymbol{T}_{4\times 4}$, $\hat{\boldsymbol{P}}_i\boldsymbol{T}$ and $\boldsymbol{T}^{-1}\hat{\boldsymbol{Q}}_p$ is an equally valid factorization of the data into projective motion and shape :

$$(\hat{\boldsymbol{P}}_i\boldsymbol{T})(\boldsymbol{T}^{-1}\hat{\boldsymbol{Q}}_p) = \hat{\boldsymbol{P}}_i\hat{\boldsymbol{Q}}_p \sim \boldsymbol{q}_{ip} \quad .$$

A consequence of this is that the factorization of $\boldsymbol{\Sigma}$ is arbitrary. For the implementation, we chose $\boldsymbol{\Sigma}' = \boldsymbol{\Sigma}'' = \boldsymbol{\Sigma}^{1/2} = \mathrm{diag}(\sigma_1^{1/2}, \sigma_2^{1/2}, \sigma_3^{1/2}, \sigma_4^{1/2})$.

3 The Algorithm

Based on the observations made above, we have developed a practical algorithm for projective reconstruction from multiple views. Besides the major two steps, determination of the scale factors λ_{ip} and factorization of the rescaled measurement matrix, the outline of our algorithm is based on some numerical considerations.

3.1 Normalization of Image Coordinates

To ensure good numerical conditioning of the method, we work with normalized image coordinates, as described in [Har95]. This normalization consists of applying a similarity transformation (translation and uniform scaling) $\boldsymbol{T}_i$ to each image, so that the transformed points are centered at the origin and the mean distance from the origin is $\sqrt{2}$.

All of the remaining steps of the algorithm are done in normalized coordinates. Since we actually compute projective motion and shape for the transformed image points $\boldsymbol{T}_i\boldsymbol{q}_{ip}$, $\hat{\boldsymbol{P}}_i\hat{\boldsymbol{Q}}_p = \lambda_{ip}\boldsymbol{T}_i\boldsymbol{q}_{ip} \sim \boldsymbol{T}_i\boldsymbol{q}_{ip}$, the resulting projection estimates $\hat{\boldsymbol{P}}_i$ must be corrected : $\hat{\boldsymbol{P}}_i{}' = \boldsymbol{T}_i^{-1}\hat{\boldsymbol{P}}_i$. The $\hat{\boldsymbol{P}}_i{}'$ and $\hat{\boldsymbol{Q}}_p$ then represent projective motion and shape corresponding to the measured image points $\boldsymbol{q}_{ip}$.

Our results show that this simple normalization drastically improves the results of the projective reconstruction.

3.2 Balancing the Rescaled Measurement Matrix

Consider the factorization of the rescaled measurement matrix W in projective motion and shape :

$$\boldsymbol{W} = \begin{pmatrix} \lambda_{11}\boldsymbol{q}_{11} & \lambda_{12}\boldsymbol{q}_{12} & \cdots & \lambda_{1n}\boldsymbol{q}_{1n} \\ \lambda_{21}\boldsymbol{q}_{21} & \lambda_{22}\boldsymbol{q}_{22} & \cdots & \lambda_{2n}\boldsymbol{q}_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{m1}\boldsymbol{q}_{m1} & \lambda_{m2}\boldsymbol{q}_{m2} & \cdots & \lambda_{mn}\boldsymbol{q}_{mn} \end{pmatrix} = \begin{pmatrix} \hat{\boldsymbol{P}}_1 \\ \hat{\boldsymbol{P}}_2 \\ \vdots \\ \hat{\boldsymbol{P}}_m \end{pmatrix} \begin{pmatrix} \hat{\boldsymbol{Q}}_1 & \hat{\boldsymbol{Q}}_2 & \cdots & \hat{\boldsymbol{Q}}_n \end{pmatrix}$$

Multiplying column l of $\boldsymbol{W}$ by a non zero scalar ν_l corresponds to multiplying $\hat{\boldsymbol{Q}}_l$ by ν_l. Analogously, multiplying the image k rows $(3k-2, 3k-1, 3k)$ by a non zero scalar μ_k corresponds to multiplying the projection matrix $\hat{\boldsymbol{P}}_k$ by μ_k. Hence, point-wise and image-wise rescalings of $\boldsymbol{W}$ do not affect the recovered projective motion and shape.

However, these considerations are only valid in the absence of noise. In presence of noise, $\boldsymbol{W}$ will only be approximately of rank 4, and scalar multiplications of $\boldsymbol{W}$ as described above *will* affect the results. We therefore aim to improve the results of the factorization by applying appropriate point- and image-wise rescalings to $\boldsymbol{W}$. The goal is to ensure good numerical conditioning by rescaling so that all rows and columns of $\boldsymbol{W}$ have on average the same order of magnitude. To do this we use the following iterative scheme :

1. Rescale each column l so that $\sum_{r=1}^{3m}(w_{rl})^2 = 1$.
2. Rescale each triplet of rows $(3k-2, 3k-1, 3k)$ so that $\sum_{l=1}^{n}\sum_{i=3k-2}^{3k} w_{il}^2 = 1$.
3. If the entries of $\boldsymbol{W}$ changed significantly, repeat 1 and 2.

Note that, since we work with normalized image coordinates $\boldsymbol{q}_{ip}$, it would be sufficient to balance only the $m \times n$ matrix (λ_{ip}) instead of $\boldsymbol{W}$.

3.3 Outline of the Algorithm

The complete algorithm is composed of the following steps.

1. Normalize the image coordinates, by applying transformations $\boldsymbol{T}_i$.
2. Estimate the fundamental matrices and epipoles with the method of [Har95].
3. Determine the scale factors λ_{ip} using equation (3).
4. Build the rescaled measurement matrix $\boldsymbol{W}$.
5. Balance $\boldsymbol{W}$ by column-wise and "triplet-of-rows"-wise scalar mutliplications.
6. Compute the SVD of the balanced matrix $\boldsymbol{W}$.
7. From the SVD, recover projective motion and shape.
8. Adapt projective motion, to account for the normalization transformations $\boldsymbol{T}_i$ of step 1.

4 Experimental Evaluation of the Algorithm

4.1 Experiments with Simulated Images

We conducted a large number of experiments with simulated images to quantify the performance of the algorithm. The simulations used three different configurations : lateral movement of a camera, movement towards the scene, and a circular movement around the scene (see figure 1). In configuration 2, the depths of points lying on the line joining the projection centers can not be recovered. Reconstruction of points lying close to this line is extremely difficult, as was confirmed by the experiments, which resulted in quite inaccurate reconstructions for this configuration.

For the circular movement, the overall trajectory of the camera formed a quarter circle, centered on the scene. For each specific experiment, the trajectory length was the same for all three configurations. The m different viewing positions were equidistantly distributed along the trajectory.

In order to simulate realistic situations, we adopted the following parameters : the camera's calibration matrix was $\mathrm{diag}(1000, 1000, 1)$. The scene was composed of points distributed uniformly in a sphere of radius 100. The distance between the camera and the center of the sphere was 200 (for configuration 2 this was the distance with respect to the view m).

For each configuration, the following experiment was conducted 50 times :

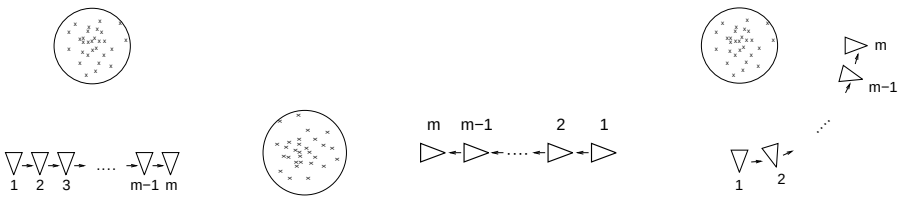


Figure 1. The 3 configurations for simulation.(1) *Lateral movement.* **(2)** *Translation towards the scene.* **(3)** *Circular movement.*

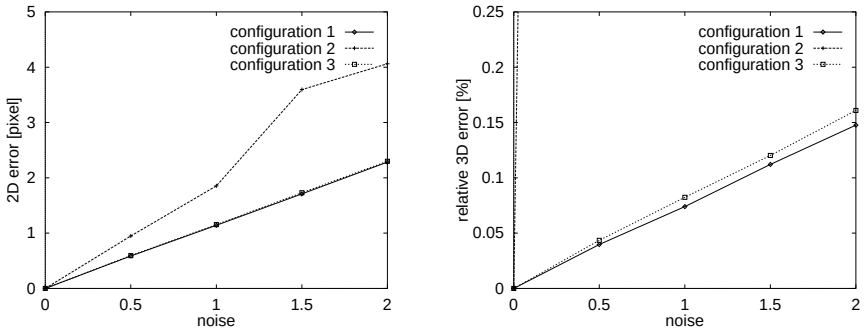
1. Determine at random 50 points in the sphere.
2. Project the points into the m views.
3. Add Gaussian noise of levels $0.0, 0.5, \ldots, 2.0$ to the image coordinates.
4. Carry out projective reconstruction with our algorithm.
5. Compute the image distance error of the backprojected points (2D error) : $\frac{1}{mn}\sum_{i=1}^{m}\sum_{p=1}^{n}||\hat{\boldsymbol{P}}_i\hat{\boldsymbol{Q}}_p - \boldsymbol{q}_{ip}||$, where $||.||$ means the Euclidean vector norm.
6. Align the projective reconstruction with the Euclidean model and compute the distance error in the Euclidean frame (3D error).

The results of these experiments were analyzed with respect to several variables, as reported in the following subsections. All values represented in the graphs are the mean result over 50 trials. To monitor the effect of outliers on the results, we also computed the median values. These gave graphs similar to those for the means, which we will not show here.

2D errors are given in pixels, whereas 3D errors are given relative to the scene's size, in percent.

4.1.1 Sensitivity to Noise

Graphs 1 and 2 show the behavior of the algorithm with respect to different noise levels for the three configurations. For this experiment, reconstruction was done from 10 views.

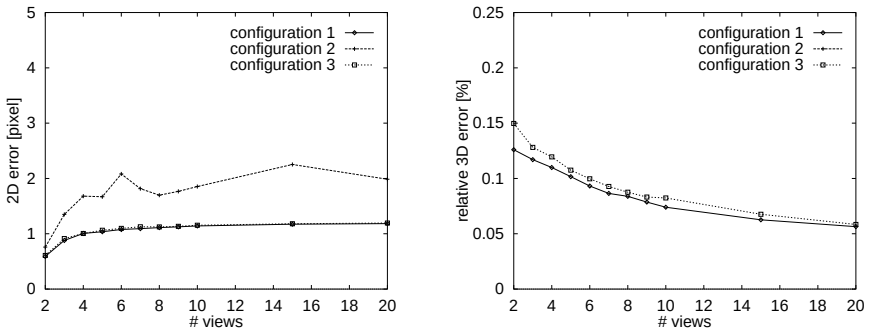


Graphs 1 and 2 : Sensitivity to noise. *The 2D error curves for the configurations 1 and 3 are nearly undistinguishable. 3D error for configuration 2 goes rapidly off scale.*

The algorithm performed almost equally well for configurations 1 and 3, whereas the 3D error for configuration 2 exceeds 100 % for 2.0 pixels noise. Considering the graphs of configuration 2, we also see that 2D and 3D error are not always well correlated. For configurations 1 and 3, the 2D error is of the same order as pixel noise. Note also the linear shape of the graphs.

4.1.2 Number of Views

The image noise for this experiment was 1.0 pixel.

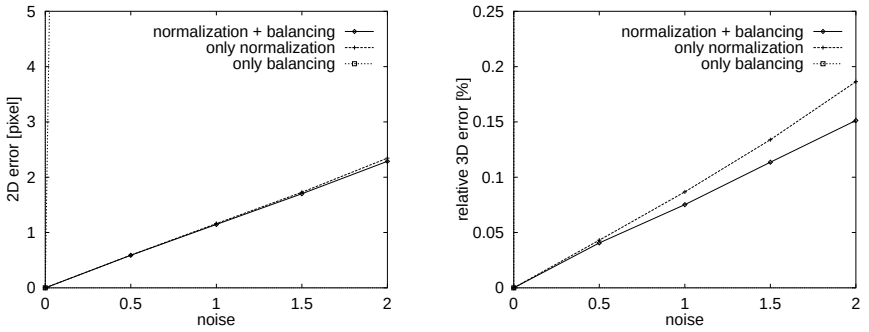


Graphs 3 and 4 : Behavior with respect to number of views. *The 2D error curves for the configurations 1 and 3 are nearly undistinguishable. The 3D error for configuration 2 lies above 5 %. The curve is thus not visible in the graph.*

The graphs show the expected behavior : when more views are used for reconstruction, the structure is recovered more accurately. Secondly, 2D error augments with increasing number of views, but shows a clearly asymptotic behavior. 1. Note that the use of 20 views reduces the 3D error to 50 % of that for 2 views.

4.1.3 Importance of Normalization and Balancing

The error values in the previous graphs were obtained with the algorithm as described in subsection 3.3. To underline the importance of using normalized image coordinates, we also ran the algorithm using unnormalized ones. The effects of not balancing the rescaled measurement matrix before factorization were also examined.



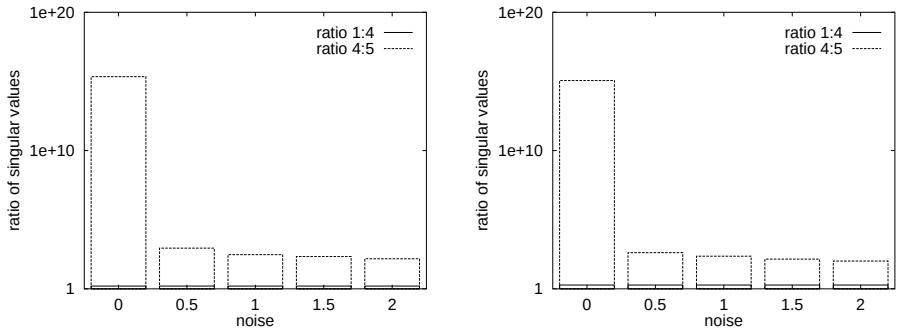
Graphs 5 and 6 : Influence of normalization and balancing. *The results presented here were obtained for configuration 1. The 2D error curve for "only balancing" goes off scale even for 0.5 pixels noise and the 3D curve is so steep that it is not even visible.*

When the image coordinates are not normalized, the error is already off scale for 0.5 pixel noise. An explanation for this is the bad conditioning of the rescaled measurement matrix (see also next paragraph). As for balancing, we see that this improves 3D errors up to 20 %, and hence should always be part of the algorithm.

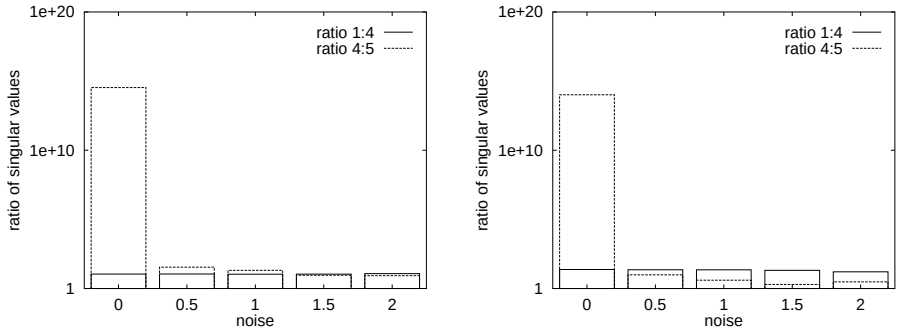
4.1.4 Robustness of the Factorization

The applicability of our factorization method is based on the rank 4-ness of the rescaled measurement matrix $\boldsymbol{W}$ (in the noiseless case). To test the robustness of this property, we evaluated how close $\boldsymbol{W}$ is to rank 4 in practice. To be close to rank 4, the ratio of the 4th and 5th largest singular values , $\sigma_4 : \sigma_5$, should be large with respect to the ratio of the 1st and 4th largest, $\sigma_1 : \sigma_4$. In the following graphs, these two ratios are represented, for configurations 1 and 2

and for 2 and 20 views. Note that the y-axes are scaled logarithmically.



Graphs 7 and 8 : Ratios of singular values for configuration 1. *The graph on the left shows the situation for 2 views, on the right for 20 views.*



Graphs 9 and 10 : Ratios of singular values for configuration 2. *The graph on the left shows the situation for 2 views, on the right for 20 views.*

We see that for configuration 1, the matrix is always very close to rank 4 : $(\sigma_1 : \sigma_4)$ is lower than 2, whereas $(\sigma_4 : \sigma_5)$ lies clearly above 100. As for configuration 2, the graphs reflect the bad performance in 3D reconstruction. $(\sigma_1 : \sigma_4)$ is about 10, while for high noise levels or many views $(\sigma_4 : \sigma_5)$ is close to 1.

4.2 Evaluation with Real Images

The algorithm has also been tested on several sequences of real images. For 2 of them we show results.

4.2.1 The House Sequence

Figure 2 shows the first and last image of a sequence of 6 images of a scene with a wooden house. 38 points were tracked over the whole sequence, but only extracted with ± 1 pixel accuracy.

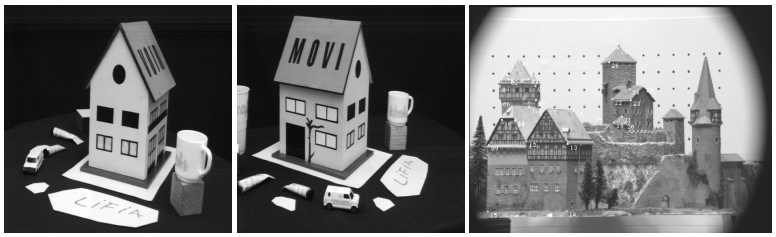


Figure 2. First and last image of the house sequence and one image of the castle sequence.

To estimate the quality of the projective reconstruction, we aligned it with an approximate Euclidean model of the scene obtained from calibrated views (see figure 3). Lines have been drawn between some of the points to aid visualization.

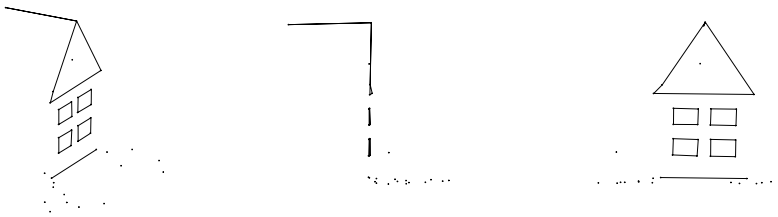


Figure 3. Three views of the reconstructed house.(1) "General view". **(2)** Side view. **(3)** Front view.

In the side and front views we see that right angles are approximately conserved, and that the windows are coplanar with the wall. The bumpiness on the left side of the roof is due to the fact that the roof stands out slightly from the house's front wall (see figure 2), thus causing occlusion in the last view of the edge point between roof and wall.

4.2.2 The Castle Sequence

28 points have been tracked through the 11 images of the scene shown in the right part of figure 2. 3D ground truth is available, and the reconstruction errors have been evaluated quantitatively. The projective reconstruction was aligned with the Euclidean model and the resulting RMS error was 4.45 mm for an object size of about 220mm $\times$ 210mm $\times$ 280mm. The RMS error of the reprojected structure with respect to the measured image points was less than 0.02 pixels.

We also applied a Levenberg-Marquardt nonlinear least-squares estimation algorithm, with the results of our method as initialization. This slightly improved the 2D reprojection error, however the 3D reconstruction error was not significantly changed.

5 Discussion and Further Work

In this paper, we have proposed a method of projective reconstruction from multiple uncalibrated images. The method is very elegant, recovering shape and motion by factorization of one matrix, containing all image points of all views. This factorization is only possible when the image points are correctly scaled. We have proposed a very simple way to obtain the individual scale factors, using only fundamental matrices and epipoles estimated from the image data.

The algorithm proves to work well with real images. Quantitative evaluation by numerical simulations shows the robustness of the factorization and the good performance with respect to noise. The results also show that it is essential to work with normalized image coordinates.

Some aspects of the method remain to be examined. In the current implementation, we recover projective depths by chaining equation (2) for pairs of views $(12), (23), \ldots, (m-1, m)$. However, it would be worth investigating whether other kinds of chaining are not more stable. Furthermore, uncertainty estimates on the fundamental matrices should be considered when choosing which of the equations (2) to use. To run the algorithm in practice, it should also be able to treat points which are not visible in all images. Finally the method could be extended to use trilinear and perhaps even quadrilinear matching tensors.

Acknowledgements. This work was partially supported by INRIA France and E.C. projects HCM and SECOND. Data for this research were partially provided by the Calibrated Imaging Laboratory at Carnegie Mellon University, supported by ARPA, NSF, and NASA (the castle sequence can be found at http://www.cs.cmu.edu/ cil/cil-ster.html).

References

[Fau92] O. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? In G. Sandini, editor, *Proceedings of the 2nd European Conference on Computer Vision, Santa Margherita Ligure, Italy*, pages 563–578. Springer-Verlag, May 1992.

[FM95] O. Faugeras and B. Mourrain. On the geometry and algebra of the point and line correspondences between n images. In *Proceedings of the 5th International Conference on Computer Vision, Cambridge, Massachusetts, USA*, pages 951–956, June 1995.

[Har93] 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.

[Har95] 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.

[HGC92] R. Hartley, R. Gupta, and T. Chang. Stereo from uncalibrated cameras. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, Urbana-Champaign, Illinois, USA*, pages 761–764, 1992.

[HS94] R. Hartley and P. Sturm. Triangulation. In *Proceedings of* ARPA *Image Understanding Workshop, Monterey, California*, pages 957–966, November 1994.

[MM95] P.F. McLauchlan and D.W. Murray. A unifying framework for structure and motion recovery from image sequences. In *Proceedings of the 5th International Conference on Computer Vision, Cambridge, Massachusetts, USA*, pages 314–320, 1995.

[MVQ93] R. Mohr, F. Veillon, and L. Quan. Relative 3D reconstruction using multiple uncalibrated images. In *Proceedings of the Conference on Computer Vision and Pattern Recognition, New York, USA*, pages 543–548, June 1993.

[PK94] C. J. Poelman and T. Kanade. A paraperspective factorization method for shape and motion recovery. In J.O. Eklundh, editor, *Proceedings of the 3rd European Conference on Computer Vision, Stockholm, Sweden*, pages 97–108, May 1994.

[Sha94] A. Shashua. Projective structure from uncalibrated images: Structure from motion and recognition. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 16(8):778–790, August 1994.

[Sha95] A. Shashua. Algebraic functions for recognition. IEEE *Transactions on Pattern Analysis and Machine Intelligence*, 17(8):779–789, August 1995.

[TK92] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: A factorization method. *International Journal of Computer Vision*, 9(2):137–154, 1992.

[Tri95a] B. Triggs. The geometry of projective reconstruction i: Matching constraints and the joint image. *International Journal of Computer Vision*, 1995. submitted.

[Tri95b] B. Triggs. Matching constraints and the joint image. In Eric Grimson, editor, *Proceedings of the 5th International Conference on Computer Vision, Cambridge, Massachusetts, USA*, pages 338–343. IEEE, IEEE Computer Society Press, June 1995.