

An Experimental Study of Projective Structure from Motion

John Oliensis (oliensis@research.nj.nec.com)

NEC Research Institute

4 Independence Way

Princeton, N.J. 08540

and

Venu Govindu

Department of Electrical Engineering and Center for

Automation Research

University of Maryland

College Park, MD 20742

Abstract

This paper studies the usefulness of the projective approach to structure from motion (SFM). We conduct an essentially algorithm-independent experimental comparison of projective versus Euclidean reconstruction. Our results show that Euclidean reconstruction is essentially as accurate as projective reconstruction, even with significant calibration error and for the pure projective structure. Thus calibration error is not a compelling motivation for the projective framework. But projective optimization is more reliable than its Euclidean equivalent: we find that it has fewer problems with local minima. We describe several techniques that enhance the convergence of our Levenberg–Marquardt optimization algorithm. Most importantly, we find that it is crucial to exploit the compactness of projective space—even in the pure Euclidean framework. We also analyze the problem of determining a stable basis set of 5 points for projective reconstruction.

Keywords Structure from motion, projective geometry, multi-frame structure from motion, optimization, Levenberg–Marquardt, calibration.

An Experimental Study of Projective Structure from Motion

Abstract

This paper studies the usefulness of the projective approach to structure from motion (SFM). We conduct an essentially algorithm-independent experimental comparison of projective versus Euclidean reconstruction. Our results show that Euclidean reconstruction is essentially as accurate as projective reconstruction, even with significant calibration error and for the pure projective structure. Thus calibration error is not a compelling motivation for the projective framework. But projective optimization is more reliable than its Euclidean equivalent: we find that it has fewer problems with local minima. We describe several techniques that enhance the convergence of our Levenberg–Marquardt optimization algorithm. Most importantly, we find that it is crucial to exploit the compactness of projective space—even in the pure Euclidean framework. We also analyze the problem of determining a stable basis set of 5 points for projective reconstruction.

Keywords Structure from motion, projective geometry, multi-frame structure from motion, optimization, Levenberg–Marquardt, calibration.

1 Introduction

The projective approach to structure from motion (SFM) was introduced originally to avoid the difficulty of calibrating cameras precisely [1, 2]. Whereas standard **Euclidean SFM** reconstructs a Euclidean 3D model of the scene assuming a known camera calibration, **projective SFM** reconstructs the projective structure without prior (linear) calibration. Projective SFM is often thought of as unphysical, but when it is applied to real motion sequences it has a perfectly physical interpretation: it is equivalent to Euclidean SFM except that the linear camera calibration is treated as unknown and potentially arbitrarily different in each image¹.

But is the difficulty of accurate calibration sufficient motivation for projective SFM? The projective approach’s assumption of complete ignorance of the calibration is unrealistic—usually some information **is** available. For instance: 1) the x and y image axes are almost always close to perpendicular; 2) we often have approximate calibration information—this can be useful since errors in some parameters (e.g., the image center) are known to have little effect on depth recovery; and especially 3) we often know that a mo-

¹In the projective framework, the image point position is determined by $I = MS$, where I is a homogeneous 3-vector representing the image, M is the 3×4 camera matrix, and S is a homogeneous 4-vector representing the structure. Projective SFM is equivalent to Euclidean SFM with unknown linear calibration (apart from singular cases) in the sense that we can interpret S as a standard Euclidean structure and M as the product of motion and linear calibration matrices. (A non-singular M has a unique decomposition into such a product.)

tion sequence was taken with a single camera, implying that all images have approximately the same calibration². The projective approach is also unrealistic in that it allows for arbitrary linear calibration errors but neglects the potentially significant nonlinear camera distortions.

The consequence of these unrealistic assumptions is a loss in accuracy—for the best reconstruction accuracy all available information must be used, including approximate information. This is true **even for the projective structure**. If the calibration is partly known or known to be fixed, Euclidean reconstruction of the projective structure using this knowledge is more accurate than projective reconstruction. For non-optimal algorithms, this loss in accuracy can translate effectively into a loss in robustness. Thus the difficulty of precise calibration should not rule out Euclidean techniques.

An additional motivation sometimes suggested for the projective framework is that it yields new reconstruction algorithms. But projective SFM is formally mathematically equivalent to Euclidean SFM with **known** calibration if the constraints on the rotation matrix are relaxed. The projective approach models the formation of an image of N 3D points by $I = MS$, where M is a 3×4 camera matrix and S a $4 \times N$ structure matrix. The columns of the structure matrix S give the homogeneous coordinates of the 3D points, and the camera matrix M summarizes the camera rotation, translation, and

²Except possibly for focal length variations, which are easy to handle in a purely Euclidean framework [9].

linear calibration parameters for the given image.

In standard Euclidean SFM, image formation can be modeled in exactly the same way; only the interpretation of the camera matrix M differs. Now the first 3 columns of the M give the rotation matrix, and the last column holds the translation. Thus the two approaches are equivalent if the Euclidean rotation matrix is allowed to be arbitrary rather than restricted to being orthogonal. Neglecting the rotation matrix constraints was already a standard technique for Euclidean algorithms, for instance, in the “8-point” computation of the essential matrix [6] or Tomasi and Kanade’s approach to orthographic SFM [14]. Though the rotation constraints must be restored eventually in the Euclidean approach, there are standard techniques for doing this, e.g., [2] or [14]. Thus any “projective” algorithm translates immediately into a Euclidean one followed by orthogonalization of the rotation matrix; the projective framework does not create the possibility of new algorithms.

In this paper, we study experimentally whether projective methods are useful in the traditional SFM context: motion sequences obtained with a single camera. Our results are effectively algorithm independent. This is important because current algorithms are far from ideal. Our aim is to focus not on the properties of imperfect algorithms, but rather on an intrinsic comparison of projective vs. Euclidean reconstruction. Unsurprisingly, we find that Euclidean SFM assuming a single camera of unknown calibration recov-

ers the projective structure more accurately than does projective SFM. But Euclidean SFM based on an assumed calibration gives reconstruction accuracies comparable to those of projective SFM, even when there are significant errors in the assumed calibration and even for the projective structure. Thus inaccurate knowledge of the calibration is not a compelling motivation for projective SFM. We also study the effects of calibration errors on Euclidean estimation.

A second purpose of our experiments is to study the optimization approach for SFM—to date the only approach that has been shown to give robust projective reconstruction. We focus particularly on the local–minimum problem, studying how reliably optimization avoids local minima and converges to the true global minimum from general starting conditions. We also describe techniques that improve the reliability and speed of optimization.

One of the important results of this paper is that the **local–minimum problem appears less severe for “projective” SFM than for Euclidean SFM**. We find that optimization converges **more stably** to the global minimum in the “projective” case than in the Euclidean case. This may be due to the fact that the extra, artificial unknowns introduced in projective SFM give high dimensional “escape routes” from the local valleys of the Euclidean error function. This phenomenon has also been observed in other contexts. For instance, it is known that introducing more hidden

variables in neural nets improves generalization [7], i.e., it eliminates false local minima.

The result also may also be due to the fact that the Euclidean approach involves nonlinear constraints on the rotation matrices. These complicate the error function and can create additional local minima. The “8-point” algorithm [6] is a good analogy. As a result of relaxing the rotation constraints, this “projective” algorithm in effect is minimizing a quadratic error function with a single minimum: there is no local–minimum problem. (But relaxing the constraints does exact a cost in accuracy.)

As we emphasize above by our use of quotes, this result does *not* imply that the projective framework is more robust than the Euclidean one. Since the projective and Euclidean frameworks are formally equivalent (neglecting the Euclidean rotation constraints), one can also think of “projective” optimization as the first, approximate stage of a Euclidean algorithm. (Such a Euclidean algorithm would recover the motion directly from the recovered camera matrices M assuming known calibration, unlike the standard projective approach which computes the calibrations as well as the motions from the M .) Thus we can equally well restate our result as: neglecting the orthogonality constraints on the rotation matrices increases the robustness of Euclidean optimization.

This paper also presents techniques which appear crucial for enhancing

the speed and reliability of optimization convergence. We find that it is particularly important to exploit the compactness of projective space, *even for Euclidean reconstruction*. We also analyze the problem of determining a stable basis set of 5 points for projective reconstruction³.

2 Formulation and Issues

We consider motion sequences taken with a single, possibly incorrectly calibrated camera moving in a fixed scene, with known correspondences. We focus on the problem of estimating structure, comparing the results for a maximum likelihood estimate (MLE) in the Euclidean and the projective frameworks. As far as possible, we factor out algorithm-dependent effects in computing the MLEs. Of course the MLE is not necessarily the “optimal” estimate. But for a severely overconstrained estimation problem like multi-frame SFM, it should give results close to those of other “optimal” estimators.

Let \mathbf{X}_i represent the 3D coordinates of the i -th structure point in the reference frame of the first camera position. R^h and \mathbf{T}^h represent the rotation and translation from this frame to the camera frame of the h -th image. In our synthetic experiments, the image points are generated by transforming

³Our experiments use both this method and Hartley’s method of fixing one of the camera matrices. In fact, it is well known in the photogrammetry literature that fixing no parameters can actually improve convergence in optimization.

to the frame of the current camera, projecting onto the ideal image plane (characterized by $\hat{\mathbf{z}} \cdot \mathbf{X} = 1$), applying a 5 parameter affine transform in the image plane to represent the effects of the calibration parameters, and then adding noise.

Let $\mathbf{X}_i^h = R^h (\mathbf{X}_i - \mathbf{T}^h)$ be the result of transforming the \mathbf{X}_i to the camera reference frame for the h -th image. The projection onto the ideal image plane is

$$\bar{\mathbf{x}}_i^h (R^h, \mathbf{T}^h, \mathbf{X}_i) \equiv \frac{1}{X_3^h} \begin{pmatrix} X_{1i}^h \\ X_{2i}^h \end{pmatrix}.$$

The shifted image coordinates are given by $\mathbf{x}_i^h (\mathbf{T}^h, R^h, \mathbf{X}_i, A, C) = \mathbf{A}\bar{\mathbf{x}}_i^h + \mathbf{C}$, where

$$A = \begin{pmatrix} (1 + \delta f)(1 + r) & S \\ 0 & (1 + \delta f)/(1 + r) \end{pmatrix}.$$

Here δf represents the deviation of the focal length from 1, r the discrepancy in scaling between the horizontal and vertical axes, S the shear distortion, and \mathbf{C} the shift in the camera center. We take the shear $S = 0$ since it is typically small for real images. The observed image coordinates \mathbf{u} are generated by adding noise: $\mathbf{u}_i^h = \mathbf{x}_i^h + \eta_i^h$, where η_i^h is a random noise.

Euclidean SFM. *Compute the structure \mathbf{X}_i and motion R^h, \mathbf{T}^h minimizing the least-squares error*

$$\sum_{h,i} \left| \bar{\mathbf{x}}_i^h \left(R^h, \mathbf{T}^h, \mathbf{X}_i \right) - \mathbf{u}_i^h \left(R_g^h, \mathbf{T}_g^h, \mathbf{X}_{gi}, A_g, \mathbf{C}_g, \eta \right) \right|^2,$$

where the subscript g denotes the fixed, ground-truth value.

Here we neglect the possibility of calibration error, i.e. $A_g \neq \mathbf{1}_2$, $C \neq 0$, where $\mathbf{1}_2$ is a two-dimensional identity matrix. The reconstruction is ambiguous up to a similarity transform $\mathbf{X}_i \rightarrow sR(\mathbf{X}_i - \mathbf{T})$.

Let \mathcal{H}_i represent the homogeneous structure coordinates: $\mathcal{H}_i^T \equiv \left(\mathbf{X}_i^T \quad \chi_4 \right)$

where we take χ_4 constant over all points. Let M^h be a 3×4 camera matrix combining the camera calibration parameters with the motion parameters for the h -th image. Define the 3D homogeneous image coordinates by $\chi_i^h \left(M^h, \mathcal{H}_i \right) = M^h \mathcal{H}_i$, up to an arbitrary scaling.

Projective SFM. *Compute the camera matrices M^h and structure \mathcal{H}_i minimizing*

$$\sum_{h,i} \left| \frac{1}{\chi_{3i}^h} \begin{pmatrix} \chi_{1i}^h \\ \chi_{2i}^h \end{pmatrix} - \mathbf{u}_i^h \right|^2. \quad (1)$$

The reconstruction is ambiguous up to a full projective transform $\mathcal{H}_i \rightarrow P\mathcal{H}_i$, where P is a 4×4 matrix.

It is sometimes stated that the maximum likelihood error function is not well defined for projective reconstruction, since imposing a Euclidean metric in the image plane seems inconsistent with abandoning it for the structure.

But, as noted in the introduction, projective SFM has a straightforward Euclidean interpretation, for which (1) is the correct error.

Fixed–Camera Reconstruction. *Compute the structure \mathbf{X}_i , motion R^h , \mathbf{T}^h , and calibration parameters A, \mathbf{C} minimizing*

$$\sum_{h,i} |\mathbf{x}_i^h (R^h, \mathbf{T}^h, \mathbf{X}_i, \mathbf{A}, \mathbf{C}) - \mathbf{u}_i^h|^2.$$

Here we assume the sequence was taken with a single camera of unknown calibration; the problem is to recover the calibration as well as to reconstruct the scene.

3 Experimental Methods

Our goal is an algorithm–independent comparison of projective vs. Euclidean SFM. Since precise ground truth is needed to compare reconstruction accuracies, and since we focus on numerical properties, our experiments are primarily synthetic. We use a large number of synthetic sequences for both random and smooth motions.

For each sequence, we wish to determine the true MLEs for the projective and Euclidean structures. The only known way to accomplish this is via brute–force minimization using a form of steepest descent. We used the well known Levenberg–Marquardt (LM) algorithm [4]. We use LM partly because it is available as a MATLAB routine and partly because it is a conservative

algorithm, likely to converge to a local minimum near the starting estimate—the ground truth, in most of our experiments. Like all such algorithms, LM can converge to incorrect local minima.

We start our algorithms at the ground truth to avoid these incorrect minima. For our multi-frame sequences, the overredundant information in the images should constrain the MLE to be near the ground truth; starting the descent from the ground truth should usually locate this nearby global minimum. In many cases, we explicitly checked that the minimum found starting from the ground truth was also obtained starting from other initial guesses. Typically, we find convergence to the same reconstruction near the ground truth from a large domain of initial guesses.

There is no guarantee that LM will find the true global minimum. But we are mainly concerned with setting lower bounds on the accuracy of projective reconstruction. Given the stability with which LM converges to a single minimum when started near the ground truth, there is probably no global minimum significantly closer to the ground truth.

We conducted several tests to confirm the correctness of our LM algorithms. Our algorithms were based on the LM implementation in MATLAB, modified in some versions following Hartley [4]. We checked that algorithms written separately by the two authors gave the same reconstructions.

We also checked that the reconstructions generated by the algorithms had

the correct invariances. The Euclidean algorithm computes the structure in the reference frame of the initial image. For this image we fixed $R = \mathbf{1}_3$ (where $\mathbf{1}_3$ is the identity matrix), $\mathbf{T} = \mathbf{0}$, and (to remove the scale ambiguity) maintained the z (or other homogeneous) coordinate of one of the 3D points at a fixed value during minimization⁴. We computed reconstructions using different base images and checked that the reconstructions were Euclidean similarity transforms of each other; this was true to about 1 part in $10^{5,6}$.

Projective SFM determines the reconstruction up to a projective transform. To eliminate this ambiguity, in some of our experiments we fixed the coordinates of 5 3D points at their ground truth values during the minimization⁵. We also used Hartley’s method of fixing one of the camera matrices, which removes only part of the ambiguity. These methods were compared and for stable 5–point bases gave identical results.

To check projective invariance, we computed reconstructions for a fixed 5–point basis but with its ground truth specified in different image coordinate systems. We found that as required the reconstructions were Euclidean transforms of each other, again to 1 part in $10^{5,6}$. We also reconstructed using different 5–point bases and confirmed that the different reconstruc-

⁴The coordinate chosen as fixed was varied during minimization, see below.

⁵We could use the 5–point technique for most of our experiments since the ground truth was known, and it was possible to select a stable 5–point basis. We did mainly use Hartley’s method for our stability experiments, since here we are simulating the realistic case where the ground truth is unknown.

tions were projectively equivalent⁶ (for bases which gave numerically stable minimization).

3.1 Techniques for LM Minimization

A number of techniques significantly improved the convergence speed and reliability of our LM algorithms. The improvement in reliability was particularly important for poor starting guesses. Some plausibly useful techniques did not improve our results.

Hartley has described a method for improving the speed and numerical stability of the LM algorithm [4]. We implemented this technique, but for our experiments it did not offer significant speed advantages. Though it is crucial for minimizing over large numbers of unknowns, for the relatively few (20–30) feature points used in our experiments its significant overhead prevented any substantial speedup.

Since computing the structure given the motion is easy, we explored a two-stage technique where the basic minimization was only over the motion variables⁷. This technique did not converge more quickly than minimizing in all variables simultaneously.

⁶They are no longer Euclidean equivalent due to noise.

⁷The objective function was computed on-line as a function of the motion variables only by minimizing over the structure variables. This differs from the standard method (e.g., [13]) that alternates structure and motion minimizations. It might be expected to be more stable, since in the alternating method the structure and motion minimizations may act at cross purposes.

A simple technique which did give important improvements was to scale the structure and translations so that the translations were of order 1. This sets all 4 homogeneous coordinates of a point to roughly the same scale. For projective SFM, it also ensures that all components of the camera matrices M are about the same scale. This is essentially a standard technique for improving numerical conditioning [3, 12].

A second important technique exploits the compactness of projective space by changing the coordinates during minimization. It is impossible to parameterize projective space by one nonsingular set of coordinates. If only one coordinate system is used throughout the minimization (e.g., the Euclidean structure coordinates), then the coordinates may go off to infinity during minimization, preventing convergence to the correct minimum and leading to long computation times.

We avoid this problem by changing the coordinate system during minimization. In projective reconstruction, for each 3D point and camera matrix, we fixed the largest components of the homogeneous parameterizations during the minimization. This change enhanced the reliability of convergence and produced significant speedups.

Note that using the homogeneous structure coordinates and the coordinate-switching strategy is crucial also for **Euclidean** SFM. With homogenous structure coordinates, we eliminate the usual difficulty that small image

changes lead to infinite or negative depths. Again, the coordinate-switching strategy compacts the search space: only a finite region need be searched to determine the correct structure.

The techniques described above significantly improved the convergence of LM for both the Euclidean and projective cases.

3.2 Representations of the Projective Structure

To compare the projective structure computed by Euclidean and projective methods, we need a well-defined comparison metric. Starting from an estimate of the homogeneous coordinates of the 3D points, we compute the projective transform that minimizes the least-squares Euclidean error between the transformed estimate and the Euclidean ground truth. Our measure of goodness for the projective structure is just the minimum value of this error: the projective minimum least-squares error (PMLSE)⁸. It is the smallest possible Euclidean structure error given the recovered projective structure. After transforming, the Euclidean structure coordinates give a redundant but noise-insensitive representation of the projective structure. Apart from intrinsic instabilities⁸, small changes in the image coordinates lead to small changes in the transformed coordinates. We will refer to these transformed Euclidean coordinates as the **best Euclidean representation** of the pro-

⁸If some of the 3D points are very distant, then the reconstructions of these points can have large errors, and it may be more appropriate to use a robust error measure.

jective structure.

To eliminate the ambiguity under projective transformations during minimization, we must fix the projective representation³. Because the best Euclidean representation is stable, minimizing in it would avoid numerical inaccuracy. But there is no easy way to avoid this representation’s redundancy. Thus we use a more computationally convenient 5–point basis for minimization, fixing the 5 points at their ground truth values. For stability, we look for a 5–point basis that gives a representation of the projective structure that is “close” to the best Euclidean representation.

The stability of the best Euclidean representation implies that the deviation of any 3D point from its ground–truth value should be small during the minimization (since minimization starts at the ground truth). The transformation from this representation to that of a 5–point basis is given by transforming the slightly perturbed coordinates of the selected 5 points back to their ground truth values. We can ensure that this transformation is small by selecting the 5 points so that **any** transformation between these points and slight perturbations of them is small. Thus for a given potential basis set of 5 points, we consider the projective transform taking this set to 5 other points as a function of the 15 coordinates of these other points. We compute the derivatives of the transform with respect to these 15 coordinates evaluated at the original set of 5 points. The sum of the squares of these

derivatives is our goodness measure for the basis. When it is small, the basis should inherit the stability of the best Euclidean representation.

We calculate the measure as follows. Let the basis set be $(x_a \ y_a \ z_a \ \chi_4)^T$, where $a = 1, 2, \dots, 5$. The fourth component χ_4 is a constant (set to the average scale of the structure), and the other components correspond to the Euclidean ground truth. The perturbation of this set is $(x'_a \ y'_a \ z'_a \ \chi_4)^T$. The projective transform P from the basis set to the perturbed set satisfies

$$0 = \begin{pmatrix} \chi_4 & 0 & 0 & -x'_a \\ 0 & \chi_4 & 0 & -y'_a \\ 0 & 0 & \chi_4 & -z'_a \end{pmatrix} P \begin{pmatrix} x_a \\ y_a \\ z_a \\ \chi_4 \end{pmatrix}, \quad (2)$$

for $a = 1, 2, \dots, 5$. The 15 constraints on the 15 degrees of freedom of P can be written in the form $A\mathbf{P} = 0$, where A is a 15×16 matrix, and \mathbf{P} is a 16×1 vector consisting of the elements of P rearranged. It is convenient to choose

$$\mathbf{P} = (P_{11} \dots P_{13}, P_{21} \dots P_{23}, P_{31} \dots P_{33}, P_{14} \dots P_{34}, P_{41} \dots P_{44})^T. \quad (3)$$

When the perturbed set equals the original basis set, P is the 4×4 unit identity matrix $\mathbf{1}_4$; we denote the vector corresponding to $P = \mathbf{1}_4$ as \mathbf{P}_0 , and the matrix A as A_0 . Generically, A_0 has just one zero singular value. We may assume this since otherwise the basis set will not be stable. When the perturbed set differs from the original, writing $A \equiv A_0 + \delta A$, $\mathbf{P} \equiv \mathbf{P}_0 + \delta \mathbf{P}$,

we have $(A_0 + \delta A)(\mathbf{P}_0 + \delta \mathbf{P}) = 0$; since $A_0 \mathbf{P}_0 = 0$ we get $\delta \mathbf{P} = -A_0^{-1} \delta A \mathbf{P}_0$, where A_0^{-1} is the pseudo-inverse of A_0 . To define δP precisely up to scale we require that $\delta P \cdot \mathbf{P}_0 = 0$, which follows automatically from the use of the pseudo-inverse.

For the derivatives with respect to $\mathbf{r}'_a \equiv (x'_a, y'_a, z'_a)^T$

$$\frac{\partial \mathbf{P}}{\partial \mathbf{r}'_{ai}} = -A_0^{-1} \frac{\partial A}{\partial \mathbf{r}'_{ai}} \mathbf{P}_0.$$

One can verify that with the choice of \mathbf{P} in (3), the matrix $\partial A / \partial \mathbf{r}'_{ai}$ is nonvanishing only for its last four columns. Since the only nonzero element from the last 4 in \mathbf{P}_0 is $\mathbf{P}_{0;16} = 1$, we need consider only the last column of $\partial A / \partial \mathbf{r}'_{ai}$. This is given by

$$\left[\frac{\partial A}{\partial \mathbf{r}'_{ai}} \right]_{k;16} = \chi_4 [\hat{1}_{ai}]_k,$$

where $\hat{1}_{ai}$ is a column vector with a 1 in the position corresponding to a, i and zeros elsewhere.

Then the sum of the norm squared of all derivatives is

$$\begin{aligned} \sum_{a,i} \left| \frac{\partial \mathbf{P}}{\partial \mathbf{r}'_{ai}} \right|^2 &= \chi_4^2 \sum_{a,i} [\hat{1}_{ai}]^T A_0^{-1T} A_0^{-1} [\hat{1}_{ai}] \\ &= \chi_4^2 \text{Tr} (A_0^{-1T} A_0^{-1}). \end{aligned}$$

This is our goodness measure. It is small when the basis set of 5 points is stable.

This goodness measure may be generally useful. We have verified experimentally that, as might be expected, a set of 5 points which is stable by this measure is strongly “generic,” in the sense that no subset of 4 points from among the 5 lies close to a plane, and no subset of 3 lies close to a line. The advantage of this measure over a measure based on coplanarity or colinearity is that its value has a well-defined meaning so that thresholds can be meaningfully defined.

4 Experiments

In our sequences the number of scene points N_p varied from 15 to 30 and the number of images N_i from 3 to 10. The 3D points were chosen uniformly in a volume $-15 \leq x, y \leq 15$, $-30 \leq z \leq 30$. The initial distance (Z in the Table) of the camera from the center of this volume was 60 in most experiments but was 90 in experiment 5.

For one motion scenario, we chose the camera positions randomly and uniformly in a cubic volume centered on the first camera position with side $2T_{\max}$, where T_{\max} varied from 2 to 8 for different sequences. The camera rotations were also random⁹. For each motion, we chose a vector $\mathbf{w} = (w_1, w_2, w_3)$

⁹The rotations should have a small effect on the MLE [8]. The effective FOV was less than 30° in our experiments.

uniformly with $-w_{\max} \leq w_i \leq w_{\max}$, with $w_{\max} = .2$. Given this vector, the rotation from the first image was obtained by rotating about the axis \mathbf{w} by an angle of $|\mathbf{w}|$ radians. The rotations varied up to about 20° .

To generate the images, we projected the ground points onto an image plane assuming focal length one. We added Gaussian or uniform image noise with $\sigma = .004$ (corresponding to 1 pixel noise assuming a 512×512 image). Finally, the images were shifted by an affine transform to simulate calibration error, as described previously. The center offset \mathbf{C} was a randomly chosen vector of size .12 (about 30 pixels). The focal length error δf was $\pm .1$, and the relative scaling factor r for the x and y axes was $\pm .05$ or 0.

We also used a smoother and perhaps more realistic motion. We simulated a smooth motion with up to third-order derivatives with respect to the frame index, allowing for accelerations and jerks. The parameters for the motion were chosen from a random distribution such that the motion was confined within a volume specified by $-T_{\max} \leq x, y, z \leq T_{\max}$, and the maximum rotation was less than about 20° .

The results for the projective errors are shown in figures 1—3. The parameters for these experiments are shown in the Table. These figures display histograms of the ratios of the PMLSE computed by the 3 algorithms (normalized to unit area). Recall that the PMLSE measures the goodness of reconstruction of the **projective** structure. The number of different sequences

Exp. #	# frames	# points	Depth (Z)	Trans (T_{\max})	Motion	x/y scaling (r)
1	10	20	60	8	Random	0.05
2	10	15	60	8	Random	0.05
3	10	20	60	4	Random	0.05
4	10	10	60	2	Random	0
5	10	20	90	4	Random	0
6	3	10	60	8	Smooth	0.05
7	10	20	60	8	Smooth	0

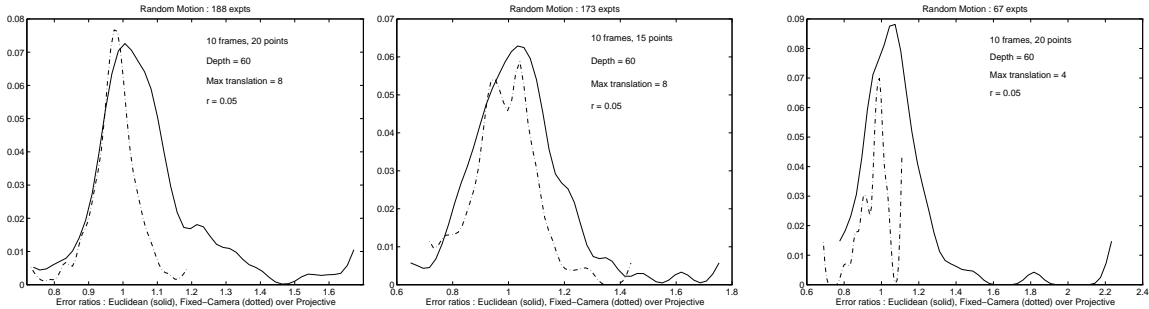


Figure 1: Experiments 1, 2 and 3: Normalized error ratio histograms. The solid and dotted line denote the ratios of Euclidean and Fixed-Camera vs. projective error respectively. Error here is the projective error.

tested for Experiments 1—7 were, respectively, 188, 173, 67, 55, 43, 42, 166.

In each experiment we “spot checked” some of the results using different algorithms, different 5 point basis sets, fixing the projective basis via the camera matrix, and starting LM with an initial guess different from the ground truth. In particular, we checked several of the projective reconstructions that gave poor reconstructions. Occasionally we did find discrepancies between reconstructions and eliminated the incorrect reconstructions, which were easy to spot because their errors were large. These incorrect reconstructions were sometimes due to our using less-than-ideal 5 point bases in

our earlier experiments. Our use of the compactness technique also helped to eliminate wrong reconstructions. Since the general trend of the data is clear, a few missed local minima cannot significantly change the results. Our convergence results in the next section also indicate this.

There is little difference among the 3 estimates of the projective structure. In agreement with our arguments, the fixed-camera estimate is always the best. The crucial point is that the Euclidean estimate is comparable to the projective estimate. It is slightly worse in experiments 1, 2, 3, but slightly better in 4, 5, and 6. In experiment 4, the translation is small, in 5, the distance of the scene from the camera is large (so that the FOV is also small), and in 6 the number of images is small. In all 3 cases, there is less information for determining the reconstruction than in 1, 2, 3.

In effect, projective SFM treats parameters that are actually known—the change in calibration from one image to the next—as additional unknowns to be estimated. In reconstruction, these parameters may be tuned away from their known correct values introducing error into the recovery of the true unknowns. This is more likely when there is relatively little information to constrain the reconstruction and may explain the results seen in these 3 experiments.

We checked this interpretation by running additional experiments with a reduced noise of $\sigma = .001$. Since in this situation the reconstruction is

better constrained, we would expect projective reconstruction to do relatively better. This is what we observed. The improvement remains marginal, however.

Note also that for experiments 4 and 5 (but not 6) there is zero error in the x/y scaling of the image axes. It is known that this scaling error affects the Euclidean reconstruction more significantly than do errors in the other calibration parameters. This also helps to explain Euclidean SFM’s relatively good performance for these experiments.

In figures 4, 5, the ratios of the Euclidean reconstruction errors (EMLSE) are shown for the Euclidean and fixed-camera estimates. The EMLSE is the minimum least-squares error between the reconstruction and ground-truth structure under all Euclidean transforms of the reconstruction to the ground truth. It is striking that the Euclidean estimate does significantly worse in estimating the Euclidean structure than does the fixed-camera estimate. This indicates that calibration error does affect the recovered Euclidean structure¹⁰.

4.1 Convergence Tests

We performed two experiments to test the convergence stability of the algorithms. For the first experiment, we created 22 sequences with the same

¹⁰Though probably not the depth.

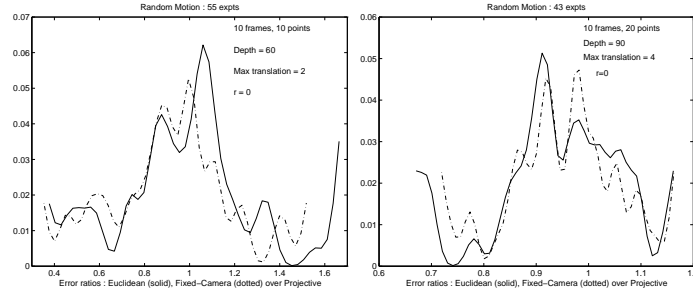


Figure 2: Experiments 4 and 5 : Normalized error ratio histograms. The solid and dotted line denote the ratios of Euclidean and Fixed-Camera vs. projective error respectively. Error here is the projective error.

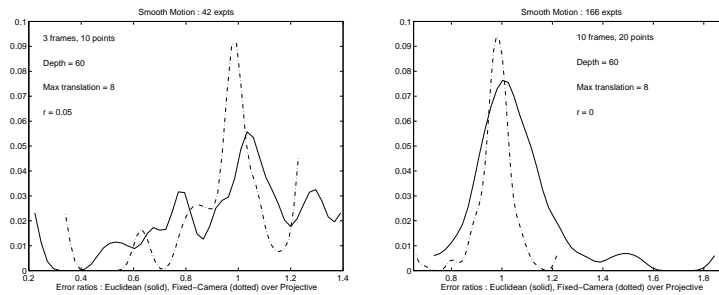


Figure 3: Experiment 6 and 7 : Normalized error ratio histograms. The solid and dotted line denote the ratios of Euclidean and Fixed-Camera vs. projective error respectively. Error here is the projective error. Note that in these experiments, the motion is smooth.

parameters as for experiment 1, but with $T_{\max} = 6$, $r = 0$.

We started the algorithm from the ground truth plus a perturbation and checked the convergence as a function of the size of the perturbation. For each sequence, only the overall scale of the perturbation was varied. The perturbation was generated in exactly the same manner as were the original structure, translations, and rotations (but the structure perturbation centers on $\mathbf{0}$).

The scales of the perturbations used were 0, .5, 1.0, 1.5, where for a scale

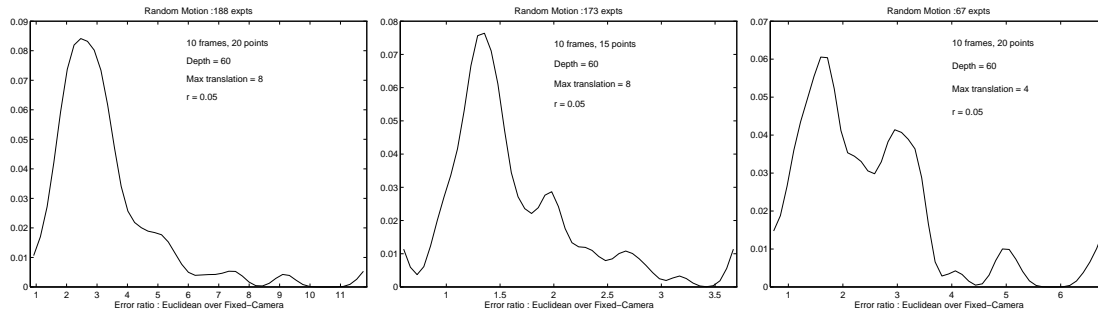


Figure 4: Experiments 1, 2 and 3 : Normalized histograms of ratios of Euclidean vs. Fixed-Camera Euclidean error.

factor of 1 the perturbation is about the same size as the original ground truth. A perturbation scaled by 1.5 is large enough that it makes some of the initial depth estimates negative.

Each of the 3 algorithms converged to the same minimum for perturbations scaled by 0, 0.5 and 1.0 in all cases, indicating that convergence to the global minimum is quite robust. For perturbations scaled by 1.5, the Euclidean algorithm failed to converge 5 times, while the projective algorithm failed just once. This implies that the Euclidean algorithm is less reliable than the projective algorithm.

These experiments were quite time consuming, with each sequence taking on the order of hours of computation time. On average, for the first 3 perturbation scales the projective algorithm required from 2 to 4 times more computation than did the Euclidean algorithm (as measured by the MATLAB flops function). The fixed-camera algorithm averaged about 3/4 of the flops used by the projective algorithm.

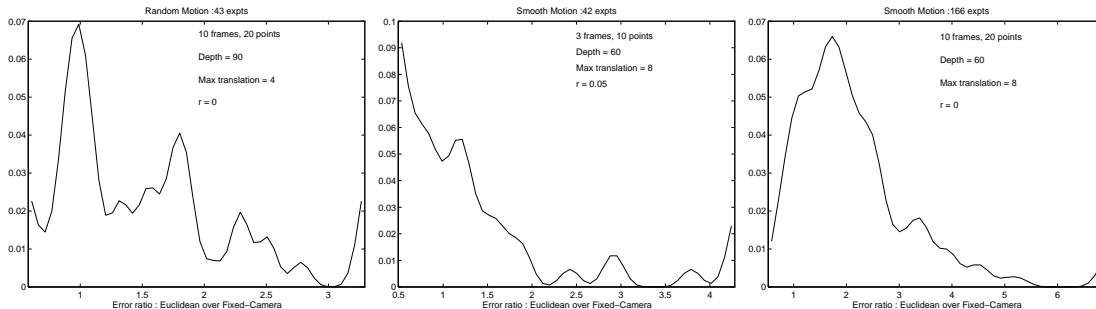


Figure 5: Experiments 5, 6 and 7 : Normalized histograms of ratios of Euclidean vs. Fixed-Camera Euclidean error.

We have observed graphically the evolution of the structure estimates over many trials of these algorithms. The structure estimates do appear to experience significant changes until convergence is nearly attained. It appears that the timings are accurate reflections of the computational cost of the algorithms.

A second experiment was run for 50 sequences under the same conditions but with $r = .05$ and $T_{\max} = 5$. The starting guesses bore no relation to the original ground truth except that they were generated similarly. The parameters used were $T_{\max} = 8$, $w_{\max} = .2$, and the structure guess was randomly chosen in a cube of side 40 centered at a distance of $Z = 50$ from the camera.

In 19 of the 50 cases, the Euclidean algorithm converged to an incorrect local minimum. Surprisingly, the typical error in recovering the projective structure for these cases was just 2 to 3 times higher than at the correct global minimum. The projective algorithm converged to the correct global minimum

in all cases. We have not fully checked the fixed-camera reconstruction, but the residual errors indicate that it too fails to converge in a number of cases.

Finally, we ran the algorithms on a real-image sequence using tracked feature points and ground truth provided to us by J. Thomas [10, 5]. One image from this sequence is shown in figure 6. There were 32 automatically tracked feature points over 16 image frames. The sequence was generated by rotating a camera attached to the end (“hand”) of a PUMA arm with length approximately 1.8 feet. Because of the small translations and large rotations this is known to be a difficult sequence. Projective SFM when started at the ground truth gave an error of 0.88 in estimating the projective structure, while the Euclidean algorithm gave an error of .95.

5 Discussion

Our experiments have shown that the difficulty of precise calibration is an inadequate motivation for projective SFM: projective SFM does not give better accuracy than Euclidean SFM when there is calibration error. Thus standard Euclidean reconstruction is a viable and simpler alternative to the projective approach for dealing with approximately calibrated single-camera sequences. Also, our experiments showed clearly that projective optimization could become unreliable without a great deal of care (e.g., in the choice of a 5-point projective basis), and the results for experiments 4, 5, and 6 are partial



Figure 6: Sample image from the “PUMA” sequence

evidence that projective SFM does worse as the information is reduced. Thus, though the experiments reveal that the MLE differs relatively little when computed in the projective or Euclidean frameworks, *non-optimal* projective algorithms are likely to be significantly less accurate and less robust than Euclidean ones, since non-optimal algorithms use just part of the available information or else weight it incorrectly [11].

On the other hand, pure optimization appears more reliable in the projective framework, or with rotation constraints relaxed in the Euclidean framework, than in the Euclidean framework enforcing the rotation constraints. The improvement is due to the reduced severity of the local-minimum problem and holds just for optimization.

References

- [1] O. Faugeras, “What can be seen in three dimensions with an uncalibrated stereo rig?” *ECCV* 1992, 563–578.

- [2] R. Hartley, “Estimation of Relative Camera Positions for Uncalibrated Cameras,” *ECCV* 1992, 579–587.
- [3] R. Hartley, “In Defence of the 8-point Algorithm,” *ICCV* 1995, 1064–1070.
- [4] R. Hartley, “Euclidean Reconstruction from Uncalibrated Views,” in *Second Workshop on Invariants*, Azores, 1993, 187–202.
- [5] R. Kumar and A.R. Hanson, “Sensitivity of the pose refinement problem to accurate estimation of camera parameters,” *ICCV* 1990, 365–369.
- [6] H. C. Longuet-Higgins, “A computer algorithm for reconstructing a scene from two projections,” *Nature*, 293:133–135, 1981.
- [7] S. Lawrence, “What Size Net Gives Optimal Generalization? Convergence Properties,” NEC TR, 1996.
- [8] J. Oliensis, “A New Structure from Motion Ambiguity,” NECI TR 1997.
- [9] J. Oliensis, “A Multi-Frame Structure from Motion Algorithm Under Perspective Projection,” submitted to *IJCV*.
- [10] J. Oliensis, “Structure from Linear of Planar Motions,” *CVPR* 1996, 335–342.

- [11] J. Oliensis, "A Critique of Structure from Motion Algorithms, panel *ICCV 98*, and <http://www.neci.nj.nec.com/homepages/oliensis.html>.
- [12] R. Szeliski and S.B. Kang, "Recovering 3D shape and motion from image streams using nonlinear least squares," **Journal of Visual Communication and Image Representation**, vol. 5, 1994, 10-28.
- [13] C.J. Taylor, D.J. Kriegman, and P. Anandan, "Structure and motion in two dimensions from multiple images: A least squares approach," *Workshop on Visual Motion*, 242-248.
- [14] C. Tomasi and T. Kanade, "Shape and motion from image streams under orthography: A factorization method," *IJCV 9*, 137-154, 1992.