Rigorous Bounds for Two–Frame Structure from Motion

J Oliensis

NEC Research Institute, Princeton NJ 08540, USA

Abstract. We analyze the problem of recovering rotation from two image frames, deriving an exact bound on the error size. With the single weak requirement that the average translational image displacements are smaller than the field of view, we demonstrate rigorously and validate experimentally that the error is small. These results form part of our correctness proof for a recently developed algorithm for recovering structure and motion from a multiple image sequence. In addition, we argue and demonstrate experimentally that in the complementary domain when the translation is large the whole motion can typically be recovered robustly, assuming the 3D points vary significantly in their depths.

1 Introduction

We have demonstrated in [12, 13] a new algorithm for structure from motion (SFM) which, in the appropriate domain, provably reconstructs structure and motion correctly. In the current paper, we present a part of the correctness proof for this algorithm, deriving tight bounds and estimates on the error in recovering rotation between two image frames when the translation is moderate. We show in [9] that these bounds typically give good estimates of the rotation error.

It has traditionally been believed that recovering rotation is difficult. We prove here that this is not true when the translation is moderate or small. Intuitively, our result is straightforward: if the translational displacements of image points are not large, then merely aligning the fields of view (FOV) of two images gives reasonable rotation estimates. We make this intuition precise in this paper.

We also argue and show experimentally that in the complementary large translation domain, if we additionally assume that perspective effects are important, then the complete motion and structure typically can be determined reliably. Experimentally, it seems that the domain where "small translation" techniques reliably determine the rotation overlaps the domain where "large translation" techniques work.

Our experimental work on the large translation domain suports our previous claim [13, 12] that this domain is essentially an easy one: any SFM algorithm, including the notoriously unreliable "8–point" [6] algorithm, typically works well in this domain. In addition, we demonstrate that often most of the error in recovering the Euclidean structure is due to a single structure component.

2 Rotation Error Bound

In this section we derive an explicit bound on the error in recovering the relative rotation between two noisy images **neglecting a nonzero translation between the two images**. The derived bound goes beyond a first order estimate; it does not require that the translation and noise be infinitesimally small.

Let the first image be represented by unit rays $\mathbf{p}_i$ and the second by unit rays $\mathbf{p}'_i$. Let there be N_p points in both images. We assume that the rotation is determined by finding R minimizing the least square error for the unit rays

$$\sum_i |\mathbf{p}'_i - \mathrm{R}\mathbf{p}|^2, \tag{1}$$

which is equivalent to maximizing

$$\sum_i \mathbf{p}'^t_i \ \mathrm{R}\mathbf{p}_i. \tag{2}$$

The rotation minimizing (1) can be found by standard techniques, e.g., using the SVD. (1) is a least square error in the image ray orientations rather than in the image plane coordinates (as is usual). For moderate FOV this makes little difference, and the ray–based error may actually better represent typical noise distributions. Probably, the rotations computed from (1) will be close to those computed by any standard algorithm and the bounds derived here will be relevant to such algorithms.

Let R_T represent the true rotation between the two images. Define $\mathrm{R} \equiv \mathrm{R}_E\mathrm{R}_T$ (R_E is the error rotation) and $\bar{\mathbf{p}}_i \equiv \mathrm{R}_T\mathbf{p}_i$. R_E is the rotation maximizing

$$\sum_i \mathbf{p}'^t_i \ \mathrm{R}_E\bar{\mathbf{p}}_i. \tag{3}$$

Let the unit vector $\hat{\omega}$ define the axis and θ the angle of rotation for R_E. $|\theta|$ measures the size of the error in the recovered rotation. We will derive a bound on $|\theta|$ starting from (3).

Proposition 1. *Define the* 3×3 *symmetric matrix* M *by*

$$M \equiv \frac{1}{N_p}(\mathbf{1}_3 \sum_i \mathbf{p}'^t_i \ \bar{\mathbf{p}}_i - \frac{1}{2}(\mathbf{p}'_i \ \bar{\mathbf{p}}^t_i + \bar{\mathbf{p}}_i \ \mathbf{p}'^t_i)), \tag{4}$$

where $\mathbf{1}_3$ *is the identity matrix, and define the 3 vector*

$$\mathbf{V} \equiv \frac{1}{N_p}\sum_i \mathbf{p}'_i \times \bar{\mathbf{p}}_i = \frac{1}{N_p}\sum_i (\mathbf{p}'_i - \bar{\mathbf{p}}_i) \times \bar{\mathbf{p}}_i. \tag{5}$$

Assume M *is positive definite. Then* $|\theta| \leq \left|M^{-1}\mathbf{V}\right|$.

Remark. As we discuss below, M will be positive definite when the translation and noise are sufficiently small. The noise is usually insignificant.

Proof. (3) can be decomposed as

$$\sum_i \mathbf{p}_i'^t \ \mathrm{R}_E \bar{\mathbf{p}}_i = \sum_i (\ \hat{\omega}^t \mathbf{p}_i')\ (\ \hat{\omega}^t \bar{\mathbf{p}}_i)$$
$$+ \sum_i (\ \hat{\omega} \times \mathbf{p}_i')^t \ \ (\cos\theta(\ \hat{\omega} \times \bar{\mathbf{p}}_i) - \sin\theta(\ \hat{\omega} \times (\ \hat{\omega} \times \bar{\mathbf{p}}_i)))\,, \quad (6)$$

where the first term on the right hand side involves the rotation only through $\hat{\omega}$ and is θ–independent. The coefficient of $\cos(\theta)$ in the second term is proportional to $\hat{\omega}^t M \hat{\omega}$ and is nonvanishing since M is positive definite. The θ_ω giving the maximum can be solved for explicitly:

$$\tan(\theta_\omega) = \frac{\sum_i (\ \hat{\omega} \times \mathbf{p}_i')^t \ \bar{\mathbf{p}}_i}{\sum_i (\ \hat{\omega} \times \mathbf{p}_i')^t \ (\ \hat{\omega} \times \bar{\mathbf{p}}_i)} \equiv \frac{A(\hat{\omega})}{B(\hat{\omega})}, \quad (7)$$

where

$$A(\hat{\omega}) \equiv \frac{1}{N_p} \sum_i (\ \hat{\omega} \times \mathbf{p}_i')^t \ \bar{\mathbf{p}}_i, \qquad B(\hat{\omega}) \equiv \frac{1}{N_p} \sum_i (\ \hat{\omega} \times \mathbf{p}_i')^t \ (\ \hat{\omega} \times \bar{\mathbf{p}}_i). \quad (8)$$

Substituting $\tan(\theta_\omega)$ back into (6) yields $N_p(A^2(\hat{\omega}) + B^2(\hat{\omega}))^{1/2}$ for the second term at the maximum. The first term in (6) can be rewritten as

$$\sum_i (\ \hat{\omega}^t \mathbf{p}_i')\ (\ \hat{\omega}^t \bar{\mathbf{p}}_i) = \sum_i \mathbf{p}_i'^t \ \bar{\mathbf{p}}_i - N_p B(\hat{\omega}). \quad (9)$$

Thus after the substitution for $\tan(\theta_\omega)$, (3) becomes up to rotation independent terms and an irrelevant factor of N_p

$$E(\ \hat{\omega}) \equiv \sqrt{A^2(\ \hat{\omega}) + B^2(\ \hat{\omega})} - B(\ \hat{\omega}) = B(\sqrt{1 + (A/B)^2} - 1). \quad (10)$$

We also have $B(\hat{\omega}) = \ \hat{\omega}^t M \ \hat{\omega}$ and $A(\hat{\omega}) = \mathbf{V}^t \ \hat{\omega}$. where M, $\mathbf{V}$ are the matrix and vector defined in the statement of the theorem.

Define $X(\hat{\omega}) \equiv (1/2)\ A^2(\hat{\omega})/B(\hat{\omega})$. As we discuss below, when the translation and noise are small $|A/B|$ will in general also be small and $E(\ \hat{\omega}) \approx X(\ \hat{\omega})$. Thus, as a first step toward maximizing E, we consider maximizing $X(\ \hat{\omega})$.

Lemma 2. *Let $\hat{\omega}_M$ be the value of the unit vector $\hat{\omega}$ maximizing $X(\hat{\omega})$. Assume M is positive definite. Then the ratio $|A(\hat{\omega}_M)/B(\hat{\omega}_M)| = |M^{-1}\mathbf{V}|$, and $\theta_M \equiv \tan^{-1}(A(\hat{\omega}_M)/B(\hat{\omega}_M))$ satisfies $|\theta_M| \leq |M^{-1}\mathbf{V}|$.*

Proof. Since M is real symmetric and positive definite, we can write $M = M^{1/2}M^{1/2}$ where $M^{1/2}$ is real symmetric and positive definite. Let $\boldsymbol{\omega}' \equiv M^{1/2}\ \hat{\omega}$ and $V' \equiv M^{-1/2}V$. Maximizing X over unit vectors $\hat{\omega}$ is equivalent to maximizing $(\hat{\omega}' \cdot \mathbf{V}')^2$, where $\hat{\omega}' \equiv \boldsymbol{\omega}'/|\boldsymbol{\omega}'|$, over the ellipsoid $\boldsymbol{\omega}'^t M^{-1} \boldsymbol{\omega}' = 1$. Clearly, the maximum occurs at one of the two points on the ellipsoid where $\boldsymbol{\omega}'$ is parallel to V', corresponding to a value for $\hat{\omega}$ of $\hat{\omega}_M \sim M^{-1/2}\mathbf{V}' = M^{-1}\mathbf{V}$. Substituting this value into the expressions for $A(\hat{\omega}_M)$, $B(\hat{\omega}_M)$ yields $|A(\hat{\omega}_M)/B(\hat{\omega}_M)| = |M^{-1}\mathbf{V}|$ and since $|\theta_M| \leq |\tan(\theta_M)|$, it follows that $|\theta_M| \leq |M^{-1}\mathbf{V}|$.

We now return to the maximization of the exact expression $E(\hat{\omega})$. In terms of $X(\hat{\omega})$ and $\rho(\hat{\omega}) \equiv A(\hat{\omega})/B(\hat{\omega})$,

$$E(\hat{\omega}) = \frac{2X(\hat{\omega})}{\rho^2(\hat{\omega})}(\sqrt{1+\rho^2(\hat{\omega})} - 1) \equiv X(\hat{\omega})\ K(\rho^2(\hat{\omega})), \tag{11}$$

where $K(x) \equiv 2(\sqrt{1+x}-1)/x$ is a monotonically decreasing function for $x \geq 0$. Recall that $\hat{\omega}_M \sim M^{-1}\mathbf{V}$ gives the maximum value X_M of $X(\hat{\omega})$. Let $\rho_M \equiv \rho(\hat{\omega}_M)$; the lemma showed that $|\rho_M| = |M^{-1}\mathbf{V}|$. The only way to achieve a value $E(\hat{\omega})$ larger than $E(\hat{\omega}_M) = X_M K(\rho_M^2)$ is via a value of $\hat{\omega}$ such that $K(\rho^2(\hat{\omega})) > K(\rho_M^2)$. But since $K(x)$ is monotonic decreasing this implies $|\rho(\hat{\omega})| < |\rho_M|$. Since $|\theta(\hat{\omega})| \leq |\tan(\theta(\hat{\omega}))| = |\rho(\hat{\omega})|$ we have the desired result: the error in the rotation is bounded by $|\theta| < |M^{-1}\mathbf{V}|$.

2.1 Estimates on the Rotation Bound

In the remainder of this section we discuss the expected magnitude of the derived bound and the conditions under which it is valid—that is, the conditions under which M is positive definite. More detail can be found in [11], where we focus especially on how the rotation error is affected by the FOV size.

Let $\delta\mathbf{p}_i \equiv \mathbf{p}'_i - \bar{\mathbf{p}}_i$. Recall that $\delta\mathbf{p}_i$ reflects the displacement due to noise and translation only since the rotation has been compensated exactly in the definition of $\bar{\mathbf{p}}_i$. We derive crude but simple bounds on the eigenvalues of M and on $|M^{-1}\mathbf{V}|$ in terms of the $\delta\mathbf{p}_i$. For extensions see [11].

Using the notation $\langle x \rangle \equiv \sum_i x_i/N_p$, where i varies over all N_p points,

$$|\mathbf{V}| = \frac{1}{N_p}|\sum_i(\delta\mathbf{p}_i \times \bar{\mathbf{p}})| \ \leq\ \langle|\delta\mathbf{p}|\rangle, \tag{12}$$

$$B(\hat{\omega}) = \frac{1}{N_p}\left(\sum_i |\ \hat{\omega} \times \bar{\mathbf{p}}_i|^2 + (\ \hat{\omega} \times \delta\mathbf{p}_i)^t(\ \hat{\omega} \times \bar{\mathbf{p}}_i)\right). \tag{13}$$

In (13), the second term is also bounded by $\langle|\delta\mathbf{p}|\rangle$,

$$\left|\frac{1}{N_p}\sum_i(\ \hat{\omega} \times \delta\mathbf{p}_i)^t(\ \hat{\omega} \times \bar{\mathbf{p}}_i)\right| \leq \frac{1}{N_p}\sum_i |\delta\mathbf{p}_i \cdot \bar{\mathbf{p}}_i| \leq\ \langle|\delta\mathbf{p}|\rangle, \tag{14}$$

while the first term is positive for arbitrary $\hat{\omega}$ (assuming the image points $\bar{\mathbf{p}}_i$ are not all collinear). We denote the first term of (13) by $B_0(\hat{\omega})$:

$$B_0(\hat{\omega}) = \frac{1}{N_p}\sum_i |\ \hat{\omega} \times \bar{\mathbf{p}}_i|^2 =\ \hat{\omega}^t M_0\ \hat{\omega}, \tag{15}$$

where

$$M_0 \equiv \mathbf{1}_3 - \frac{1}{N_p}\sum_i \bar{\mathbf{p}}_i\bar{\mathbf{p}}_i^t. \tag{16}$$

Denote the eigenvalues of M_0 by e_i and those of M by e'_i, ordered from greatest to least. e_3 depends just on the initial observed image points and is typically of order 1 if the FOV is moderate or large [11]. Elementary linear algebra implies

$$B(\hat{\omega}) \equiv \hat{\omega}^t M \, \hat{\omega} \; \geq e'_3 \geq \; e_3 - \langle |\delta \mathbf{p}| \rangle. \tag{17}$$

If $\langle |\delta \mathbf{p}| \rangle$ is sufficiently small, $e_3 - \langle |\delta \mathbf{p}| \rangle > 0$ and $B(\hat{\omega}) > 0$ for arbitrary $\hat{\omega}$, implying that M is positive definite. Moreover, $e_3 - \langle |\delta \mathbf{p}| \rangle > 0$ implies

$$|\theta| \leq \left| M^{-1} \mathbf{V} \right| \leq \frac{|\mathbf{V}|}{e'_3} \leq \frac{\langle |\delta \mathbf{p}| \rangle}{e_3 - \langle |\delta \mathbf{p}| \rangle}. \tag{18}$$

When $e_3 \approx 1$ and $\langle |\delta \mathbf{p}| \rangle \ll 1$, the rotation error (in radians) is bounded approximately by $\langle |\delta \mathbf{p}| \rangle$, the average displacement of an image point due to the translation or noise. Recall that $\delta \mathbf{p}_i$ is a difference of unit vectors; $\langle |\delta \mathbf{p}| \rangle \approx 1$ would imply a large average translational shift in the image of about 45°.

2.2 An Improved Estimate

(18) is a pessimistic bound on $\left| M^{-1} \mathbf{V} \right|$ since the overlap of $\mathbf{V}$ with the least eigendirection of M—the one corresponding to e'_3—is typically small. We can improve this bound based on the propositions below (proofs omitted). Propositions similar to these appear in [2] and [16] but we obtain slightly better bounds since they apply to the **least** eigenvalues.

Definition 3. Let H', H be $N \times N$ symmetric matrices, and let H be positive definite. Define the perturbation $\Delta H \equiv H' - H$. Let e'_i, e_i be the eigenvalues respectively of H' and H, ordered from greatest to least, and let $\mathbf{E}'_i$ and $\mathbf{E}_i$ be the corresponding unit eigenvectors. Define $P_{\perp N}$ to be the projection operator $P_{\perp N} \equiv \mathbf{1}_N - \mathbf{E}_N \mathbf{E}^t_N$, where $\mathbf{1}_N$ is the $N \times N$ identity matrix. Define

$$D \equiv P_{\perp N} \left(H' - \mathbf{1}_N (e_N + \Delta H_{NN}) \right) P_{\perp N}, \tag{19}$$

where $\Delta H_{NN} \equiv \mathbf{E}^t_N \; \Delta H \; \mathbf{E}_N$, and also define $\Delta H_{\perp\perp} \equiv P_{\perp N} \Delta H P_{\perp N}$. Let a be a lower bound on the eigenvalues of the matrix D (excluding the zero eigenvalue associated with $\mathbf{E}_N$) and let $b \equiv |P_{\perp N} \Delta H \mathbf{E}_N|$. Finally, define

$$B_2(\xi) = \frac{\sqrt{1 + 4\xi^2} - 1}{2} \leq \xi^2.$$

Proposition 4 Eigenvalue bound. *Assume that* $a > 0$. *Then the perturbation of the least eigenvalue is bounded by*

$$\Delta H_{NN} \geq e'_N - e_N \geq \Delta H_{NN} - a B_2(b/a). \tag{20}$$

Definition 5. With the definitions above and assuming that $\mathbf{E}'_i \cdot \mathbf{E}_i \neq 0$, define $\delta \mathbf{E}_i$ by

$$\mathbf{E}'_i \equiv \frac{\mathbf{E}_i + \delta \mathbf{E}_i}{|\mathbf{E}_i + \delta \mathbf{E}_i|}, \qquad\qquad \mathbf{E}_i \cdot \delta \mathbf{E}_i = 0.$$

Proposition 6 Eigenvector bound. *Assume that $a > 0$. Then*

$$|\delta \mathbf{E}_N| \leq \frac{2b}{a}.$$

Based on these propositions, a detailed theoretical analysis [11] shows that, even for moderate to small FOV,

1. M is positive definite and the derived bound is valid unless the average image displacement due to the translation or noise is comparable to the FOV.
2. Larger rotation errors can be expected for translations parallel to the image plane and for 3D points lying on a plane approximately perpendicular to the image plane.
3. The rotation error due to neglecting the translation is bounded roughly by the ratio of the average translational image displacement to the FOV.

2.3 Experiments

We have experimentally verified these theoretical expectations. The results of our synthetic experiments appear in Table 1, where each entry summarizes 100 trials. For each trial, feature points were selected randomly within the specified FOV (shown in the Table), and values of $|\mathbf{T}|/r_i$ (the inverse radial depth scaled by the translation magnitude) were chosen randomly varying uniformly from 0 up to the input maximum value. (Note that using the uniform distribution for the **inverse** depths gives a stringent test of our claims. Typically, the radial **depths** r_i are distributed uniformly; assuming this instead for the inverse depths causes the 3D points to cluster at small depths, with probability density $P(r_i) \sim 1/r_i^2$.) A second image was generated from the first using the specified translation and randomly computed structure; then Gaussian image noise with 2 pixel standard deviation was added to each image point in the second image.

The maximum computed rotation error over all trials and all cases is 23.4° $\sim$ .41 radians. In our structure from motion algorithm [13, 12], the first order effects of rotation error are eliminated. Thus the effect of rotation error on the estimates of structure and motion are in the worst case scaled by $.41^2 \approx 17\%$. As shown in the last column, expecially at moderate to large FOV the upper bound B is a very good estimate of the actual rotation error $|\theta|$. It can be shown that a lower bound on $|\theta|$ exists which at such FOV is very close to the upper bound. Since for 3D planar scenes the rotation errors are expected to be relatively large, we also ran experiments explicitly for this case. The results were similar to those reported in the Table.

We also report results on two real image sequences: the rocket field sequence [1, 17, 10] and the PUMA sequence [14, 15] (Figure 7b). The rocket sequence consists of nine images obtained by approximately forward motion in an outdoor environment. The maximum translation magnitude, between the first and ninth image, was 7.4 feet, while the depth of the nearest point relative to the first (farthest) camera position was 17.8 feet. The field of view was about 45°. We

N_{pts}	FOV°	T	$\max(\lvert T\rvert/r_i)$	range($\lvert\theta\rvert$)	range(B)	$\bar{m}, \sigma(B)$	range($\lvert\theta\rvert/B$)
20	90	$(1\,0\,0)$	.3	5.5°–10.8°	5.5–11.0	8.4°, 1.2°	.984–.997
20	40	$(1\,0\,0)$	.3	6.3–12.5	6.3–13.3	9.3, 1.4	.94–.996
20	20	$(1\,0\,0)$	.3	5.6–19.0	5.7–28.3	11.5, 4.0	.62–.995
20	20	$(1\,0\,1)$	.3	4.5–13.7	4.5–18.1	8.8, 2.4	.75–.995
20	20	$(1\,0\,2)$	.3	3.3–10.9	3.3–11.6	5.6, 1.5	.92–.998
20	20	$-(1\,0\,2)$	.3	2.3–7.7	2.3–8.1	4.1, 1.2	.93–.9995
20	40	$(1\,0\,0)$	.5	9.7–18.1	9.8–22.9	15.3, 2.6	.78–.989
20	30	$(1\,0\,0)$	.5	9.5–23.4	9.6–45.0	17.8, 5.3	.39–.988
20	30	$(1\,0\,1)$	.5	7.9–20.5	7.9–31.0	15.0, 3.5	.66–.992
20	30	$-(1\,0\,1)$	.5	6.0–14.9	6.0–16.6	9.4, 1.9	.88–.995
40	30	$(1\,0\,1)$	.5	10.2–17.6	10.3–21.0	14.1, 2.1	.81–.988
40	30	$(1\,0\,0)$	.5	10.8–17.8	10.9–23.5	15.7, 2.6	.73–.986
20	30	$(1\,0\,2)$	.5	5.8–16.4	5.8–20.1	10.2, 2.1	.81–.994

Table 1. Measured rotation error computed by SVD and bound $B \equiv M^{-1}\mathbf{V}$. Each entry summarizes the results of 100 trials.

computed the rotations neglecting the translation. The rotation was recovered with errors increasing nearly linearly with the translation size from .33° to 4.3°.

The PUMA sequence consists of 16 images of 32 points, with predominantly rotational motion of the camera at the rate of about 4° per image. The maximum translation was 1.5 feet, while the depth of the nearest point to the first camera position was 13.4 feet. The maximum rotation from the first image was 60.5°. The FOV was about 40°. Neglecting the translation, the rotation was recovered with error increasing almost perfectly linearly with the translation magnitude. The error varied from .37° to a maximum of 4.3°.

3 Large Translations

We now turn to the case when the translation is of the same order or large compared to the distances of the 3D points. We argue and demonstrate experimentally that if the feature points vary significantly in depth then typically both the rotation and translation can be recovered accurately[1]. The structure can also be recovered accurately except for very distant points.

The reasons why the structure and motion can be recovered robustly for large translation from 2 images are straightforward. Most importantly, because of the significant depth variation, it is easy to distinguish translations from rotations. For a rotation, the image displacements depend smoothly on the image point positions, while for translations the displacements for near and far points are very different and typically uncorrelated with image position. A large translation amplifies these differences allowing easy disambiguation. This leads to accurate recovery of the entire motion since it is well known that the rotation/translation

[1] Unless the 3D points lie close to a critical surface.

ambiguity is the main error source. With accurate motion, the structure computation is also typically robust; the nonlinearities in computing the depth are minimized because for large translations image rays from the two images tend to intersect at large angles.

We conducted a set of 100 synthetic experiments to test these claim. In each trial, 20 3D points were chosen randomly in a 40° FOV, with depths varying uniformly from 2 to 40 in units of the translation size. The translation was always in the x–direction (translation parallel to the image plane is traditionally believed to be a difficult case), and the rotation was chosen randomly up to a maximum of about 25°. The image noise was Gaussian with a σ of one pixel. The results are displayed in the figures. The angular error in the translation direction as computed by the "8-point" algorithm is shown in Figure 1a, while an improved error as computed by Levenberg–Marquardt starting from the result of the "8–point" algorithm is shown in Figure 1b. Even though the "8–point" algorithm is notorious for its unreliability, it reconstructs the translation direction accurately when the translation is large. Similar results for the rotation error (in degrees) are shown in Figures 2a and 2b.

Figures 3a and b show 3D histograms of the number of points reconstructed for given values of the depth and depth error; the results for the percentage depth error appear in Figure 4a[2]. Clearly the error is small for depths less than about 20 units.

The results for the depths are actually better than indicated. It can be shown that the structure component that is typically least accurately recovered is the constant component of the **inverse** depths (e.g., [5, 8, 7, 9]). In Figures 4b—6a, this component has been corrected to the ground truth for each of the 100 trials and the histogram shows the residual depth error. In Figures 6b and 7a, the fraction of points of a given depth with depth errors greater than 10% are shown before and after correction of the constant component.

Finally, we computed the rotation directly assuming that the translation is zero as in the earlier sections. Since the minimum 3D depth in these experiments is 2, the "large translation" domain considered here actually overlaps the "small translation" domain previously discussed, where the maximum inverse depth was .5. The maximum and average rotation errors obtained were 12° and 6.4°. These experiments can be considered successfully as either large translation or small translation.

We have also obtained similar results for translations in the directions (1 0 1), (1 0 2), and (0 0 1). In these cases the single constant component accounts for less of the error (as expected [9])) but the motion is recovered more accurately.

Lastly, we ran trials similar to those above but using the ground truth structure obtained from the real PUMA image sequence (Figure 7b and [14, 15]). The results for the "8–point" algorithm are shown in Table 2, where each entry summarizes 500 trials. 1 pixel uniform noise was added to each image point, and

[2] Note that in these histograms the bins corresponding to the largest depth error also include all hits with errors larger than the nominal value.

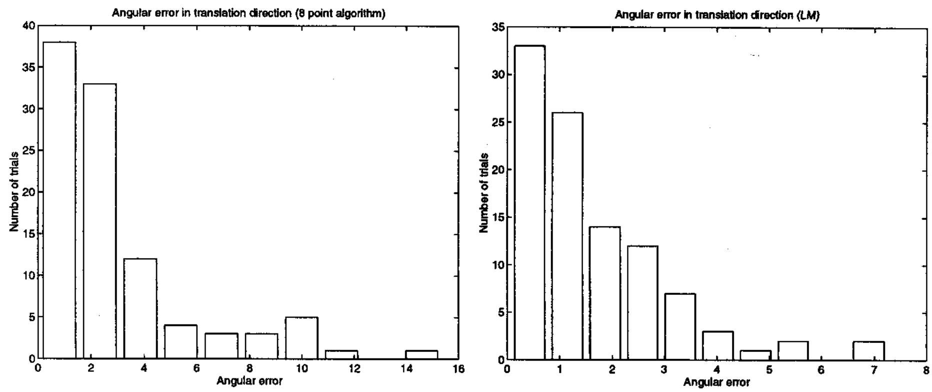


Fig. 1. a, b

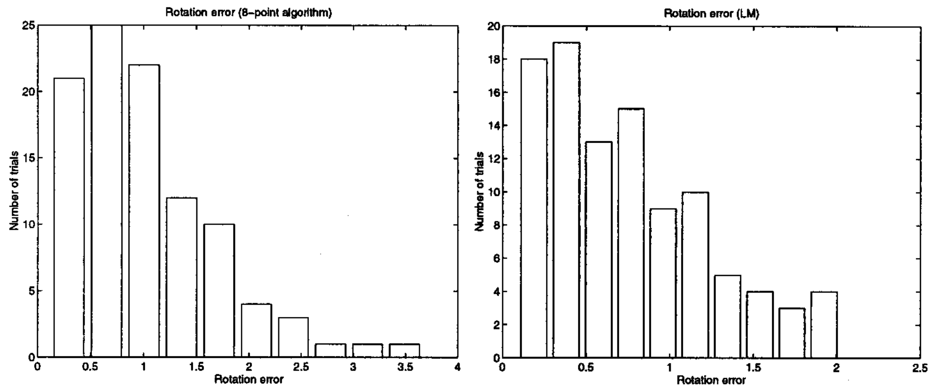


Fig. 2. a, b

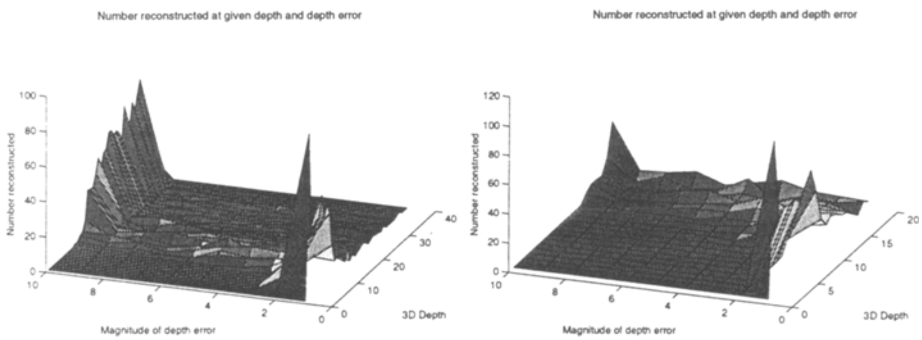


Fig. 3. a, b

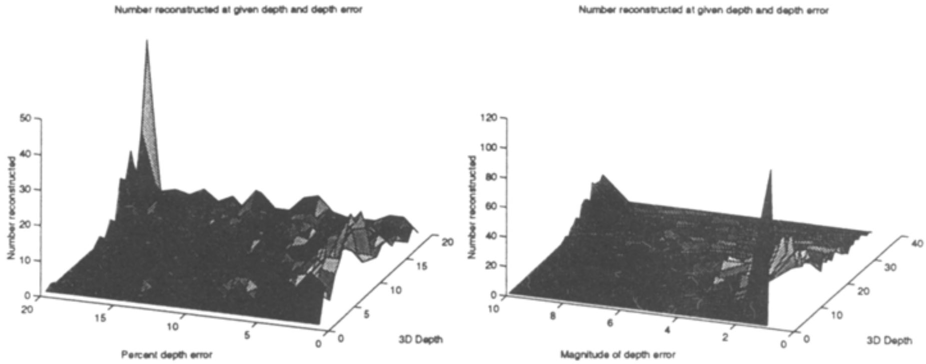


Fig. 4. a, b

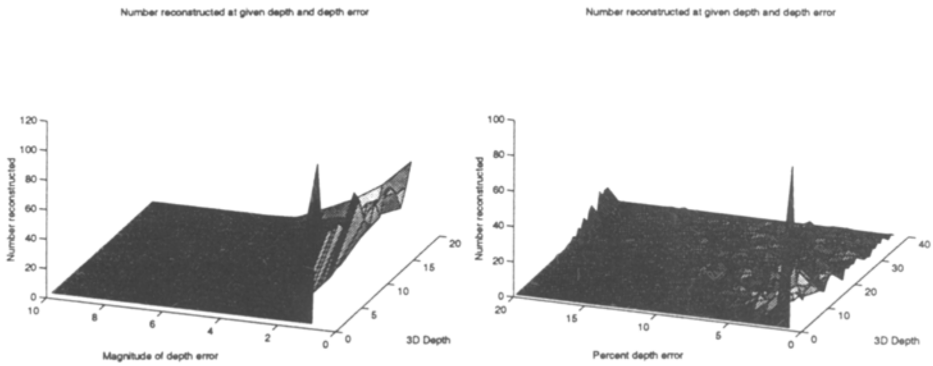


Fig. 5. a, b

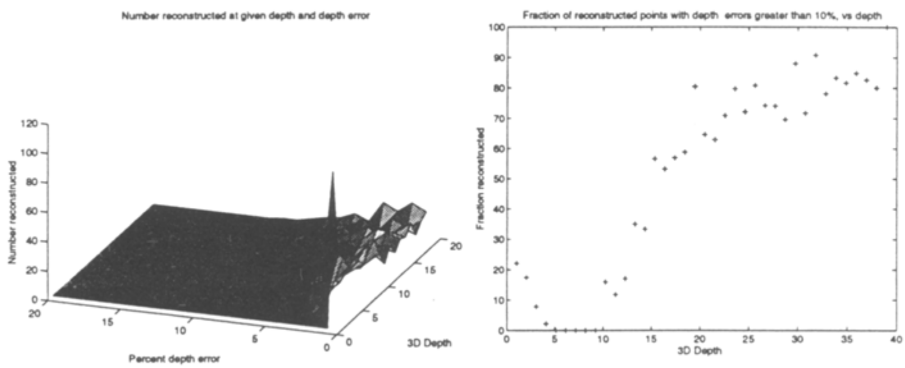
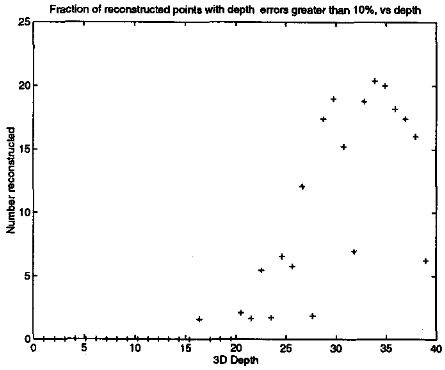


Fig. 6. a, b



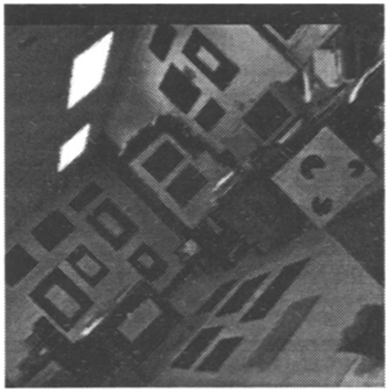


Fig. 7. a, b

the translation was selected randomly under the constraints indicated.

The results of Levenberg–Marquardt starting from the 8–point estimates on 250 trials with $Tz = 0$ and $|T| = 12$ were: the translation direction was determined with an average angular error of 1.6°, standard deviation of 1.2°, and maximum angular error of 8.6°. Also, the average depth error (in feet), its standard deviation, and the maximum depth error were recorded for each trial. The averages (and standard deviations over trials σ_T) of these results over the 250 trials were .3 feet ($\sigma_T = .2$), .2 feet ($\sigma_T = .1$), and .7 feet ($\sigma_T = .4$), and their maximum values were 1.3 feet, .6 feet, and 2.5 feet. Thus the largest depth error for any feature point over all 250 trials was 2.5 feet. Clearly, these results support our claim that the depth can be recovered accurately from just two image frames when the translational displacement is large. These experiments were repeated for forward motion. As expected, for forward motion the translation is better determined than for sideways motion. The depths are determined with only slightly less accuracy, presumably because the depths of points near the FOE are recovered relatively badly.

References

1. R. Dutta, R. Manmatha, L.R. Williams, and E.M. Riseman, "A data set for quantitative motion analysis," *CVPR*, 159-164, 1989.
2. G. Golub and C. F. Van Loan, *Matrix Computations*, John Hopkins Press, Baltimore, Maryland,1983.
3. D.J. Heeger and A.D. Jepson, "Subspace methods for recovering rigid motion I: Algorithm and implementation," **IJCV** 7, 95-117, 1992.
4. R. Hummel and V. Sundareswaran, "Motion parameter estimation from global flow field data," **PAMI** 15, 459-476, 1993.
5. A.D. Jepson and D.J. Heeger, "Subspace methods for recovering rigid motion II: Theory," University of Toronto Technical Report RBCV-TR-90-36, 1990.
6. H. C. Longuet–Higgins, "A computer algorithm for reconstructing a scene from two projections," *Nature*, 293:133–135, 1981.
7. S. Maybank, *Theory of Reconstruction from Image Motion*, Springer, Berlin, 1992.

Table 2. Two Frame Motion Recovery for Large Translation Using the 8–point algorithm. The angular error in determining the translation direction is displayed (in degrees) for different magnitudes of the input 3D translation. The mean of the angular error, its standard deviation, the maximum angular error, and the number of trials with errors respectively greater than 10° and 20° are shown. The angular error quoted is $\min(|\Delta\theta|, |180° - \Delta\theta|)$ since both T and $-T$ are correct solutions.

	Baseline (ft)	4	6	8	10	12	14	16
	Mean($\|\Delta\theta\|$)	41	14	7	6	5	5	4
	$\sigma_{\|\Delta\theta\|}$	25	12	6	5	4	3	3
$T_z = 0$	max($\|\Delta\theta\|$)	90	81	32	29	20	19	14
	$N_{>10}$	437	253	130	84	45	37	27
	$N_{>20}$	370	114	17	5	1	0	0
	Mean($\|\Delta\theta\|$)	22	7	4	3	2		
	$\sigma_{\|\Delta\theta\|}$	18	6	4	4	2		
Forward T	max($\|\Delta\theta\|$)	89	42	23	20	19		
	$N_{>10}$	369	106	39	21	5		
	$N_{>20}$	191	19	3	1	0		
	Mean($\|\Delta\theta\|$)	32	18	12	9	8	7	7
	$\sigma_{\|\Delta\theta\|}$	16	10	7	5	5	4	4
Backward T	max($\|\Delta\theta\|$)	88	86	42	35	25	22	22
	$N_{>10}$	476	411	287	210	171	123	112
	$N_{>20}$	393	174	66	22	14	3	6

8. S. Maybank, "A Theoretical Study of Optical Flow," Doctoral Dissertation, University of London, 1987.
9. J. Oliensis, in preparation.
10. J. Oliensis, "Provably Correct Algorithms for Multiframe Structure from Motion: Constant Translation Direction," *ECCV* 1996.
11. J. Oliensis, "Rigorous Bounds for Two–Frame Structure from Motion," NEC TR October 1995, http://www.neci.nj.nec.com/homepages/oliensis.html.
12. J. Oliensis, "A Linear Solution for Multiframe Structure from Motion," *IUW*, 1225–1231, November 1994.
13. J. Oliensis, "Multiframe Structure from Motion in Perspective," *Workshop on the Representations of Visual Scenes*, 77–84, June 1995.
14. H. S. Sawhney, J. Oliensis, and A. R. Hanson, A "Description and Reconstruction from Image Trajectories of Rotational Motion", *ICCV*, 494-498, 1990.
15. H.S. Sawhney and A.R. Hanson, "Comparative results of some motion algorithms on real image sequences," *IUW*, 307-313, 1990.
16. G.W. Stewart and J. G. Sun, *Matrix Perturbation Theory*, Academic Press, Boston, 1990.
17. J. Inigo Thomas, A. Hanson, and J. Oliensis, "Refining 3D reconstructions: A theoretical and experimental study of the effect of cross–correlations", **CVGIP:IU**, Vol. 60, 359–370, 1994.