# The Least-Squares Error for Structure from Infinitesimal Motion

John Oliensis

Computer Science Department
Stevens Institute of Technology
Hoboken, NJ 07030, USA
`oliensis@cs.stevens-tech.edu`
`http://www.cs.stevens-tech.edu/~oliensis/`

**Abstract.** We analyze the least–squares error for structure from motion (SFM) with a single infinitesimal motion ("structure from optical flow"). We present approximations to the noiseless error over two, complementary regions of motion estimates: roughly forward and non–forward translations. Experiments show that these capture the error's detailed behavior over the entire motion range. They can be used to derive new error properties, including generalizations of the bas–relief ambiguity. As examples, we explain the error's complexity for epipoles near the field of view; for planar scenes, we derive a new, double bas–relief ambiguity and prove the absence of local minima. For nonplanar scenes, our approximations simplify under reasonable assumptions. We show that our analysis applies even for large noise, and that the projective error has less information for estimating motion than the calibrated error. Our results make possible a comprehensive error analysis of SFM.

## 1  Introduction

A structure–from–motion (SFM) algorithm has two tasks: *matching* the 3D features across different images, and *estimating* the camera motion and 3D structure. This paper reports progress toward a comprehensive analysis of *estimation*.

Under standard assumptions, the goal of an "optimal" estimation algorithm is to find the minimum of the least–squares image–reprojection error [8], and the shape of this error as a function of the estimates determines the intrinsic problem that the algorithm solves. Here, we analyze this shape for SFM with a single infinitesimal motion ("structure from optical flow").

Little is known about the least–squares error. Yet, without understanding it, one can't predict when algorithms will succeed or fail—for instance, when bundle adjustment [24] will find the optimal least–squares estimate rather than a bad estimate at a false local minimum. Given some understanding, algorithms can avoid local minima and compute estimates more reliably, as shown in [18][3].

Previous research on estimation (as opposed to geometry) in SFM focussed on the bas–relief ambiguity [1][4][22][10] [14] [7][18][3][23][9][6]. Other results include the proof in [3] that the error is singular when the epipole estimate coincides

with an image point, and a semi–quantitative description [18] of the error over a linear slice through the plane of all epipole estimates. None of this work comes close to giving a detailed picture of the least–squares error.

In this paper, we present approximations to the noiseless error over two, complementary regions of motion estimates: roughly forward and non–forward translations. Together, these approximations describe the whole error. They reproduce its detailed shape, yet are simple enough to be useful for understanding it. We believe that they make it possible to study the least–squares error in depth, and we illustrate this by deriving several new properties of the error.

As in many previous analyses, e.g., [7][14][3], our theoretical discussion assumes infinitesimal motion and zero noise. Experiments show that the theory also works for large noise. We study calibrated cameras, taking the focal length as 1 without loss of generality, and also present results for projective SFM. For lack of space, all proofs are omitted. They can be found in [17].

## 1.1    Preliminaries

The standard least–squares error for infinitesimal motion (or optical flow) is [14]

$$
E_{\mathrm{LS}}\left(\mathbf{T}, \omega, \{Z\}\right) \equiv \sum_{m=1}^{N_p} \left| \mathbf{d}_m - Z_m^{-1}\left(T_z \mathbf{p}_m - [T_x; T_y]\right) - \sum_{a \in \{x,y,z\}} \omega_a \mathbf{r}^{(a)}\left(\mathbf{p}_m\right) \right|^2 .
$$

(1)

Here $N_p$ is the total number of scene points, $\mathbf{p}_m \equiv \mathbf{p}_{1m} \equiv (x_m; y_m)$ is the $m$th image point in the first image, $\mathbf{d}_m \equiv \mathbf{p}_{2m} - \mathbf{p}_{1m}$ is the $m$th measured flow from image 1 to 2, the $Z_m$ are the 3D depth estimates, $\mathbf{T}$ is the translation estimate, $\omega \equiv \left(\omega^{(x)}; \omega^{(y)}; \omega^{(z)}\right)$ is the estimate of the infinitesimal rotation, and the $\mathbf{r}^{(x)}(\mathbf{p})$, $\mathbf{r}^{(y)}(\mathbf{p})$, $\mathbf{r}^{(z)}(\mathbf{p})$ are the rotational flows at the image point $\mathbf{p}$ due to unit rotations around the $x$, $y$, or $z$ axes: $\mathbf{r}(\mathbf{p}) \equiv$

$$
\left[\mathbf{r}^{(x)}(\mathbf{p}), \mathbf{r}^{(y)}(\mathbf{p}), \mathbf{r}^{(z)}(\mathbf{p})\right] \equiv \left[\begin{pmatrix} -xy \\ -\left(1+y^2\right) \end{pmatrix}, \begin{pmatrix} 1+x^2 \\ xy \end{pmatrix}, \begin{pmatrix} -y \\ x \end{pmatrix}\right] \in \Re^{2\times 3}.
$$

(2)

We study an effective error $E(\mathbf{e}) \equiv \min_{\{Z\},\omega} E_{\mathrm{LS}}\left(\mathbf{T}, \omega, \{Z\}\right)$, with $\mathbf{e}$ the epipole.

**Definition 1.**

*Define the cross–product for vectors* $\mathbf{v}$, $\mathbf{v}' \in \Re^2$ *by* $\mathbf{v} \times \mathbf{v}' \equiv v_x v_y' - v_y v_x'$.
*Define the error vector* $\epsilon \in \Re^{N_p}$ *by*

$$
\epsilon_m(\mathbf{e}) \equiv \frac{\mathbf{p}_m - \mathbf{e}}{|\mathbf{p}_m - \mathbf{e}|} \times \mathbf{d}_m.
$$

(3)

*Define the 3 rotational contributions to* $\epsilon$, $\Psi^{(a)}(\mathbf{e}) \in \Re^{N_p}$, $a \in \{x, y, z\}$: *Let* $\Psi_m^{(a)} \equiv ((\mathbf{p}_m - \mathbf{e})/|\mathbf{p}_m - \mathbf{e}|) \times \mathbf{r}^{(a)}(\mathbf{p}_m)$ *and* $\Psi(\mathbf{e}) \equiv \left[\Psi^{(x)}, \Psi^{(y)}, \Psi^{(z)}\right] \in \Re^{N_p \times 3}$. *Define the projection* $\Pi(\mathbf{e}) \equiv \mathbf{1}_{N_p} - \Psi(\mathbf{e})\left(\Psi^T(\mathbf{e})\Psi(\mathbf{e})\right)^{-1}\Psi^T(\mathbf{e}) \in \Re^{N_p \times N_p}$, *where* $\mathbf{1}_{N_p}$ *denotes the* $N_p \times N_p$ *identity matrix.*

**Proposition 1.** *[19][20]  Assume the candidate epipole* $\mathbf{e}$ *does not coincide*[1] *with any image point* $\mathbf{p}_m$. *Then*

$$E\left(\mathbf{e}\right) \equiv \min_{\{Z\},\omega} E_{\mathrm{LS}}\left(\mathbf{T}, \omega, \{Z\}\right) = \epsilon^T\left(\mathbf{e}\right) \Pi\left(\mathbf{e}\right) \epsilon\left(\mathbf{e}\right). \tag{4}$$

*Remark 1.* The definition of $\Pi\left(\mathbf{e}\right)$ shows that it cancels the rotational contributions to $\epsilon$. Thus, for noiseless data $E\left(\cdot\right)$ does not depend on the value of the true rotation $\omega_{\mathrm{true}}$, and we are free to take $\omega_{\mathrm{true}} = \mathbf{0}$ in analyzing it. We do this for the rest of the paper, without loss of generality.

For noiseless images, we get a more explicit expression for $E\left(\mathbf{e}\right)$ by substituting the ground truth for the flow $\mathbf{d}_m$ into the result of **Proposition 1**:

**Proposition 2.** *Assume* $\forall m$, $\mathbf{e} \neq \mathbf{p}_m$, *as in Proposition 1. Then*

$$E\left(\mathbf{e}\right) = T_{\mathrm{true},z}^2 \sum_m \left(\hat{\Delta}_{m,\mathbf{e}}^\perp \cdot \left(Z_m^{-1}\mathbf{1}_2 - \mathbf{r}_m\Omega\right)\left(\mathbf{e} - \mathbf{e}_{\mathrm{true}}\right)\right)^2, \tag{5}$$

with $$\Omega \equiv \left(\sum_n \mathbf{r}_n^T\left(\hat{\Delta}_{n,\mathbf{e}}^\perp \hat{\Delta}_{n,\mathbf{e}}^{\perp T}\right)\mathbf{r}_n\right)^{-1} \left(\sum_m Z_m^{-1}\mathbf{r}_m^T\left(\hat{\Delta}_{m,\mathbf{e}}^\perp \hat{\Delta}_{m,\mathbf{e}}^{\perp T}\right)\right) \in \Re^{3\times 2},$$
$$\tag{6}$$

*and* $\Delta_{m,\mathbf{e}} \equiv \mathbf{p}_m - \mathbf{e}$, $\hat{\Delta}_{m,\mathbf{e}} \equiv \left(\mathbf{p}_m - \mathbf{e}\right)/\left|\mathbf{p}_m - \mathbf{e}\right|$, $\mathbf{v}^\perp \equiv \left(-v_y; v_x\right)$.

*Remark 2.* Each summand in (5) is proportional to $\left|\mathbf{e} - \mathbf{e}_{\mathrm{true}}\right|^2$, so $E\left(\mathbf{e}\right)$ is continuous at $\mathbf{e} = \mathbf{e}_{\mathrm{true}}$. This gives a direct proof of the result of [3].

## 2    Forward Motion: e in or near the Image

We first analyze $E\left(\mathbf{e}\right)$ for candidate epipoles in or near the image, with $\left|\mathbf{e}\right| \stackrel{\sim}{<} \theta_{\mathrm{FOV}}/2$ radians, where $\theta_{\mathrm{FOV}}$ gives the angular extent of the image points. We refer to this as the *forward region*. The true epipole is *not* constrained.

*Previous results.* For $\mathbf{e}$ near the image points, [18][3] show that $E\left(\mathbf{e}\right)$ typically is complex and has local minima. Also, [3] proved: $E\left(\mathbf{e}\right)$ is singular when $\mathbf{e}$ coincides with an image point and $\mathbf{e} \neq \mathbf{e}_{\mathrm{true}}$; $E\left(\mathbf{e}\right)$ is continuous at $\mathbf{e} = \mathbf{e}_{\mathrm{true}}$.

The singularity is not enough to explain the minima: The error can be singular at an image point and yet behave smoothly a short distance away (Figure 1d). To explain them, one must understand what causes the singular effects to extend far from the image points, so that effects from different points can interact.

To state this another way, the error's singularity at an image point reflects the known $\sin^2\theta$ dependence on the angle between the hypothesized epipolar direction and the observed translational flow. Thus, the singularity at the image point comes from a known property of the error around an image point and does not give a new explanation of the error's behavior around the point.

---

[1] If it does, the expression for $\min_{\{Z\},\omega} E_{\mathrm{LS}}\left(\mathbf{T}, \omega, \{Z\}\right)$ must be modified slightly. Strictly, our formulas in $\mathbf{e}$ aren't valid at $T_{\mathrm{true},z} = 0$, but they are easily extended.
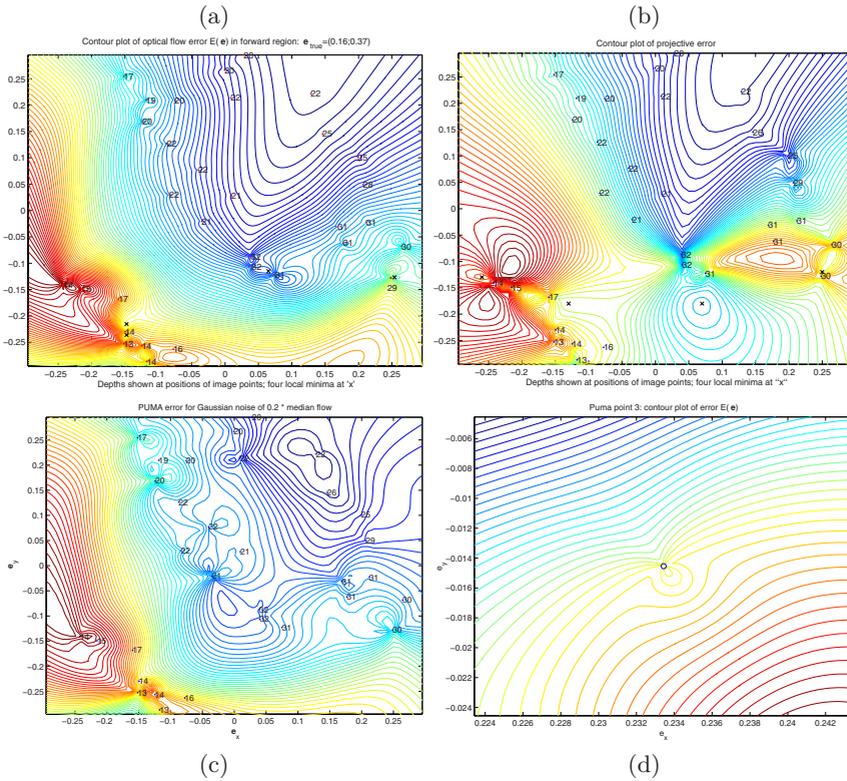
**Fig. 1.** Contour plot of error for **e** in the field of view. The structure comes from PUMA [12] and $\mathbf{e}_{\text{true}} = (0.16; 0.37)$. The 3D depths are shown at the image points. (a): $E(\mathbf{e})$, with 4 minima at 'x'; (b): Projective error, with 4 marked minima; (c): $E(\mathbf{e})$ for noisy images, with $\sigma_{\text{noise}} = 0.2 d_{\text{med}}$ (see Fig. 7); (d) Closeup of $E(\mathbf{e})$ around an image point, showing that the singularity quickly becomes invisible.

## 2.1   Forward Analysis

*Remark 3.* The singularity of $E(\mathbf{e})$ at an image point $\mathbf{p}_k$ causes it to have two local minima on an infinitesimal circle around the point. In a region where $E(\mathbf{e})$ behaves smoothly, it has a single minimum on an infinitesimal circle. Thus, we analyze $E(\mathbf{e})$, and its minima, on small circles centered on the image points. Let $\rho_k$ be the radius of the circle $\mathcal{C}_k$ around $\mathbf{p}_k$. A particular limit turns out to give a useful approximation ($E_{\text{near}}$ in Proposition 3).

**Definition 2 (Near–point limit).** *Define the limit* $\rho_k \xrightarrow{\text{near}} 0$ *by* $(\rho_k \longrightarrow 0, N_p \longrightarrow \infty$ *for fixed* $\rho_k N_p$ *and* $\theta_{\text{FOV}})$, *where we stipulate that image–point sums* $\sum_{m=1}^{N_p}$ *are* $\Theta(N_p)$ *as* $N_p \longrightarrow \infty$.

*Remark 4.* For real images, the stipulation amounts to assuming that sums over the image points have no unexpected cancellations. This holds unless the image points cluster near a line or the 3D depths have a few outliers at small depths.

**Proposition 3** $(E_{\text{near}})$. *In the limit* $\rho_k \xrightarrow{\text{near}} 0$, *we have the asymptotic estimate* $E(\mathbf{e}) \approx E_{\text{near}}(\mathbf{e}) + O\left(\rho_k, N_p^{-1}\right)$ *on* $\mathcal{C}_k$, *where* $E_{\text{near}}(\mathbf{e}) \equiv T_{\text{true},z}^2 |\mathbf{p}_k - \mathbf{e}_{\text{true}}|^2 \times$

$$
\left( cN_p + (\rho_k N_p) \left( \frac{\mathbf{a}}{|\mathbf{p}_k - \mathbf{e}_{\text{true}}|} + N_p^{-1} \sum_{m \neq k} \frac{\tilde{\mathbf{a}}_m}{|\mathbf{p}_k - \mathbf{p}_m|} \right) \hat{\Delta}_{k,\mathbf{e}} - \hat{\Delta}_{k,\mathbf{e}}^T Q \hat{\Delta}_{k,\mathbf{e}} \right).
$$
(7)

*The* $O(1)$ $c \in \Re$, $\mathbf{a}, \tilde{\mathbf{a}}_m \in \Re^{1 \times 2}$, $Q \in \Re^{2 \times 2}$ *don't depend on* $\mathbf{e}$, $|\mathbf{p}_k - \mathbf{e}_{\text{true}}|$, $\rho_k N_p$.

We rewrite our approximation as $E_{\text{near}} = \gamma + \alpha \cos(\theta - \phi_1) + \beta \cos^2(\theta - \phi_2)$, where $\alpha$, $\beta$ give the linear and quadratic terms in (7), and $(\cos\theta; \sin\theta) \equiv \hat{\Delta}_{k,\mathbf{e}}$.

**Lemma 1.** *Let* $f(\theta) \equiv a \cos^2(\theta - \phi_1) + \cos(\theta - \phi_2)$. *For any values of* $\phi_1$, $\phi_2$, *the function* $f(\theta)$ *has one minimum for* $|a| < 1/2$ *and two minima for* $|a| > 1$.

Thus, the value of $|\beta/\alpha|$ determines how many minima $E_{\text{near}}$ has on the circle $\mathcal{C}_k$ and, from Remark 3, the rate of decrease of $|\beta/\alpha|$ with $\rho_k$ (i.e., with the distance from $\mathbf{p}_k$) determines how far the singular effects due to $\mathbf{p}_k$ extend.

*Experiments.* We compared $E_{\text{near}}$ with the true error's behavior for 1200 synthetic flows generated from real structures. We measured the singularity of $E(\mathbf{e})$ on a circle by: the number of its local minima, and the ratio of its second fundamental (3rd Fourier coefficient) to its standard deviation. This second measure indicates the singularity's size. Figures 2a,b verify that all but a small fraction (3%) of the one–minimum results have $|\beta/\alpha| \leq 1$, and all but 1.7% of the two–minimum results have $|\beta/\alpha| \geq 1/2$. Figure 2c shows that the "size" of the singularity grows roughly linearly with $|\beta/\alpha|$ until it saturates. These results demonstrate that our analysis predicts the error's behavior very well.

One can use $E_{\text{near}}$ to understand the factors causing the error's complexity [17]. Figure 3 confirms our predictions from (7): the error behaves smoothly near image points close to $\mathbf{e}_{\text{true}}$ (Fig. 3a), and is more likely to have a complex behavior near an isolated image point (Fig. 3b) or one with extreme 3D depth (Fig. 3c). Also, experiments show that the fraction of "singular" results decreases roughly like $N_p^{0.5}$, and the size of the singular fluctuations in the error decreases roughly like $N_p^{-1}$, in agreement with the behavior of $|\beta/\alpha|$, see [17].

## 3    Sideways Motion: $|\mathbf{e}| > \theta_{\text{FOV}}/2$

*Preliminaries.* Define $A \equiv T_{\text{true},z}(\hat{\mathbf{e}} \times \mathbf{e}_{\text{true}})$, $B \equiv T_{\text{true},z}(1 - \hat{\mathbf{e}} \cdot \mathbf{e}_{\text{true}}/|\mathbf{e}|)$, with $\hat{\mathbf{e}} \equiv \mathbf{e}/|\mathbf{e}|$. The $A$, $B$ capture all dependence on the true translation $\mathbf{T}_{\text{true}}$.

It is convenient to use an image coordinate system that rotates around the image center with the candidate epipole $\mathbf{e}$ such that $\mathbf{e} = (|\mathbf{e}|; 0)$.
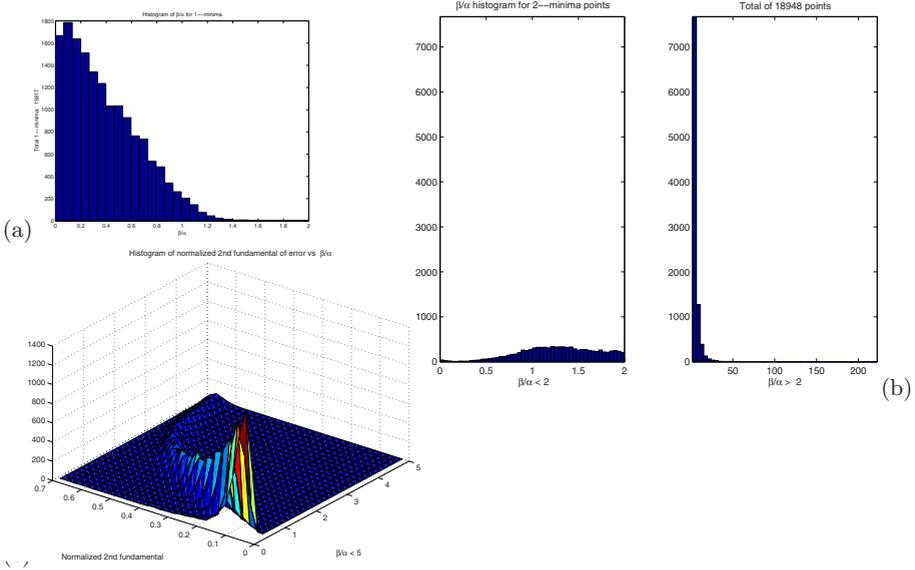
**Fig. 2.** Histograms of two measures of the complexity of $E(\mathbf{e})$. (a) $|\beta/\alpha|$, one minimum results; (b) Two minima results, separately for $|\beta/\alpha| < 2$ and $|\beta/\alpha| > 2$. (c) Normalized second fundamental of $E(\mathbf{e})$ for $|\beta/\alpha| < 5$.
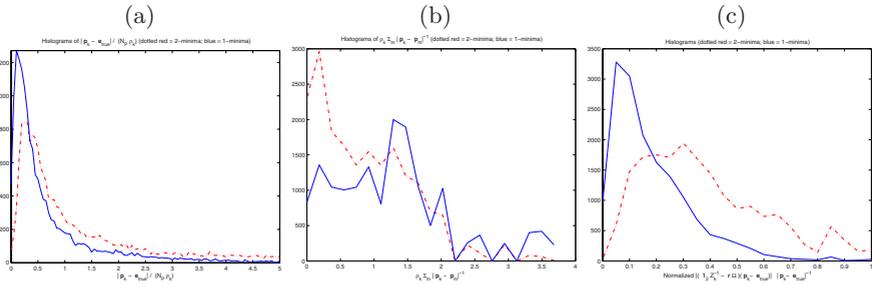


**Fig. 3.** Histograms, plotted separately for circles $\mathcal{C}_k$ where $E(\mathbf{e})$ had two minima (dotted curves) and one minimum. (a) Epipolar–distance ratio $(N_p\rho_k)^{-1}|\mathbf{p}_k - \mathbf{e}_{\text{true}}|$; (b) Isolation measure $\rho_k \sum_{m \neq k} |\mathbf{p}_m - \mathbf{p}_k|^{-1}$; (c) 3D depth ratio $|\mathcal{Z}_k \hat{\mathbf{d}}_{\text{true}}| / \max_m |\mathcal{Z}_m \hat{\mathbf{d}}_{\text{true}}|$, where $\mathcal{Z}_k \equiv (Z_k^{-1}\mathbf{1}_2 - \mathbf{r}_k\Omega)$ and $\hat{\mathbf{d}}_{\text{true}} \equiv (\mathbf{p}_k - \mathbf{e}_{\text{true}})/|\mathbf{p}_k - \mathbf{e}_{\text{true}}|$

We represent the inverse depths $Z_m^{-1}$ as a sum of a *linear component* and a *nonlinear component*. We write $Z_m^{-1} \equiv n_z + n_x x_m + n_y y_m + Z_{\text{NL},m}^{-1}$, where $Z_{\text{NL},m}^{-1}$ is the nonlinear and $Z_{L,m}^{-1} \equiv n_z + n_x x_m + n_y y_m$ is the linear component of the structure, and where we define these components uniquely from

$$0 = \sum_m Z_{\text{NL},m}^{-1} = \sum_m x_m Z_{\text{NL},m}^{-1} = \sum_m y_m Z_{\text{NL},m}^{-1}. \tag{8}$$

We refer to $\mathbf{n} \equiv (n_x; n_y; n_z)$ above as the *planar normal*, since $Z_m^{-1} = n_z + n_x x_m + n_y y_m$ for a planar scene neglecting noise. We define the *planar epipole* $\underline{\mathbf{n}} \equiv (n_x; n_y)/n_z$ by analogy with the epipole $\mathbf{e}$, and $\tilde{n}_z \equiv n_z - n_x/|\mathbf{e}|$.

**Definition 3 (Limit of zero field of view (FOV)).** *Let $\theta_{\mathrm{FOV}}$ be the angular extent of the region spanned by the image points. We define the zero–FOV limit by writing the image points as $\mathbf{p}_m = \lambda_{\mathrm{FOV}} \mathbf{p}_m^*$ and taking $\lambda_{\mathrm{FOV}} \longrightarrow 0$ keeping the $\mathbf{p}_m^*$ and $Z_m^{-1}$ fixed. We denote the limit by $\theta_{\mathrm{FOV}} \longrightarrow 0$ or $\lambda_{\mathrm{FOV}} \longrightarrow 0$.*

The classical result is on the bas–relief ambiguity [7][14][15].

**Theorem 1 (Jepson/Heeger/Maybank (JHM)).** *Assume the image points do not lie on a line, and that $\mathbf{e}$ is finite and $|\mathbf{e}| > 0$. In the limit of zero field of view, the noiseless least–squares error for infinitesimal motion is given by* $E(\mathbf{e}) = T_{\mathrm{true},z}^2 (\hat{\mathbf{e}} \times \mathbf{e}_{\mathrm{true}})^2 \sum_{m=1}^{N_p} Z_{\mathrm{NL},m}^{-2}.$

*Remark 5 (Limitations of the JHM Theorem).*

*The JHM result models the error only for $\theta_{\mathrm{FOV}}^{-1} \gg |\mathbf{e}| \gg \theta_{\mathrm{FOV}}$ and does not capture any of the error's dependence on $e \equiv |\mathbf{e}|$ (it cannot be used to analyze the minima); It gives no information about the error on the line $\mathbf{e} = t\mathbf{e}_{\mathrm{true}}, \quad t \in (-\infty, \infty)$—despite the fact that the true epipole lies on it; It says nothing about the error for $|\mathbf{e}_{\mathrm{true}}| \sim O(\theta_{\mathrm{FOV}})$; It says nothing about the error for planar scenes or the effect on $E$ from the linear scene component, which is always important.*

### 3.1   New Analysis

**Definition 4 (Sideways limit).** *The sideways limit $e \overset{sideways}{\longrightarrow} \infty$ is defined by*

$$\left( \lambda_{\mathrm{FOV}} \longrightarrow 0, e \longrightarrow \infty \text{ for fixed } \kappa \equiv e\lambda_{\mathrm{FOV}}, \mathbf{p}_m^*, A, B, \tilde{n}_z, n_y, Z_{\mathrm{NL},m}^{-1} \right),$$

*where the zero–FOV limit $\lambda_{\mathrm{FOV}} \longrightarrow 0$ and the $\mathbf{p}_m^*$ are given in Definition 3.*

**Theorem 2 (Main Theorem).** *The approximation $E_{\mathrm{side}}(\mathbf{e})$ in (13) gives an asymptotic estimate of $E$ in the sideways limit:*

$$E(\mathbf{e}) - E_{\mathrm{side}}(\mathbf{e}) = O(E/e) \quad \left( e \overset{sideways}{\longrightarrow} \infty \right). \tag{9}$$

*Remark 6.* The sideways limit fixes $\tilde{n}$ and $B \equiv T_{\mathrm{true},z}(1 - e_{\mathrm{true},x}/e)$ which depend on $e$. We do this since we want $E_{\mathrm{side}}(\mathbf{e})$ to remain a good approximation when $|\underline{\mathbf{n}}|$ and $|\mathbf{e}_{\mathrm{true}}|$ are as large as $e$, and since this simplifies our approximation and makes it display the two–fold ambiguity of SFM for planes [14].

To derive $E_{\text{side}}(\mathbf{e})$, we neglect effects suppressed by factors of $\theta_{\text{FOV}}$ and $\theta_{\text{FOV}}/|\mathbf{e}|$. First, we must "pre-subtract" the leading rotational contribution from $\epsilon$ in (3). This is necessary since $E(\cdot)$ is given by a subtraction of two terms (due to the rotation cancellation from $\Pi$) and we need to ensure that its leading dependence comes from the leading dependencies of the individual terms. After this "pre–subtraction," $\Pi$ must be replaced by $\Pi_\perp$, where the latter annihilates the remaining subspace of rotational contributions. Define

$$L_\Pi^{a,b} \equiv \left\{ \frac{e}{|\mathbf{p}-\mathbf{e}|}(y^a - \langle y^a\rangle)\right\}^T \Pi_\perp^{(\text{side})} \left\{\frac{e}{|\mathbf{p}-\mathbf{e}|}(y^b - \langle y^b\rangle)\right\}, \tag{10}$$

$$Z_\Pi^{a,b} \equiv \left\{ \frac{e}{|\mathbf{p}-\mathbf{e}|}y^a Z_{\text{NL}}^{-1}\right\}^T \Pi_\perp^{(\text{side})} \left\{\frac{e}{|\mathbf{p}-\mathbf{e}|}(y^b - \langle y^b\rangle)\right\}, \tag{11}$$

$$\tilde{Z}_\Pi^{a,b} \equiv \left\{ \frac{e}{|\mathbf{p}-\mathbf{e}|}y^a Z_{\text{NL}}^{-1}\right\}^T \Pi_\perp^{(\text{side})} \left\{\frac{e}{|\mathbf{p}-\mathbf{e}|}y^b Z_{\text{NL}}^{-1}\right\}, \tag{12}$$

where $\Pi_\perp^{(\text{side})}$ equals $\Pi_\perp$ evaluated in the sideways limit, $\{V\}$ denotes a vector in $\Re^{N_p}$ with entries $V_m$, $\alpha \equiv An_y - B\tilde{n}_z$, $\beta \equiv Bn_y + A\tilde{n}_z$. Then

$$
\begin{aligned}
E_{\text{side}} \equiv\ & \alpha^2 L_\Pi^{(1,1)} + \beta^2 L_\Pi^{(2,2)} - 2\alpha\beta L_\Pi^{(1,2)} \\
& -2A\beta Z_\Pi^{(0,2)} + 2A\alpha Z_\Pi^{(0,1)} + 2B\beta Z_\Pi^{(1,2)} - 2B\alpha Z_\Pi^{(1,1)} \\
& +A^2 \tilde{Z}_\Pi^{(0,0)} + B^2 \tilde{Z}_\Pi^{(1,1)} - 2AB\tilde{Z}_\Pi^{(1,0)}.
\end{aligned} \tag{13}
$$

[17] gives explicit formulas for the $L_\Pi^{a,b}$, $Z_\Pi^{a,b}$, and $\tilde{Z}_\Pi^{a,b}$ in terms of the image and structure moments

$$\mu_{a,b} \equiv e^2 \sum_m x_m^a y_m^b / |\mathbf{p}_m - \mathbf{e}|^2, \qquad \sigma_{a,b} \equiv e^2 \sum_m Z_{\text{NL},m}^{-1} x_m^a y_m^b / |\mathbf{p}_m - \mathbf{e}|^2,$$

$$S_{a,b} \equiv e^2 \sum_m \left(x_m^a y_m^b Z_{\text{NL},m}^{-2}\right) / |\mathbf{p}_m - \mathbf{e}|^2.$$

*Discussion.* Our result nicely separates the dependencies on the various parameters. For example, $E_{\text{side}}$ depends on $\mathbf{e}_{\text{true}}$ only through the quantities $A$ and $B$. It depends on $e$ just through $B$, $\tilde{n}_z$, and the dot products $L_\Pi^{a,b}$, $Z_\Pi^{a,b}$, and $\tilde{Z}_\Pi^{a,b}$, where the first two are linear in $e^{-1}$ and the dot products can be approximated by simple ratios of quadratic expressions in $e$. One can easily read off from our formulas which contributions dominate at small FOV. For planar scenes (or the linear scene contribution), all the structure/motion unknowns appear in the leading factors; the $L_\Pi^{a,b}$ depend only on the known image coordinates. $E_{\text{side}}$ can be shown to respect the planar two–fold ambiguity.

Our result depends on the nonlinear part of the scene through the structure moments $\sigma_{a,b}$, for $2 \le a + b \le 3$, and on $S_{c,d}$. Thus, the error's dependence on the scene can be approximated using 15 parameters to describe the scene. Just 6 are usually enough. Our expression for $E_{\text{side}}$ often simplifies dramatically. This is because our approximation works for many types of scenes and motions, and we can often neglect most of the terms for a particular scene/motion.
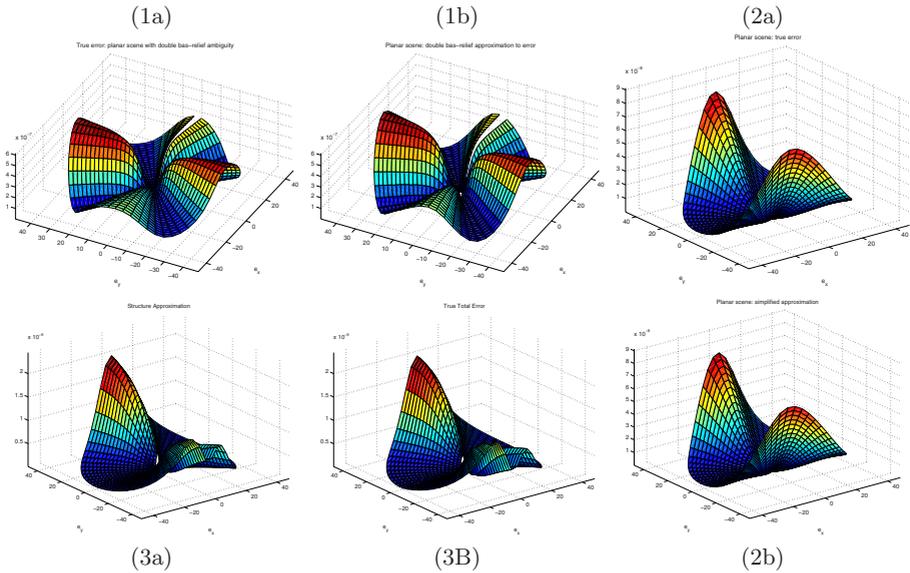
(1a) (1b) (2a)



(3a) (3B) (2b)

**Fig. 4.** Sideways error. (1): Planar example with double bas–relief ambiguity, $\mathbf{e}_{\text{true}} = (-6.9; 7.2)$ $\underline{\mathbf{n}} = (0.62; 0.64)$; (2): Planar example showing lack of minima, $\mathbf{e}_{\text{true}} = (-0.69; 0.72)$ $\underline{\mathbf{n}} = (0.81; 0.74)$; (3): Rocket structure [5], $\mathbf{e}_{\text{true}} = (-0.14; -0.045)$. (a): True $E(\mathbf{e})$; (b): Simple planar approximation $\alpha^2 L_{\Pi}^{(1,1)}$; (3B): Simple approx. (15).

### 3.2 Some Examples of Consequences

*Planar Scene, Non–forward true motion; large planar slant.* Assume an image pair with $\theta_{\text{FOV}} \ll 1$ (small FOV), $|\mathbf{e}_{\text{true}}| \gg 1$ (sideways true motion), and $|\underline{\mathbf{n}}| \equiv |(n_x; n_y)/n_z| \gg 1$ (large slant). We assume $e \gg 1$, excluding $e < |\mathbf{e}_{\text{true}}|$ and $e < |\underline{\mathbf{n}}|$ (large–$e$ assumption). Then $E(\mathbf{e}) \approx \alpha^2 L_{\Pi}^{(1,1)}$, where $\alpha^2 \approx T_{\text{true},z}^2 (\hat{\mathbf{e}} \times \mathbf{e}_{\text{true}})^2 (\hat{\mathbf{e}} \times \underline{\mathbf{n}})^2$ and $L_{\Pi}^{(1,1)}$ is roughly constant. The error is small on two lines—a *double* bas–relief ambiguity. Figure 4(1) illustrates this new effect.

*Planar Scene, Non-sideways true motion; small planar slant.* Assume $\theta_{\text{FOV}} \ll 1$ (small FOV), $1 \ll e$ (large $e$), $|\mathbf{e}_{\text{true}}| \ll 1$ (forward true motion), and $|\underline{\mathbf{n}}| \ll 1$ (small slant). Under these conditions, our approximation has no false local minima in a region $e \geq e_{\text{thresh}}$, where $e_{\text{thresh}} \sim O(1)$.

One can show that the derivative with respect to $e$ of our asymptotic estimate gives an asymptotic estimate of the derivative of $E(\mathbf{e})$. Thus, in the sideways limit $E(\mathbf{e})$ has no false local minima for sufficiently large $e$. Figure 4(2) shows an example, comparing the true error to our simple approximation above.

*Symmetry.* The image moments $\mu_{a,b}$ divide into two categories: the *even* moments, such as $\mu_{2,0}$, $\mu_{0,2}$, $\mu_{2,2}$, that involve sums over even powers and nonnegative terms only, and the *odd* moments. For randomly distributed image points, the odd $\mu_{a,b}$ are suppressed by roughly $1/\sqrt{N_p}$ compared to the even $\mu_{a,b}$.

With many correspondences, the image usually has some rotational symmetry and we can neglect the odd $\mu_{a,b}$ to a good approximation. Then, $E \approx E_{\text{side}}$

$$
\approx \alpha^2 \frac{e^2 \mu_{0,2} \mu_{2,2}}{e^2 \mu_{2,2} + \mu_{0,2}} + \beta^2 \left( \mu_{0,4} - \frac{\mu_{0,2}^2}{N_p} \right) \tag{14}
$$
$$
- 2A \left( \beta \sigma_{0,2} + \alpha \frac{e \mu_{0,2}}{e^2 \mu_{2,2} + \mu_{0,2}} \sigma_{1,1} \right) + 2B \left( \beta \sigma_{0,3} - \alpha \frac{e^2 \mu_{2,2} \sigma_{0,2} - e \mu_{0,2} \sigma_{1,2}}{e^2 \mu_{2,2} + \mu_{0,2}} \right)
$$
$$
+ A^2 S_{0,0} + B^2 S_{0,2} - 2AB S_{0,1} - \frac{(Ae\sigma_{1,1} - B(e\sigma_{1,2} + \sigma_{0,2}))^2}{e^2 \mu_{2,2} + \mu_{0,2}} - B^2 \frac{\sigma_{1,1}^2}{\mu_{2,0}}.
$$

Also, symmetry makes the even $\mu_{a,b}$ depend weakly on the epipolar direction $\hat{\mathbf{e}}$ [17], which give a further simplification.

In the same way as for the $\mu_{a,b}$, we can estimate the relative sizes of the structure–dependent moments $S_{0,a}$ and $\sigma_{a,b}$. All the $\sigma_{a,b}$ will be small, for any direction $\hat{\mathbf{e}}$, if the $Z_{\text{NL}}^{-1}$ have no good approximation in terms of a cubic polynomial (*noncubic condition*). Also, [17] argues that the mixed terms combining the nonlinear and linear structure components can often be neglected. Assuming this and the noncubic condition, we get the simple estimate

$$
E(\mathbf{e}) \approx E_{\text{side}}(\mathbf{e}) \approx \alpha^2 L_{\Pi}^{(1,1)} + \beta^2 L_{\Pi}^{(2,2)} - 2\alpha\beta L_{\Pi}^{(2,1)} + A^2 S_{0,0} + B^2 S_{0,2}. \tag{15}
$$

Our experiments show that (15) accurately describes the error for our sequences.

In addition to the conclusions above, [17] uses our estimate to generalize the JHM theorem [7][14] and to extend the results of [18] to planar scenes.

*Experiments.* We tested $E_{\text{side}}$ against the true error, using synthetic structures and structures extracted from five real sequences (PUMA[12], Rocket [5], CMU CASTLE, and two of our indoor sequences). We show only a few results.

Figure 5 compares $E(\mathbf{e})$ to our simplest approximation (15), which has slight problems only for the PUMA example. Figure 5 also shows $E_{\text{side}}$ for this example; it is indistinguishable from the true error. For the Rocket structure, we compared the global minimum positions for the true error, $E_{\text{side}}$, and (15). Within measurement error, they were identical. Fig. 6(1,2) shows that our symmetry–based approximation (14) gives good results with just 192 and 132 image points.

## 4   Projective Geometry

Suppose one fixes the camera matrix for image 1 to be $(\mathbf{1}_3, \mathbf{0}_3)$. The projective transforms that leave this camera matrix unaltered change the structure by adding an arbitrary plane to it or scaling it [16]. The real projective ambiguity is just this freedom to scale or add a plane. Since scenes that differ only in their planar component are equivalent in projective geometry, the linear component $Z_{\text{L},m}^{-1}$ of the scene cannot make a contribution to the least–squares error. In effect, only the nonlinear terms contribute to $E_{\text{side}}$.

(1a)                          (1b)                          (2a)



(3a)                          (3b)                          (2b)

(4a)                          (4b)                          (4C)
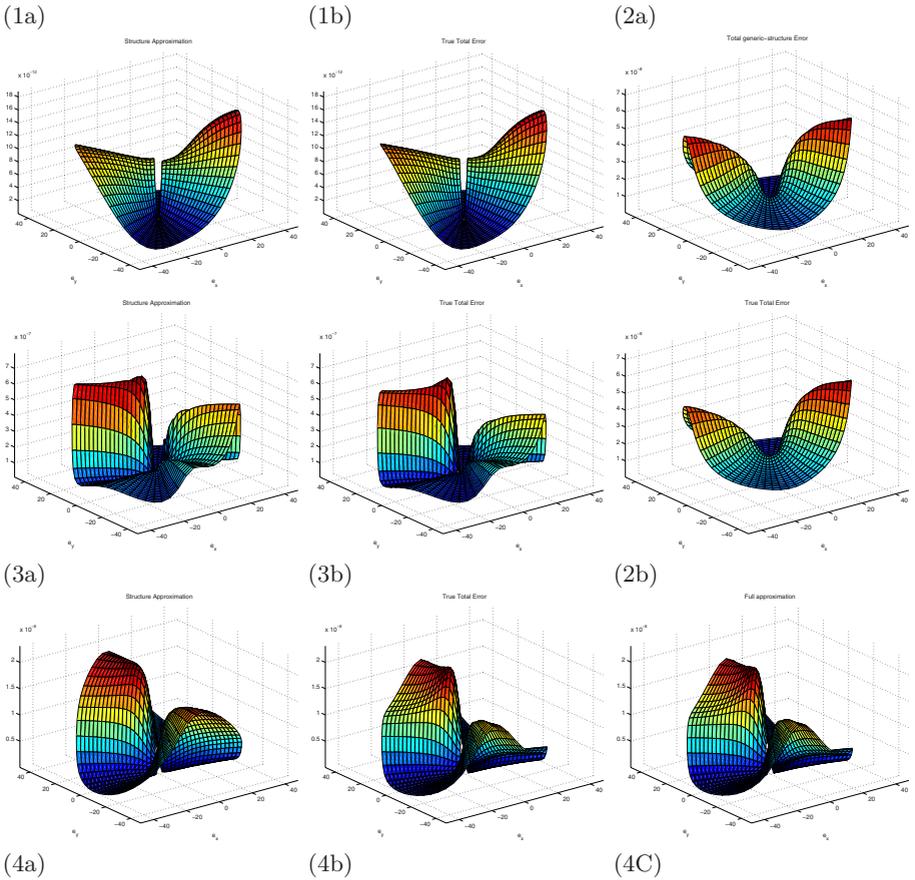
**Fig. 5.** (1): CMU Castle, $\mathbf{e}_{\text{true}} = (5.78; 8.16)$; (2): Indoor 1, $\mathbf{e}_{\text{true}} = (0.11; 0.16)$; (3): Indoor 2, $\mathbf{e}_{\text{true}} = (-0.125; -1.49)$; (4): PUMA, $\mathbf{e}_{\text{true}} = (10; -.025)$. (a): Simplified approximation (15); (b): True error; (C): $E_{\text{side}}$.

[17] shows this directly. For projective SFM and infinitesimal motion, one can define a projective error $E_{\text{proj}}(\mathbf{e}) \equiv \epsilon^T \Pi_{\text{proj}} \epsilon$ as for the Euclidean case, where $\epsilon$ is the same as in (3). The same arguments as before give a sideways asymptotic estimate: $E_{\text{proj}}(\mathbf{e}) \approx A^2 \{Z_{\text{NL}}\}^T \Pi_{\perp \text{proj}} \{Z_{\text{NL}}\}$

$$-2AB\{Z_{\text{NL}}\}^T \Pi_{\perp \text{proj}} \{yZ_{\text{NL}}\} + B^2 \{yZ_{\text{NL}}\}^T \Pi_{\text{proj}} \{yZ_{\text{NL}}\}. \qquad (16)$$

Thus, the error for projective SFM is simpler than the Euclidean error. This simplicity comes at a cost [18]. At large $e$, one can show that the projective error on the line $\mathbf{e} = t\mathbf{e}_{\text{true}}$ has no quadratic growth with $e$ as in the Euclidean error. This implies that the projective error gives less information to estimate the epipole. Figure 6(3B) compares the Euclidean and projective errors for the same image pair.
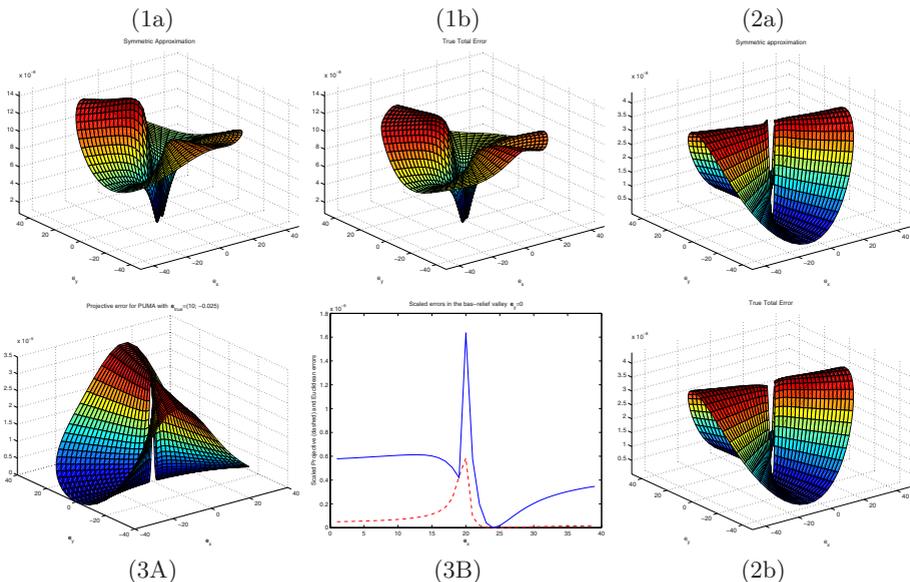
**Fig. 6.** (1): Extended PUMA sequence, $\mathbf{e}_{\text{true}} = (0.13; -0.08)$; (2): Extended Rocket, $\mathbf{e}_{\text{true}} = (-1.0; -3.9)$; (3): Projective error for PUMA, same images as in Figure 5(4). (a): Approximation (14); (b): True error; (3A): True projective error; (3B): Projective (dashed) and Euclidean errors on the line $e_y = 0$, the "bas–relief valley" in (3A).

In the forward region, the projective analysis is similar to the Euclidean one. Experiments confirm that the results are also similar, see Figure 1.

## 5   Noise

We report experiments on *noisy* images. We ran a standard two–image algorithm to estimate the structure/motion and used the result to compute $E_{\text{side}}$.[2] $E_{\text{side}}$ continues to model the true error well, despite our using a larger than normal noise. The noise is large enough that our two–image routine usually returns bad **T** estimates and the noisy error looks quite different from the noiseless one.

For noisy images, we cannot assume without loss of generality that the true rotation is zero. Fortunately, rotation has a small effect on the error [18][25].

We have not studied the forward noisy error carefully, but experiments (e.g., Figure 1) indicate that noise increases its complexity, as might be expected [17].

## 6   Conclusion

We studied the least–squares error for infinitesimal motion, giving two simple asymptotic estimates of the error which capture its detailed behavior over the

---

[2] Our rationale is that the error depends on the observed flow, which is modelled better by the estimated structure and motion than by the ground truth.
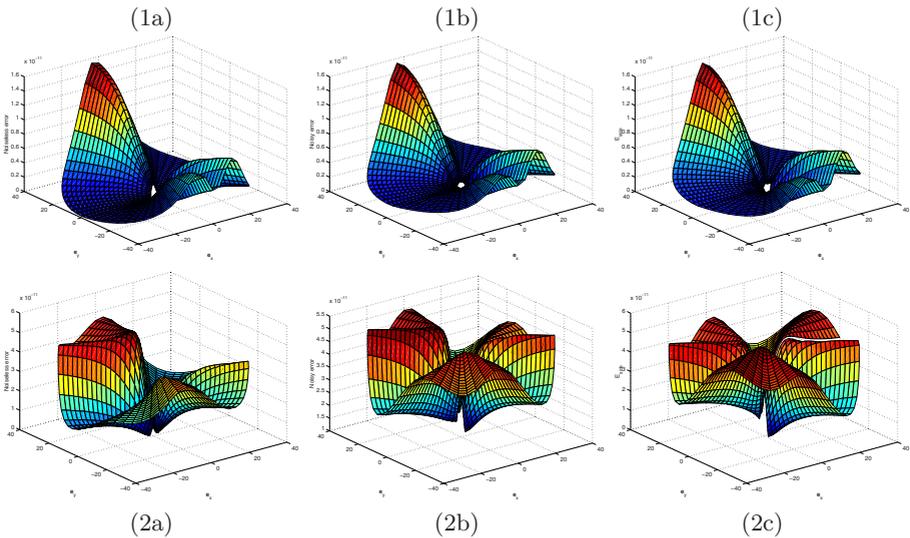
**Fig. 7.** Noise results. $\sigma_{\text{noise}}$ gives the noise standard deviation, $d_{\text{med}}$ the median size of the true flow, and $\theta_{\mathbf{T}\text{err}}$ the angular error in in the initial $\mathbf{T}$ estimate. (1): Rocket, $\mathbf{e}_{\text{true}} = (-0.01; -0.05)$, $\sigma_{\text{noise}} = 0.4d_{\text{med}}$, $\theta_{\mathbf{T}\text{err}} = 77°$. (2): PUMA, $\mathbf{e}_{\text{true}} = (0.5; -0.2)$, $\sigma_{\text{noise}} = 0.3d_{\text{med}}$, $\theta_{\mathbf{T}\text{err}} = 38°$. (a) Noiseless error; (b) True noisy error; (c) $E_{\text{side}}$.

entire range of motions. We illustrated the use of these estimates by deriving new error properties.

For roughly forward translation estimates, we showed by theory and experiment that the error tends to be complex for candidate epipoles near image points, and that this is more likely when: the true epipole is far from the point; and/or the point is isolated in the image; and/or the corresponding 3D depth is small; and/or the number of image points is small. Our experiments show that the complexity near image points produces local minima, confirming [3][18]. We pointed out that the previous arguments of [3][18] do not explain the error's complexity or local minima.

For non–forward translation estimates, we gave a simple model of the error for planar scenes. For two special cases, we derived a new double bas–relief ambiguity and proved the absence of local minima at large $|\mathbf{e}|$. For nonplanar scenes, we simplified our approximations under various assumptions, including rough rotational symmetry of the image and a reasonably "generic" distribution of 3D depths. Our simplest approximation gives a good model of the least–squares error in all our noiseless experiments. We analyzed the error for projective SFM, pointing out that it is flatter than the Euclidean error, and showed by experiments that our analysis remains useful for noisy images.

We believe that our results will lead to an in–depth understanding of the least–squares error. For example, our sideways asymptotic estimate depends on just 29 parameters, and often 13 are enough. This suggests that a semi–

exhaustive search through the space of least–squares errors may be feasible to determine the pitfalls that algorithms could encounter.

# References

1. G. Adiv, "Inherent ambiguities in recovering 3-D motion and structure from a noisy flow field," **PAMI** 11, 477-489, 1989.
2. P. Belhumeur, D. Kriegman, and A. Yuille "The Bas–Relief Ambiguity," **IJCV** 35:1, 33–44, 1999.
3. A. Chiuso, R. Brockett, and S. Soatto, "Optimal Structure from Motion: Local Ambiguities and Global Estimates," **IJCV** 39: 3, 95-228, 2000.
4. K. Daniilidis and H.H. Nagel, "Analytical results on error sensitivity of motion estimation from two views," **IVC** 8 297–303, 1990.
5. R. Dutta, R. Manmatha, L. R. Williams, and E. M. Riseman, "A data set for quantitative motion analysis," *CVPR*, 159-164, 1989.
6. C. Fermüller, Y. Aloimonos, "Observability of 3D Motion," **IJCV** 37:1, 43–63, 2000.
7. A. D. Jepson and D. J. Heeger, "Subspace methods for recovering rigid motion II: Theory," U. of Toronto TR RBCV-TR-90-36, 1990.
8. R. Hartley and A. Zisserman, "Multiple View Geometry in Computer Vision," Cambridge, 2000.
9. J. Hornegger, C. Tomasi, "Representation Issues in the ML Estimation of Camera Motion," **ICCV** 640–647, 1999.
10. K. Kanatani, *Geometric Computation for Machine Vision,* Oxford, 1993.
11. J. J. Koenderink and A. J. Van Doorn, "Affine structure from motion," **JOSA** 8:2 377–385, 1991.
12. R. Kumar and A. R. Hanson, "Sensitivity of the Pose Refinement Problem to Accurate Estimation of Camera Parameters," *ICCV*, 365-369, 1990.
13. Y. Ma, J. Kosecka, and S. Sastry, "Optimization criteria and geometric algorithnms for motion and structure estimation," **IJCV** 44:3, 219–249, 2001.
14. S. Maybank, *Theory of Reconstruction from Image Motion*, Springer, Berlin, 1993.
15. S. Maybank, "A Theoretical Study of Optical Flow," Doctoral Dissertation, University of London, 1987.
16. J. Oliensis and Y. Genc, "Fast and Accurate Algorithms for Projective Multi–Image Structure from Motion," **PAMI** 23:6 546–559, 2001.
17. J. Oliensis, "The Least–Squares Error for Structure from Infinitesimal Motion," **IJCV**, to appear.
18. J. Oliensis, "A New Structure from Motion Ambiguity," **PAMI** 22:7, 685–700, 2000 and *CVPR* 185–191, 1999.
19. J. Oliensis, "Computing the Camera Heading from Multiple Frames," *CVPR* 203–210, 1998.
20. S. Soatto and R. Brocket, "Optimal Structure from Motion: Local Ambiguities and Global Estimates," *CVPR* 282–288, 1998.
21. S. Soatto and P. Perona, "Recursive 3-D Visual Motion Estimation Using Subspace Constraints," **IJCV** 22:3 235–259, 1997.
22. M. Spetsakis and Y. Aloimonos, "Optimal visual motion estimation: a note," **PAMI** 14:959–964.
23. R. Szeliski and S.B. Kang, "Shape ambiguities in structure from motion," **PAMI** 19 506–512, 1997.

24. B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon, "Bundle Adjustment—A Modern Synthesis," *Workshop on Vision Algorithms: Theory and Practice* 298–372, 1999.
25. T. Y. Tian, C. Tomasi, and D. J. Heeger, "Comparison of Approaches to Egomotion Computation" *CVPR*, 315–320, 1996.
26. C. Tomasi and J. Shi, "Direction of Heading from Image Deformations," *CVPR* 422–427, 1993.
27. T. Xiang; L.-F.Cheong, "Understanding the Behavior of SFM Algorithms: A Geometric Approach" **IJCV** 51(2): 111-137; 2003
28. Weng, J., Huang, T.S., and Ahuja, N., "Motion and Structure from Two Perspective Views: Algorithms, Error Analysis, and Error Estimation," **PAMI** 11:5, 451-476, 1989.