Error Analysis of 3-D Motion Estimation Algorithms in the Differential Case

Tao Xiang
Department of Computer Science
Queen Mary, University of London
txiang@dcs.qmul.ac.uk

Loong-Fah Cheong
Department of Electrical & Computer Engineering
National University of Singapore
eleclf@nus.edu.sg

Abstract

We put forth in this paper a geometrically motivated 3-D motion error analysis which is capable of supporting investigation of global effect such as inherent ambiguities. The error expression that we derive allows us to predict the exact conditions likely to cause ambiguities and how these ambiguities vary with motion types such as lateral or forward motion. Our formulation, though geometrically motivated, is employed to model the effect of noise.

1 Introduction

Structure From Motion (SFM) problem has become a central topic of computer vision and received increasing attention since 1980s. SFM problem is usually treated as two subproblems, namely, the measurement of 2-D image displacement (correspondences) or velocity (optical flow), and the extraction of 3-D relative motion and structure information using as input the 2-D image measurements. Due to the ill-conditioned nature of the first subproblem, the input to 3-D motion estimation algorithms inevitably contains errors. In view of such errors, most of the previous error analysis on 3-D motion estimation [5, 6] related the errors of the estimated 3-D motion parameters to the measurement errors in the first subproblem. The errors are typically expressed as a high variance or a bias in the motion parameters through some statistical analysis or given as empirical figures through some simulations. However, few works have been contributed to a systematical characterization of the topology of the cost functions, which is important if we are to understand the full range of the algorithms' behaviour. Recently several studies have emerged in this direction. Soatto and Brockett [9] attempted to achieve optimal SFM (in differential approaches) by understanding the error surface configuration of the cost functions. They noted the existence of a minimum at the opposite end of the bas-relief valley (termed as rubbery ambiguity in their paper), and attributed it to the presence of noise, although their simulation results showed that the minimum persisted with noiseless input. The work of Fermüller and Aloimonos [4] presented a geometrical-statistical investigation of the observability of 3-D motion. They studied the conditions on errors in the motion estimates for the local minima on error surface to arise.

In this paper, we propose an approach that lends itself towards understanding the full behavior of 3-D motion estimation algorithms using optical flow as input. Instead of dealing with specific algorithms each using different optimization techniques, we study one class of algorithms based on the weighted differential epipolar constraint. This class includes most of the existing differential 3-D motion estimation algorithms. What permits an unifying view of these different algorithms is a new optimization criterion to be presented in this paper. This criterion allows us to develop a simple and explicit expression for the residual errors of the optimization functions in terms of the errors in the 3-D motion estimates and enables us to predict the exact conditions likely to cause ambiguities. We also extend our analysis on inherent ambiguities to include the effect of noise in the image measurements. Our investigation unravels the impact a realistic anisotropic noise distribution can have on the topology of the cost functions.

2 Differential Reprojection Criterion

In this paper, we denote the estimated parameters with the hat symbol ($\hat{}$) and errors in the estimated parameters with the subscript e. Given a n-vector $\mathbf{s}$, $[\mathbf{s}]_m$ is defined as the m-vector which consist of the first m $(m < n)$ components of $\mathbf{s}$, $\underline{\mathbf{s}}$ is defined as the (n+1)-vector with 0 added as the last component, and $\bar{\mathbf{s}}$ is the associated skew-symmetric matrix of $\mathbf{s}$. For any vector $\mathbf{s} = (s_1, s_2)^T$, $\mathbf{s}^\perp$ represents the vector $(s_2, -s_1)^T$ which is perpendicular to $\mathbf{s}$ with the same magnitude. A pinhole camera model with perspective projection is assumed as shown in Fig. 1; it is moving with a translational velocity $\mathbf{v} = (U, V, W)^T$ and a rotational velocity $\mathbf{w} = (\alpha, \beta, \gamma)^T$. A point P in the world produces an image point p in the image plane which is f pix-

Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'03)
1063-6919/03 $17.00 © 2003 IEEE

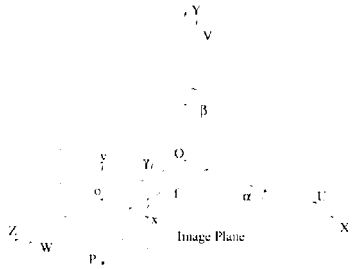

Figure 1. The image formation model.

els away from the optical center; if $\mathbf{P} = (X, Y, Z)^T$ and $\mathbf{p} = (x, y, f)^T$ are the co-ordinates corresponding to P and p respectively, we have: $\mathbf{p} = f\frac{\mathbf{P}}{Z}$. The focal length f is assumed to be known since we are dealing with calibrated motion in this paper.

The image velocity due to camera motion is given by the following familiar equation [6]:

$$\dot{\mathbf{p}} = -\mathbf{Q_p}\left(\frac{\mathbf{v}}{Z} + \bar{\mathbf{w}}\mathbf{p}\right) \tag{1}$$

where $\dot{\mathbf{p}} = (u, v, 0)^T$, $\mathbf{Q_p} = \begin{bmatrix} f & 0 & -x \\ 0 & f & -y \\ 0 & 0 & 0 \end{bmatrix}$.

Eq. (1) can alternatively be written in terms of its components as:

$$\begin{aligned} u &= (x - x_0)\tfrac{W}{Z} + \tfrac{\alpha x y}{f} - \beta\left(\tfrac{x^2}{f} + f\right) + \gamma y \\ v &= (y - y_0)\tfrac{W}{Z} + \alpha\left(\tfrac{y^2}{f} + f\right) - \tfrac{\beta x y}{f} - \gamma x \end{aligned} \tag{2}$$

where $(x_0, y_0) = (f\frac{U}{W}, f\frac{V}{W})$ is the focus of expansion (FOE). We define $\dot{\mathbf{p}}_{tr} = (u_{tr}, v_{tr})^T$ and $\dot{\mathbf{p}}_{rot} = (u_{rot}, v_{rot})^T$, where $\frac{\dot{\mathbf{p}}_{tr}}{Z}$ and $\dot{\mathbf{p}}_{rot}$ are the components of the flow due to translation and rotation respectively. Since only the direction of the translation can be recovered from the flow field, we can set $W = 1$ for the case of general motion; the case of pure lateral motion ($W = 0$) will be discussed separately where required.

A 3-D motion estimation algorithm based on the differential epipolar constraint trys to find motion estimates that minimize the following cost function [1]:

$$J_E = \sum_{i=1}^{n}\left(\mathbf{p}_i^T\hat{\bar{\mathbf{v}}}\dot{\mathbf{p}}_i + \mathbf{p}_i^T\hat{\bar{\mathbf{v}}}\bar{\mathbf{w}}\mathbf{p}_i\right)^2 \tag{3}$$

where n is the number of image velocity measurement. We first focus on the case where the optical flow input was noise-free; thus we have used the term $\dot{\mathbf{p}}_i$ in (3). Minimizing J_E amounts to a linear optimization problem which can be solved by a linear least square method. A well-studied bias of the linear algorithms is that the estimated translation will be biased towards the image center [5, 8]. In view of this bias, a variety of non-linear cost functions have been proposed, which are basically different weighted versions of J_E each driven by slightly different considerations based on statistical analysis of noise [6, 8].

From Eq. (1), the reconstructed depth can be expressed using the estimated motion parameters:

$$\hat{Z} = -\frac{\hat{\mathbf{v}}^T{\mathbf{Q_p}}^T\underline{\mathbf{n}}}{\left(\dot{\mathbf{p}}^T - \mathbf{p}^T\hat{\bar{\mathbf{w}}}{\mathbf{Q_p}}^T\right)\underline{\mathbf{n}}} \tag{4}$$

where $\mathbf{n}$ is an unit vector in the image plane representing a direction. Substituting the recovered depth in Eq. (4) into Eq. (1), we obtain the reprojected (estimated) flow field, denoted by $\hat{\dot{\mathbf{p}}}$, as follows:

$$\hat{\dot{\mathbf{p}}} = \frac{\mathbf{Q_p}\hat{\mathbf{v}}\left(\dot{\mathbf{p}}^T - \mathbf{p}^T\hat{\bar{\mathbf{w}}}{\mathbf{Q_p}}^T\right)\underline{\mathbf{n}}}{\hat{\mathbf{v}}^T{\mathbf{Q_p}}^T\underline{\mathbf{n}}} - \mathbf{Q_p}\bar{\mathbf{w}}\mathbf{p} \tag{5}$$

The difference between the original optical flow and the reprojected flow can thus be expressed as:

$$\dot{\mathbf{p}}_e = \frac{\left(\mathbf{C_1} - {\mathbf{C_1}}^T + \mathbf{C_2} - {\mathbf{C_2}}^T\right)\underline{\mathbf{n}}}{\hat{\mathbf{v}}^T{\mathbf{Q_p}}^T\underline{\mathbf{n}}} = \frac{\mathbf{C}\underline{\mathbf{n}}}{\hat{\mathbf{v}}^T{\mathbf{Q_p}}^T\underline{\mathbf{n}}} \tag{6}$$

where $\mathbf{C_1} = \dot{\mathbf{p}}\hat{\mathbf{v}}^T\mathbf{Q_p}$, $\mathbf{C_2} = \mathbf{Q_p}\hat{\bar{\mathbf{w}}}\mathbf{p}\hat{\mathbf{v}}^T{\mathbf{Q_p}}^T$ and $\mathbf{C} = \left(\mathbf{C_1} - {\mathbf{C_1}}^T + \mathbf{C_2} - {\mathbf{C_2}}^T\right)$. We can define a cost function J_R based on the reprojected flow difference:

$$J_R = \sum_{i=1}^{n}\left(\frac{\mathbf{p}_i^T\hat{\bar{\mathbf{v}}}\dot{\mathbf{p}}_i + \mathbf{p}_i^T\hat{\bar{\mathbf{v}}}\bar{\mathbf{w}}\mathbf{p}_i}{\hat{\mathbf{v}}^T{\mathbf{Q}_{\mathbf{p_i}}}^T\underline{\mathbf{n}_i}}\right)^2 \tag{7}$$

A comparison of Eq. (3) with Eq. (7) reveals that J_R is a weighted version of J_E with the weight given by the projection of $\hat{\dot{\mathbf{p}}}_{tr_i}$ on the direction $\mathbf{n}_i$. It can be shown that various weighted differential epipolar constraints differ mainly in the choice of $\mathbf{n}$ [10]. For instance, the choice of the "epipolar reconstruction" direction ($\mathbf{n} = \frac{\hat{\dot{\mathbf{p}}}_{tr}}{\|\hat{\dot{\mathbf{p}}}_{tr}\|}$) results in the cost function presented in [8]. It follows that J_R can be seen as a scheme unifying the various weighted epipolar constraints. Therefore, to understand the behavior of these SFM algorithms based on weighing the epipolar constraint, one can focus on studying the differential reprojection criterion. All these algorithms inherit properties from the differential reprojection criterion; in particular, much of the ambiguity conditions of these algorithms are common and can be studied by looking at the numerator of J_R.

3 Inherent Ambiguities

We express J_R in terms of the various component errors in the 3-D motion estimates. Substituting $[\dot{\mathbf{p}}_i]_2 = (u_i, v_i)^T = (\frac{x_i - x_0}{Z_i} + u_{rot_i}, \frac{y_i - y_0}{Z_i} + v_{rot_i})^T$, $\hat{\dot{\mathbf{p}}}_{tr_i} =$

Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'03)
1063-6919/03 $17.00 © 2003 IEEE

$(x_i - \hat{x_0}, y_i - \hat{y_0})^T$ and $\dot{\mathbf{p}}\hat{}_{rot_i} = (u\hat{}_{rot_i}, v\hat{}_{rot_i})^T$ into (7), we have:

$$J_R = \sum \left(\frac{(x - \hat{x_0}, y - \hat{y_0}) \cdot (v_{rot_e} - \frac{y_{0_e}}{Z}, \frac{x_{0_e}}{Z} - u_{rot_e})}{(x - \hat{x_0}, y - \hat{y_0}) \cdot \mathbf{n}} \right)^2 \tag{8}$$

For notational convenience, we omit the subscript i in the expression of J_R; We denote the expression contained in the outer bracket of (8) as $\dot{\mathbf{p}}_e(\hat{\mathbf{v}}, \mathbf{v}, \mathbf{w}_e)$, and the vectors $(x - \hat{x_0}, y - \hat{y_0})^T$ and $(v_{rot_e} - \frac{y_{0_e}}{Z}, \frac{x_{0_e}}{Z} - u_{rot_e})^T$ as $\mathbf{t_1}$ and $\mathbf{t_2}$ respectively. We also adopt the terminology that for the vectors $\mathbf{t_1}$ and $\mathbf{t_2}$, $\mathbf{t_{1,n}}$ and $\mathbf{t_{2,n}}$ denote the n^{th} order component with respect to x and y; thus we have:

$$\begin{cases} \mathbf{t_1} &= \mathbf{t_{1,0}} + \mathbf{t_{1,1}} \\ \mathbf{t_2} &= \mathbf{t_{2,0}} + \mathbf{t_{2,1}} + \mathbf{t_{2,2}} + \mathbf{t_{2,Z}} \end{cases} \tag{9}$$

where $\mathbf{t_{1,0}} = (-\hat{x_0}, -\hat{y_0})^T$, $\mathbf{t_{1,1}} = (x, y)^T$, $\mathbf{t_{2,0}} = (\alpha_e f, \beta_e f)^T$, $\mathbf{t_{2,1}} = (-\gamma_e x, -\gamma_e y)^T$ and $\mathbf{t_{2,2}} = (\alpha_e \frac{y^2}{f} - \frac{\beta_e xy}{f}, -\frac{\alpha_e xy}{f} + \beta_e \frac{x^2}{f})^T$. The last item $\mathbf{t_{2,Z}}$ in the above equation denotes the depth dependent term $(-\frac{y_{0_e}}{Z}, \frac{x_{0_e}}{Z})^T$.

To visualize the residual error surface, it is easier to deal with a 3-D surface. We use for this purpose the translation error surface, which is described parametrically with two free variables, the estimated FOE $(\hat{x_0}, \hat{y_0})$. We know that given this hypothesized FOE, the rotation variables can be solved in terms of the estimated FOE so as to minimize J_R. The residual error J_R can thus be obtained for each FOE candidate, describing the entire residual surface completely. We first make some assumptions on the distribution of feature points and depth. We assume that the feature points are evenly distributed in the image plane, as is the distribution of the "depth-scaled feature points" $(\frac{x}{Z}, \frac{y}{Z})$. An analysis on how the error surface will be affected when these assumptions do not hold can be found in [10].

Eq. (8) shows that for any given data set (x, y, Z), the residual error is a function of the true FOE, the estimated FOE and the error in the rotation estimates. Evidently, ambiguities would arise when the errors in the motion estimates satisfy the following conditions to make the numerator of $\dot{\mathbf{p}}_e(\hat{\mathbf{v}}, \mathbf{v}, \mathbf{w}_e)$ vanish: a) making $\|\mathbf{t_2}\|$ small and b) making $\mathbf{t_1}$ and $\mathbf{t_2}$ perpendicular to each other. The second condition is generally not satisfiable at all points of the image; thus making $\|\mathbf{t_2}\|$ small (condition a) contributes towards ambiguity. Notice that Making $\|\mathbf{t_1}\|$ small does not contribute towards ambiguity if we have suitably normalized J_R with the term in the denominator. We thus can make the following observations:

1). When the estimated FOE moves towards infinity, the direction of $\mathbf{t_1}$ approaches that of $\mathbf{t_{1,0}}$, which is constant. Pointing towards a constant direction represents a necessary condition for $\mathbf{t_1}$ and $\mathbf{t_2}$ to be perpendicular to each other.

2). From the expression of $\mathbf{t_2}$, we can see that $\mathbf{t_{2,0}}$ and $\mathbf{t_{2,Z}}$ are pointing towards constant directions for all the feature points. Intuitively, $\mathbf{t_2}$ will be more perpendicular to $\mathbf{t_1}$ when both $\mathbf{t_{2,0}}$ and $\mathbf{t_{2,Z}}$ are perpendicular to $\mathbf{t_{1,0}}$. This relationship can be illustrated with the diagram shown in Fig. 2. The vector $\mathbf{t_{1,1}}$ can be regarded as a perturbation to the vector $\mathbf{t_{1,0}}$, and similarly, $\mathbf{t_{2,1}}$ and $\mathbf{t_{2,2}}$ can be regarded as perturbations to $\mathbf{t_{2,0}}$ and $\mathbf{t_{2,Z}}$. However, if the feature points are sufficiently evenly distributed (such that the vectors $\mathbf{t_{1,1}}$ are evenly spread on either side of $\mathbf{t_{1,0}}$ and the sum of vectors $\mathbf{t_{2,1}}$ and $\mathbf{t_{2,2}}$ are evenly spread on either side of $\mathbf{t_{2,0}}$ and $\mathbf{t_{2,Z}}$), and the distribution of depth Z is symmetrical with respect to the $\mathbf{t_{1,0}}$ direction, making $\mathbf{t_{2,0}}$ and $\mathbf{t_{2,Z}}$ perpendicular to $\mathbf{t_{1,0}}$ is a reasonable choice for the minimization of J_R. Thus we have

$$\frac{x_0}{y_0} = \frac{\hat{x_0}}{\hat{y_0}} \tag{10}$$

$$\frac{\alpha_e}{\beta_e} = -\frac{\hat{y_0}}{\hat{x_0}} \tag{11}$$

Eq. (10) imposes a constraint on the direction of the estimated translation, namely, the three points $(\hat{x_0}, \hat{y_0})$, (x_0, y_0) and $(0, 0)$ should lie on a straight line. We henceforth refer to this constraint as the Translation Direction (TDir) constraint. Eq. (11) imposes a constraint on the direction of $\mathbf{w}_e$, which shall be henceforth referred to as the Rotation Error Direction (RDir) constraint.

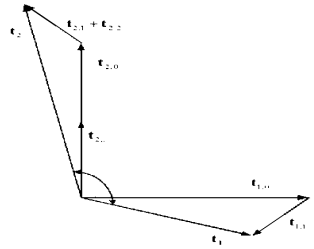

Figure 2. Geometry of $\mathbf{t_1}$ and $\mathbf{t_2}$.

3). Since $\mathbf{t_1}$ cannot be made exactly perpendicular to $\mathbf{t_2}$, small $\|\mathbf{t_2}\|$ will help to reduce the numerator term of J_R. Obviously, small $\|\mathbf{t_2}\|$ can be achieved by having small errors in the motion estimates. Alternatively, since $\mathbf{t_{2,0}}$ and $\mathbf{t_{2,Z}}$ are pointing towards constant direction, they can be made to approximately cancel off each other by an appropriate choice in the errors in the motion estimates:

$$\begin{cases} \alpha_e &= \frac{y_{0_e}}{Z_{avg} f} \\ \beta_e &= -\frac{x_{0_e}}{Z_{avg} f} \end{cases} \tag{12}$$

where Z_{avg} is the average scaled depth of the scene in view. In view of its constraint on the magnitude of $\mathbf{w}_e$, we refer to this constraint as the Rotation Magnitude (RMag) constraint, although it also implies directional constraint given by $\frac{\alpha_e}{\beta_e} = -\frac{y_{0_e}}{x_{0_e}}$. Comparing with (11), it is evident that RDir and RMag can hold simultaneously only if TDir also holds.

4). $\mathbf{t_{2,1}}$ and $\mathbf{t_{2,2}}$ are determined by γ_e and α_e, β_e respectively. Under general scene, there is no way for $\mathbf{t_{2,1}}$ and $\mathbf{t_{2,2}}$ to be canceled off with other terms; $\|\mathbf{w}_e\|$ has to be small for $\mathbf{t_{2,1}}$ and $\mathbf{t_{2,2}}$ to be small. In this sense, $\mathbf{t_{2,1}}$ and $\mathbf{t_{2,2}}$ contribute to accurate estimation of rotation; unfortunately, their effects are weak unless the field of view is large. In view of the subsidiary role of the constraint it exerts on the magnitude of the rotation errors, we term it as RMag2 constraint. Another important fact about $\mathbf{t_{2,2}}$ is that it will be exactly perpendicular to $\mathbf{t_1}$ (independent of the feature points co-ordinates) when $\mathbf{t_{1,0}} = (0, 0)$. Therefore, RMag2 constraint will be ineffective on the magnitude of α_e and β_e when the estimated FOE coincides with the origin.

One of the well known phenomenon in motion perception is the bas-relief ambiguity [9]. It amounts to a valley on the translation error surface, along a straight line defined by the true FOE and the image center. We term this straight line the bas-relief line and this valley the bas-relief valley. From the above observations, apparently TDir is the direct reason for the formation of such a valley. Since the motion estimates are more likely to reside along the bas-relief valley, a detailed profiling of the shape of the bas-relief valley is important. Due to the constraint of space, a detailed analysis is not presented here, which can be found in [10]. The general finding is that the coupling of translational error and rotational error results in the formation of a local minimum on the opposite side of the true FOE within the bas-relief valley. The exact location and the "depth" of the local minimum depends on various factors, including the type of the true translation, FOV, focal length, etc. The major findings obtained so far are summarized as follows:

Rotation error —The rotation errors satisfy the following constraints: $\gamma_e = 0$ and $\frac{\alpha_e}{\beta_e} = -\frac{\hat{y_0}}{\hat{x_0}}$ when motion ambiguities arise. The magnitude of the rotation error may be further subject to Eq. (12), but this constraint weakens as the true and the estimated FOE approach infinity. In particular, when the estimated translation approaches infinity, the RMag constraint is not needed anymore and only the RMag2 constraint is operative, which tends to make the rotation estimates close to the true solution. Another influential factor is FOV. Under large FOV, accurate rotation estimation is expected. On the other hand, when FOV is small, the rotation parameters are estimated with difficulties.

Translation error —Bas-relief valley is the major characteristic of the error surface; Along the bas-relief valley, there is a local minimum at the opposite side of the global minimum with respect to the origin, which we called the opposite minimum. The location and the depth of the opposite minimum is determined by several factors. In particular, the opposite minimum will be further away from the image center and its residual value smaller when the FOV is small and the true translation is largely lateral.

Type of translation —The type of translation has important influence on the configuration of the residual error surface. Under largely forward translation, the estimation of both translation and rotation is relatively accurate unless the feature points are locally clustered resulting in strong local minima within the image plane. In contrast, the SFM algorithms are more likely to give erroneous motion estimation when the true translation is largely lateral. However, as we point out in [2], as far as the depth recovery is concerned, translation that is largely lateral results in depth distortion that has nice properties such as the preservation of relief.

We perform simulations on synthetic images to both visualize and verify the predictions obtained from the preceding theory. These simulations were carried out based on the "epipolar reconstruction" scheme. As discussed in the preceding section, we use the translation error surface for visualization purpose. The residuals were plotted over the whole FOE search space, subtending the entire hemisphere in front of the image plane. We used visual angle in degree rather than pixel as the FOE search step; thus the co-ordinates in the plots were not linear in the pixel unit. Unless otherwise stated, the synthetic experiments have the following parameters: the focal length was 512 pixels which meant a FOV of approximately 53^o; there were 200 object points whose depths ranged randomly from 512 to 1536 pixels; feature points were also distributed randomly over the image plane; true rotational parameters were $(0, 0.001, 0.001)$.

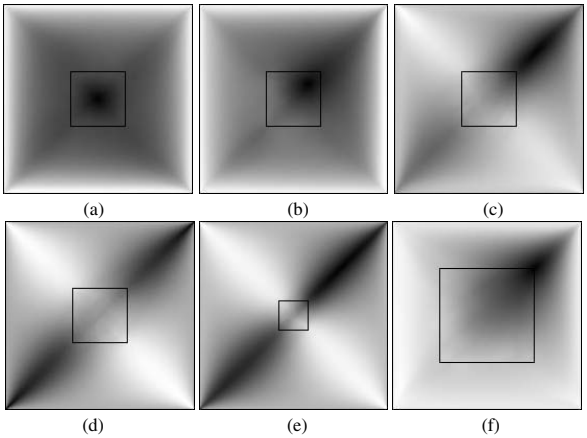

Figure 3. Influence of true translations and FOV on the residual error images.

Fig. 3 shows the residual error images for different translations. Translational parameters (U, V, W) for Fig. 3(a), (b), and (d) are $(0, 0, 2)$, $(0.5, 0.5, 2)$ and $(1, 1, 0)$ respectively and for Fig. 3(c), (e) and (f) are all $(1, 1, 1)$. FOV was 28^o for (e), 90^o for (f) and 53^o for the others. It can be seen that the bas-relief valley becomes more obvious when the translation changed from being purely forward to being purely lateral. Since the feature points distribution

were (roughly) even, the TDir constraint was a line passing through the image center and the true FOE. This can be clearly seen from Fig. 3(b)–(f) where the translation was not purely forward. Distinct local minima were centered around the true FOE and somewhere on the opposite side of the image center. The opposite minimum disappeared (or merged with the global minimum) for pure forward motion (Fig. 3(a)). As the global minimum moved towards the infinity, so did the opposite minimum. The "false" minimum on the opposite side was much shallower than the "true" minimum in the case of non-lateral motion and can hardly be seen from Fig. 3(f) (numerical data do show the existence of the opposite minimum), but in Fig. 3(d) under lateral motion they are almost equal in depth. By looking at the numerical values of the simulation data, we also found that for all the FOE candidates, the errors in the estimated rotational parameters were such that Eq. (11) held, while for the rotational estimates around the opposite minimum, we further have their magnitudes satisfying Eq. (12). Under lateral motion (Fig. 3(d)), those candidates with the smallest residuals were either the true translation or the translation in the opposite direction, while the estimated rotation satisfied $\frac{\alpha_e}{\beta_e} = -\frac{V}{U}$. Not surprisingly, we found that for these candidates, the magnitudes of $\mathbf{w}_e$ were quite arbitrary (though small) since the RMag constraint is ineffective. Fig. 3 also shows the influence of FOV on the residual error images. In our simulations, larger FOV was obtained by fixing the image size and decreasing the focal length. Thus, under larger FOV the true FOE $(f\frac{U}{W}, f\frac{V}{W})$ would be closer to the image center with the same translational velocity (U, V, W). It can be seen that the opposite minimum in Fig. 3(e) was prominent with small residual value, while in Fig. 3(f), the opposite minimum was almost invisible.

4 Influence of Noise

In practice, optical flow is always estimated with some noise. We express the noise-corrupted flow $\breve{\dot{\mathbf{p}}}$ as:

$$\breve{\dot{\mathbf{p}}} = \dot{\mathbf{p}} + \dot{\mathbf{p}}_n = (u + u_n, v + v_n, 0)^T \tag{13}$$

where $\dot{\mathbf{p}}_n$ is the flow component caused by noise. It follows that the noise-corrupted cost function, denoted as J_{Rn}, can be obtained as follows:

$$J_{Rn} = J_R + \sum \left(\frac{(x - \hat{x_0}, y - \hat{y_0}) \cdot (-v_n, u_n)}{(x - \hat{x_0}, y - \hat{y_0}) \cdot \mathbf{n}} \right)^2 + 2 \sum \left(\begin{array}{l} \frac{(x-\hat{x_0}, y-\hat{y_0}) \cdot (v_{rote} - \frac{y_{0_e}}{Z}, \frac{x_{0_e}}{Z} - u_{rote})}{(x-\hat{x_0}, y-\hat{y_0}) \cdot \mathbf{n}} \\ \times \frac{(x-\hat{x_0}, y-\hat{y_0}) \cdot (-v_n, u_n)}{(x-\hat{x_0}, y-\hat{y_0}) \cdot \mathbf{n}} \end{array} \right) \tag{14}$$

J_{Rn} consists of three terms. The first term is that of the noise-free case. The second term is always zero or positive, whereas the last term can be positive or negative. Next, we investigate the behavior of 3-D motion estimation algorithms under the effects of specific noise types.

Isotropic noise models have been frequently used for noise analysis in the computer vision community. However isotropic noise models are often unrealistic. We thus focus on anisotropic noise models. A simple anisotropic noise model is one where the noise added to each flow depends on the flow itself. Specifically, for each noiseless flow we add a noise whose horizontal and vertical components are Gaussian with standard deviations proportional to the horizontal and vertical components of the noiseless flow respectively. With this model, the noise tends to point towards the same direction as the noiseless flow. Such a model receives partial theoretical support from [3].

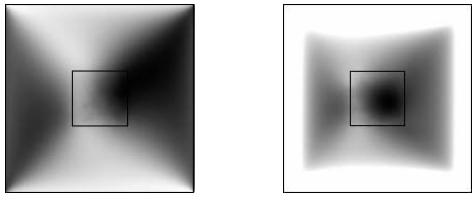

Figure 4. Residual error images for the flow fields with anisotropic noise.

The effect of noise under this model shows a strong directional anisotropy; this is especially so for the case where the noiseless flow field itself is also predominant in certain direction. Let us denote this direction as $\mathbf{n}_n$. This effect of such anisotropy is most significant at the periphery of the plots where the FOE estimates are far away from image center and have $(x - \hat{x_0}, y - \hat{y_0})$ approximately pointing in the same direction. When this direction is parallel to $\mathbf{n}_n$, the contribution of the second term in Eq. (14) would be small, and vice versa. The effect of the third term in Eq. (14) is also strongly direction-dependent, although the dependence is more complex. Suffice it to say that the resultant residual error images have their local minima being pulled towards the $\mathbf{n}_n$ direction and the periphery of the plots. This is illustrated in Fig. 4 where numbers of feature points were 2000 for (a) and 800 for (b) and $\mathbf{v} = (1, 1, 1)$ for (a) and $\mathbf{v} = (0, 0, 2)$ for (b). $\mathbf{w} = (0, 0.0005, 0.001)$ and noise level was 50% for all the images in Fig. 4. It can be seen from Fig. 4 that the influence of a 50% anisotropic noise is significant. For the case of Fig. 4(a), the noise was biased towards the direction $\mathbf{n}_n = (0.82, 0.56)$. We can see that the bas-relief valley was "pulled" towards the $\mathbf{n}_n$ direction. For the case of forward motion (Fig. 4(b)), the value of $\mathbf{n}_n$ is $(0.77, 0.64)$. We can see a clear minima strip formed outside the image plane with the global minimum perturbed to $(90, -28)$. The implication of the above is that while true rotation parameters do not explicitly appear in the expression of J_{Rn}, their values can influence the performance

of 3-D motion estimation algorithms by indirectly affecting the distribution of noise. Such phenomenon is often observed in practice, the result being that the FOE cannot be reliably estimated when the rotation is dominant.

We carry out experiments on computer generated and real images so as to verify the various predictions made. Three familiar image sequences were used. The parameters of these sequences are listed in Table 1. Relatively dense optical flow fields (around 3000 feature points for each sequence) were obtained using Lucas's method [7]. Again, the estimated epipolar direction was adopted as the direction for depth reconstruction.

Table 1. The parameters of image sequences

	Size	f	Trans	Rot ($\times 10^{-3}$)
COKE	300×300	439.4	$(x_0, y_0) = (-25, 25)$	(0.6,0.6,4)
YNC	252×316	337.5	$(x_0, y_0) = (0, 59)$	(0.2,1.6,-0.2)
SOFA1	256×256	309.0	$(U, V, W) = (0.8, 0.6, 0)$	(-20.3,28.4,0)

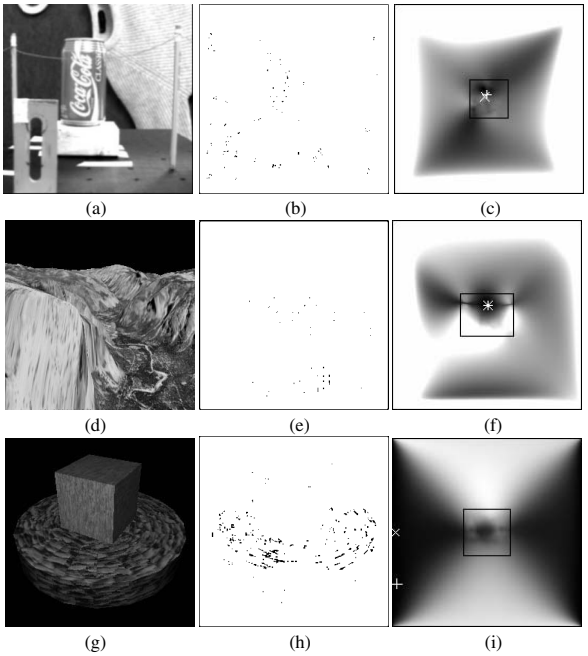

Figure 5. Typical images, optical flow fields and residual error images.

The residual error images were shown in Fig. 5 where each row corresponds to one image sequence (from top to bottom: COKE, YosemiteNoCloud and SOFA1). For each row, an image of the sequence, the optical flow field and the residual error image are shown from left to right. True FOEs and global minima of the residual error surfaces were highlighted by "+" and"×" on the residual error images respectively. Fig. 5(c) shows local minima strips along the average optical flow direction, especially outside the image plane. This was due to the effect of anisotropic noise as discussed above. As for the case of lateral motion in SOFA1, the anisotropic noise also influences the direction of the bas-relief valley. As can be seen in Fig. 5(i), the bas-relief valley was pulled towards the average optical flow direction, which was roughly horizontal. While the numerical values of the simulation data with synthetic images in Section 3 showed that under all configurations except the planar case, γ was invariably estimated with high accuracy for all the FOE candidates, this was not the case for real images. As far as α_e and β_e were concerned, the numerical values also showed that the corresponding RDir and RMag constraints were significantly modified, possibly due to the presence of non-isotropic noise.

5 Conclusion

We have developed a geometrically motivated motion error analysis method which is capable of depicting the topological structures of the various optimization cost functions. The motion error configurations likely to cause ambiguities were made clear, under noiseless and noisy conditions and under different motion types.

References

[1] M.J. Brooks, W. Chojnacki, and L. Baumela Determining the egomotion of an uncalibrated camera from instantaneous optical flow. *J. Opt. Soc. of America A*, 14(10):2670–2677, 1997.

[2] L-F. Cheong and T. Xiang. Characterizing Depth Distortion under Different Generic Motions. *IJCV*, 44:199–217, 2001.

[3] C. Fermüller, D. Shulman, and Y. Aloimonos. The statistics of Optical Flow. *CVIU*, 82:1–32, 2001.

[4] C. Fermüller, and Y. Aloimonos, 2000. Observability of 3D Motion. *IJCV* 37(1):43–63, 2000.

[5] D.J. Heeger and A.D. Jepson. Subspace methods for recovering rigid motion I: Algorithm and implementation. *IJCV* 7:95–117, 1992.

[6] K. Kanatani. 3-D Interpretation of Optical Flow by Renormalization. *IJCV* 11(3):267–282, 1993.

[7] B.D. Lucas. Generalized Image Matching by the Method of Differences. PhD Thesis, Carnegie-Mellon University, 1984.

[8] Y. Ma, J. Košecká, and S. Sastry. Linear Differential Algorithm for Motion Recovery: A Geometric Approach. *IJCV* 36(1):71–89, 2000.

[9] S. Soatto and R. Brockett. Optimal structure from motion: local ambiguities and global estimates. In *CVPR*, pp. 282–288, 1998.

[10] T. Xiang. *Understanding the Behavior of Structure From Motion Algorithms: A Geometric Approach*. Ph. D. dissertation, National University of Singapore, 2001.

Proceedings of the 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'03)
1063-6919/03 $17.00 © 2003 IEEE