International Journal of Computer Vision 51(2), 111–137, 2003

Understanding the Behavior of SFM Algorithms: A Geometric Approach

TAO XIANG AND LOONG-FAH CHEONG
Electrical and Computer Engineering Department, National University of Singapore, 10 Kent Ridge Crescent, Singapore 119260
txiang@dcs.qmul.ac.uk
eleclf@nus.edu.sg

Received November 11, 2001; Revised August 8, 2002; Accepted August 8, 2002

Abstract. We put forth in this paper a geometrically motivated motion error analysis which is capable of supporting investigation of global effect such as inherent ambiguities. This is in contrast with the usual statistical kinds of motion error analyses which can only deal with local effect such as noise perturbations, and where much of the results regarding global ambiguities are empirical in nature. 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. Given the erroneous 3-D motion estimates caused by the inherent ambiguities, it is also important to study the behavior of the resultant distortion in depth recovered under different motion-scene configurations. Such an investigation may alert us to the occurrence of ambiguities under different conditions and be more careful in picking the solution. Our formulation, though geometrically motivated, was also put to use in modeling the effect of noise and in revealing the strong influence of feature distribution. Experiments on both synthetic and real image sequences were conducted to verify the various theoretical predictions.

Keywords: structure from motion, error analysis, epipolar constraint, inherent ambiguity, depth distortion

1. Introduction

The estimation of the 3-D motion and structure is notorious for its noise sensitivity; a small amount of error in the image measurements can lead to very different solutions. Structure from motion (SFM) algorithms proposed in the past two decades faced this problem to varying extent which has led to many error analyses (Adiv, 1989; Daniilidis and Spetsakis, 1997; Weng et al., 1991; Young, 1992). To date, however, few of them have ever attempted to give a topological characterization of the residuals associated with different optimization criteria which would make explicit the configuration of the error surface, especially the distribution of the local minima of the cost functions. Ideally, such a characterization should consider the ambiguities under a full range of motion-scene configurations. The rationale for such a comprehensive description of the ambiguities is that since most SFM algorithms perform well only in restricted domains, it was important to evaluate the limits of applicability of these algorithms. That is, each algorithm should be evaluated specifically against likely problem conditions. If such understanding could be achieved, it then becomes possible to fuse the results of several SFM algorithms or to fuse the visual motion cues with other cues such as vestibular signals. This viewpoint has been expressed by Oliensis (2000a).

In this paper, we propose an approach that lends itself towards understanding the full behavior of SFM algorithms. 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 SFM algorithms using optical flow as input. What permits an unifying view of these

different algorithms is a new optimization criterion to be presented in this paper. It is based on the difference between the original optical flow and the reprojected flow obtained via a backprojection of the reconstructed depth, analogous to the distance between the observation and the reprojection of the reconstructed point in the discrete case (Zhang, 1998). We showed that the different weighted differential epipolar constraints used in the literature correspond to the different ways of reconstructing depth using the optimization criterion presented in this paper. Thus this criterion also lends a geometric interpretation to the various weights used. More importantly, it 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. The result is that the inherent ambiguities in both translation and rotation estimates are identified; how the likelihood of these ambiguities varies with the scene and the motion types such as lateral or forward motion is also made apparent. To round off this section on the investigation of motion ambiguity, we extend our analysis to include the effect of noise in the image measurements, using both the isotropic and the anisotropic noise models. Our investigation unravels the impact a realistic anisotropic noise distribution can have on the topology of the cost functions.

The behavioral description of SFM algorithms would not be complete without saying something about how the depths would be recovered given such motion ambiguities. As a consequence of the motion ambiguities, the estimated 3-D motion parameters contain errors; thus the reconstructed depth would be a distorted version of the physical depth. The need to characterize such depth distortion arising from errors in the motion estimates prompted the work of Cheong et al. (1998), which gave an account of the systematic nature of the errors in the depth estimates via the so-called iso-distortion framework. It showed that the most general description of such a transformation from the physical to the perceived space is very complicated, belonging to the family of Cremona transformations. The work of Cheong and Xiang (2001) built upon that framework and considered the depth distortion under two generic types of motion, namely, lateral and forward motion, with a view to obtain robust recovery of depth information. In this paper, we use the same framework to study the special properties of depth distortion when the spurious motion estimates are caused by the inherent ambiguities associated with any general motion. The consequence of these distortion enables us to explain some well-known human perceptual illusions such as the "rotating cylinder illusion". Correlatively, if human suffers from such distortion and yet can perform many tasks efficiently, it is hoped that a deeper understanding of the distortion would help to emulate human in these performances. The understanding of such distortion may also be useful for other purposes such as alerting one to the occurrence of ambiguities, thereby allowing us to pick up the true solution.

1.1. Relation to Previous Work

The 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 the 3-D motion estimation algorithms inevitably contains errors. In view of such errors, most of the previous error analysis on 3-D motion estimation (Adiv, 1989; Daniilidis and Spetsakis, 1997; Weng et al., 1991; Young, 1992; Heeger and Jepson, 1992; Maybank, 1993) 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 (Adiv, 1989; Daniilidis and Spetsakis, 1997; Weng et al., 1991; Young, 1992; Heeger and Jepson, 1992; Maybank, 1993), or given as empirical figures (Dutta and Snyder, 1990) through some simulations. A comprehensive survey of such analysis was given by Daniilidis and Spetsakis (1997). Several results have been established by such analysis:

- Maybank (1993) and Heeger and Jepson (1992) established the result that the plane defined by the true translation and the optical axis can be determined by most SFM algorithms reliably. They obtained this result based on strict assumptions such as infinitesimal field of view (FOV). The finding is closely related to the bas-relief ambiguity obtained in this paper, although we do not need the field of view assumption.
- If the field of view is small or depth variation is insufficient, rotation about an axis parallel to the image

plane can easily be confounded with lateral translations. This has been demonstrated through both theoretical work and experimental study (Daniilidis and Spetsakis, 1997).
- The estimated translation is biased towards the viewing direction if the error metric is not appropriately normalized.

Little work has been contributed to a systematical characterization of the topology of the cost functions. Recently, however, several studies have emerged in this direction. Soatto and Brockett (1998) and Chiuso et al. (2000) 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 papers), and attributed it to the presence of noise, although the simulation results showed that the minimum persisted with noiseless input. Ma et al. (2001), adopting the discrete approach, unified different optimization criteria under an "optimal triangulation" procedure and analyzed the impact of noise on the ambiguities for different optimization criteria. They characterized the behavior of the critical points under noise by making use of the properties of the so-called "essential manifold".

The major difference between our work and the preceding work lies in the fact that we highlight the importance of the inherent ambiguities of the SFM problem itself, without considering the effect of noise initially. Indeed, all the major ambiguities identified in the literature can be accounted for by such noiseless consideration. We argue that while dealing with the statistical adequacy of the various criteria is important, it is equally important to understand the detailed nature of the inherent ambiguities which is caused by the geometry of the problem itself and thus cannot be removed by any statistical schemes (relieved, yes). In this respect, the work of Fermüller and Aloimonos (2000) and Oliensis (2000b, 2001) are the closest in spirit to our work. Not surprisingly, there are many common findings, though there are some aspects that are different too.

The work of Fermüller and Aloimonos (2000) presented a geometrical-statistical investigation of the observability of 3-D motion. They studied the conditions on the errors in the motion estimates for the local minima on error surface to arise. The cost functions are expressed in terms of the true motion parameters and the errors in the estimated motion parameters. Our work adopted similar notations but used very different method of analysis. Various assumptions were required in their work such as random distribution of feature points over the image plane and random depths over the 3-D space. They also assumed small FOV and neglected all the second order flow terms caused by the rotational parameters. The epipolar constraint considered in their work was unweighted, which, together with those assumptions led to some results that were different from ours. In particular, it was shown that when all the motion parameters are estimated simultaneously, the solution for the focus of expansion (FOE) can have a local minima at the image center, which is obviously due to the unweighted epipolar constraint. Our result shows that if the epipolar constraint is properly weighted, this minima should not occur, unless some specific motion-scene configuration such as forward motion arises, in which case the true minimum will also be at the image center. Indeed, our paper considers a variety of motion-scene configurations which are not studied in Fermüller and Aloimonos (2000). Finally, the kind of noise considered in Fermüller and Aloimonos (2000) was found to have no influence on the overall structure of the cost functions. Our theoretical analysis and experimental results show that a more realistic noise model often has significant impact on the cost function behavior.

The objective of Oliensis' work (Oliensis, 2000b, 2001) is very similar to ours, that is, to characterize the overall error surface via an explicit analytical model. However, the means through which we achieve the end are quite different. Our approach is much simpler, allowing us to achieve an intuitive grasp of the geometric nature of the ambiguities. For instance, we explain the formation of the local minimum at the opposite end of the bas-relief valley (termed as flipped minimum in Oliensis' work) through the coupling of the rotational and the translational motions. Our more intuitive formulation renders it more suitable for analysis under a wider range of motion-scene configurations; specifically, we focus on different types of translational motions, ranging from purely forward motion to purely lateral motion. Other factors that influence the error surface, such as the distribution of feature points and the scene structure, are also studied in a more systematic and detailed manner. Oliensis' work concentrated only on the error surface for translation estimation; the corresponding ambiguities in the rotation estimates are

not made explicit. Our work fully characterizes the error surfaces for both the rotational and the translational parameters.

Finally, in contrast to our work, all the preceding works focus on motion estimation and devote much less attention on the closely related problem of depth estimation. While some of the works (Weng et al., 1991; Szeliski and Kang, 1997; Grossmann and Victor, 2000) predicted the sensitivity of the depth estimates to small amounts of image noise, the situation where the errors in the depth estimates arise from the erroneous 3-D motion parameters has not been dealt with, except in the case of critical surface pairs (Horn, 1987; Negahdaripour, 1989).

1.2. Organization

The organization of this paper is as follows. First, we briefly review in Section 2 the optimization criteria in both the discrete and the differential cases. We then introduce the notions of the iso-distortion framework and discuss how it can be used to address the reliability of depth recovery. In Section 3, the differential reprojection criterion is proposed to unify the various criteria based on differential epipolar constraints. We then seek to characterize the various inherent ambiguities in 3-D motion estimation and the corresponding depth distortion properties under these ambiguity configurations. We employ a cost function visualization method to visualize the topology of the cost functions, so as to both verify the various theoretical predictions and to reveal further properties of the cost functions. Based on such understanding, we are able to explain some well-known human visual illusions. We characterize the role of the measurement noise on the behavior of SFM algorithms in Section 4. In particular, the global effects of isotropic and anisotropic noise distribution are studied. These are followed by experiments on real images to verify the various predictions made and to study the feasibility of a more robust algorithm based on the topology of the cost functions. The paper ends with the conclusions of the work.

2. Background and Prerequisite

2.1. Model and Notations

In this paper, we denote the estimated parameters with the hat symbol ($\hat{}$) and errors in the estimated parameters with the subscript e (where error of any estimate s is defined as $s_e = s - \hat{s}$). We use bold lower-case character to denote vector and bold upper-case character to denote matrix. Unless otherwise stated, vectors are column vectors. 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 $\mathbf{s}$. The symbol $(\cdot)$ represents the dot product of vectors. 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. The symbol $(\|\|)$ represents the Euclidean norm of a vector and the symbol $(|\ |)$ the absolute value of a variable.

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 pixels 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 (Longuet-Higgins, 1981):

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

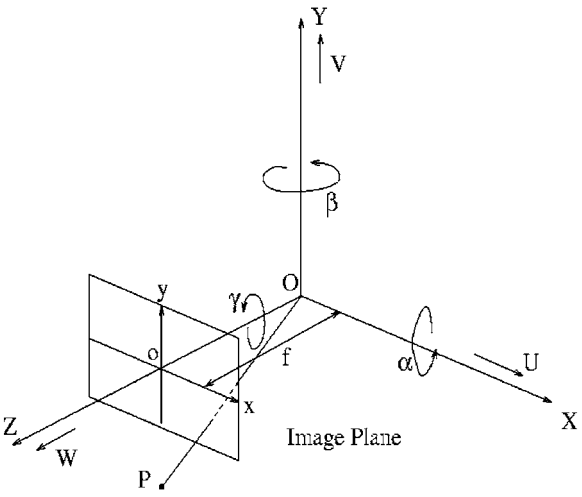

Figure 1. The image formation model.

where

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

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

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

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.

2.2. 3-D Motion Estimation

2.2.1. Discrete Case. The SFM problem in the discrete case amounts to the estimation of the fundamental matrix **F** (or the Essential matrix **E** if the camera is calibrated) based on a sufficiently large set of point correspondences. The geometry of the discrete two-image motion analysis has been well studied and is succinctly captured by the epipolar equation:

$$\mathbf{p}_1 \mathbf{F} \mathbf{p}_2 = 0 \tag{3}$$

where $\mathbf{p}_1$ and $\mathbf{p}_2$ are the corresponding points on the two images and **F** is the fundamental matrix. The geometric meaning of the epipolar equation is that $\mathbf{p}_1$ must lie on the epipolar line of $\mathbf{p}_2$ given by $\mathbf{F}\mathbf{p}_2$. Directly minimizing Eq. (3) leads to a closed-form solution whose results are sensitive to noise. It was also argued (Zhang, 1998) that this optimization criterion does not totally reflect the epipolar geometry and thus is not physically meaningful. A couple of non-linear optimization criteria were thus proposed. Three criteria are derived based on: distance between points and epipolar lines, gradient-weighted epipolar errors and distances between points and their reprojections. The corresponding cost function are denoted as J_{D1}, J_{D2} and J_{D3} respectively. The geometric meanings of J_{D1} and J_{D3} are obvious from their names, whereas J_{D2} is obtained based on the following statistical consideration: When independent and identically distributed Gaussian noise is assumed, minimizing the Mahalanobis distance between points and epipolar lines gives rise to J_{D2}. While J_{D3} has obvious geometric meaning, it also possesses a statistical interpretation: It corresponds to the case of optimizing based on the maximum *a posterior* (MAP) principle under the same noise model. Zhang (1998) studied the relationship between these three criteria under different motion configurations. J_{D2} was recommended since it is equivalent to J_{D3} under most configurations and yet is computationally more efficient. Ma et al. (2001) investigated the behavior of different criteria with and without noise. Similar to Zhang's results, these criteria were shown to be intimately related and were unified under a new "optimal triangulation" procedure (our proposed unifying scheme for the differential case is similar in idea).

2.2.2. Differential Case. A motion estimation algorithm based on the differential epipolar constraint can be developed analogous to the discrete case, from which the following cost function can be obtained (Brooks et al., 1997):

$$J_{E1} = \sum_{i=1}^{n} \left(\mathbf{p}_i^T \bar{\hat{\mathbf{v}}} \dot{\mathbf{p}}_i + \mathbf{p}_i^T \bar{\hat{\mathbf{v}}} \bar{\hat{\mathbf{w}}} \mathbf{p}_i\right)^2 \tag{4}$$

where n is the number of image velocity measurement. As discussed before, 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 (4). The actual case of the optical flow containing noise component would be addressed in later sections.

The geometric meaning of J_{E1} is that in the image plane the vector $[\dot{\mathbf{p}}_i]_2 - \hat{\dot{\mathbf{p}}}_{rot_i}$ (the de-rotated flow) should be parallel to the vector $\hat{\dot{\mathbf{p}}}_{tr_i}$, or equivalently, perpendicular to $\hat{\dot{\mathbf{p}}}^{\perp}_{tr_i}$. Thus the cost function J_{E1} can

also be expressed as:

$$J_{E1} = \sum_{i=1}^{n} \left(\left([\dot{\mathbf{p}}_i]_2 - \hat{\dot{\mathbf{p}}}_{rot_i} \right) \cdot \hat{\dot{\mathbf{p}}}_{tr_i}^{\perp} \right)^2 \tag{5}$$

Minimizing the preceding amounts to a linear optimization problem which can be solved by a linear least square method. However, it faces the same problem as its discrete counterpart. A well-studied bias of the linear algorithms is that the estimated translation will be biased towards the image center. In view of this bias, a statistically more adequate implementation of the differential epipolar constraint should be:

$$\begin{aligned} J_{E2} &= \sum_{i=1}^{n} \left(\frac{\left([\dot{\mathbf{p}}_i]_2 - \hat{\dot{\mathbf{p}}}_{rot_i} \right) \cdot \hat{\dot{\mathbf{p}}}_{tr_i}^{\perp}}{\left\| [\dot{\mathbf{p}}_i]_2 - \hat{\dot{\mathbf{p}}}_{rot_i} \right\| \left\| \hat{\dot{\mathbf{p}}}_{tr_i}^{\perp} \right\|} \right)^2 \\ &= \sum_{i=1}^{n} (\sin \theta_i)^2 \end{aligned} \tag{6}$$

where θ_i is the angle between $\hat{\dot{\mathbf{p}}}_{tr_i}$ and $[\dot{\mathbf{p}}_i]_2 - \hat{\dot{\mathbf{p}}}_{rot_i}$.

The counterpart of J_{E2} in the discrete case would be J_{D1}. Both of them are non-linear and involve heavy computation to obtain their solutions. Like the discrete case, there are a variety of other non-linear methods which are basically different weighted versions of J_{E1} each driven by slightly different considerations. Some of these methods derived the weight based on a statistical analysis of noise (Kanatani, 1993; Ma et al., 2000). For instance, Ma et al. (2000) presented a normalized epipolar constraint which would yield MAP estimates given an independent and identically distributed Gaussian noise:

$$J_{E3} = \sum_{i=1}^{n} \left(\frac{\mathbf{p}_i^T \bar{\hat{\mathbf{v}}} \dot{\mathbf{p}}_i + \mathbf{p}_i^T \bar{\hat{\mathbf{v}}} \bar{\hat{\mathbf{w}}} \mathbf{p}_i}{\left\| \hat{\dot{\mathbf{p}}}_{tr_i} \right\|} \right)^2 \tag{7}$$

Others like Brooks et al. (1998) presented an algebraic-geometric point of view: the hyperpoint $[\mathbf{p}_i, \dot{\mathbf{p}}_i]^T$ should lie on the hypersurface $\mathbf{p}_i^T \bar{\hat{\mathbf{v}}} \dot{\mathbf{p}}_i + \mathbf{p}_i^T \bar{\hat{\mathbf{v}}} \bar{\hat{\mathbf{w}}} \mathbf{p}_i = 0$. Like the gradient-weighted epipolar error in the discrete case, they proposed a cost function which minimizes the first order approximation of the Euclidean distance between the hyperpoint and the hypersurface:

$$J_{E4} = \sum_{i=1}^{n} \frac{\left(\mathbf{p}_i^T \bar{\hat{\mathbf{v}}} \dot{\mathbf{p}}_i + \mathbf{p}_i^T \bar{\hat{\mathbf{v}}} \bar{\hat{\mathbf{w}}} \mathbf{p}_i \right)^2}{\| 2 \bar{\hat{\mathbf{v}}} \bar{\hat{\mathbf{w}}} \mathbf{p}_i + \bar{\hat{\mathbf{v}}} \dot{\mathbf{p}}_i \|^2 + \| \bar{\hat{\mathbf{v}}} \mathbf{p}_i \|^2} \tag{8}$$

This cost function is very similar to Kanatani's renormalization criterion (Kanatani, 1993) which is based on statistical consideration.

As in the discrete case, one can ask what is the geometric meaning of these various criteria, beside their statistical interpretation? The differential reprojection criterion to be developed later allows us to answer this question.

2.3. Depth Estimation

3-D motion estimation is regarded as the first step towards the full recovery of 3-D shape information from 2-D measurements. Therefore any error in the 3-D motion estimates will systematically affect the perceived space. However, the reliability of the depth estimates could have quite different behavior from that of 3-D motion estimates. That is, motion-scene configuration that allows robust motion recovery may yield less than desirable depth estimates, and vice versa. Another substantive question is of course, whether there is any interaction between the errors in the motion estimates and the corresponding distortion in the recovered depth. That is, would the distortion in the perceived space in turn affect motion estimation? Partially to address these questions, the iso-distortion framework was introduced in Cheong et al. (1998). Let us first revisit some notations that would be useful for this paper.

2.3.1. Iso-Distortion Framework. The iso-distortion framework seeks to understand the geometric laws under which the recovered scene is distorted due to some errors in the estimated motion parameters. The distortion in the perceived space is visualized by looking at the locus of constant distortion, known as the iso-distortion surfaces.

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

$$\hat{Z} = -\frac{\hat{\mathbf{v}}^T \mathbf{Q}_{\mathbf{p}}^T \underline{\mathbf{n}}}{\left(\dot{\mathbf{p}}^T - \mathbf{p}^T \bar{\hat{\mathbf{w}}} \mathbf{Q}_{\mathbf{p}}^T \right) \underline{\mathbf{n}}} \tag{9}$$

where $\mathbf{n}$ is a unit vector in the image plane representing a direction. In general, when the estimated motion parameters contain errors, different choices of $\mathbf{n}$ will give rise to different reconstructions. One possibility is to recover depth by setting $\mathbf{n}$ to be along the estimated epipolar direction, which is the direction pointing from the

image feature point to the estimated FOE; this scheme is heretoforth named as the "epipolar reconstruction" scheme. It is based on the intuition that the epipolar direction contains the strongest translational flow and hence represents the best direction for depth recovery. Another possibility is to let $\mathbf{n}$ be the image intensity gradient direction, based on the intuition that the normal flow can be recovered reliably. Various other possibilities exist, each can be given different geometric or statistical interpretation.

Substituting the expression for the true flow $\dot{\mathbf{p}}^T$ (using Eq. (1) again) into Eq. (9), we have:

$$\hat{Z} = Z\left(\frac{-\hat{\mathbf{v}}^T\mathbf{Q}_\mathbf{p}^T\underline{\mathbf{n}}}{-\mathbf{v}^T\mathbf{Q}_\mathbf{p}^T\underline{\mathbf{n}} + Z\left(\mathbf{p}^T\bar{\mathbf{w}}_e\mathbf{Q}_\mathbf{p}^T\underline{\mathbf{n}}\right)}\right) \tag{10}$$

From the above equation we can see that $\hat{Z}$ is related to Z through a multiplicative factor given by the terms inside the bracket, which we denote by D and term as the distortion factor:

$$D = \frac{-\hat{\mathbf{v}}^T\mathbf{Q}_\mathbf{p}^T\underline{\mathbf{n}}}{-\mathbf{v}^T\mathbf{Q}_\mathbf{p}^T\underline{\mathbf{n}} + Z\left(\mathbf{p}^T\bar{\mathbf{w}}_e\mathbf{Q}_\mathbf{p}^T\underline{\mathbf{n}}\right)} \tag{11}$$

For specific values of $\mathbf{v}$, $\hat{\mathbf{v}}$ and $\mathbf{w}_e$ and for any fixed distortion factor D, Eq. (11) describes a surface $g(x, y, Z) = 0$ in the xyZ-space, which we call an iso-distortion surface. This iso-distortion surface has the obvious property that points lying on it are distorted in depth by the same multiplicative factor D. The systematic nature of the distortion can then be made clear by looking at the organization of these iso-distortion surfaces.

The geometric laws for distortion can also be characterized algebraically as a distortion transformation from the physical space to the perceived space. In general, the resulting distortion transformation is a Cremona transformation (Cheong and Ng, 1999) whose properties are quite complex. However, under special cases, as are some of the ambiguity cases to be discussed later, the transformation reduces to that of the projective transformation with some nice depth properties.

2.3.2. Depth Error Sensitivity Under Different Motion Configurations. The iso-distortion framework has been used to seek some generic motion types that rendered depth recovery more robust and reliable (Cheong and Xiang, 2001). Lateral and forward motions were compared both under calibrated and uncalibrated scenarios. The fundamental conclusions are that under lateral movement (possibly coupled with rotation) and certain conditions, while it might be very difficult to resolve the ambiguity between translation and rotation, ordinal depth can be recovered with robustness, whereas for forward motion, the depth recovery is too sensitive to errors to admit meaningful scene reconstruction.

The preceding conclusions were established without imposing any constraints on the motion error. However, it is evident that the values of these motion errors are not arbitrary. Rather, ambiguities inherent in the SFM algorithms are likely to impose further constraints on these motion errors. Given these errors, what can be said about the distortion in depth given any types of translational motion? This will be addressed in the next section.

3. Differential Reprojection Criterion and its Error Surface

3.1. Differential Reprojection Criterion

From the geometric standpoint, the differential epipolar constraint in Eq. (4) is a "weak" constraint in the sense that it can be satisfied by any two vectors $([\dot{\mathbf{p}}_i]_2 - \hat{\dot{\mathbf{p}}}_{rot_i})$ and $\hat{\dot{\mathbf{p}}}_{tr_i}$ that are parallel to each other. There is no requirement on the magnitudes of the two vectors although the true estimate should satisfy: $[\dot{\mathbf{p}}_i]_2 - \frac{\hat{\dot{\mathbf{p}}}_{tr_i}}{\hat{Z}} - \hat{\dot{\mathbf{p}}}_{rot_i} = 0$. Thus a "stronger" and more adequate criterion based on the idea of reprojection is proposed; it is based on the difference between the original optical flow and the reprojected flow obtained via a backprojection of the reconstructed depth. It is thus analogous to J_{D3} in the discrete case. Furthermore, similar to Ma et al.'s "optimal triangulation" in the discrete case (Ma et al., 2001), we will see in this section that this criterion unifies the various weighted versions of the differential epipolar constraint and also lends a geometric interpretation to the weights used. More importantly it allows us to develop a geometric treatment of the motion ambiguity conditions, as we shall see in Section 3.2.

Substituting the recovered depth in Eq. (9) into Eq. (1), we obtain the reprojected (estimated) flow field,

denoted by $\hat{\dot{\mathbf{p}}}$, as follows:

$$\hat{\dot{\mathbf{p}}} = \frac{\mathbf{Q_p}\hat{\mathbf{v}}\left(\dot{\mathbf{p}}^T - \mathbf{p}^T\bar{\hat{\mathbf{w}}}\mathbf{Q_p}^T\right)\underline{\mathbf{n}}}{\hat{\mathbf{v}}^T\mathbf{Q_p}^T\underline{\mathbf{n}}} - \mathbf{Q_p}\bar{\hat{\mathbf{w}}}\mathbf{p} \tag{12}$$

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

$$\begin{aligned}\dot{\mathbf{p}}_e &= \dot{\mathbf{p}} - \hat{\dot{\mathbf{p}}} \\ &= \frac{\left(\mathbf{C_1} - \mathbf{C_1}^T + \mathbf{C_2} - \mathbf{C_2}^T\right)\underline{\mathbf{n}}}{\hat{\mathbf{v}}^T\mathbf{Q_p}^T\underline{\mathbf{n}}} \\ &= \frac{\mathbf{C}\underline{\mathbf{n}}}{\hat{\mathbf{v}}^T\mathbf{Q_p}^T\underline{\mathbf{n}}}\end{aligned} \tag{13}$$

where $\mathbf{C_1} = \dot{\mathbf{p}}\hat{\mathbf{v}}^T\mathbf{Q_p}$, $\mathbf{C_2} = \mathbf{Q_p}\bar{\hat{\mathbf{w}}}\mathbf{p}\hat{\mathbf{v}}^T\mathbf{Q_p}^T$, $\mathbf{C} = (\mathbf{C_1} - \mathbf{C_1}^T + \mathbf{C_2} - \mathbf{C_2}^T)$ and $\mathbf{C_1}, \mathbf{C_2}, \mathbf{C}$ are all 3×3 matrices. It can be easily shown that $\mathbf{C}$ is a skew-symmetrical matrix. Thus we have:

$$\underline{\mathbf{n}}^T\mathbf{C}\underline{\mathbf{n}} = 0$$

The above equation implies that along $\mathbf{n}$, the direction of depth recovery, the reprojected flow $\hat{\dot{\mathbf{p}}}$ has exactly the same component as the original optical flow $\dot{\mathbf{p}}$. As a consequence:

$$\|\dot{\mathbf{p}}_e\| = \left|\dot{\mathbf{p}}_e^T\underline{\mathbf{n}}^\perp\right| \tag{14}$$

If we define a cost function J_R based on the reprojected flow difference, it can be written as:

$$J_R = \sum_{i=1}^{n}\left(\frac{\|\mathbf{C}_i\underline{\mathbf{n}_i}\|}{\hat{\mathbf{v}}^T\mathbf{Q_{p_i}}^T\underline{\mathbf{n}_i}}\right)^2$$

Using Eq. (14), J_R can be written as:

$$\begin{aligned}J_R &= \sum_{i=1}^{n}\left(\frac{\underline{\mathbf{n}_i}^T\mathbf{C}_i\underline{\mathbf{n}_i}^\perp}{\hat{\mathbf{v}}^T\mathbf{Q_{p_i}}^T\underline{\mathbf{n}_i}}\right)^2 \\ &= \sum_{i=1}^{n}\left(\frac{\mathbf{p}_i^T\bar{\hat{\mathbf{v}}}\dot{\mathbf{p}}_i + \mathbf{p}_i^T\bar{\hat{\mathbf{v}}}\bar{\hat{\mathbf{w}}}\mathbf{p}_i}{\hat{\mathbf{v}}^T\mathbf{Q_{p_i}}^T\underline{\mathbf{n}_i}}\right)^2\end{aligned} \tag{15}$$

A comparison of Eq. (4) with Eq. (15) reveals the relationship between J_R and J_{E1}: J_R is a weighted version of J_{E1} with the weight given by the projection of $\hat{\dot{\mathbf{p}}}_{tr_i}$ on the direction $\mathbf{n}_i$. It follows that J_R can also be written as:

$$J_R = \sum_{i=1}^{n}\left(\frac{\hat{\dot{\mathbf{p}}}_{tr_i}\cdot\left([\dot{\mathbf{p}}_i]_2 - \hat{\dot{\mathbf{p}}}_{rot_i}\right)^\perp}{\hat{\dot{\mathbf{p}}}_{tr_i}\cdot\mathbf{n}_i}\right)^2 \tag{16}$$

Various weighted differential epipolar constraints differ mainly in the choice of $\mathbf{n}$. Possible choices of $\mathbf{n}$ include the "epipolar reconstruction" direction ($\mathbf{n} = \frac{\hat{\dot{\mathbf{p}}}_{tr}}{\|\hat{\dot{\mathbf{p}}}_{tr}\|}$), which results in J_{E3}, gradient direction (the normal flow approach), or the Linear Least Square Reconstruction (LLSR) direction.[1] Other more simplistic choices include constant direction, random direction, etc.

It follows that while the formulation of the differential reprojection criterion J_R is motivated by the need to have a stronger geometric constraint, it often has statistical meaning too. Furthermore J_R can be seen as a scheme unifying the various weighted epipolar constraints. It follows that 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.

3.2. Analyzing from a Geometric Point of View

To analyze how various factors, such as motion types, field of view, feature and depth distribution, govern the formation of motion ambiguities (or equivalently, the local minima in the error surface described by J_R), we need to express J_R in terms of the various component errors in the 3-D motion estimates. This allows us to obtain a more obliging form for analyzing in more specific details the ambiguity behavior over a wide range of conditions. Substituting $[\dot{\mathbf{p}}_i]_2 = (u_i, v_i)^T = (\frac{x_i - x_0}{Z_i} + u_{rot_i}, \frac{y_i - y_0}{Z_i} + v_{rot_i})^T$, $\hat{\dot{\mathbf{p}}}_{tr_i} = (x_i - \hat{x}_0, y_i - \hat{y}_0)^T$ and $\hat{\dot{\mathbf{p}}}_{rot_i} = (\hat{u_{rot_i}}, \hat{v_{rot_i}})^T$ into (16), we have:

$$\begin{aligned}&J_R \\ &= \sum\left(\frac{(x - \hat{x}_0, y - \hat{y}_0)\cdot\left(v_{rot_e} - \frac{y_{0_e}}{Z}, \frac{x_{0_e}}{Z} - u_{rot_e}\right)}{(x - \hat{x}_0, y - \hat{y}_0)\cdot\mathbf{n}}\right)^2\end{aligned} \tag{17}$$

where

$$\begin{aligned}(x_{0_e}, y_{0_e}) &= (x_0 - \hat{x}_0, y_0 - \hat{y}_0)\\ (u_{rot_e}, v_{rot_e}) &= \left(\frac{\alpha_e xy}{f} - \beta_e\left(\frac{x^2}{f} + f\right) + \gamma_e y,\right.\\ &\qquad \left.\alpha_e\left(\frac{y^2}{f} + f\right) - \frac{\beta_e xy}{f} - \gamma_e x\right)\end{aligned}$$

For notational convenience, we omit the subscript i in the expression of J_R; it is understood that the summation runs over all feature points. To facilitate discussion, we also introduce the following notations. We denote the expression contained in the outer bracket of (17) as $\dot{\mathbf{p}}_e(\hat{\mathbf{v}}, \mathbf{v}, \mathbf{w}_e)$ (where the dependence of $\dot{\mathbf{p}}_e$ on the motion errors has been made explicit), 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 (it is indeed the interaction between $\mathbf{t_1}$ and $\mathbf{t_2}$ that accounts for much of the inherent motion ambiguities). 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 nth order component with respect to x and y; thus we have:

$$\begin{cases}\mathbf{t_1} &= \mathbf{t_{1,0}} + \mathbf{t_{1,1}}\\ \mathbf{t_2} &= \mathbf{t_{2,0}} + \mathbf{t_{2,1}} + \mathbf{t_{2,2}} + \mathbf{t_{2,Z}}\end{cases} \tag{18}$$

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 $(-\frac{y_{0_e}}{Z}, \frac{x_{0_e}}{Z})^T$. The depth Z may be dependent on x and y in a complex manner; thus we use the notation $\mathbf{t_{2,Z}}$ and leave the order unspecified.

To visualize the residual error surface, it is easier to deal with a 3-dimensional 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. Unless otherwise stated, the error surface in this paper refers to this type of error surface.

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" $(\frac{x}{Z}, \frac{y}{Z})$. The latter assumption generally requires that the distribution of depths are independent of the corresponding image co-ordinates x and y. Later we will see how the error surface will be affected when these assumptions do not hold.

3.2.1. Several General Observations. Equation (17) shows that for any given data set (x, y, Z), the residual error is a function of the true FOE (x_0, y_0), the estimated FOE $(\hat{x}_0, \hat{y}_0)$ and the error in the rotation estimates $(\alpha_e, \beta_e, \gamma_e)$. 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: (1) making $\|\mathbf{t_2}\|$ small and (2) 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 one) contributes towards ambiguity. 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, then 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

$$\frac{x_0}{y_0} = \frac{\hat{x}_0}{\hat{y}_0} \tag{19}$$

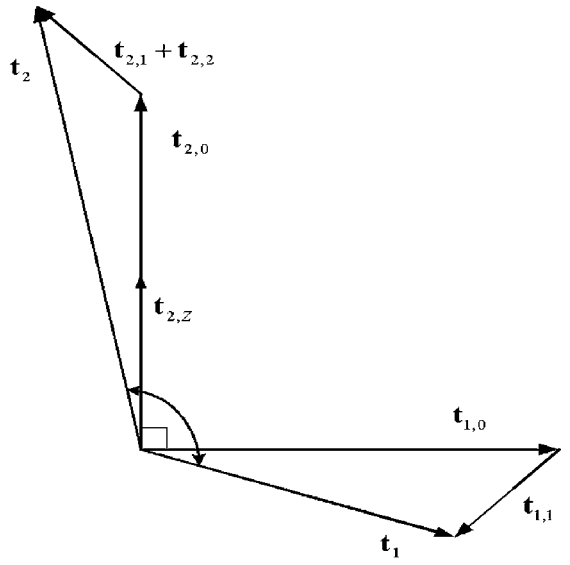

Figure 2. Geometry of $\mathbf{t_1}$ and $\mathbf{t_2}$.

and

$$\frac{\alpha_e}{\beta_e} = -\frac{\hat{y}_0}{\hat{x}_0} \tag{20}$$

Equation (19) 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. Equation (20) imposes a constraint on the direction of $\mathbf{w}_e$, which shall be henceforth referred to as the Rotation Error Direction (RDir) constraint.

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, which is:

$$\begin{cases} \alpha_e = \dfrac{y_{0_e}}{Z_{avg} f} \\ \beta_e = -\dfrac{x_{0_e}}{Z_{avg} f} \end{cases} \tag{21}$$

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 (20), it is evident that RDir and RMag can hold simultaneously only if TDir also holds. Note that an exact cancelation of $\mathbf{t_{2,0}}$ and $\mathbf{t_{2,Z}}$ is impossible unless the scene can be modeled as a frontal-parallel plane.

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.

From the preceding observations, we can establish the following conclusions:

1. *Translation estimates*. One of the well known phenomenon in motion perception is the bas-relief ambiguity. Basically it amounts to a valley on the translation error surface, along a straight line that is 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. TDir is the direct reason for the formation of such a valley.
2. *Rotation estimates*. Of the three rotational estimates, any error in $\hat{\gamma}$ would have purely deleterious effect on the minimization of $\|\dot{\mathbf{p}}_e(\hat{\mathbf{v}}, \mathbf{v}, \mathbf{w}_e)\|$. Thus, in the case of noiseless flow field, we expect accurate estimation of γ (those experienced in the art of the SFM algorithms will know that this is often not the case in numerical practice). The effects of α_e and β_e on the residual error are more complex. On the one hand, given a FOE error, α_e and β_e that satisfy the RMag constraint can make $\|\mathbf{t_2}\|$ small, thus leading to small residual error. On the other hand, the RMag2 constraint would prefer α_e and β_e to be small. Furthermore,

a FOE estimate that is close to the origin weakens the RMag2 constraint on α_e and β_e. These effects will determine the values of the rotation estimates that minimize J_R, and the rotation estimates will in turn influence the shape of the bas-relief valley.

3.2.2. Error Profile Along the Bas-Relief Valley. We have established in the preceding section the existence of the bas-relief valley; the variation of the error along the bas-relief valley itself is the subject of this section. First and foremost, the location of the true FOE has a critical influence on the shape of the bas-relief valley and this in turn has implication for any motion algorithm trying to deal with a wide range of translational motion. We will use an example to elucidate the influence of the true FOE location. Figure 3(a) illustrates the case where the true FOE (x_0, y_0) lies somewhere in between the image center and infinity. The dotted line in the figure corresponds to the bas-relief valley. As the estimated FOE leaves the true FOE and moves along the bas-relief valley, we can identify several factors that influence the outcome of Eq. (17) for J_R.

1. *Translational error*. Consider first the effects of the translational terms by setting the rotational errors to zero. As we vary the estimated FOE along the bas-relief valley, we study the $\|\dot{\mathbf{p}}_e(\hat{\mathbf{v}}, \mathbf{v}, \mathbf{w}_e)\|^2$ curve for a single feature point (x, y). As shown by the dashed curve in Fig. 3(b), there would be two

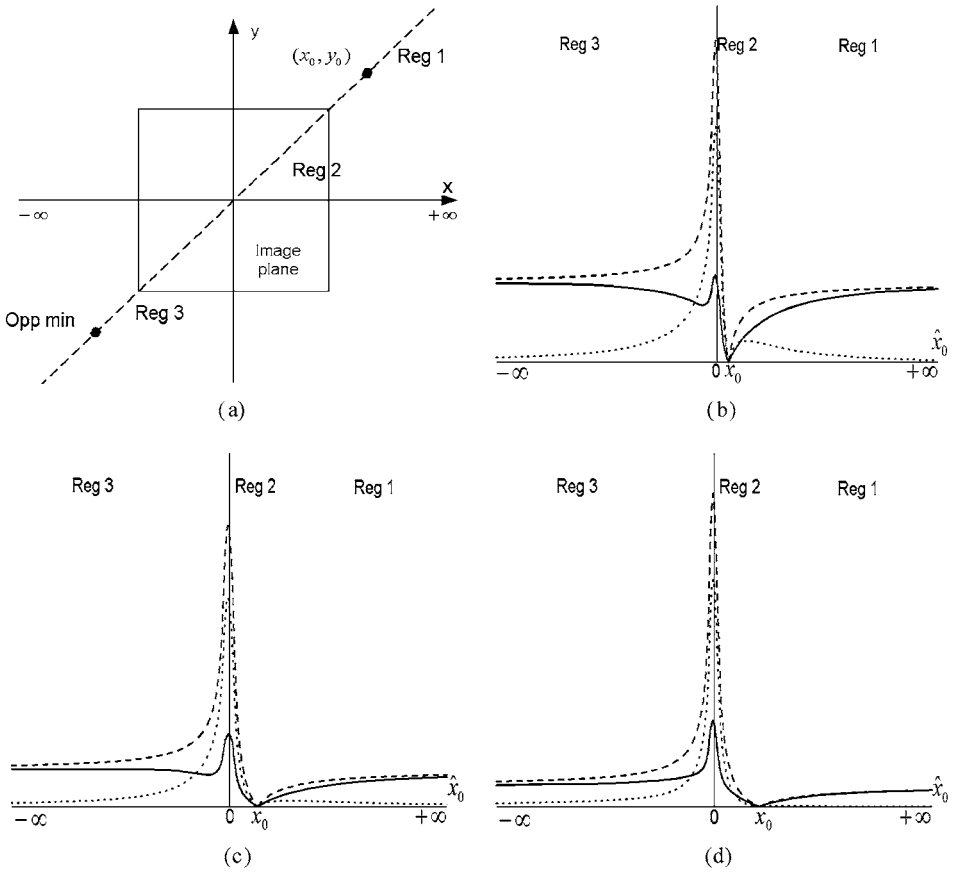

Figure 3. Configuration within the bas-relief valley when the true FOE is out of the image plane. (a) Overall configuration in the image plane; along the bas-relief valley, we divide it into regions 1, 2, and 3 as shown. (b), (c), and (d) show the residual errors along the bas-relief valley with increasing amount of lateral translation. The dashed line represents the $\|\dot{\mathbf{p}}_e(\hat{\mathbf{v}}, \mathbf{v}, \mathbf{w}_e)\|^2$ curve for a particular point, the dotted line the RotComp curve, and the solid line the overall J_R curve.

turning points on the curve; they correspond to a minimum where the estimated FOE coincides with the true FOE, and a maximum whose location depends on the value of (x, y). In particular, it can easily be shown (Xiang, 2001) that the location of the maximum depends on the position of the projection of (x, y) on the bas-relief line relative to the true FOE; if it falls on the left side of the true FOE, the maximum will be on the left, and vice versa. Finally, it is also clear that as the estimated FOE approaches infinity on either end of the bas-relief line, the curve would approach asymptotically towards a constant value. The total effect of the translational terms would be obtained by summing up all the $\|\dot{\mathbf{p}}_e(\hat{\mathbf{v}}, \mathbf{v}, \mathbf{w}_e)\|^2$ curves for each feature point. Let us denote this summed curve as the TrErr (Translational Error) curve. In this particular example of Fig. 3(a) (and for most true FOE not near the image center), all (or most of) the feature points have their projections on the bas-relief line lying on one side of the true FOE. Thus, all the individual curves would have maxima lying on one side of the true FOE. As a result, the TrErr curve would have shape similar to that of the individual curve. That is, it would have an overall maximum on the opposite side of the true FOE with respect to the origin, and when the estimated FOE approaches infinity on either end of the bas-relief line, the residual value would approach asymptotically towards a constant value.

The asymptotic value at infinity will be largely determined by the types of true translation. Predominantly lateral motion causes low asymptotic value (see dashed curves in Fig. 3(c) and (d)), with the latter approaching zero as the translational motion approaches that of pure lateral motion.

2. *Rotational error*. How do the rotation parameters enter the picture? If these parameters could be estimated accurately, the SFM problem would be simple. The error profile along the bas-relief valley would be represented by the TrErr curve. In particular, there would be no local minimum within the bas-relief valley. However, *it is precisely the coupling of the rotation with the translation that results in local minima within the bas-relief valley.* By coupling, we mean that the residual error caused by the translational errors can be compensated for by a suitable choice of α_e and β_e. Figure 3(b) to (d) show this compensating capability of α_e and β_e along the bas-relief line by the dotted curves, where high values on the curves indicate that the compensation is highly effective. We denote these dotted curves as the RotComp (Rotation Compensation) curves.

   Referring to Fig. 3, as the estimated FOE departs from the true FOE and enters region 1, the RMag constraint is operative and works towards compensating the translation error. However, as $\|(x_{0_e}, y_{0_e})\|$ increases, two factors restrict the applicability of the RMag constraint. Firstly, the corresponding increase of α_e and β_e means that $\|\mathbf{t}_{\mathbf{2},\mathbf{2}}\|$ increases, that is, RMag2 constraint comes to the fore, which works against the minimization of J_R. Secondly, as $\mathbf{t_1}$ approaches more and more towards the direction perpendicular to $\mathbf{t}_{\mathbf{2},\mathbf{0}}$ and $\mathbf{t}_{\mathbf{2},\mathbf{Z}}$, there is less advantage to be gained from the cancelation of the $\mathbf{t}_{\mathbf{2},\mathbf{0}}$ and $\mathbf{t}_{\mathbf{2},\mathbf{Z}}$ terms, as long as the directional constraints TDir and RDir are observed. The resulting RotComp curve in region 1 is such that it increases firstly, then decreases asymptotically towards zero. As the estimated FOE departs from the true FOE in the other direction and enters region 2, RotComp would increase first, as in the case of the beginning of region 1. As long as RMag2 is not operative yet, the RotComp is able to follow in tandem the rapid increase of the TrErr curve and therefore to compensate the latter. As the estimated FOE enters region 3, RMag2 takes effect again. Furthermore $\mathbf{t_1}$ becomes more perpendicular to $\mathbf{t}_{\mathbf{2},\mathbf{0}}$ and $\mathbf{t}_{\mathbf{2},\mathbf{Z}}$ thus obviating the need for the RMag constraint. Thus the RotComp curve again decreases asymptotically towards zero.

3. *Formation of local minimum*. The final J_R curve as a result of this coupling between the rotation and the translation would be equal to the subtraction of RotComp curve from the TrErr curve as shown in Fig. 3(b) to (c) (the solid curve). Clearly, it is due to the sharp drop-off of the TrErr curve as well as the compensatory effect of the rotational terms that a local minimum forms on the opposite side of the true FOE. We call this minimum the opposite minimum because it always lies on the opposite side of the true FOE. It is located around where the RotComp curve starts to enter region 3, that is, where the RotComp curve is no longer able to follow the TrErr curve. Figure 3(d) illustrates the case where the true FOE is further out (though not at infinity); here the opposite minimum has already been

pushed to infinity at the other end of the bas-relief valley.

4. *Factors affecting local minimum*. The exact location and the "depth" of the local minimum depends on various factors discussed below:

   - The type of the true translation affects the shape of the TrErr curve, specifically its maximum location and the asymptotic values. In general, largely lateral translations present a more difficult scenario for most SFM algorithms, because at the opposite end of the bas-relief valley, $\mathbf{t}_1$ and $\mathbf{t}_2$ will be almost entirely perpendicular, resulting in small asymptotic value of J_R. The location of the opposite minimum will approach infinity with residual value approaching zero. A large part of the bas-relief valley becomes very flat, thus presenting a highly ambiguous situation (Fig. 4(b)). Furthermore, as far as rotation estimates are concerned, in the limiting case of pure lateral motion, the RMag constraint fails to exert any constraint on the magnitude of α_e and β_e. The reason is as mentioned before: $\mathbf{t_1}$ and $\mathbf{t_2}$ can be in this case made perpendicular to each other at all points through the TDir and RDir constraints; the RMag constraint, which makes $\mathbf{t_2}$ small, becomes redundant.

     Conversely, if the true FOE approaches that of the pure forward motion case, the opposite minimum will merge with the true solution and disappear as shown in Fig. 4(a).

   - Field of view determines the effectiveness of the RMag2 constraint. With other conditions fixed, small FOV is a favorable condition for the formation of the opposite minimum. The later the RMag2 constraint sets in, the longer the RotComp curve are able to follow the TrErr curve. Thus the opposite minimum will form further away from the origin, and the residual error J_R will be smaller in value (though the valley around this local minimum may be less steep), making it more likely for the opposite minimum to be picked up as the solution.
   - Focal length. A change of focal length brings about several effects. Firstly, the values of α_e and β_e as dictated by the RMag constraint will be smaller. This in turn means that all the $\mathbf{t_{2,2}}$ terms will be reduced in magnitude, both due to the smaller α_e and β_e and the larger f value in the denominator. Lastly, the true FOE ($f\frac{U}{W}$, $f\frac{V}{W}$) will edge towards infinity with a larger f. All these factors will push the opposite minimum further, as well as making it more conducive for the opposite minimum to be picked up as the solution.
   - Unweighted epipolar constraint. Finally, it should be mentioned that if the epipolar constraint is unweighted, it can be shown that there will no maximum in the TrErr curve. Clearly, the resultant coupling with the RotComp curve would yield a minimum in the error surface near the center of the image, a result well-known in the literature.

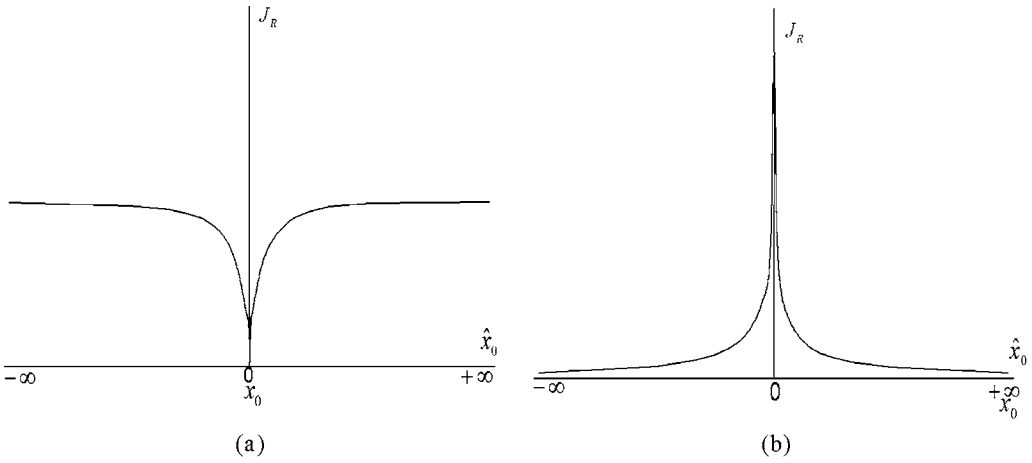

Figure 4. Error profiles of the bas-relief valley in the limiting cases. (a) The true FOE coincides with the origin and (b) The true FOE lies at infinity.

3.2.3. Relaxation of the Assumptions on the Distribution of Feature Points and Depth. In the beginning of this section, we made some assumptions on the distribution of feature points and depths, which are necessary for the derivation of the preceding results, especially the formation of the bas-relief valley. It is important to know how the error surface would be affected when these assumptions do not hold.

1. *Uneven feature point distribution*. We first consider the case where the feature points are not evenly distributed (with the assumption on the depth distribution still valid). We can model one aspect of uneven distribution by a shift of the centroid of the feature points from the image center to the point $(\tilde{x}, \tilde{y})$. Since $\mathbf{t_{2,1}}$ is always parallel to $\mathbf{t_{1,1}}$ and $\mathbf{t_{2,2}}$ is always perpendicular to $\mathbf{t_{1,1}}$, the contributions of these two components of $\mathbf{t_2}$ to the error surface will not be affected by the distribution of the feature points. However, to make the remaining components of $\mathbf{t_2}$, which are $(\mathbf{t_{2,0}} + \mathbf{t_{2,Z}})$, perpendicular to $\mathbf{t_1}$, the TDir and RDir constraints need to be modified near the region where the feature points are clustered as follows:

$$\frac{(x_0 - \tilde{x})}{(y_0 - \tilde{y})} = \frac{(\hat{x}_0 - \tilde{x})}{(\hat{y}_0 - \tilde{y})} \tag{22}$$

and

$$\frac{\alpha_e}{\beta_e} = -\frac{(\hat{y}_0 - \tilde{y})}{(\hat{x}_0 - \tilde{x})} \tag{23}$$

respectively. As a consequence, the bas-relief valley is attracted towards the new centroid when it is in the vicinity of the feature points. However, when the estimated FOE approaches infinity, the direction of $\mathbf{t_1}$ will be mainly determined by $\mathbf{t_{1,0}}$; in other words, the bas-relief valley will be very little affected by the shift of the feature centroid. The resultant bas-relief valley is illustrated in Fig. 5.

2. *Feature point grouped in local clusters*. Another aspect of unevenly distributed feature points is that the feature points are grouped in local clusters. Here it is well to mention the presence of critical points on the error surface due to the vanishing of the denominator of $\dot{\mathbf{p}}_e(\hat{\mathbf{v}}, \mathbf{v}, \mathbf{w}_e)$. In particular, these critical points are formed when the estimated FOE coincides with (x, y), at which both the numerator and the denominator vanish. These critical points may also be caused by the $\mathbf{n}$ term in the denominator being perpendicular to $(x - \hat{x}_0, y - \hat{y}_0)$, the value of $\mathbf{n}$ being dependent on the adopted reconstruction schemes. Such critical points are the possible sources of local minima or maxima. However, our simulations in the next section show that they do not have a significant effect on the overall error surface as long as the feature points are evenly distributed. However, when feature points are grouped in local clusters, the surrounding error surface can be significantly affected. In particular, each cluster may cause shallow local minimum to form around the cluster.

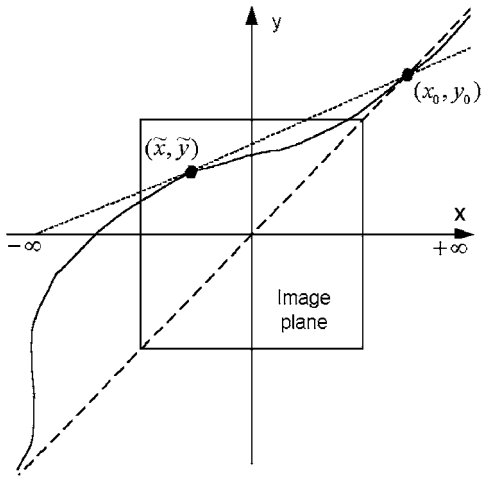

Figure 5. Geometry of the bas-relief valley when the distribution of feature points is uneven, with centroid at $(\tilde{x}, \tilde{y})$. The solid line corresponds to the bas-relief valley; the dashed line corresponds to the bas-relief valley if the feature points are evenly distributed; the dotted line is the line passing through $(\tilde{x}, \tilde{y})$ and (x_0, y_0).

3. *Uneven depth distribution*. Another factor to consider is that the depth is usually dependent on the feature co-ordinates (x, y) in a complex way. Referring to Fig. 2, this depth dependency means that the perturbation terms $\mathbf{t_{1,1}} = (x, y)$ and $\mathbf{t_{2,Z}} = (-\frac{y_{0_e}}{Z}, \frac{x_{0_e}}{Z})$ are correlated, or equivalently, $(\frac{x}{Z}, \frac{y}{Z})$ and $(-y_{0_e}, x_{0_e})$ are correlated. Thus we can model this dependency as a shift of centroid of the depth-scaled feature points $(\frac{x}{Z}, \frac{y}{Z})$. It can be shown that the effect of this shift on the bas-relief valley is very similar to that caused by a shift of the feature points' centroid. For example, if the depths associated with the feature points at the upper-left corner of the image plane are smaller than those

of other feature points, the centroid of the depth-scaled feature points will move toward the upper-left corner. The location of the resulting bas-relief valley will change as if the centroid of the feature points were shifted to the upper-left corner (see Fig. 5).

4. *Special case of planar scene*. One special case of depth dependency on the feature co-ordinates is that of a plane, a case well studied by numerous researchers such as Longuet-Higgins (1984). It is simple but nevertheless capable of modeling many real scene configurations. Longuet-Higgins showed that if all the object points come from a plane which is expressed as:

$$LX + MY + NZ = 1$$

the motion field caused in the image plane will be identical to the motion field caused by another plane:

$$UX + VY + WZ = 1$$

moving with the translation (L, M, N) and rotation $(\alpha + VN - WM, \beta + WL - UN, \gamma + UM - VL)$. That is, there now exists on the translation error surface another minimum at the location $(f\frac{L}{N}, f\frac{M}{N})$. One can relate this planar case to the case discussed in the preceding paragraph, where depth dependency results in a shift of the centroid of the depth-scaled feature points $(\frac{x}{Z}, \frac{y}{Z})$. Substituting the planar equation for Z into $(\frac{x}{Z}, \frac{y}{Z})$, one can show that the centroid of these depth-scaled feature points is now lying along the direction given by $(f\frac{L}{N}, f\frac{M}{N})$. Thus, we would expect the original bas-relief valley to be pulled towards this effective centroid $(f\frac{L}{N}, f\frac{M}{N})$. Indeed, as we will show through simulation (Fig. 8) later, along the direction defined by the origin and the alternative FOE $(f\frac{L}{N}, f\frac{M}{N})$, the residual error surface also forms a valley similar to the bas-relief valley (the bas-relief valley of the alternative motion-scene configuration, as it were). As expected, the two bas-relief valleys will influence each other when they are near each other. The estimates of the rotation will also be affected by the existence of the alternative solution. The RMag2 constraint will be modified since it is now possible to partially cancel the $\mathbf{t_{2,1}}$ and $\mathbf{t_{2,2}}$ terms with the $\mathbf{t_{2,Z}}$ term. The result is that while large field of view still improves the estimation of rotation due to the RMag2 constraint, we expect larger errors to remain in the rotation estimates, including that for the rotation around the optical axis.

3.3. *Depth Distortion at the Opposite Minimum Solution*

In this section, we attempt to investigate how the reconstructed structure would be distorted when the opposite-side minimum is picked up as the solution. In particular, we investigate depth distortion under configurations favourable for the formation of such spurious opposite minimum, namely, the FOV is small and the translation is largely lateral. Rewriting the expression of the distortion factor in Eq. (11) in terms of its component error terms, we obtain:

$$D = \frac{(x - \hat{x}_0, y - \hat{y}_0) \cdot \mathbf{n}}{(x - x_0, y - y_0) \cdot \mathbf{n} + Z\left(u_{rot_e}, v_{rot_e}\right) \cdot \mathbf{n}} \quad (24)$$

Under the aforementioned favourable conditions, we are able to make two simplifications. Firstly, the condition of small FOV allows us to ignore the second order terms in u_{rot_e} and v_{rot_e}. We also further assume $\gamma_e = 0$ (under noiseless condition, we expect accurate estimation of γ). Secondly, given the large values of $x_0, y_0, \hat{x}_0, \hat{y}_0$ in this configuration, we make these approximations: $(x - \hat{x}_0, y - \hat{y}_0) \approx (-\hat{x}_0, -\hat{y}_0)$ and $(x - x_0, y - y_0) \approx (-x_0, -y_0)$. Equation (24) can then be expressed as:

$$D = \frac{(-\hat{x}_0, -\hat{y}_0) \cdot \mathbf{n}}{(-x_0, -y_0) \cdot \mathbf{n} + Z(-\beta_e f, \alpha_e f) \cdot \mathbf{n}} \quad (25)$$

We know that at the opposite minimum solution, the TDir and the RDir constraints hold. These constraints mean that the numerator and the denominator of the above expression are parallel, and thus D is independent of $\mathbf{n}$, the direction of depth reconstruction. We can now write D as

$$D = \frac{1}{\lambda_1 + \lambda_2 Z} \quad (26)$$

where λ_1 and λ_2 are constant, with $\lambda_1 = -\frac{\|(x_0, y_0)\|}{\|(\hat{x}_0, \hat{y}_0)\|}$ (since (x_0, y_0) and $(\hat{x}_0, \hat{y}_0)$ are opposite in direction) and $\lambda_2 = \pm\frac{\|(-\beta_e f, \alpha_e f)\|}{\|(\hat{x}_0, \hat{y}_0)\|}$.

A distortion factor with the form $\frac{1}{\lambda_1+\lambda_2 Z}$ generates iso-distortion surfaces which are frontal-parallel planes. The resulting distortion transformation is that of a relief transformation which has some nice properties (Koenderink and van Doorn, 1995). In particular, consider two points in space with depths $Z_1 > Z_2$. Given $\lambda_1 < 0$, it can be shown (Cheong and Ng, 1999) that depth order will be preserved when

$$(\lambda_1 + \lambda_2 Z_1)(\lambda_1 + \lambda_2 Z_2) < 0$$

Equivalently, since $\frac{1}{\lambda_1+\lambda_2 Z}$ is the distortion factor, the above means that depth order will be preserved when $\hat{Z}_1\hat{Z}_2 < 0$, and conversely, depth order will be reversed when $\hat{Z}_1\hat{Z}_2 > 0$. In other words, if $\hat{Z}_1\hat{Z}_2 > 0$, we need to perform a depth order inversion to obtain the correct depth order. Therefore, given the signs of two recovered depths, we can always determine the correct depth order. Of course, it remains open to question if human actually performs the required depth order inversion.

What can we say about the sign of λ_2? If the RMag constraint holds, then

$$\begin{cases} sign(-\beta_e f) = sign(-\hat{x}_0) \\ sign(\alpha_e f) \quad = sign(-\hat{y}_0) \end{cases} \tag{27}$$

which means that λ_2 is positive. Under such condition, the iso-distortion surfaces have the following additional properties. The $D = 1$ distortion surface divides the whole space into two parts: the near field in which the space is expanded ($D > 1$) and the far field in which the space is compressed ($D < 1$), with negative distortion factor in the region $0 < Z < -\frac{\lambda_1}{\lambda_2}$. However, the sign of λ_2 may be indeterminate when the true FOE moves towards infinity (as does the opposite minimum). Here the RMag constraint weakens and the RMag2 constraint is ineffective given the small FOV. Under such limiting case, we cannot determine the sign of λ_2.

Before we close this section, a brief remark on the case of translations close to the forward direction is warranted. When the opposite-side minimum is picked up as the solution, the distortion factor can be expressed as:

$$D = \frac{(x - \hat{x}_0, y - \hat{y}_0) \cdot \mathbf{n}}{(x - x_0, y - y_0) \cdot \mathbf{n} + Z(-\beta_e f, \alpha_e f) \cdot \mathbf{n}} \tag{28}$$

In such case, the distortion shows complicated behavior described by the Cremona transformation. This is in accordance with the view presented in Cheong and Xiang (2001) that depth recovery is less reliable when forward motion is executed.

3.4. Summary of Results and Discussion

Equation (17) has been critical in our analysis; its simple form renders possible the geometric treatment of the error surface via a consideration of the two vectors $\mathbf{t}_1$ and $\mathbf{t}_2$. The error surface configuration and in particular, the local minima on the surface which are the cause of inherent ambiguity of SFM algorithms, are identified. More importantly, the underlying mechanisms for the formation of such local minima are also investigated in a geometric way which is helpful towards obtaining an intuitive grasp of the problem. The major findings obtained so far are summarized as follows:

1. *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 the constraint of $\alpha_e = \frac{y_{0_e}}{Z_{avg} f}$ and $\beta_e = -\frac{x_{0_e}}{Z_{avg} f}$, 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. 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.
2. *Translation error*. Bas-relief valley is the major characteristic of the error surface; it is a line defined by the true FOE and the centroid of the feature points. The distribution of the feature points and the depth-scaled feature points will also affect the location of the bas-relief valley. 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 residual error along the bas-relief valley also tends to have a local maximum somewhere near the origin and approaches an asymptotic value as the estimated FOE moves towards infinity. 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. This opposite minimum in the bas-relief valley poses severe problem to most SFM algorithms.

3. *Depth distortion*. If the SFM algorithms return the opposite minimum as the solution, a distorted structure will be recovered. The behavior of such distortion depends on the location of the opposite minimum. If it is far away from the origin, the distortion transformation between the physical and reconstructed space belongs to relief transformation. Depth order can be preserved if we perform the necessary depth order inversion. In contrast, when the opposite minimum is close to the origin and returned as the solution, the depth distortion is complex and can only be described by a full Cremona Transformation. However it should be noted that if the features and depths are evenly distributed, a SFM algorithm taking proper precaution should be able to avoid this kind of opposite minimum due to its large residual error.
4. *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 far as the depth recovery is concerned, translation that is largely lateral results in depth distortion that has nice properties such as preservation of relief.
5. *Type of cost function*. Different cost functions are obtained by setting $\mathbf{n}$ to different directions. Since the bas-relief ambiguity occurs when $\mathbf{t}_1$ and $\mathbf{t}_2$ are roughly perpendicular to each other, making the numerator of $\dot{\mathbf{p}}_e(\hat{\mathbf{v}}, \mathbf{v}, \mathbf{w}_e)$ vanish, different choices of $\mathbf{n}$ has little influence on the formation of the bas-relief valley on the error surface. However the shape of the error surface, and in particular, the error profile along the bas-relief valley, could be affected by the different choices of $\mathbf{n}$. Indeed, if $\mathbf{n}$ is not set as the "epipolar reconstruction" direction, new local extrema would be introduced on the error surface when $\mathbf{t}_1$ is perpendicular to $\mathbf{n}$ for any feature point, making the denominator of $\dot{\mathbf{p}}_e(\hat{\mathbf{v}}, \mathbf{v}, \mathbf{w}_e)$ vanish. For instance, when $\mathbf{n}$ is set as a constant direction, there will be a large number of local maxima and minima, forming bands running roughly along the $\mathbf{n}^{\perp}$ direction. For the cases of $\mathbf{n}$ equal to constant direction and $\mathbf{n}$ equal to Linear Least Square Reconstruction direction, it can be shown that the opposite minimum on the bas relief valley still persists. Last but not least, it is also clear that the choice of $\mathbf{n}$ would directly affect the properties of the recovered depth.
6. *Minimization strategy*. There are many variants of SFM algorithms based on the differential epipolar constraint: Some estimates translation first and then the rotation (Zhang and Tomasi, 1999; Horn, 1990; Adiv, 1985; Heeger and Jepson, 1992), some estimates rotation first and then the translation (Prazdny, 1980), and others estimate all motion parameters simultaneously (Ma et al., 2000; Kanatani, 1993; Brooks et al., 1998). As long as the various algorithms are purely based on the differential epipolar constraint, the results of our analysis is applicable. For the algorithms that estimate the translation first based on other constraints such as the motion parallax (Rieger and Lawton, 1985), we need to first characterize the error likely to exist in the estimated translation which is beyond the scope of this paper. However, assuming an erroneous FOE has been obtained, we know that the corresponding rotational errors will satisfy $\frac{\alpha_e}{\beta_e} = -\frac{\hat{V}}{\hat{U}}$ and $\gamma_e = 0$ if the feature points and the depth-scaled feature points are distributed evenly.

3.5. *Experimental Analysis*

3.5.1. Visualization of the Cost Functions. In this section, we perform simulations on synthetic images to both visualize and verify the predictions obtained from the preceding theory. We also make additional observations along the way, for instance, regarding the influence of density of feature points on the residual errors. 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. At each point on the plot, the FOE are fixed; then J_R can be expressed as:

$$J_R = \sum_{i=1}^{n} \left(\frac{c_{1_i} - \left(c_{2_i}\hat{\alpha} + c_{3_i}\hat{\beta} + c_{4_i}\hat{\gamma}\right)}{\delta_i} \right)^2 \qquad (29)$$

where

$$
\begin{aligned}
c_{1_i} &= u(y - \hat{y}_0) - v(x - \hat{x}_0) \\
c_{2_i} &= \frac{xy}{f}(y - \hat{y}_0) - \left(\frac{y^2}{f} + f\right)(x - \hat{x}_0) \\
c_{3_i} &= \frac{xy}{f}(x - \hat{x}_0) - \left(\frac{x^2}{f} + f\right)(y - \hat{y}_0) \\
c_{4_i} &= x(x - \hat{x}_0) + y(y - \hat{y}_0) \\
\delta_i &= \sqrt{(x - \hat{x}_0)^2 + (y - \hat{y}_0)^2}
\end{aligned}
$$

from which the rotation variables can be solved by a typical linear least squares fitting algorithm such as the SVD (singular value decomposition) method. We performed this fitting for each fixed FOE candidate over the whole 2-D search space and obtained the corresponding reprojected flow difference J_R. The residual values were then plotted in such a way that the image intensity encoded the relative value of the residual (bright pixel corresponded to high residual value and vice versa). Furthermore, to illustrate the variation of J_R along the bas-relief valley in details, we also plot the cross-section of the residual error surface along the bas-relief line. Three types of curves were plotted for this purpose, namely, the residual error curve, TrErr, and RotComp as defined before, respectively drawn in solid line, dashed line and dotted line.

The imaging surface was a plane with a dimension of 512×512 pixels; its boundary was delineated by a small rectangle in the center of the plots. 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°; 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).

We conducted experiments under the following conditions: (1) varying amount of forward translation, ranging from head-on to lateral motion; (2) small versus large FOV; (3) feature points distributed evenly over the whole image plane versus those clustered at a corner; (4) depth-scaled feature points distributed evenly over the whole image plane versus those clustered at a corner; (5) sparse versus dense flow field; and (6) planar scene versus random scene.

Figure 6 shows the residual error images for different translations. 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. 6(b)–(d) 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. We also plotted the residual profiles along the bas-relief valleys (note that the residual profiles were plotted in terms of pixels, whereas the residual error surfaces were plotted in terms of visual angles). Apparently, the types of true translation had a significant influence on the formation of the opposite minimum. The opposite minimum disappeared (or merged with the global minimum) for pure forward motion (Fig. 6(e)). 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, as can be seen from Fig. 6(f), but in Fig. 6(h) under lateral motion they are almost equal in depth. The residual profiles also show clearly how the opposite minimum was formed by the coupling of the RotComp curve and the TrErr curve. 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. (20) held (RDir constraint), while for the rotational estimates around the opposite minimum, we further have their magnitudes satisfying Eq. (21) (RMag constraint). Under lateral motion (Fig. 6(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. It may be noted in passing that from Fig. 6 the maxima on the residual error image tended to form a strip perpendicular to the minima strip, and was more prominent for the case of lateral motion.

Figure 7 shows the influence of FOV on the residual error images. While examining these plots, it should be kept in mind that in our simulations, larger FOV was obtained by fixing the image size and decreasing the focal length. Thus, under larger FOV the true FOE

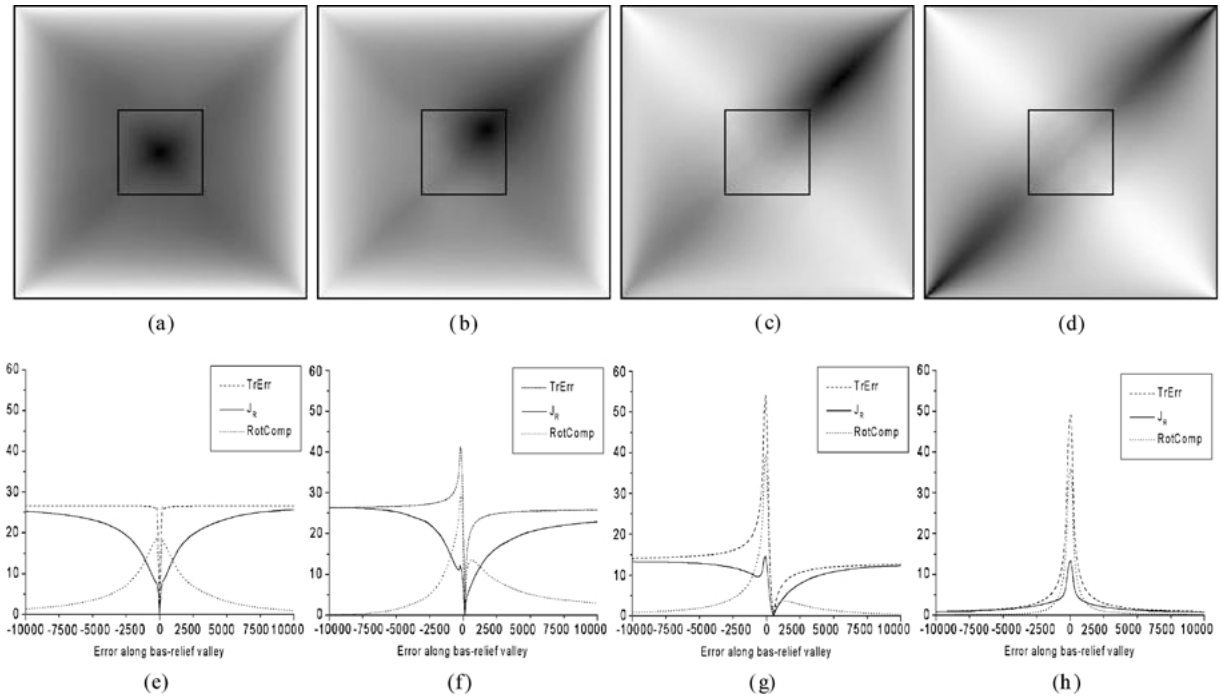

Figure 6. Residual error for different translations. Translational parameters (U, V, W) for (a), (b), (c), and (d) are (0, 0, 2), (0.5, 0.5, 2) (1, 1, 1) and (1, 1, 0) respectively. Figures (e), (f), (g), and (h) correspond to the residual profiles of the bas-relief valleys in (a), (b), (c), and (d) respectively. The dashed, dotted, and solid curves respectively represent the TrErr, RotComp and Jr curves. The residual error surfaces were plotted in terms of visual angles, whereas the residual profiles were plotted in terms of pixels.

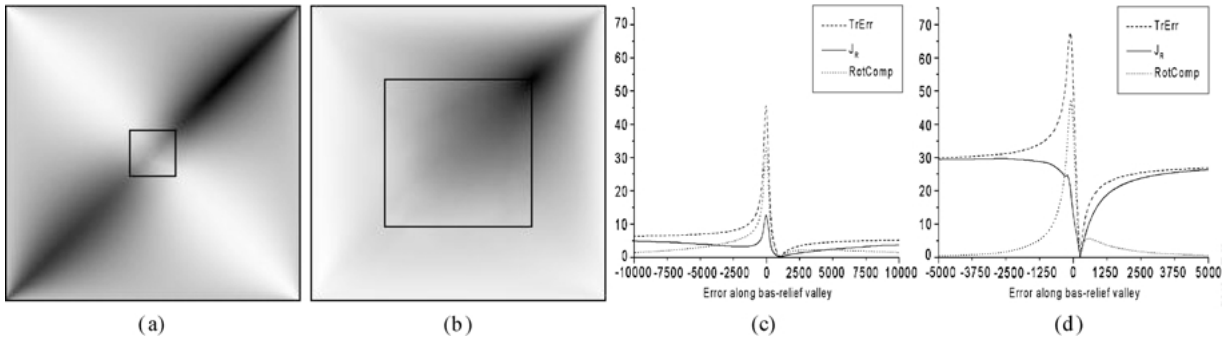

Figure 7. Influence of the FOV on the residual error images. $\mathbf{v} = (1, 1, 1)$ for both (a) and (b); FOV was 28° for (a) and 90° for (b). Figures (c) and (d) correspond to the residual profiles of the bas-relief valleys in (a) and (b) respectively. Notations as before.

$(f\frac{U}{W}, f\frac{V}{W})$ would be closer to the image center with the same translational velocity (U, V, W). The opposite minimum in Fig. 7(a) was prominent with small residual value, while in Fig. 7(b), the opposite minimum was almost invisible (it was barely visible in Fig. 7(d), being located at where the solid curve was just rising towards the asymptotic value in the opposite direction).

Figure 8 shows how the distribution of feature points and depth-scaled feature points affect the residual error images. When the feature points were clustered around different parts of the image, it was as if the bas-relief valley were pulled by the centroid of the feature points. This was illustrated in Fig. 8(a) where the feature points were clustered in the upper left corner of the image plane. In Fig. 8(b), with sparse and randomly

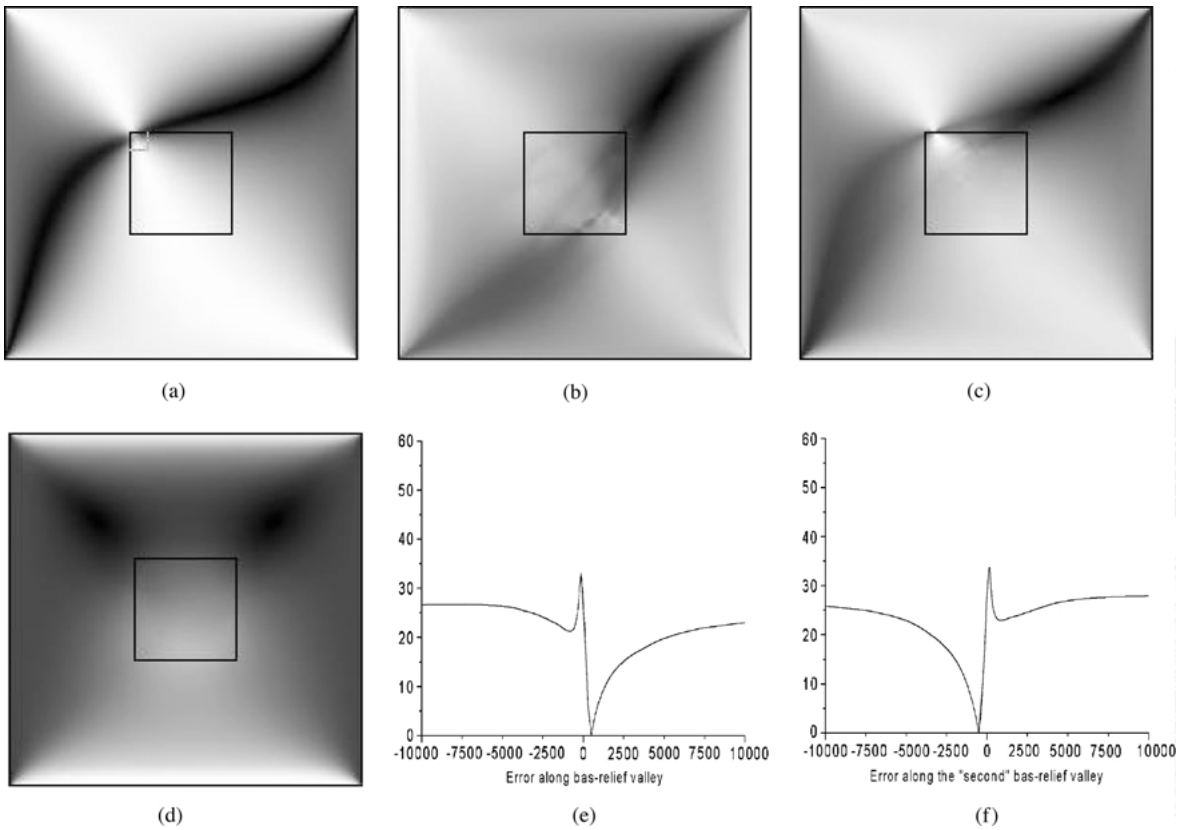

Figure 8. Influence of the distribution of feature points and depth-scaled feature points on the residual error images. (a) Feature points were clustered within a corner highlighted by a gray rectangle; (b) 20 feature points were distributed randomly over the image plane. (c) Feature point distribution was random over the image plane, whereas the depth distribution was such that the depth-scaled feature points were clustered at the upper-left corner of the image plane; (d) All the object points came from a plane with $(L, M, N) = (-0.002, 0.002, 0.002)$. Figure 8(e) and (f) correspond to the residual profiles of the bas-relief valley and the second bas-relief valley in (d) respectively. $\mathbf{v} = (1, 1, 1)$ for all the figures.

distributed feature points, the extrema caused by these feature points can be seen to form around these feature points. Such local extrema always exist but their effect was not significant if the feature points were sufficiently dense, as can be seen from all the other residual error images (with 200 feature points) in this section. Figure 8(c) shows that the location of the centroid of the depth-scaled feature points would also affect the formation of the bas-relief valley. Here, the centroid of the depth-scaled feature points was located at the upper left corner of the image plane, resulting in an error surface as if the feature points were centered in that region. The case of a planar scene was illustrated in Fig. 8(d). There were two clear minima, corresponding to the true and the alternative solutions for the planar scene. In addition, as shown by the cross-section of the residual error surface along the bas-relief valley in Fig. 8(e), the opposite minimum still exists. Another bas-relief valley was also apparent along the direction defined by the alternative FOE and the origin. Figure 8(f) shows that this second bas-relief valley also has similar profile, that is, it also has a local minimum on the opposite side of the alternative FOE. Finally, the numerical values of the simulation data show that, as predicted, the rotational estimates no longer observed the RMag and the RMag2 constraints.

3.5.2. Inherent Ambiguities and Visual Illusions. Our analyses in the preceding sections show that there are various ambiguities inherent to the SFM problem which may cause erroneous 3-D motion estimation and distorted 3-D space reconstruction. The case of lateral motion was particularly studied because it possesses some unique properties. It is also a motion often found in the biological world and a case heavily studied by visual psychophysicists. In this section, we attempt to

explain a well-known human visual illusion, the "rotating cylinder illusion" which is commonly observed in psychophysical experiments. We attribute this illusion to the imprecise estimation of the 3D motion caused by the inherent ambiguities and the corresponding distorted structure recovered from the erroneous motion estimates.

3.5.2.1. Rotating Cylinder Illusion. In the psychophysical experiments, the rotating cylinder illusion amounts to the following situation: Dynamic random-dot display representing a rotating cylinder occupies a small portion of the visual field and rotates around a vertical axis passing through the center of the cylinder, as shown in Fig. 9. It was found that sometimes the cylinder (as well as other curved objects) was perceived as rotating in a direction opposite to the true one and the correspondingly perceived structure underwent a change in the sign of curvature; that is a convex object was perceived as concave and vice versa (Hoffman, 1998).

3.5.2.2. Explanations. Some computational models have been proposed to explain the rotating cylinder illusion (Koenderink and van Doorn, 1991; Soatto and Brockett, 1998). Koenderink and van Doorn (1991) argued that since the solution of SFM under perspective projection cannot explain the effect (there can be only one solution), it could be that the human visual system somehow adopts an affine projection model. Soatto and Brockett (1998) attributed these illusions to the presence of noise.

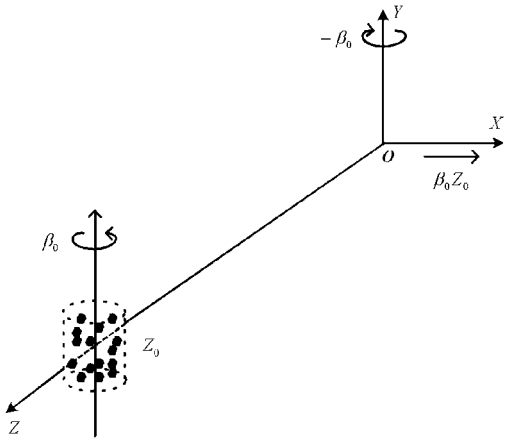

Figure 9. Configuration of the "rotating cylinder illusion". Dynamic random-dot display representing a rotating cylinder rotates about a vertical axis through its centroid with speed β_0. The equivalent egomotion for an observer positioned at o is given by $(\alpha, \beta, \gamma) = (0, -\beta_0, 0)$ and $(U, V, W) = (\beta_0 Z_0, 0, 0)$ where Z_0 is the distance between the optical center and the centroid of the cylinder.

A different view can be inferred from the standpoint of our theory. Given the conditions present in that of the rotating cylinder illusion (small FOV, lateral translation and small depth range), all conducive towards the formation of the opposite minimum, rotating cylinder illusion can be attributed to the erroneous motion estimates caused by the opposite minimum, or more precisely, the error configuration with $\hat{\mathbf{v}} = -\mathbf{v}$, $\frac{\alpha_e}{\beta_e} = -\frac{V}{U}$ and $\gamma_e = 0$. Consider the case where $\beta_0 > 0$, we have under the error configuration $\beta = -\beta_0 < 0, \hat{U} = -\beta_0 Z_0 < 0$. Since the object was perceived as rotating opposite to the veridical direction, we further have: $\hat{\beta} > 0$. Thus it holds that $\beta_e < -\beta_0 < 0$. According to the results in section 3.3, when the relative translation is perceived as opposite to the veridical motion such that $\lambda_1 < 0$, the depth order relationship of any two depths Z_1 and Z_2 would depend on the signs of the perceived depths $\hat{Z}_1$ and $\hat{Z}_2$. We can determine the sign of $\hat{Z}_1$ and $\hat{Z}_2$ as follows. When the rotating cylinder illusion took place, the subject often reported a perceived rotation which had roughly the same speed as the veridical one, that is $\hat{\beta} = \beta_0$ and $\hat{U} = -\beta_0 Z_0$. With all depths under view greater than $\frac{Z_0}{2}$ under the experimental configuration, we immediately obtain the distortion factor $\frac{1}{(-1+\lambda_2 Z)} > 0$ for all the perceived depths. This means that all depths are perceived as positive and thus all depth orders are reversed; it follows that convex object is perceived as concave and vice versa. Since the erroneous 3-D motion at the opposite minimum would be perceived with equal likelihood as the accurate 3-D motion under pure lateral motion, it also explains why the illusion was reported to occur only intermittently.

4. Role of Noise on 3-D Motion Estimation

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

$$\begin{aligned}\breve{\mathbf{p}} &= \dot{\mathbf{p}} + \dot{\mathbf{p}}_n \\ &= (u + u_n, v + v_n, 0)^T \end{aligned} \tag{30}$$

where $\dot{\mathbf{p}}_n$ is the flow component caused by noise. If we replace $\dot{\mathbf{p}}$ by $\breve{\mathbf{p}}$ in Eq. (13), Eq. (14) still holds, that is, the reprojected flow difference along the $\mathbf{n}$ direction is still zero. It follows that the noise-corrupted cost function,

denoted as J_{Rn}, can be obtained as follows:

$$J_{Rn} = J_R + 2\sum\left(\begin{array}{l}\frac{(x-\hat{x}_0, y-\hat{y}_0)\cdot(v_{rot_e}-\frac{y_{0_e}}{Z}, \frac{x_{0_e}}{Z}-u_{rot_e})}{(x-\hat{x}_0, y-\hat{y}_0)\cdot\mathbf{n}} \\ \times\frac{(x-\hat{x}_0, y-\hat{y}_0)\cdot(-v_n, u_n)}{(x-\hat{x}_0, y-\hat{y}_0)\cdot\mathbf{n}}\end{array}\right) + \sum\left(\frac{(x-\hat{x}_0, y-\hat{y}_0)\cdot(-v_n, u_n)}{(x-\hat{x}_0, y-\hat{y}_0)\cdot\mathbf{n}}\right)^2 \quad (31)$$

J_{Rn} consists of three terms. The first term is that of the noise-free case. The second term can be positive or negative, whereas the last term is always zero or positive. These are the most that can be said about the effect of noise without further introducing assumptions about the noise. Next, we investigate the behavior of SFM algorithms under the effects of specific noise types.

4.1. Isotropic Noise Model

Isotropic noise model has been frequently used for noise analysis in the computer vision community. The isotropic noise is defined as an independent Gaussian noise with identical covariance matrix $K = diag\{\sigma^2, \sigma^2, 0\}$. Under this noise model, the effect of noise on the periphery of the search space are small. Referring to the expression $(\frac{(x-\hat{x}_0, y-\hat{y}_0)\cdot(-v_n, u_n)}{(x-\hat{x}_0, y-\hat{y}_0)\cdot\mathbf{n}})$ contained in both the second and the third terms of Eq. (31), it is basically a projection of the random noise on the vectors $(x-\hat{x}_0, y-\hat{y}_0)$ which are approximately constant in the periphery of the search space. The net effect of noise in the periphery is therefore rather benign as shown in Fig. 10 where an isotropic noise with a standard deviation fixed at 50% of the average flow speed ($SNR = 7.08dB$) has been added to each component of the flow vector. The same noise has a "stronger" and more complex effect on the topology of the residual in the center of the image. This effect is especially obvious when the features are sparse, which can be seen by comparing Fig. 10(b) with Fig. 10(c).

Some researchers demonstrated the robustness of their algorithms by conducting experiments using dense flow field with isotropic noise. However such noise model is often unrealistic. We show in the next section what happens when the noise model is non-isotropic.

4.2. Anisotropic Noise Model

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 (Fermüller et al., 2001) in which a more complicated model was presented compared to the one adopted here.

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 third term in Eq. (31) would be small, and vice versa. The

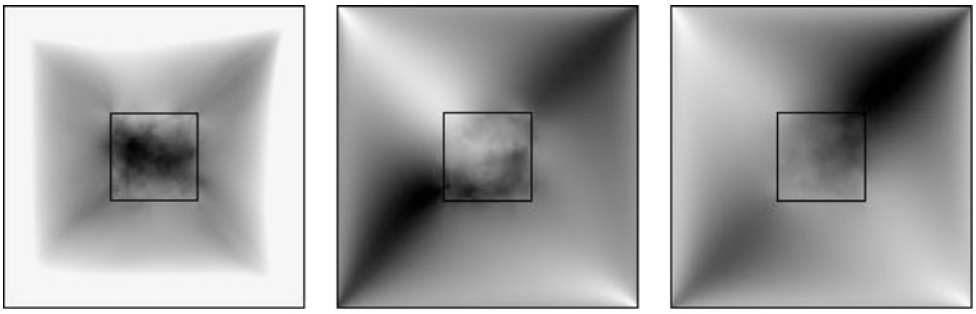

Figure 10. Residual error images for flow fields with isotropic noise. Number of feature points were 200 for (a) and (b) and 2000 for (c). $\mathbf{v} = (0, 0, 2)$ for (a) and $\mathbf{v} = (1, 1, 1)$ for both (b) and (c).

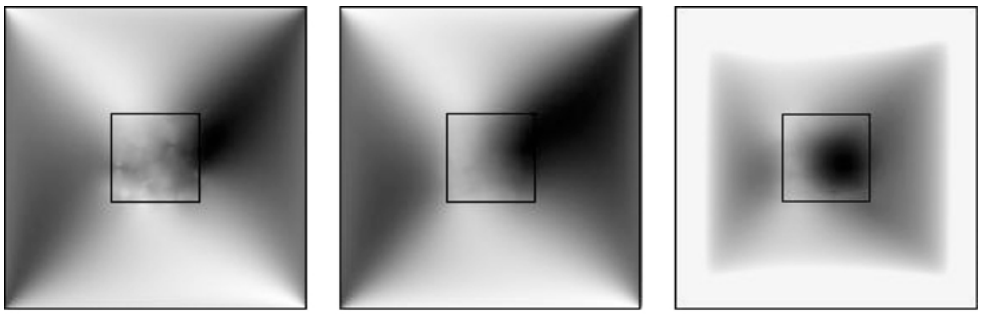

Figure 11. Residual error images for the flow fields with anisotropic noise. Numbers of feature points were 200 for (a), 2000 for (b) and $\mathbf{v} = (1, 1, 1)$ for (a) and (b). For (c), number of feature points was 800, and $\mathbf{v} = (0, 0, 2)$. $\mathbf{w} = (0, 0.0005, 0.001)$ and noise level was 50% for all the images.

effect of the second term in Eq. (31) 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. 11.

It can be seen from Fig. 11 that the influence of a 50% anisotropic noise is quite significant. For the case of $\mathbf{v} = (1, 1, 1)$ and $\mathbf{w} = (0, 0.0005, 0.001)$ under the scene in view (Fig. 11(a) and (b)), 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. This effect persisted as we increased the number of feature points, as shown in Fig. 11(b). For the case of forward motion (Fig. 11(c)), 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 SFM algorithms by indirectly affecting the distribution of noise. For instance, a strong rotation around the Y-axis would result in a strong horizontal flow, which will pull the bas-relief valley towards the horizontal direction due to the aforementioned anisotropic noise. Such phenomenon is often observed in practice, the result being that the FOE cannot be reliably estimated when the rotation is dominant. Figure 12 shows the influence of the true rotation on the residual error images under anisotropic noise model.

In real images, features found in one surface patch may have different optical flows from those in another surface patch. According to our anisotropic noise model, the average noise directions in these two patches would also be different. If we were to perform SFM separately from these two patches, we would expect the minima strip to be pulled along different directions; however, the region around the true solution would be less affected. This was illustrated in Fig. 13(a) and (b). One can capitalize on this characteristic, say, by performing a simple thresholding on the residual values of Fig. 13(a) and (b) and intersecting the resulting binary maps. The result was illustrated in Fig. 13(c), where the centroid of the "common area" was at $(-12, 0)$, closer to the true FOE estimate than the global minimum obtained by using all the feature points in the image simultaneously. This observation can be used to formulate a plausible strategy to improve motion estimation; its feasibility will be further tested on real images in the next section.

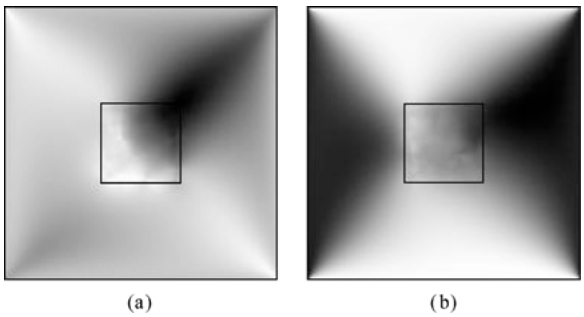

Figure 12. Influence of true rotation. Translational parameters were $(1, 1, 1)$ for both (a) and (b). Rotational parameters were $(0, 0, 0)$ for (a) and $(0, 0.001, 0)$ for (b). Noise level is 50% for all the images.

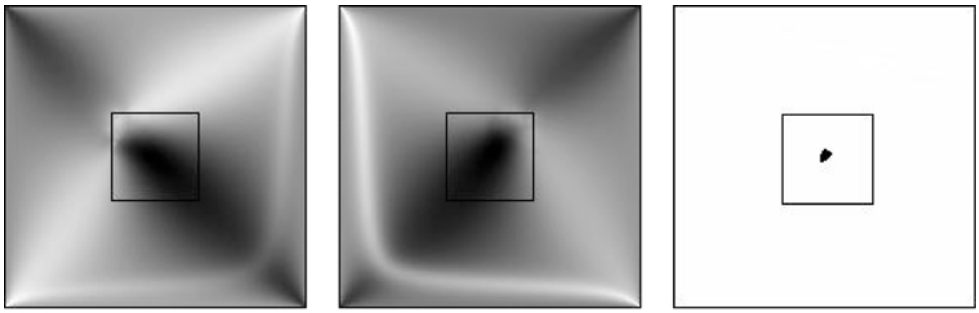

Figure 13. Residual error images using feature points from different patches. All the parameters were the same as Fig. 11(c) except in the way we utilized the feature points. The feature points used in (a) and (b) were those found in the top-left and the top-right quarter of the image plane respectively. Figure (c) was the binary image obtained by intersecting the thresholded versions of (a) and (b).

4.3. Experiments on Real Images

The aim of this section is to carry out experiments on real images so as to verify the various predictions made, and to study the feasibility of a better algorithm based on the knowledge of the topology of the error surfaces.

Three familiar real image sequences were used. The parameters of these sequences are listed in Table 1. Among them, only the COKE sequence is genuinely "real", while the other two are computer generated sequences. The YosemiteNoCloud sequence is the well-known Yosemite sequence minus the cloud in the top portion of the images, whereas the SOFA1[2] sequence describes a simple indoor scene with constant lateral translation and quite significant rotational components. The optical flow was obtained using Lucas's method (Lucas, 1984) with a temporal window of 15 frames. Relatively dense optical flow fields (around 3000 feature points for each sequence) were obtained. Again, the estimated epipolar direction was adopted as the direction for depth reconstruction.

The residual error images were shown in Fig. 14, from which several observations can be made.

- Figure 14(c) shows local minima strips along the average optical flow direction, especially outside the image plane. This might be 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. 14(i), the bas-relief valley was pulled towards the average optical flow direction, which was roughly horizontal.
- The effect of the clustered feature points was obvious for each case. Specifically, prominent edges on the image resulted in a clustered feature distribution, as in the case of YosemiteNoCloud. Local minima were formed along the edge, as shown in Fig. 14(f).
- Rotation estimates. While the numerical values of the simulation data with synthetic images in Section 3.5 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.

Table 1. The parameters of three real image sequences.

	Image size	f	Translation	Rotation
COKE	300×300	439.4	$(x_0, y_0) = (-25, 25)$	(0.0006, 0.0006, 0.004)
YosemiteNoCloud	252×316	337.5	$(x_0, y_0) = (0, 59)$	(0.0002, 0.0016, −0.0002)
SOFA1	256×256	309.0	$(U, V, W) = (0.814, 0.581, 0)$	(−0.0203, 0.0284, 0)

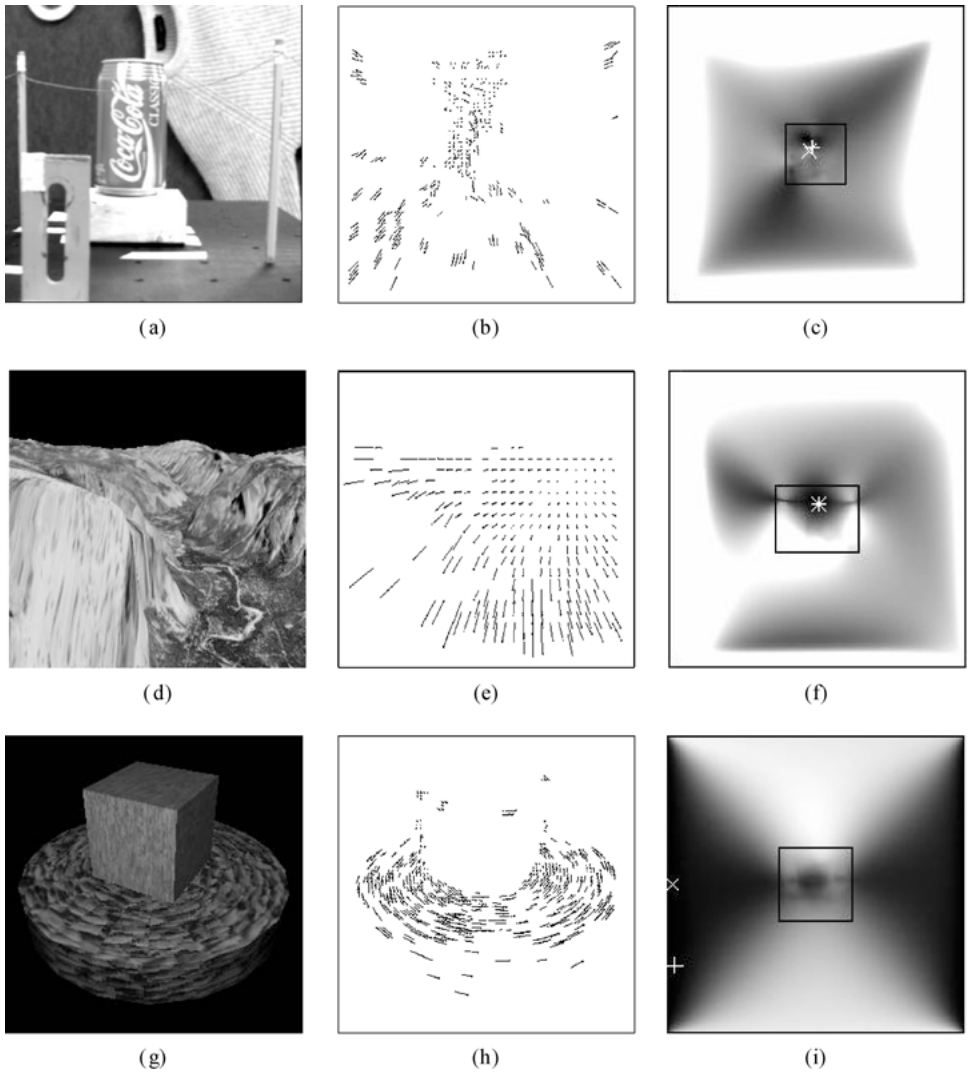

Figure 14. Residual error images for real image sequences. Each row corresponds to one image sequence. From top to bottom: COKE, YosemiteNoCloud and SOFA1. For each row, an actual image of the sequence, the optical flow field and the residual error image are shown from left to right. True FOEs and global minima of the residual error surfaces were highlighted by "+" and "×" on the residual error images respectively.

Figure 15 demonstrates how we can make use of the knowledge of the topology to design a better algorithm. In the case of anisotropic noise, instead of using all the features simultaneously in an image, we performed SFM separately, each time using features found in different patches. The average optical flows found in different patches will be usually pointing to different directions. The resultant residual error images would thus be pulled differently according to the average flow directions. By combining the separate residual error images, one can obtain a better and more robust FOE estimate. Figure 15(c) was the binary image obtained by performing an intersection of the thresholded version of Fig. 15(a) and (b) (the residual error images resulting from using different patches). It can be seen that the uncertainty area of the FOE estimate has been much reduced, thereby illustrating the feasibility of the idea. Better strategy for combining the estimation results from different patches can be devised so that factors such as feature number and flow configuration in each patch can be taken into account.

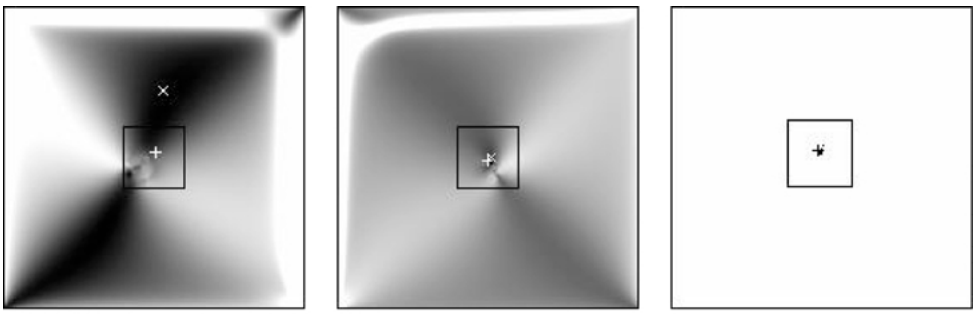

Figure 15. Residual error images resulting from using features in different regions in the COKE image. Figure (a) was obtained by using features from the bottom-left and (b) from the bottom-right quarter of the image plane respectively. Figure 15(c) was the binary image obtained by performing an intersection of the thresholded version of (a) and (b).

5. Conclusions and Future Directions

Understanding the inherent motion ambiguities is critical for addressing the SFM problem. To this end, 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. Other conditions that may affect the location of the ambiguities such as feature distribution and density were also considered. This in turn was followed by an analysis of the depth distortion caused by these motion ambiguities using the iso-distortion framework. The analysis shed light on a well-known human perceptual illusion—the rotating cylinder illusion. Experiments on both synthetic and real image sequences were carried out. Results obtained on real image sequences seemed to confirm the anisotropic noise model. Results also showed that factors such as sparseness of features, coupled with anisotropic noise distribution, may have great impact on the residual error distribution.

This work represents part of the ongoing study towards fully understanding the behavior of SFM algorithms. More work needs to be done to extend our understanding in areas such as uncalibrated motion ambiguities. Though we focus on the calibrated case throughout the paper, our approach can be readily adopted to analyzing the behavior of SFM algorithms under uncalibrated case. More details and some preliminary results can be found in Xiang (2001). Some partial analyses with a view towards such understanding has been carried out in Cheong and Peh (2000) using the depth-is-positive constraint. Similarly, the preceding depth distortion analysis should lead to the more important question of what can be done with such distorted depth. Finally, by understanding how the topology changes, it opens up the possibility of a more robust algorithm. The important conclusion of this work is that the SFM algorithms assume different behaviors under different motion-scene configurations, corroborating the view that current SFM algorithms can perform well only in restricted domains (Oliensis, 2000a). It follows that if we can characterize and identify the different behaviors and domains, it then becomes possible to propose better algorithms or to fuse the results of several existing algorithms.

Notes

1. LLSR direction refers to $\mathbf{n} = \frac{[\dot{\mathbf{p}}]_2 - \hat{\dot{\mathbf{p}}}_{rot}}{\|[\dot{\mathbf{p}}]_2 - \hat{\dot{\mathbf{p}}}_{rot}\|}$. Depth recovered along this direction is the standard linear least square estimate of depth from Eq. (2), which minimizes the "estimated measurement error" $\|\hat{\dot{\mathbf{p}}}_{tr} - \hat{Z}([\dot{\mathbf{p}}]_2 - \hat{\dot{\mathbf{p}}}_{rot})\|$. Details of the properties of this depth reconstruction scheme can be found in Cheong and Xiang (2001).
2. courtesy of the Computer Vision Group, Heriot-Watt University (http://www.cee.hw.ac.uk/∼mtc/sofa).

References

Adiv, G. 1985. Determining 3-D motion and structure from optical flow generated by several moving objects. *IEEE Trans. PAMI*, 7:384–401.

Adiv, G. 1989. Inherent ambiguities in recovering 3-D motion and structure from a noisy flow field. *IEEE Trans. PAMI*, 11:477–489.

Brooks, M.J., Chojnacki, W., and Baumela, L. 1997. Determining the egomotion of an uncalibrated camera from instantaneous optical

flow. *Journal of the Optical Society of America* A, 14(10):2670–2677.

Brooks, M.J., Chojnacki, W., Hengel, A.V.D., and Baumela, L. 1998. Robust techniques for the estimation of structure from motion in the uncalibrated case. In *Proc. Conf. ECCV*, pp. 283–295.

Cheong, L.-F., Fermüller, C., and Aloimonos, Y. 1998. Effects of errors in the viewing geometry on shape estimation. *Computer Vision and Image Understanding*, 71(3):356–372.

Cheong, L.-F. and Ng, K. 1999. Geometry of distorted visual space and cremona transformation. *International Journal of Computer Vision*, 32(2):195–212.

Cheong, L.-F. and Peh, C.H. 2000. Characterizing depth distortion due to calibration uncertainty. In *Proc. Conf. ECCV*, Dublin, Ireland, vol. I, pp. 664–677.

Cheong, L.-F. and Xiang, T. 2001. Characterizing depth distortion under different generic motions. *International Journal of Computer Vision*, 44(3):199–217.

Chiuso, A., Brockett, R., and Soatto, S. 2000. Optimal structure from motion: Local ambiguities and global estimates. *International Journal of Computer Vision*, 39(3):195–228.

Daniilidis, K. and Spetsakis, M.E. 1997. Understanding noise sensitivity in structure from motion. In *Visual Navigation: From Biological Systems to Unmanned Ground Vehicles*, Y. Aloimonos (Ed.), Lawrence Erlbaum: Hillsdale, NJ.

Dutta, R. and Snyder, M.A. 1990. Robustness of correspondence-based structure from motion. In *Proc. Int. Conf. on Computer Vision*, Osaka, Japan, pp. 106–110.

Fermüller, C. and Aloimonos, Y. 2000. Observability of 3D motion. *International Journal of Computer Vision*, 37(1):43–63.

Fermüller, C., Shulman, D., and Aloimonos, Y. 2001. The statistics of optical flow. *Computer Vision and Image Understanding*, 82:1–32.

Grossmann, E. and Victor, J.S. 2000. Uncertainty analysis of 3D reconstruction from uncalibrated views. *Image and Vision Computing*, 18(9):686–696.

Heeger, D.J. and Jepson, A.D. 1992. Subspace methods for recovering rigid motion I: Algorithm and implementation. *International Journal of Computer Vision*, 7:95–117.

Hoffman, D.D. 1998. Visual intelligence, W.W. Norton and Company, Inc.

Horn, B.K.P. 1987. Motion fields are hardly ever ambiguous. *International Journal of Computer Vision*, 1:259–274.

Horn, B.K.P. 1990. Relative orientation. *International Journal of Computer Vision*, 4:59–78.

Kahl, F. 1995. Critical motions and ambiguous Euclidean reconstructions in auto-calibration. In *Proc. Int. Conf. on Computer Vision*, pp. 469–475.

Kanatani, K. 1993. 3-D interpretation of optical flow by renormalization. *International Journal of Computer Vision*, 11(3):267–282.

Koenderink, J.J. and van Doorn, A.J. 1991. Affine structure from motion. *J. Optic. Soc. Am.*, 8(2):377–385.

Koenderink, J.J. and van Doorn, A.J. 1995. Relief: Pictorial and otherwise. *Image and Vision Computing*, 13(5):321–334.

Longuet-Higgins, H.C. 1981. A computer algorithm for reconstruction of a scene from two projections. *Nature*, 293:133–135.

Longuet-Higgins, H.C. 1984. The visual ambiguity of a moving plane. *Proc. R. Soc. Lond.*, B233:165–175.

Lucas, B.D. 1984. Generalized image matching by the method of differences. Ph.D. Dissertation, Carnegie-Mellon University.

Ma, Y., Košecká, J., and Sastry, S. 2000. Linear differential algorithm for motion recovery: A geometric approach. *International Journal of Computer Vision*, 36(1):71–89.

Ma, Y., Košecká, J., and Sastry, S. 2001. Optimization criteria and geometric algorithms for motion and structure estimation. *International Journal of Computer Vision*, 44(3):219–249.

Maybank, S.J. 1993. *Theory of Reconstruction from Image Motion*. Springer: Berlin.

Negahdaripour, S. Critical surface pairs and triplets. *International Journal of Computer Vision*, 3:293–312.

Oliensis, J. 2000a. A critique of structure-from-motion algorithms. *Computer Vision and Image Understanding*, 80:172–214.

Oliensis, J. 2000b. A new structure from motion ambiguity. *IEEE Trans. PAMI*, 22(7):685–700.

Oliensis, J. 2001. The error surface for structure from motion. NECI Technical Report, available at http://www.neci.nec.com/homepages/oliensis/.

Prazdny, K. 1980. Egomotion and relative depth map from optical flow. *Biological Cybernetics*, 36:87–102.

Rieger, J.H. and Lawton, D.T. 1985. Processing differential image motion. *J. Optic. Soc. Am.* A, 2:354–359.

Soatto, S. and Brockett, R. 1998. Optimal structure from motion: Local ambiguities and global estimates. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 282–288.

Spetsakis, M.E. 1994. Models of statistical visual motion estimation. *CVGIP*, 60:300–312.

Sturm, P. 1997. Critical motion sequences for monocular self-calibration and uncalibrated Euclidean reconstruction. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 1100–1105.

Szeliski, R. and Kang, S.B. 1997. Shape ambiguities in structure from motion. *IEEE Trans. PAMI*, 19(5):506–512.

Weng, J., Huang, T.S., and Ahuja, N. 1991. *Motion and Structure from Image Sequences,* Springer-Verlag: Berlin.

Xiang, T. 2001. Understanding the behavior of structure from motion algorithms: A geometric approach. Ph. D. Dissertation, National University of Singapore.

Young, G.S. and Chellapa, R. 1992. Statistical analysis of inherent ambiguities in recovering 3-D motion from a noisy flow field. *IEEE Trans. PAMI*, 14:995–1013.

Zhang, Z. 1998. Understanding the relationship between the optimization criteria in two-view motion analysis. In *Proc. Conf. ICCV*, pp. 772–777.

Zhang, T. and Tomasi, C. 1999. Fast, robust, and consistent camera motion estimation. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 164–170.