Computer Vision and Image Understanding 115 (2011) 16–30

Contents lists available at ScienceDirect

Computer Vision and Image Understanding

journal homepage: www.elsevier.com/locate/cviu

Behaviour of SFM algorithms with erroneous calibration

Loong-Fah Cheong *, Xu Xiang

Electrical and Computer Engineering, National University of Singapore, 4 Engineering Drive 3, Singapore 117576, Singapore

ARTICLE INFO

Article history:
Received 23 April 2009
Accepted 31 August 2010
Available online 8 September 2010

Keywords:
Motion analysis
Camera calibration
Error analysis

ABSTRACT

This paper presents an algorithm-independent geometrical analysis of the behavior of differential Structure from Motion (SFM) algorithms when there are errors in intrinsic parameters of the camera. We demonstrate both analytically and in simulation how uncertainty in the calibration parameters gets propagated to motion estimates in a differential setting. In particular, we studied how erroneous focal length and principal point estimates affect the behavior of the bas-relief ambiguity and introduce additional biasing to the translation estimate in a non-simple manner not revealed by previous analyses. Our formulation allows us to characterize the influence of various factors such as different scene-motion configurations and field of views in an analytically tractable manner. Guidelines are given as to whether one should err on the low or the high side in the estimation of the focal length depending on various operating conditions such as the feature density and the noise level. Simulations with synthetic data and real images were conducted to support our findings.

1. Introduction

Highly accurate camera calibration with special calibration object is not always possible when the camera system is changing its intrinsic parameters dynamically. In response to this, the computer vision community has developed schemes for self-calibration and investigated what structures can be recovered. Indeed a full-fledged Euclidean reconstruction is not always necessary, for instance in the task of visual servoing or in image-based rendering. Projective approaches aim to perform Structure from Motion (SFM) without calibration, that is, all the calibration information is neglected and the intrinsic camera parameters are assumed to vary freely from frame to frame. Oliensis [20] questioned whether the projective approach might not be too general to a fault. The projective approach assumes zero knowledge of the calibration. In practice, there are always something known about the intrinsic camera parameters, for instance, the skew factor being zero, or we do have a rough estimate of certain parameter even though it might not be exact. It is questionable whether such total neglect of available information leads to an increased or decreased robustness. This question remains largely unanswered despite the enormous amount of work done on developing projective algorithms. To answer this better, we need to know how using partial or approximate calibration might affect the estimation of the camera's egomotion and accordingly scene recovery, and whether these influences are large enough in practice to affect the goal of tasks to be carried out by the camera system. This paper aims at unraveling the errors introduced to the extrinsic motion parameters when there are uncertainties in the intrinsic parameters. Our geometrical approach reveals a clearer insight into the nature of this error propagation than what has been discovered by first order statistical analysis [24,27].

1.1. Related works

Early observations on the impact of erroneous calibration seem to run along the following line as typified by Bougnoux [1], who noted the difficulty of obtaining focal length in self-calibration but suggested from empirical evidence that part of the structure can be recovered despite error in self-calibration. The ground for this view, in so far as can be ascertained, seems to be based on the empirically good results of depth reconstructed as part of the self-calibration process. Zhang [26] performed self-calibration with a moving stereo rig, thus achieving redundancy compared to monocular sequence. They also empirically found that depths reconstructed are of good quality despite error in focal length estimate. Both did not analytically address the accuracy of depth reconstructions under general motion-scene configurations using those erroneous intrinsic parameters. Thus, despite these works that suggest that depth reconstruction is stable against error in calibration, there is still a paucity of theoretical evidence that this notion is true for all motion-scene configurations.

Beside the work of Bougnoux [1], there have been various other works [8,10] which report on the fact that self-calibration algorithms are sensitive to noise and attempt to improve the algorithms robustness and reliability. For instance, Hartley and Silpa-Anan [8] found that it is difficult to obtain the principal point

* Corresponding author. Fax: +65 67791103.
E-mail address: eleclf@nus.edu.sg (L.-F. Cheong).

1077-3142/$ - see front matter © 2010 Elsevier Inc. All rights reserved.
doi:10.1016/j.cviu.2010.08.004

and the focal length from the fundamental matrix, even though the matched points used are of high quality and the fundamental matrix is obtained using the bundle adjustment method. Other authors [9,13,21,22] analyzed the critical motion sequence in which no unique calibration can be obtained. Some of these general results have practical importance for certain motions and the special case of two-frame situation, which is also analyzed by [9,18]. The significance of these works lies in that those configurations near to the critical motion sequence would yield unstable intrinsic estimates. However, how these uncertainties in estimating intrinsic parameters would in turn affect egomotion estimates is not made clear in these papers.

Both [24,27] studied how uncertainty in the calibration parameters gets propagated to the motion parameters. Our work is closest in spirit to these works in that it examines the effects of errors in the intrinsic parameters on the estimation of the extrinsic parameters. However, instead of a statistical approach that examines the covariances like in [24], we adopt a geometrical approach. The conclusions from [24] regarding the impact on the rotational component of the egomotion estimates are not clear, though it seems that the rotational estimates can be quite badly affected. The authors noted that the influence of the precision of the calibration parameters on the motion parameters estimation depends on the types of camera motion and the scene type. However, their formulation did not allow an intuitive grasp of this scene-motion dependency and its interaction with the calibration errors. Simulations were conducted for some scene-motion configurations and error was introduced in the individual calibration parameter one by one. The piecemeal nature of the analysis means that the general applicability of the results is hard to ascertain. We investigate this dependency in a geometric manner and reveal clear insights into this dependence. Zucchelli and Košecká [27] also looked at the relationship between the errors in the intrinsic parameters and the egomotion estimates; their approach is based on perturbation analysis which relates the bias of the translation solution to the eigenvectors of matrices. With some first order approximations, it reveals that the calibration errors introduce an additional bias in the direction of the optical axis and that the bias produced by erroneous calibration increases in magnitude for increasing field of view (FOV). The experiments conducted belonged to the special type of motions whereby the camera's viewpoint was fixated at the centroid of the point cloud. We showed later that the additional bias can in fact move in a variety of ways depending on the motion-scene configuration, and explained why in the aforementioned fixating motions, the additional bias is constrained in its direction. Thus our approach offers a much more comprehensive picture than the first order numerical analysis of [27].

Another work that investigated the coupling between the intrinsic and the extrinsic parameters from another perspective is the recent work by González et al [7]. They have shown experimentally that there exists a strong coupling between the intrinsic and the extrinsic parameters. Most calibration methods, even those using static camera and calibration objects, suffer instability in the sense that the set of intrinsic parameters returned by a calibration method suffered important variations even though the camera only changes its extrinsic pose relative to the calibration pattern. Similar results have been obtained for the extrinsic parameters when the camera only changed its internal configuration (i.e. when it zooms in or out) and not its relative position to the calibration pattern. Although the error functions minimized by these different calibration techniques (usually minimizing the reprojection errors in the image or the reconstruction errors of the reference points in the 3-D space) yield similar error levels, it does not guarantee that the parameter estimates converge to the ground truth values, which is a serious problem if we want to use the calibrated camera in mobile applications. The main practical implication of this fact is that, when a camera is calibrated with any of these methods, we are "calibrating" the camera with just that pose. When subsequently the camera extrinsic parameters change, as in mobile applications, can we assume the same "calibrated" intrinsic values to carry out calibrated SFM analysis? In other words, one can get jointly optimal camera parameters (intrinsic and extrinsic) in a calibration algorithm but it is only optimal with respect to the cost function without necessarily meaning that these camera parameters are correct. Thus modeling the errors in the intrinsic parameters by simple Gaussian distributions—a typical approach in many error analyses—may not be appropriate, since the errors may depend on the extrinsic parameters. Clearly, there is a need to investigate how errors in the intrinsic parameters affect the estimation of extrinsic motion parameters, and how depth reconstruction would be affected as it critically depends on accurate egomotion estimation, especially in certain scene-motion configuration such as forward translation [3]. It is abundantly plausible that the task of depth reconstruction in the face of calibration uncertainty is a more complicated task than might be thought at first [1,26].

Last but not least, the bas-relief ambiguity is a well-known ambiguity for the case of calibrated two-frame SFM [6,16,19]. Oliensis [19] showed that unknown focal length variations strengthen the effects of the bas-relief ambiguity. This is attributed to the simple fact that the zoom flow is essentially not recoverable from the forward translation component. Together with the original bas-relief ambiguity, this new coupling renders all directions of the translation not accurately recoverable. The paper also went on to note that the motion errors depend simply on the estimated focal length and image center (e.g. the estimated translation differs from the true translation by a factor of the unknown focal length), but this is based on various assumptions such as the non-translational terms can still be annihilated in the proposed algorithm and that the second order terms are small. We look at this relationship between the extrinsic motion and the estimated focal length in detail, assuming that the camera is not undergoing zoom motion, and we found that errors in the focal length estimates modify the phenomenon of bas-relief ambiguity in a non-simple way (depending on the rotation and the scene depths). For instance, when the focal length is under-estimated, the bas-relief valley that marks the error surface could become truncated and the feasibility of the so-called flipped minimum solution [20] might diminish.

1.2. Summary of our contribution

The contribution of this paper lies in a geometrical, algorithm-independent analysis of the differential SFM cost function with some errors in the calibration parameters. In contrast to works that investigate the ambiguities in self-calibration under the so-called critical motion sequence [9,13,21,22], we are more interested to know how uncertainties in the calibration parameters get propagated to the egomotion estimates. In the preliminary version of this work [4], we analysed the case of erroneous estimate in the focal length and carried out simulation on synthetic data only. Here we showed how errors in both focal length and principal point estimates affect the behavior of the bas-relief ambiguity and introduce additional biasing to the translation estimate in a non-simple manner. We also obtained an approximate bound to the maximum amount of bias in the translation estimate. Other changes in the error surface, such as the length of the bas-relief valley and the number of local minima, were also noted. We provide experimental results on both synthetic and real data, with the latter revealing the crucial tradeoff between the bias and the variance of the motion estimates under realistic operating conditions. Our differential formulation characterizes the influence of various factors such as different scene-motion configurations and field of views in an analytically tractable manner—the biggest virtue of our approach. It is thus able to offer a much more comprehensive analysis of the various phenomenon than other approaches such as [19,24,27].

In terms of practical insights, it allows us to predict the type of errors in the egomotion estimates and to put a bound on them, as a function of the errors in the intrinsic parameters, as well as the scene and motion types. This allows us to partially answer the question raised in [20]: is it better to perform self-calibration or to use an existing calibrated setting despite small error in the parameters? If self-calibration is desired, our results also provide guideline as to whether to err on the low or the high side in the estimation of the focal length. This would result in different tradeoff between the bias and the variance of the motion estimates and could be crucial depending on the expected operating conditions such as the feature density and the noise level. Lastly, while our analysis is for the simpler differential case, the apparent differential/discrete dichotomy in the SFM problem is not really a fundamental one. Much of the ambiguities analyzed here are clearly also applicable to the discrete case.

The organization of this paper is as follows. First, we briefly review in Section 2 the various requisite background and introduce the notations used in this paper. In Section 3, we seek to characterize the various inherent ambiguities in 3-D motion estimation under erroneous calibration parameters. 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 compare our results against those obtained in previous works. These theoretical investigations are followed by experiments on synthetic and real images to verify the various predictions made. The paper ends with the conclusions of the work.

2. Background and prerequisite

2.1. Notations

In this subsection, we explain some of the mathematical notations that a reader will frequently encounter when reading the paper.

1. Unless otherwise stated, we use bold lower-case character to denote vector and bold upper-case character to denote matrix. Vectors are column vectors.
2. $[\mathbf{s}]_m$ 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 $\mathbf{s}$.
3. $\bar{\mathbf{s}}$ is the associated skew-symmetric matrix of $\mathbf{s}$.
4. 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.
5. $\hat{\mathbf{s}}$ and $\mathbf{s}_e$ We denote any estimate of the parameter s as $\hat{s}$, and error in this estimate $\hat{s}$ as $s_e = s - \hat{s}$.

2.2. Models

We now briefly introduce the various notations used in our error analysis via a review of the calibrated SFM scenario. A pinhole camera model with perspective projection is assumed as shown in Fig. 1. In the figure, the camera is moving with a translational velocity $\mathbf{v} = (U, V, W)^T$ and a rotation velocity $\mathbf{w} = (\alpha, \beta, \gamma)^T$. A point $\mathbf{P} = (X, Y, Z)^T$ in the world produces an image point $\mathbf{p} = (x, y, f)^T$ in the image plane, where f is the focal length. For a fixed focal length, the derivative $\dot{\mathbf{p}}$ is given by $(u, v, 0)^T$ where (u, v) is the optical flow. The latter is related to the 3D motion parameters by the following [14]:

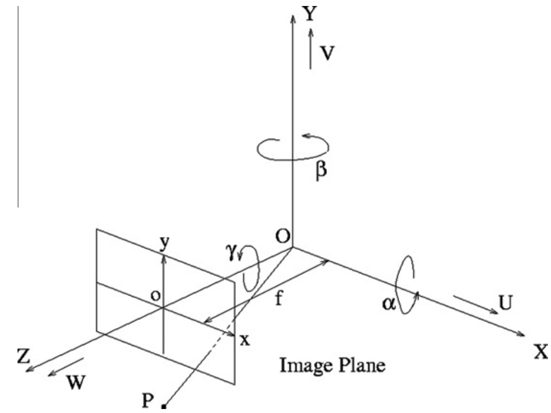

Fig. 1. The pinhole camera model.

$$u = \frac{u_{tr}}{Z} + u_{rot} = (x - x_0)\frac{W}{Z} + \frac{\alpha xy}{f} - \beta\left(\frac{x^2}{f} + f\right) + \gamma y$$
$$v = \frac{v_{tr}}{Z} + v_{rot} = (y - y_0)\frac{W}{Z} - \frac{\beta xy}{f} + \alpha\left(\frac{y^2}{f} + f\right) - \gamma x \quad (1)$$

In the preceding equation, $(x_0, y_0) = (f\frac{U}{W}, f\frac{V}{W})$ is the focus of expansion (FOE), and $\frac{\dot{\mathbf{p}}_{tr}}{Z} = \left(\frac{u_{tr}}{Z}, \frac{v_{tr}}{Z}\right)^T$ and $\dot{\mathbf{p}}_{rot} = (u_{rot}, v_{rot})^T$ are the flows components due to translation and rotation respectively. Since only the translational direction can be recovered from the flow field, we can set $W = 1$ without loss of generality.

2.3. Optimization criteria for SFM

Most of the existing cost functions for SFM are based on some forms of the epipolar constraint which was proposed by Longuet-Higgins [14]. The epipolar constraint relates the 3-D motion estimates with the image displacements in a manner independent of depth. In the differential case, the epipolar equation can be written in terms of the 3D motion estimates $\hat{\mathbf{v}}$ and $\hat{\mathbf{w}}$ as [2]:

$$\mathbf{p}^T\bar{\hat{\mathbf{v}}}\dot{\mathbf{p}} + \mathbf{p}^T\bar{\hat{\mathbf{v}}}\bar{\hat{\mathbf{w}}}\mathbf{p} = 0 \quad (2)$$

from which one can minimize the following cost function

$$J_{E1} = \sum_{i=1}^{n}\left(\mathbf{p}_i^T\bar{\hat{\mathbf{v}}}\dot{\mathbf{p}}_i + \mathbf{p}_i^T\bar{\hat{\mathbf{v}}}\bar{\hat{\mathbf{w}}}\mathbf{p}_i\right)^2 \quad (3)$$

where n is the number of image velocity measurement. The constraint J_{E1}can also be written in the following equivalent form:

$$J_{E1} = \sum_{i=1}^{n}\left(\left([\dot{\mathbf{p}}_i]s_2 - \hat{\dot{\mathbf{p}}}_{rot_i}\right)\cdot\hat{\dot{\mathbf{p}}}_{tr_i}^{\perp}\right)^2 \quad (4)$$

It says that in the image plane the derotated flow vector $[\dot{\mathbf{p}}_i]_2 - \hat{\dot{\mathbf{p}}}_{rot_i}$ should be parallel to the epipolar direction $\hat{\dot{\mathbf{p}}}_{tr_i}$, or equivalently perpendicular to $\hat{\dot{\mathbf{p}}}_{tr_i}^{\perp}$. However, a bias of the estimated translation is well-known to be present when a linear algorithm based on (4) is applied. In view of this bias, various weighted version of J_{E1} have been proposed as a statistically more adequate implementation of the differential epipolar constraint, and it has been shown in [25] that all of them can be expressed in the following form:

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

where $\mathbf{n}_i$ is a unit vector in the image plane representing a particular direction associated with the ith image point. Various weighted differential epipolar constraints differ mainly in the choice of this unit vector **n**. It was also shown that the key properties of the various cost functions used in different algorithms are determined by the angle between the two vectors involved in the dot product in the numerator; the choice of **n** in the denominator might affect the detailed numerical properties but has little influence on key properties such as the formation of the bas-relief valley on the error surface (see [25] for more details). Thus we can study the properties of the various algorithms by looking at the generic cost function given by Eq. (5).

3. Behavior of motion estimation algorithms with erroneous estimated focal length

The preceding section reviewed the key factors affecting the behavior of motion estimation algorithms for the calibrated case. This section will investigate the behavior of extrinsic motion estimation under erroneous camera calibration. In particular, we consider how extrinsic motion estimation would be affected by fixed errors in the estimates of the focal length and the principal point offset.

First, we need to express the cost function J_R in terms of the various component errors in the 3-D motion estimates together with terms arising from errors in the estimates for the intrinsic parameters. We reiterate that it is a geometrical and algorithm-independent approach. Thus we ignore those errors that arise from stochastic noise in the image measurements but directly consider the errors in the camera parameters made by *any* motion estimation algorithms. For clarity of presentation, we first consider the case where the only intrinsic parameter with error is the focal length, leaving the full case to Section 3.3 and Appendix A.2. Substituting $\hat{\dot{\mathbf{p}}}_{tr_i} = (x_i - \hat{x}_0, y_i - \hat{y}_0)^T$, $[\dot{\mathbf{p}}_i]_2 = (u_i, v_i)^T = \left(\frac{x_i - x_0}{Z_i} + u_{rot_i}, \frac{y_i - \hat{y}_0}{Z_i} + v_{rot_i}\right)^T$ and $\hat{\dot{\mathbf{p}}}_{rot_i} = (\hat{u}_{rot_i}, \hat{v}_{rot_i})^T$ into Eq. (5) we have:

$$J_R = \sum_{i=1}^{n} \left(\frac{(x_i - \hat{x}_0, y_i - \hat{y}_0) \cdot \left(v_{rot_e} - \frac{y_{0e}}{Z_i}, \frac{x_{0e}}{Z_i} - u_{rot_e} \right)}{(x_i - \hat{x}_0, y_i - \hat{y}_0) \cdot \mathbf{n_i}} \right)^2 \tag{6}$$

where the various error terms are expanded as follows:

$$\begin{aligned}
u_{rot_e} &= -\left(\beta f - \hat{\beta}\hat{f}\right) + \left(\frac{\alpha}{f} - \frac{\hat{\alpha}}{\hat{f}}\right) x_i y_i - \left(\frac{\beta}{f} - \frac{\hat{\beta}}{\hat{f}}\right) x_i^2 + \gamma_e y_i \\
v_{rot_e} &= \left(\alpha f - \hat{\alpha}\hat{f}\right) + \left(\frac{\alpha}{f} - \frac{\hat{\alpha}}{\hat{f}}\right) y_i^2 - \left(\frac{\beta}{f} - \frac{\hat{\beta}}{\hat{f}}\right) x_i, y_i - \gamma_e x_i \\
(x_{0_e}, y_{0_e}) &= (x_0 - \hat{x}_0, y_0 - \hat{y}_0)
\end{aligned} \tag{7}$$

Note that u_{rot_e} and v_{rot_e} now contain terms due to the inaccurate focal length estimate $\hat{f}$ for our uncalibrated scenario. For notational convenience, we shall henceforth omit the subscript i in the expression of J_R, although it is understood that the summation runs over all feature points. Like the calibrated case analyzed in [25], it is the angular relationship between the two terms in the numerator of Eq. (6) $(x - \hat{x}_0, y - \hat{y}_0)^T$ and $\left(v_{rot_e} - \frac{y_{0e}}{Z}, \frac{x_{0e}}{Z} - u_{rot_e}\right)^T$ that governs the behaviour of the error surface of J_R, and in particular, the formation of the bas-relief valley. We denote these two terms as $\mathbf{t}_1$and $\mathbf{t}_2$ respectively, and will analyze how the angular relationship between them changes in the light of calibration errors. Furthermore, we also want to investigate the individual contribution of the terms in $\mathbf{t}_1$and $\mathbf{t}_2$. For that purpose, we adopt the terminology that $\mathbf{t}_{1,n}$ and $\mathbf{t}_{2,n}$ denote the nth order component (with respect to x and y) in $\mathbf{t}_1$ and $\mathbf{t}_2$ respectively. Accordingly, we have:

$$\begin{aligned}
J_R &= \sum \left(\frac{\mathbf{t}_1 \cdot \mathbf{t}_2}{\mathbf{t}_1 \cdot \mathbf{n}} \right)^2 \\
\mathbf{t}_1 &= \mathbf{t}_{1,0} + \mathbf{t}_{1,1} \\
\mathbf{t}_2 &= \mathbf{t}_{2,0} + \mathbf{t}_{2,1} + \mathbf{t}_{2,2} + \mathbf{t}_{2,Z}
\end{aligned} \tag{8}$$

where

$$\begin{aligned}
\mathbf{t}_{1,0} &= (-\hat{x}_0, -\hat{y}_0)^T \\
\mathbf{t}_{1,1} &= (x, y)^T \\
\mathbf{t}_{2,0} &= \left(\left(\alpha f - \hat{\alpha}\hat{f}\right), \left(\beta f - \hat{\beta}\hat{f}\right) \right)^T \\
\mathbf{t}_{2,1} &= (-\gamma_e x - \gamma_e y)^T \\
\mathbf{t}_{2,2} &= \left(\left(\frac{\alpha}{f} - \frac{\hat{\alpha}}{\hat{f}}\right) y^2 - \left(\frac{\beta}{f} - \frac{\hat{\beta}}{\hat{f}}\right) xy, - \left(\frac{\alpha}{f} - \frac{\hat{\alpha}}{\hat{f}}\right) xy + \left(\frac{\beta}{f} - \frac{\hat{\beta}}{\hat{f}}\right) x^2 \right)^T \\
\mathbf{t}_{2,Z} &= \left(-\frac{y_{0e}}{Z}, \frac{x_{0e}}{Z} \right)^T
\end{aligned} \tag{9}$$

Since the depth Z may be dependent on x and y in a complex manner, we use the notation $\mathbf{t}_{2,Z}$ without explicitly specifying the order of this term.

Eqs. (6) and (7) show 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 $(\hat{x}_0, \hat{y}_0)$, the error in the rotation estimates $(\alpha_e, \beta_e, \gamma_e)$ and the estimated focal length $\hat{f}$. In comparison with the calibrated case (by setting $\hat{f} = f$), we immediately note the following:

1. The estimation of γ is quite independent of camera calibration since the γ_e term is not coupled with the intrinsic parameters in any meaningful way. Thus, geometrically speaking, γ can be estimated well, like in the case of calibrated SFM.
2. The true rotational parameters play a part in the formation of the error surface. This is unlike the calibrated case where the cost function only depends on errors in the rotational parameters and not the true rotational parameters themselves (note that for the calibrated case, $\mathbf{t}_{2,0} = (\alpha_e f, \beta_e f)^T$ and $\mathbf{t}_{2,2} = \left(\alpha_e \frac{y^2}{f} - \frac{\beta_e xy}{f}, -\frac{\alpha_e xy}{f} + \beta_e \frac{x^2}{f}\right)^T$).

3.1. Changes to the bas-relief valley

The properties of the motion estimation algorithms depend on the angular relationship between the terms in the numerator of J_R in Eq. (8) in the following sense. If there exists a class of motion solutions that make the dot product in the numerator of J_R vanish, then ambiguities exist. We recapitulate the two conditions discussed in [25] that should be satisfied to make the numerator of the cost function J_R vanish:

$$\begin{aligned}
&(1)\ \text{making } \mathbf{t}_1 \text{ and } \mathbf{t}_2 \text{ perpendicular to each other, and} \\
&(2)\ \text{making } \|\mathbf{t}_2\| \text{ small.}
\end{aligned} \tag{10}$$

The requirement for condition (1) is clear from Eq. (8). Condition (2) helps because condition (1) can never be completely satisfied at every image point under general motion-scene configuration with depth Z not a constant value. Making $\|\mathbf{t}_1\|$ small does not help since it appears in both the numerator and the denominator.

We now examine the implications of the two conditions in (10) for the uncalibrated case. From the expressions of $\mathbf{t}_1$ and $\mathbf{t}_2$ in Eq. (8), we can see that $\mathbf{t}_{1,0}$, $\mathbf{t}_{2,0}$ and $\mathbf{t}_{2,Z}$ are pointing towards constant directions for all the feature points. If we consider $\mathbf{t}_{1,1}$ as a perturbation to the constant-direction vector $\mathbf{t}_{1,0}$ and $(\mathbf{t}_{2,1} + \mathbf{t}_{2,2})$ as a perturbation to $(\mathbf{t}_{2,0} + \mathbf{t}_{2,Z})$,[1] then making the constant-direction vectors $(\mathbf{t}_{2,0} + \mathbf{t}_{2,Z})$ and $\mathbf{t}_{1,0}$ perpendicular to each other is a reasonable choice for the minimization of J_R (see Fig. 2). Thus we have

[1] This statement means that we require the feature points to be 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 $\mathbf{t}_{2,0} + \mathbf{t}_{2,Z}$, and the distribution of depth Z is symmetrical with respect to the $\mathbf{t}_{1,0}$ direction.

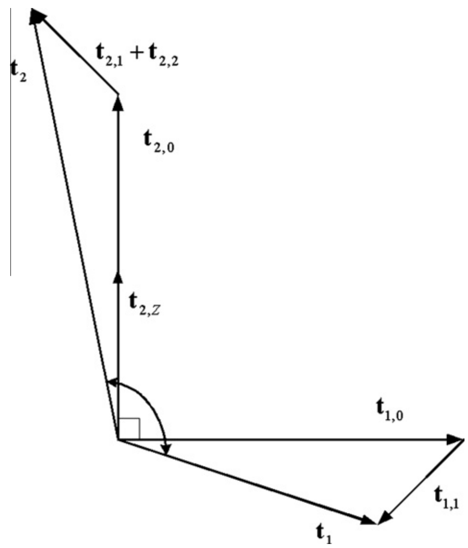

Fig. 2. Geometry of $\mathbf{t}_1$ and $\mathbf{t}_2$. To create conditions conducive for the formation of the bas-relief ambiguity, the vectors ($\mathbf{t}_{2,0} + \mathbf{t}_{2,Z}$) and $\mathbf{t}_{1,0}$ should be perpendicular to each other.

$$\frac{y_{0e} - \alpha f Z + \hat{\alpha}\hat{f}Z}{x_{0e} + \beta f Z - \hat{\beta}\hat{f}Z} = \frac{\hat{y}_0}{\hat{x}_0} \tag{11}$$

or equivalently

$$\frac{y_{0e} - \alpha_e f Z - \hat{\alpha} f_e Z}{x_{0e} + \beta_e f Z + \hat{\beta} f_e Z} = \frac{\hat{y}_0}{\hat{x}_0} \tag{12}$$

The last equation shows that, in the case $\hat{f} = f$, the last terms in both the numerator and the denominator on the left hand side vanish. The equation reduces to the calibrated case, and as discussed in [25] it can be satisfied by obeying two independent constraints, the first one relating to the translational parameters $\frac{x_{0e}}{y_{0e}} = \frac{\hat{x}_0}{\hat{y}_0}$ (which implies $\frac{x_{0e}}{y_{0e}} = \frac{x_0}{y_0}$), and the second one relating to the rotational parameters $\frac{\alpha_e}{\beta_e} = -\frac{\hat{y}_0}{\hat{x}_0}$. The first constraint characterizes the valley that gives rise to the bas-relief ambiguity found in calibrated SFM algorithms. Any motion estimate whose FOE lies along the straight line passing through the true FOE and the origin will have small J_R as long as the second constraint is also satisfied. However, in the uncalibrated case, when the error in the focal length f_e is significant, α_e and β_e cannot be freely varied to satisfy the second constraint $\frac{\alpha_e}{\beta_e} = -\frac{\hat{y}_0}{\hat{x}_0}$. Rather, if there is significant error in the estimate $\hat{f}$, the term $\mathbf{t}_{2,2}$ can no longer be treated as second order effect and be ignored relative to $\mathbf{t}_{2,0}$. Comparing terms in $\mathbf{t}_{2,0}$ and $\mathbf{t}_{2,2}$, we observe that even if the FOV is small (i.e. f is large) such that $\frac{\alpha}{f} \ll \alpha f$ and $\frac{\beta}{f} \ll \beta f$, the corresponding relationships for the estimated terms $\frac{\hat{\alpha}}{\hat{f}} \ll \hat{\alpha}\hat{f}$ and $\frac{\hat{\beta}}{\hat{f}} \ll \hat{\beta}\hat{f}$ may not be valid, and thus we cannot assert $\|\mathbf{t}_{2,2}\| \ll \|\mathbf{t}_{2,0}\|$. This is the case when $\hat{f}$ is under-estimated such that the estimated FOV is large. Under such circumstance, making $\|\mathbf{t}_{2,2}\|$ small is just as important towards minimizing the cost function J_R. Clearly this gives rise to the following constraint on the rotational estimates:

$$\frac{\alpha}{f} = \frac{\hat{\alpha}}{\hat{f}}, \quad \frac{\beta}{f} = \frac{\hat{\beta}}{\hat{f}} \tag{13}$$

Note that in the above, the quantity $\hat{f}$ is fixed (since we are considering fixed focal length estimate); thus the two equations in (13) fully specify $\hat{\alpha}$ and $\hat{\beta}$. With this constraint, α_e and β_e cannot be freely varied such that the original calibrated constraint $\frac{\alpha_e}{\beta_e} = -\frac{\hat{y}_0}{\hat{x}_0}$ is satisfied. Thus to satisfy constraint (11), we cannot decompose it into two independent constraints like in the calibrated case. Rather, to satisfy both (11) and (13) at the same time, we substitute (13) into (11) and obtain a single constraint:

$$\frac{y_{0e} - \alpha f Z\left(1 - \left(\frac{\hat{f}}{f}\right)^2\right)}{x_{0e} + \beta f Z\left(1 - \left(\frac{\hat{f}}{f}\right)^2\right)} = \frac{\hat{y}_0}{\hat{x}_0} \tag{14}$$

which can also be written as

$$\frac{y_0 - \alpha f Z\left(1 - \left(\frac{\hat{f}}{f}\right)^2\right)}{x_0 + \beta f Z\left(1 - \left(\frac{\hat{f}}{f}\right)^2\right)} = \frac{\hat{y}_0}{\hat{x}_0} \tag{15}$$

The above expresses a constraint on the direction of the estimated FOE $(\hat{x}_0, \hat{y}_0)$ that dictates the formation of the bas-relief valley. Compared to the original bas-relief constraint in the calibrated case $\frac{\hat{y}_0}{\hat{x}_0} = \frac{y_0}{x_0}$, which is a straight line passing through the true FOE and the origin, this modified constraint indicates a "bas-relief" valley that has a different slope in general given by $\frac{y_0 - \alpha f Z\left(1 - \left(\frac{\hat{f}}{f}\right)^2\right)}{x_0 + \beta f Z\left(1 - \left(\frac{\hat{f}}{f}\right)^2\right)}$. In particular, consider the shift in the FOE estimate $(\hat{x}_0, \hat{y}_0)$ caused by the term $\beta f Z\left(1 - \left(\frac{\hat{f}}{f}\right)^2\right)$ and $-\alpha f Z\left(1 - \left(\frac{\hat{f}}{f}\right)^2\right)$. One can also interpret this shift as an additional bias to the FOE estimate caused by the error in the focal length estimate, over and above the well-known bias towards the optical center. This bias was also investigated in [27], but their approach has difficulty in analytically deriving the bias as a function of the various factors. Using simulation, they seemed to obtain the result (though without offering any explanation) that under-estimation of focal length results in a larger bias than over-estimation of focal length. We confirm and explain later that the bias is indeed larger for under-estimation of focal length, but our approach also allows us to show how the direction of the FOE bias is a function of the actual translation and rotation.

Furthermore, recall from the conditions stated in (10) that ambiguity is more likely to arise if $\|\mathbf{t}_2\|$ is also small. Of the terms in $\|\mathbf{t}_2\|$, the rotational errors α_e and β_e in $\|\mathbf{t}_2\|$ can no longer be freely varied due to Eq. (13); thus $\hat{x}_0$ and $\hat{y}_0$ are clearly constrained in magnitude in order to make $\|\mathbf{t}_{2,Z}\|$ and thus $\|\mathbf{t}_2\|$ small. In other words, $(\hat{x}_0, \hat{y}_0)$ is not only just constrained in direction but also in magnitude; this is unlike the small field calibrated case, where any residual error caused by the translational errors can be compensated for by a suitable choice of α_e and β_e. Accordingly, we expect that in general, the bas-relief valley might not straddle across the entire visual field. In particular, the feasibility of the flipped minimum solution (i.e. $(\hat{x}_0, \hat{y}_0) = -(x_0, y_0)$) that exists under calibrated scenario [20] would be diminished. On the other hand, due to the presence of the Z term in the constraint (14), we expect the shape of this bas-relief valley to be markedly affected by the way the scene points are distributed. For a cluttered scene with non-smooth depth distribution, the valley will be less well-defined. That is, instead of a narrow and elongated valley that stretches across the entire visual field, it would be broader and rather reduced in length to a local quadrant. We also expect more local minima in the solution space due to the non-smooth Z term in the constraint (14), which could pose convergence problem for a Euclidean SFM algorithm assuming erroneous calibration parameters. As a result, using a projective SFM algorithm under such situation might have the advantage of facing less of a local-minimum problem.

In sum, the Euclidean SFM algorithms assuming erroneous calibration parameters exhibit different behavior from the error-free case, and these deviations are more distinct when either the actual FOV or the estimated FOV is large, because then the constraint on $\hat{\alpha}$ and $\hat{\beta}$ (Eq. (13)) is stronger. Fig. 3 corroborates the prediction that over- and under-estimating focal length f by the same amount (i.e. same $|f_e|$) has different degree of influence on the estimation of FOE. There are 50 trials for over-estimating f and 50 trials for under-estimating f. An isotropic random noise of 30% is added to the optical flow on each trial. The motion parameters are estimated using the procedure described in Section 3.2. The results show that under-estimating f gives rise to more pronounced shift of the estimated FOE compared to over-estimating f (given the same magnitude in f_e). This is consistent with the somewhat paradoxical finding of [27] that larger FOV gives rise to larger bias in the translation estimate. Note, however, that over-estimating f results in a larger variance in the FOE estimate under the influence of random image noise. Eq. (13) also means that we can recover the ratio of α to β with better accuracy when the FOV or the estimated FOV is large. This can be seen in Fig. 4, where the motion parameters are again estimated using the procedure described in Section 3.2 under noiseless condition. With a FOV of 53°, the curves $\frac{\hat{f}}{f}$, $\frac{\hat{\alpha}}{\alpha}$ and $\frac{\hat{\beta}}{\beta}$ increase approximately in tandem, thus bearing out the predictions set out in Eq. (13), which also means that the ratio of α to β can be recovered relatively well.

3.2. Visualizing the error surface J_R

Further properties of the motion estimation process under calibration errors will be visualized through plotting the residual of the cost function J_R. Before doing so, let us discuss briefly the plotting of this surface. For easier visualization, we consider a 3-dimensional surface, where each point on the surface represents a FOE hypothesis, with the height representing the residue J_R. Given a particular FOE hypothesis and a fixed (possibly erroneous) focal length estimate, the rotation variables can be solved via a linear algorithm while minimizing J_R (for more details of this rotation estimation procedure, please refer to Appendix A.1). Thus for each FOE candidate, we have an associated residual value J_R. These residual values J_R were then plotted over the 2D solution space for the FOE, in such a way that the image intensity encoded the

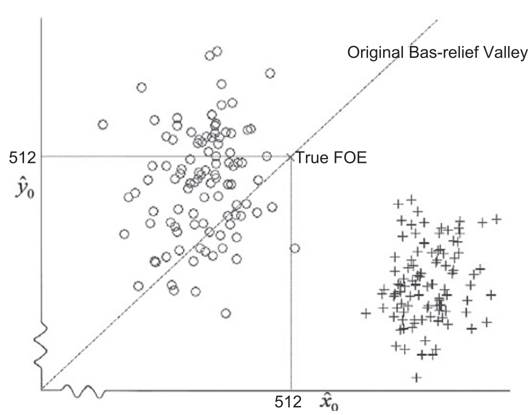

Fig. 3. Over- and under-estimating focal length f by the same amount (i.e. same $|f_e|$) has different degree of influence on the estimation of FOE. The true FOE is marked with "×". Estimated FOEs with under- and over-estimated focal length are marked with "+" and "∘" respectively. There are 50 trials for over-estimating f and 50 trials for under-estimating f. The scene depths are randomly distributed between 512 and 1536 pixel units and the translational motion is given by (1,1,1) pixel units per second. An isotropic random noise of 30% is added to the optical flow on each trial. Under-estimating f ("+") gives rise to more pronounced shift of the estimated FOE compared to over-estimating f ("∘"); however, the latter displays a larger variance in the estimate under the influence of random image noise.

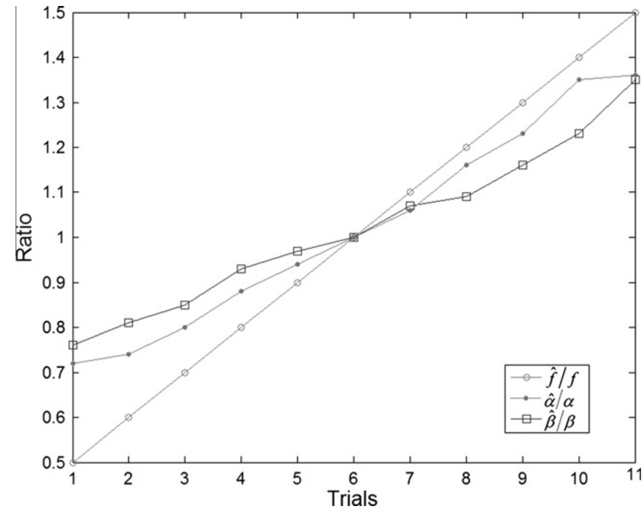

Fig. 4. With a relatively wide FOV of 53°, the constraint exerted on the rotational estimates $\hat{\alpha}$ and $\hat{\beta}$ is strong. The curves $\frac{\hat{f}}{f}$, $\frac{\hat{\alpha}}{\alpha}$ and $\frac{\hat{\beta}}{\beta}$ increase approximately in tandem with increasing $\hat{f}$, which means that the ratio of α to β can be recovered well. Conditions simulated are the same as in Fig. 3 but without any image noise.

relative value of the residual (bright pixels corresponded to high residual values and vice versa).

Some assumptions are made regarding the distribution of the feature points and the depths. 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. As for the intrinsic and extrinsic parameters of the camera, different combinations of translation and rotation with over- and under-estimation of f are simulated and plotted separately.

To represent the entire hemisphere in front of the camera, we used visual angle in degree rather than pixel when stepping through the FOE search space; thus the co-ordinates in the plots from Figs. 5–10 were not linear in the pixel unit. The imaging surface was a plane with a dimension of 512 × 512 pixels; its boundary was represented by the small rectangles in the center of the plots. The synthetic experiments have the following parameters: unless otherwise stated, the focal length was 512 pixels which meant a FOV of approximately 53°; the estimated focal length is either halved (256 pixels) for under-estimation or doubled (1024 pixels) for over-estimation. There were 200 feature points distributed randomly over the image plane, with depths

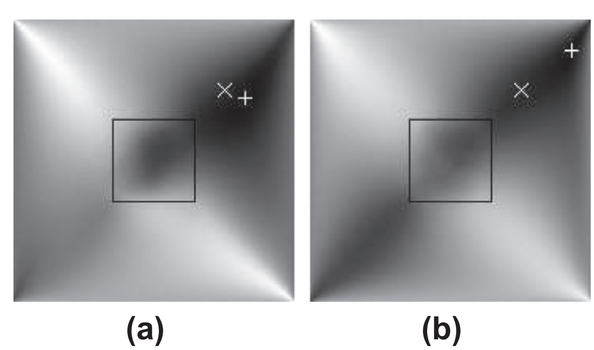

Fig. 5. The bas-relief valley is rotated if there is an error in the focal length estimate (50% under-estimated here). $\mathbf{v}$ = (1,1,1), $\mathbf{w}$ = (0.001,0.001,0.001). (a) FOV = 53°. (b) FOV = 28°. For all figures, true FOEs and global minima are highlighted by "×" and "+" respectively. Comparison between (a) and (b) reveals the influence of FOV on the amount of bas-relief valley rotation. Larger FOV results in larger rotation and the bas-relief valley becomes less well-defined and less elongated.

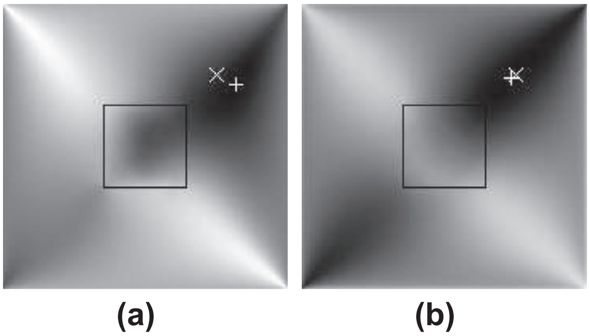

Fig. 6. The influence of estimate $\hat{f}$ (with $f = 512$) on the amount of bas-relief rotation. (a) $\hat{f} = 256$, focal length under-estimated by 50%, with distinct rotation of the bas-relief valley, (b) $\hat{f} = 1024$, focal length over-estimated by 100%, but rotation of the bas-relief valley not conspicuous. Bas-relief valley also becomes less well-defined under large estimated FOV in (a).

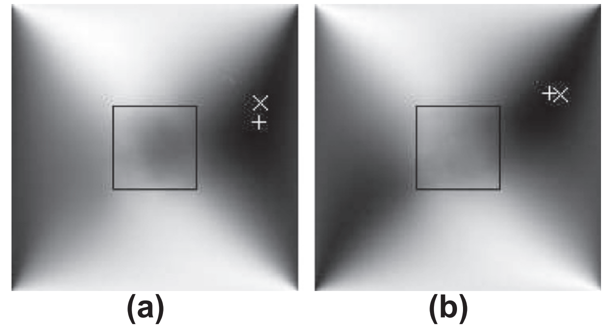

Fig. 8. Rotation of bas-relief valley when the "directions" of (x_0, y_0) and (α, β) are in adjacent quadrants. $(U, V, W) = (3, 1, 1)$, $f = 512$, and $\hat{f} = 256$. Residual error maps are plotted with (a) $(\alpha, \beta, \gamma) = (0.003, -0.001, 0)$, and (b) $(\alpha, \beta, \gamma) = (0.001, -0.007, 0)$. The direction of rotation is clockwise for (a) and anti-clockwise for (b).

ranging from one to three times the focal length (i.e. 512–1536 pixel units). The camera was undergoing a general translation with $\mathbf{v} = (1, 1, 1)$ pixel units per second. The rotation is such that the translational flow and the rotational flow are approximately equal in magnitude. For details, please see the individual figures.

3.3. Further properties of motion estimation with calibration errors

We use the next few figures (Figs. 5–10) to corroborate predictions made in the preceding subsection as well as making further observations. For all figures, true FOEs and the estimated FOEs are indicated by "×" and "+" respectively.

1. **Influence of FOV**. Fig. 5 illustrates the influence of visual field. Under large FOV (53°), the second order flow field $\mathbf{t}_{2,2}$ exerts a stronger influence through Eq. (13), which constrains the value of $\hat{\alpha}$ and $\hat{\beta}$. As discussed above, this constraint on $\hat{\alpha}$ and $\hat{\beta}$ in turn reduces the length of the valley formed by the bas-relief ambiguity, while at the same time the rotation of the bas-relief valley is made more pronounced, although the valley itself becomes more "diffused" and shallow (Fig. 5a). In small FOV (28°), the constraint (13) is less effective; the constraint in (11) can be broken down into two independent constraints like in the calibrated case, resulting in a bas-relief valley that stretches across almost the entire visual field, with little rotation in the direction of this valley compared to the calibrated case (Fig. 5b).
2. **Error in the estimate** $\hat{f}$. The relative importance of $\mathbf{t}_{2,2}$ is also affected by the estimated focal length $\hat{f}$. This can be seen by pitting the magnitude of the various terms of $\mathbf{t}_{2,2}$ against those of $\mathbf{t}_{2,0}$ (see the definitions in (9)), which include among others,

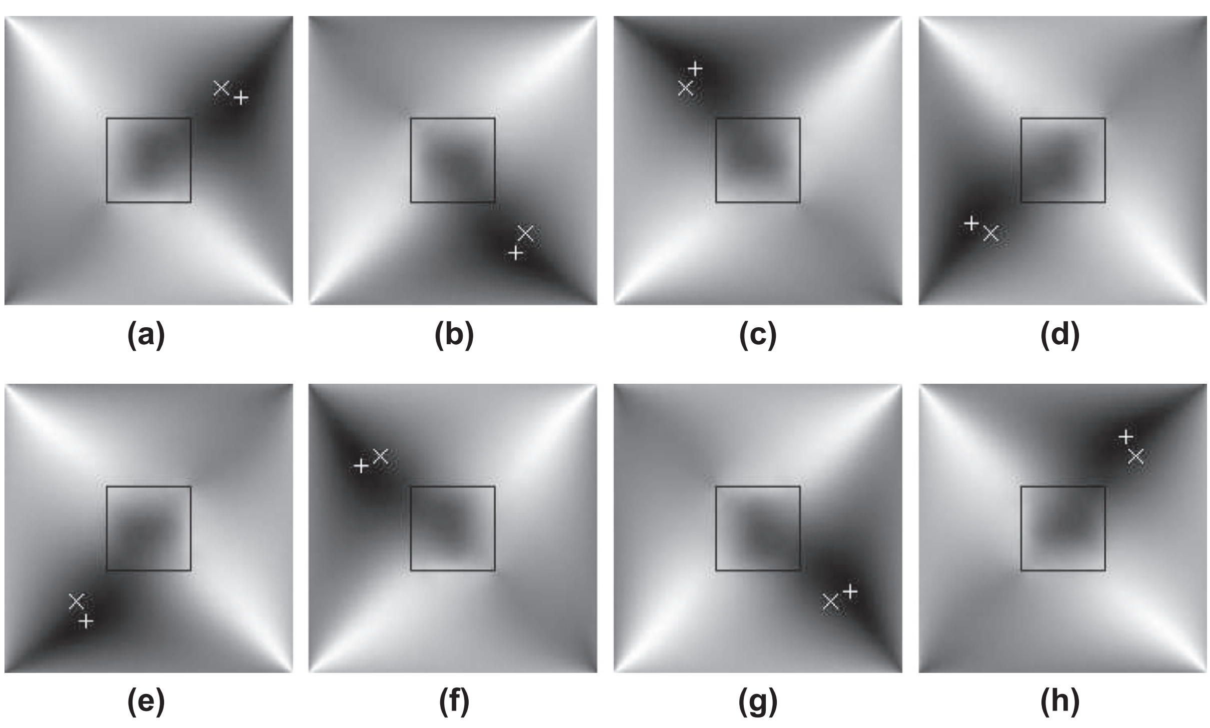

Fig. 7. Rotation of the bas-relief valley for (x_0, y_0) and (α, β) in different quadrants, with under-estimated focal length. In the first row, where (x_0, y_0) and (α, β) are in the same quadrant, the bas-relief valley experiences a clockwise rotation; whereas in the second row, where (x_0, y_0) and (α, β) are in diametrically opposite quadrants, the bas-relief valley rotates in an anti-clockwise direction. $W = 1$, $\gamma = 0.001$ $f = 512$ and $\hat{f} = 256$ for all figures. The (U, V) and (α, β) are respectively (a) $(1, 1)$, $(0.001, 0.001)$ (b) $(1, -1)$, $(0.001, -0.001)$ (c) $(-1, 1)$, $(-0.001, 0.001)$ (d) $(-1, -1)$, $(-0.001, -0.001)$ (e) $(-1, -1)$, $(0.001, 0.001)$ (f) $(-1, 1)$, $(0.001, -0.001)$ (g) $(1, -1)$, $(-0.001, 0.001)$ (h) $(1, 1)$, $(-0.001, -0.001)$.

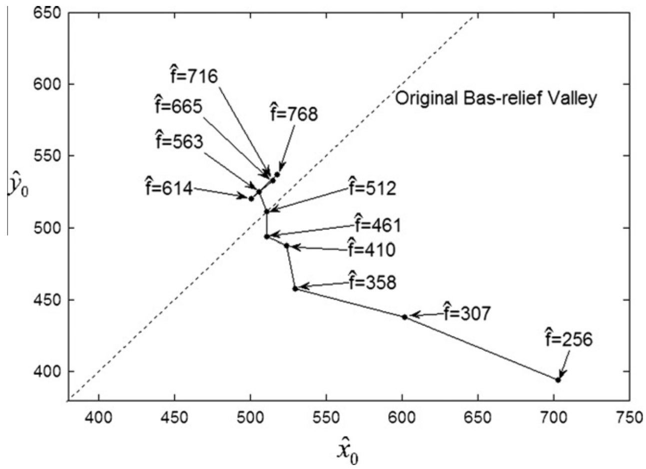

Fig. 9. The amount of shift in the estimated FOE with different errors in the estimated focal length. The true focal length is 512, whereas the estimated focal length vary from 256 (50% under-estimation) to 768 (50% over-estimation), with a step size of 10% error. The translational and rotational parameters are $(U, V, W) = (1, 1, 1)$ and $(\alpha, \beta, \gamma) = (0.001, 0.001, 0.001)$ respectively. True FOE lies at the point (512, 512) on the bas-relief valley. The estimated FOEs deviate very little away from the true solution for the case of over-estimation in $\hat{f}$. For the case of under-estimation in $\hat{f}$, the amount of shift in the FOE is more significant. However, even with a rather large under-estimation error of 50% in $\hat{f}$, the relative shift in the estimate $\hat{x}_0$ is only about 37%.

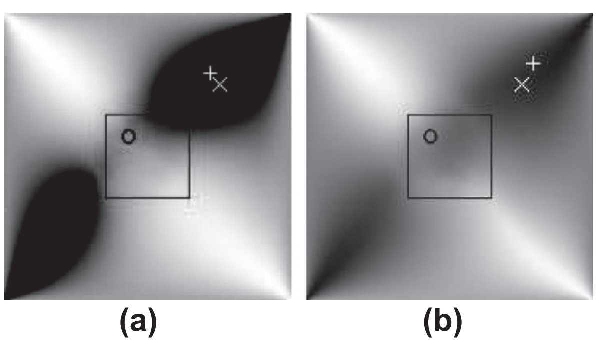

Fig. 10. The bas-relief valley with erroneous principal point estimate $(\hat{O}_x, \hat{O}_y) = (0, 0)$. The entire bas-relief valley is shifted by a constant amount and passes through the true principal point at $(100, -100)$ (indicated by "°"). The bas-relief valleys appear bent because we have used visual angle in degree rather than pixel as the FOE search step and thus the co-ordinates in the plots were not linear in the pixel unit. $(U, V, W) = (3, 1, 1)$, $(\alpha, \beta, \gamma) = (0.003, -0.001, 0)$, and $f = 512$. (a) $\hat{f} = 512$ (b) $\hat{f} = 256$ (50% under-estimation).

$\frac{\alpha}{f}$ versus αf, $\frac{\beta}{f}$ versus βf, $\frac{\hat{\alpha}}{\hat{f}}$ versus $\hat{\alpha}\hat{f}$, and $\frac{\hat{\beta}}{\hat{f}}$ versus $\hat{\beta}\hat{f}$. Given a particular f, under-estimating f (i.e. $\hat{f}$ becomes small) has the effect of enhancing the second order effect through raising $\frac{\hat{\alpha}}{\hat{f}}$ compared to $\hat{\alpha}\hat{f}$ and $\frac{\hat{\beta}}{\hat{f}}$ compared to $\hat{\beta}\hat{f}$. Thus under-estimating f would in general produce a stronger modification to the bas-relief valley compared to over-estimating f. This is clearly illustrated in Fig. 6, where even the amount of f over-estimation (100%) is larger than the amount of under-estimation (50%), the tilting of the bas-relief valley for the former (Fig. 6b) is much less than that of the latter (Fig. 6a). What this means is that if we want to recover the true FOE, it is better to over-estimate f than to under-estimate f. Note also that due to the larger estimated FOV in Fig. 6a, there is a shortening of the bas-relief valley like Fig. 5a; its more diffused character is also clear. To a certain extent, the effects of $\hat{f}$ and f (related to the FOV effect discussed in the preceding section) can be swapped. This "symmetry" can be seen from the various terms found in $\mathbf{t}_{2,2}$ and $\mathbf{t}_{2,0}$: $\frac{\alpha}{f}$, αf, $\frac{\beta}{f}$, βf, $\frac{\hat{\alpha}}{\hat{f}}$, $\hat{\alpha}\hat{f}$, $\frac{\hat{\beta}}{\hat{f}}$ and $\hat{\beta}\hat{f}$: The effect of under-estimating $\hat{f}$ can be replicated by an increase in the FOV (i.e. lowering f). However, in general, when the rotational estimates are not equal to their true values ($\hat{\alpha} \neq \alpha$, $\hat{\beta} \neq \beta$), this swapping between $\hat{f}$ and f may not yield identical effects.

3. **Direction of valley rotation**. Referring to Eq. (15), the direction in which the bas-relief valley rotates depends on a variety of factors such as the sign of f_e and the angle between (α,β) and (x_0,y_0). We illustrate the relationship by first looking at the case when $\alpha > 0, \beta > 0$, $x_0 > 0$ and $y_0 > 0$. The direction of rotation depends on the sign of f_e in the following way. If $f_e > 0$ (under-estimation), the signs of the terms $\alpha f Z\left(1 - \left(\frac{\hat{f}}{f}\right)^2\right)$ and $\beta f Z\left(1 - \left(\frac{\hat{f}}{f}\right)^2\right)$ in Eq. (15) are both positive. It is then clear that the new slope of the bas-relief valley $\frac{\hat{y}_0}{\hat{x}_0} = \frac{y_0 - \alpha f Z\left(1-\left(\frac{\hat{f}}{f}\right)^2\right)}{x_0 + \beta f Z\left(1-\left(\frac{\hat{f}}{f}\right)^2\right)}$ deviates from the original direction $\frac{y_0}{x_0}$ (when $f_e = 0$) in a clockwise manner (Fig. 6a). Conversely, when $f_e < 0$ (over-estimation), the rotation in the bas-relief valley is in an anti-clockwise direction. However the amount of rotation is not so conspicuous compared to the case of $f_e > 0$ (Fig. 6b). The reason for this anisotropy with respect to the sign of f_e has been explained earlier by their respective effects on the importance of the $\mathbf{t}_{2,2}$ term. To aid further discussion for all the other cases, we define the direction of various vectors as follows. For instance, when $\alpha > 0$ and $\beta > 0$, we say that the vector (α,β) is in the first quadrant. Carrying out the analysis for all the other cases, we find that the bas-relief valley rotates as follows. For the case of under-estimation of f, if (α,β) is in the same quadrant as (x_0,y_0), the bas-relief valley rotates in a clockwise direction (Fig. 7, first row). Conversely, if the two vectors (α,β) and (x_0,y_0) reside in diametrically opposite quadrants, the bas-relief valley rotates in an anti-clockwise direction (Fig. 7, second row). The amount of angular deviation from the true FOE is about 17° on the average. For the case of over-estimation of f, this relationship is exactly reversed but the amount of deviation is much less. If the two vectors (α,β) and (x_0,y_0) are in adjacent quadrants (e.g. quadrants 1 and 2), the direction of valley rotation can be clockwise or anti-clockwise or there can be no rotation, depending on the relative magnitudes of the various terms. For instance, in Fig. 8, the "directions" of (x_0,y_0) and (α,β) are in the first and fourth quadrant respectively and f is under-estimated. The bas-relief valley rotates in different directions depending on the relative magnitude of α and β. If we regard the movement of the bas-relief valley as an indication of the amount of bias in the FOE estimate, caused by an error in the focal length estimate, we can see that the bias is not necessarily towards the image center (as have been claimed by [27] from their experimental results using fixating motions), but depends on a variety of factors discussed above. If the egomotion is a fixating motion, i.e. an off-centered rotation about a point on the Z-axis, the translation is then purely induced by the off-centered rotation. Under such a scenario, one can say more about the relationship between (x_0,y_0) and (α,β); indeed it can be shown that (x_0,y_0) and (α,β) are in adjacent quadrants and that the offset term $\left(\alpha f\left(1 - \left(\frac{\hat{f}}{f}\right)^2\right), -\beta f\left(1 - \left(\frac{\hat{f}}{f}\right)^2\right)\right)$ in Eq. (14) is along the same direction as the original bas-relief direction. Thus, irrespective of the error in the focal length estimate, the bas-relief direction remains unchanged. The fixation constraint also allows us to show that the bias of the FOE estimate is always towards the optical center, along the direction of the original bas-relief valley. Thus our model

confirms the result of [27] under this specific motion configuration but also predicts other bias directions under more general motion configurations.

4. **Amount of FOE shift**. Having looked at the direction of the bias in the FOE estimate, we next examine the quantitative aspect of this bias, given different amount of error in the focal length estimate $\hat{f}$. Fig. 9 illustrates the error surface for varying amount of error in the estimate $\hat{f}$, and for a relatively large FOV of 53° under which we expect the effect of bias caused by the error in the estimate $\hat{f}$ would be more keenly felt. It can be seen that even with a rather large under-estimation error of 50% in $\hat{f}$ (the rightmost point of Fig. 9), the relative shift in the estimate $\hat{x}_0$ is only about 37%. For the case of over-estimation in $\hat{f}$, the FOE estimate deviates very little away from the calibrated case. This anisotropy has been explained before and is due to effect of $\hat{f}$ on the relative importance of the $\mathbf{t}_{2,2}$ term, which in turn gives rise to Eq. (15). Thus, to the extent that Eq. (15) is operative, we can then characterize the maximum amount of shifts in x_0 and y_0 respectively by the two terms $\beta fZ\left(1-\left(\frac{f}{\hat{f}}\right)^2\right)$ and $\alpha fZ\left(1-\left(\frac{f}{\hat{f}}\right)^2\right)$ in that equation. To pin down the value for such a bound, we assume that the effect of Z in the above two terms can be represented by some average depth Z_{ave} (for a scene sufficiently smooth). Then in relative terms, the changes to x_0 can be expressed as follows:

$$\frac{x_0-\left(x_0+\beta fZ_{ave}\left(1-\left(\frac{f}{\hat{f}}\right)^2\right)\right)}{x_0} = \frac{\beta f}{fU/Z_{ave}}W\left(1-\left(\frac{\hat{f}}{f}\right)^2\right) \approx \frac{u_{pan}}{u_{trans-x}}W\left(1-\left(\frac{\hat{f}}{f}\right)^2\right) \quad (16)$$

where u_{pan} and $u_{trans-x}$ are respectively the horizontal flow components due to panning rotation β and lateral translation U with some average depth Z_{ave}. Similar expression can be obtained for the relative change in the estimate for y_0. It can be seen that the relative change is affected by the ratio of the rotational flow u_{pan} and the translational flow $u_{trans-x}$; which is in turn moderated by a multiplicative factor $W\left(1-\left(\frac{f}{\hat{f}}\right)^2\right)$. Thus, for the simulation conducted in Fig. 9, where the translational flow and rotational flow are approximately equal in magnitude and $W = 1$, a large under-estimation error of 50% in $\hat{f}$ would result in a shift of 75% in the FOE shift. That this figure is much larger than the actual shift (37%) obtained could be due to violation of the two assumptions made in deriving this figure: (1) the $\mathbf{t}_{2,2}$ term is maximally effective and (2) scene points at different depths play an equal role such that their effect can be represented by some average depth Z_{ave}. Despite the looseness and approximate nature of the bound, we can use Eq. (16) as a guide in assessing whether the resulting bias in FOE is acceptable when using an approximate value of the focal length in a calibrated SFM algorithm, or it is better to face the tricky problem of estimating the focal length (as discussed in [1,8,10]) using a general uncalibrated SFM algorithm. As an illustrative example, consider a more typical error of 10% in the estimate $\hat{f}$ and under the same motion-scene configuration as above: the bound obtained via Eq. (16) for the relative FOE shift would be 19% (for under-estimation of f). Furthermore, this is likely to be a very loose bound; the actual shift obtained in Fig. 9 is only 4%. Thus we might want to proceed with a calibrated SFM algorithm even though the focal length estimate has small error.

5. **Effect of erroneous principal point**. Besides being affected by error in the focal length estimate, the bas-relief valley is also changed by error in the principal point estimate. With some approximation which is detailed in Appendix A.2, we obtained the following constraint:

$$\frac{y_{0e}-\alpha fZ\left(1-\left(\frac{f}{\hat{f}}\right)^2\right)}{x_{0e}+\beta fZ\left(1-\left(\frac{f}{\hat{f}}\right)^2\right)} = \frac{\hat{y}_0-O_{y_e}}{\hat{x}_0-O_{x_e}}$$

where (O_{x_e}, O_{y_e}) is the error in the principal point estimate. The constraint differs from Eq. (14) in that the bas-relief valley has been translated by an uniform amount (O_{x_e}, O_{y_e}) and passes through the true principal point. Fig. 10 illustrates the changes caused by $(O_{x_e}, O_{y_e}) = (100, -100)$, for (a) when there is no error in $\hat{f}$ and (b) when there is an under-estimation error of 50%. The bas-relief valleys appear bent because we have used visual angle in degree rather than pixel as the FOE search step; as the co-ordinates in the plots were not linear in the pixel unit, the uniform shift in (O_{x_e}, O_{y_e}) pixels would result in a non-linear bending of the bas-relief valley.

3.4. Summary of results and implication for various tasks

Eq. (6) 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 local minima on the surface which are the cause of inherent ambiguity of SFM algorithms are identified in the form of a constraint given by (15). The various contributing factors towards 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. As a result of error in the estimate $\hat{f}$, the bas-relief valley is rotated in a direction that depends on the relationship between the translation and the rotation. Under-estimating the focal length would have the effect of shortening the bas-relief valley and making it less well-defined in character. The feasibility of the flipped minimum solution that exists under calibrated scenario would be diminished. It also gives rise to a larger bias in the FOE estimate though with a smaller variance. On the other hand, over-estimating the focal length results in less change to the bas-relief valley and the FOE estimate would have smaller bias but larger variance.
2. We showed that the FOE is biased in a direction that depends on a variety of factors such as the sign of f_e and the angle between (α,β) and (x_0,y_0). We also obtain an analytical bound that quantifies the magnitude of this bias. For a typical figure of 10% error in the estimate $\hat{f}$ and given certain generic motion-scene conditions (such as rotation not too dominant), the bound obtained for the relative FOE shift might turn out to be acceptable. Furthermore, this bound is likely to be conservative as the actual shift obtained in simulation is consistently much smaller.
3. On the other hand, if the scene is very cluttered with very non-smooth depth distribution, then we expect the shape of this bas-relief valley to be markedly affected by the way the scene points are distributed, due to the presence of the Z term in the constraint (14). We also expect more local minima in the solution space which could pose convergence problem for a Euclidean SFM algorithm assuming erroneous calibration parameters. As a result, using a projective SFM algorithm under such situation might have the advantage of facing less of a local-minimum problem. This partially answers Oliensis's question [20] about whether projective approach is the right tool for

dealing with calibration uncertainty. If calibration is to be performed, our results suggest that one should err on the low side in the estimation of the focal length, since this would reduce the variance in the FOE estimate.

4. Error in the principal point estimate is shown to result in a simple change to the error surface. The entire bas-relief valley is shifted by a constant amount such that it passes through the true principal point.

The above results explain the various phenomenon often observed empirically in uncalibrated SFM but taken as true without theoretical explanation, such as the non-significant effect of small calibration error in f on FOE estimation. Our analysis also predicts what will happen when the errors are more severe, and if the scene depths are non-smooth, in which cases there might be significant impact on the extrinsic motion estimation. Such cases of severe calibration errors, while not common in SFM applications, is indeed pertinent for the perceptual experience of a viewer in a cinematic theatre or in a virtual reality system. For instance, for a cinema viewer seated at a general position, it can be shown that the visual system of the viewer experiences an altered optical flow resulting from changed intrinsic parameters [5]. The resulting bias in the FOE estimate might be an important factor for a navigating user wearing a head-mounted virtual reality system. As for the implications for depth perception from motion under such situation, we would like to refer the readers to [5] for a more complete analysis.

Here, we would like to examine briefly how these errors in the intrinsic parameters affect metric depth recovery. In [3], we have shown that the type of motion executed is crucial for depth recovery. Under lateral movement, while it might be very difficult to resolve the ambiguity between translation and rotation, depth orders of scene points can be recovered with robustness. Conversely, under forward translation, it is difficult to recover structure unless favorable conditions such as large field of view exist, because under this motion configuration, small error in the FOE estimate can introduce large distortion in the depth recovered. In the case of uncalibrated motion, in spite of uncertainty in the focal length, the qualitative aspect of the depth recovery process is not affected, regardless of whether it is a lateral or a forward motion. That is, under lateral motion, despite possible rotation of the bas-relief valley, the depth orders of scene points are shown in [3] to be preserved. Conversely, under forward motion, the inherent difficulty in depth recovery would have been compounded by the errors in the intrinsic parameters, as we have shown earlier that errors in the intrinsic parameters introduce additional bias to the FOE estimate.

Let us explore the ecological implications even we do suffer from depth distortion when we are executing forward motions. Such motions are mainly used in moving towards an object or for navigating through an environment. In the context of such tasks, we might only need partial aspects of structural information to successfully complete the tasks, rather than acquiring a comprehensive metric scene reconstruction. For instance, the ability to estimate the time-to-collision (TTC) is important for avoiding collision. It has been argued [12,17,23] that TTC can be recovered directly from the first order derivatives of the optical flow, without going through the step of 3D motion recovery. As a consequence, the TTC estimate would not be affected by the aforementioned depth distortion, which stems from errors in the 3D motion recovery. Nevertheless, calibration errors do affect the TTC estimate even it is recovered directly from the optical flow. In the calibrated case, the TTC estimate is not exact but bounded by some deformation terms [23] depending on the amount of lateral translation and the surface slant. If there now exists some error in the principal point estimate, the TTC bound would be affected by this error too. The detailed examination of how such task-specific structural information is affected by calibration errors, while interesting, is beyond the scope of this paper.

4. Experiments and discussion

To verify the theoretical findings just set out, we perform a series of experiments on both the Yosemite sequence and the Coke sequence. We also exploit the Brown range image database [11] to generate complex forest scenes.

For the two image sequences, the optical flow was obtained using the Lucas-Kanade algorithm [15] with a temporal window of 11 frames. Relatively dense optical flow fields were obtained. Despite the stochastic noise that is now present in the optical flow measurement, we demonstrate that given fairly dense and uniform distribution of scene points, our predictions about the changes to the bas-relief valley and the bias in the FOE estimate due to erroneous focal length hold true.

In the first experiment, the computer generated Yosemite sequence (Fig. 11a) was used. The average FOV is 46°, the true focal length is 337.5 pixels, the true FOE is located at (0,59.5), and $(\alpha,\beta,\gamma) = (0.0002319, 0.001625, -0.0002341)$. Fig. 11b shows the

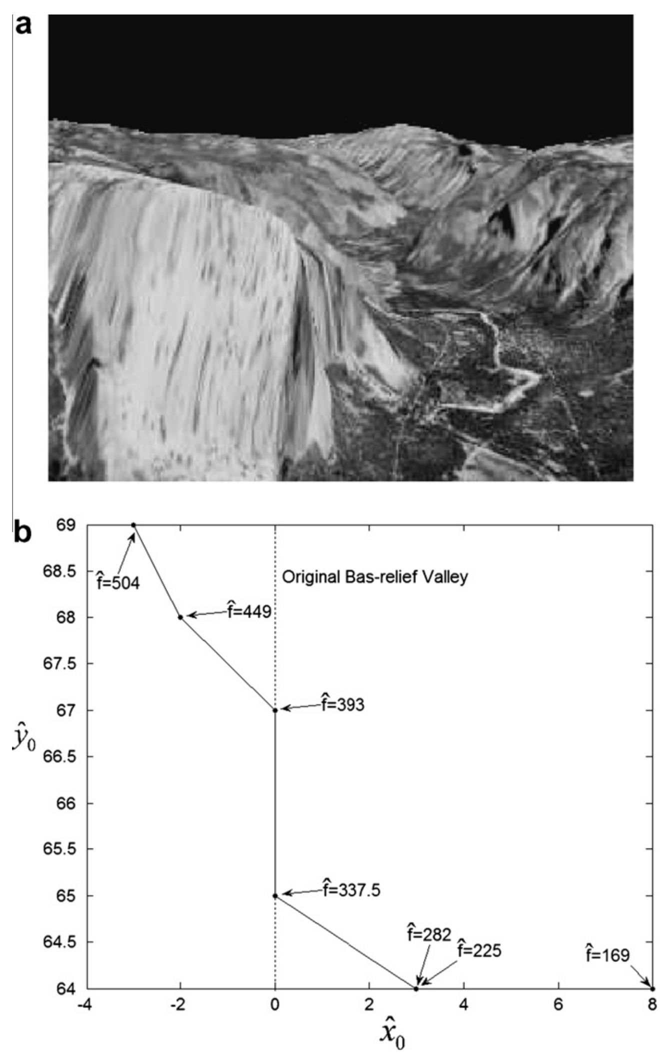

Fig. 11. (a) Yosemite sequence. (b) Shift of the FOE estimate as a result of erroneous focal length estimate $\hat{f}$. The true focal length of the image sequence is 337.5, the true FOE is at (0,59.5), and $(\alpha,\beta,\gamma) = (0.0002319, 0.001625, -0.0002341)$. Estimated FOEs are plotted for $\hat{f}$ having errors of 0%, ±16%, ±33%, and ±50% respectively.

estimated FOE locations for $\hat{f}$ having errors of 0%, ±16%, ±33%, and ±50%. According to our prediction, the bas-relief valley should shift in the clockwise direction for the case of under-estimating f, and vice versa. This is borne out by the result.

In the second experiment, similar analysis was conducted on the Coke image sequence (Fig. 12a). The parameters of this sequence are FOV = 28°, f = 620 pixels, the true FOE at (65, 73), and $(\alpha, \beta, \gamma) = (-0.00025, -0.00013, 0)$. The experimental results are shown in Fig. 12b for the same range of error in $\hat{f}$. According to our prediction, the bas relief valley should now shift in the anti-clockwise direction for the case of under-estimating f, and vice versa. This is again borne out by the result.

For the range images, we choose the forest scene from the Brown range image database (Fig. 13). This scheme allows us to experiment with realistic scenes with its clustered depth distribution and varying degree of feature density, and yet able to control the exact amount of noise added to the image. In this experiment, we add to the flow field zero-mean Gaussian noise, isotropic in direction and with standard deviation equal to the Noise-to-Signal Ratio (NSR) times the average flow speed. In particular, a 10% noise was added to the optical flow. To make the setting compatible with those of the synthetic experiments, we only selected subsets of scene features that lie in the visual field of a camera (FOV = 53°), and we scaled these selected points such that the depths are within a range of 512–1536 pixels. The true focal length of the camera is 512 pixels. We endowed the scene with a translation of (1, 1, 1) and a rotation of (0.001, 0.001, 0.001) as usual, and projected the 3D scene points and their flows onto the camera's image plane. The true FOE is thus at the image location (512, 512). Two sets of scene features are tested, one with 200 features (very sparse) and the other with 5000 features (reasonably dense). We then carried out the whole procedure of motion estimation under the same range of calibration errors in $\hat{f}$. The resultant error cost functions are depicted in Figs. 14 and 15 for the sparse and dense set of features respectively.

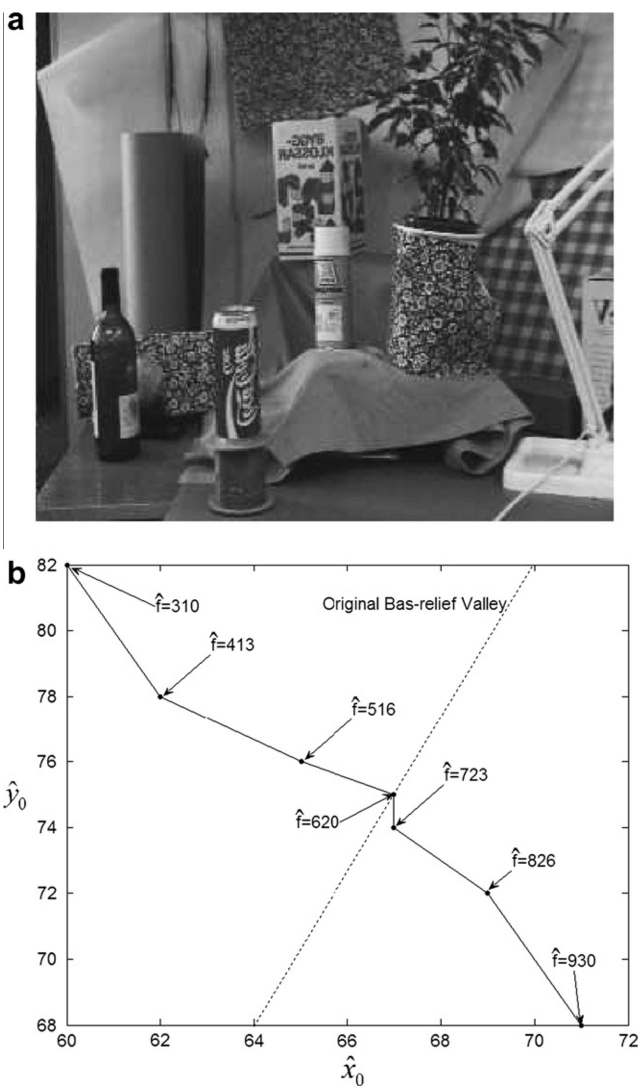

Fig. 12. (a) Coke sequence. (b) Shift of the FOE estimate as a result of erroneous focal length estimate $\hat{f}$. The true focal length of the image sequence is 620, the true FOE is at (65, 73), and $(\alpha, \beta, \gamma) = (-0.00025, -0.00013, 0)$. Estimated FOEs are plotted for $\hat{f}$ having errors of 0%, ±16%, ±33%, and ±50% respectively.

Fig. 13. The original range image of a forest scene. Intensity represents depth values, with distant object looking brighter.

For the former case with sparse feature set, Fig. 14 clearly shows the bas-relief rotation and the corresponding shift of the estimated FOE when the focal length is under-estimated. The shift is relatively pronounced, with an angular deviation of 18° (which is still consistent with the direction and magnitude obtained in our theoretical prediction). Fig. 14a and c depict the cases of correct and over-estimated f respectively. Here, the accuracy of the FOE estimation is heavily influenced by the local minima caused by the sparse and clustered feature distribution. As we have predicted for the case of over-estimation of f, the FOE estimate might have smaller bias but the variance is also larger. This larger variance proved fatal when the features are sparse and the noise is large, resulting in the FOE estimate being trapped in a local minimum. In Fig. 14, we plot the estimated FOE locations for various erroneous $\hat{f}$ values. Again, it can be seen clearly that most of the FOE estimates are trapped in a local minimum region, except when the focal length is significantly under-estimated.

Compare this with Fig. 15 with dense feature set. It can be seen that the FOE estimate is much less affected by local minima. Thus, our theoretical predictions about the changes to the bas-relief valley and the bias in the FOE estimate due to erroneous focal length remain largely true.

The results obtained seem to corroborate the various predictions made in this paper. In all the sequences, the direction of bias in the FOE estimate is consistent with the predictions made in the preceding section. We also predicted that the bias will be less pronounced for over-estimating rather than under-estimating f, though this prediction is apparently not borne out by the results, with the case of over-estimation sometimes exhibiting comparable or even larger amount of FOE shift as that of under-estimation. However, this is not surprising as we can see from Fig. 3 that in the case of over-estimation, the FOE estimate, while displaying a smaller bias, suffers from a larger variance under the influence of noise. With the presence of noise in real images and the significant effect of local minima introduced by non-uniform feature distribution (in particular, the depths are not smooth), this high variance term becomes important, thus contributing to the larger-than-expected FOE errors seen in the results. In fact, as can be seen from Fig. 11b, Fig. 12b, and Fig. 14a, these non-ideal effects also hamper the FOE recovery under perfect calibration, with the direction of FOE errors lying along the bas-relief valley.

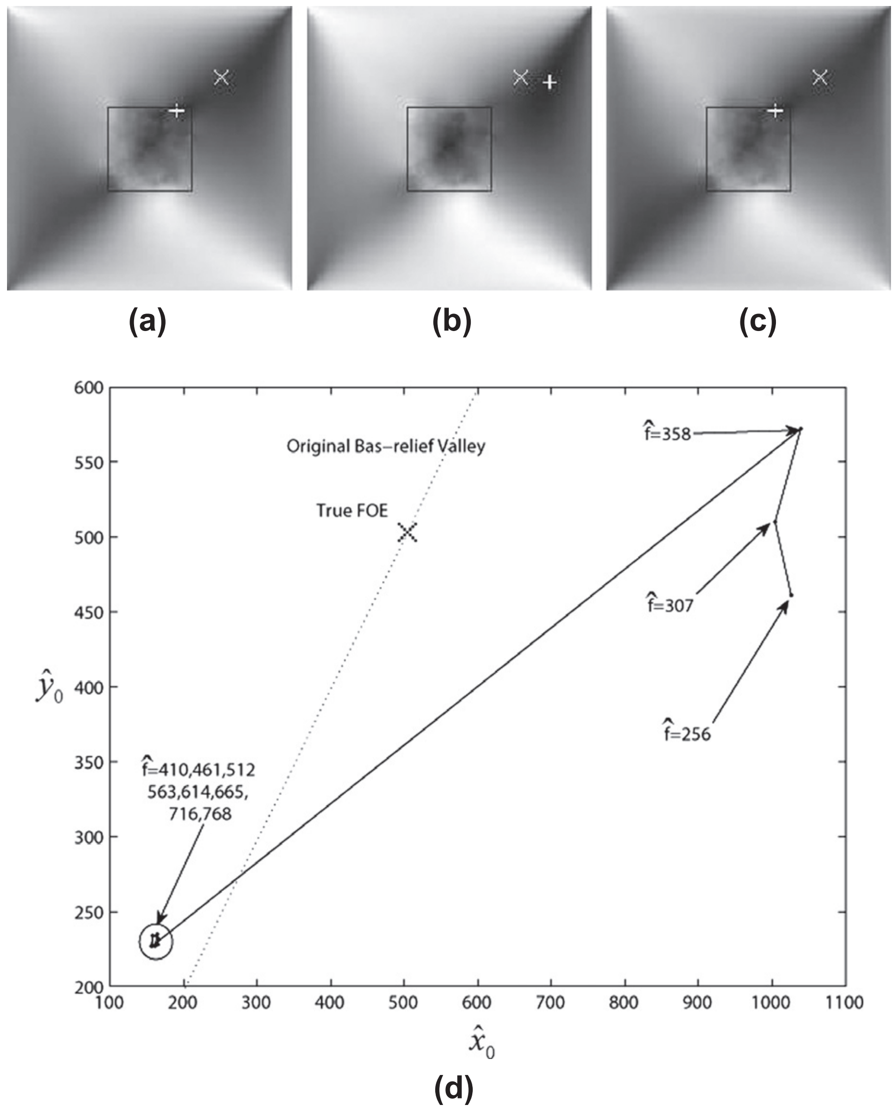

Fig. 14. (a)–(c) The bas-relief valley in the residual error surface for the forest image with 200 feature points, with various errors in the focal length estimate. (a) $\hat{f} = f = 512$, (b) $\hat{f} = 256$, (c) $\hat{f} = 1024$. The true motion is given by **v** = (1, 1, 1) and **w** = (0.001, 0.001, 0.001). The true FOEs and the respective global minima are indicated by "×" and "+" respectively. The plot in (d) shows the shift of the FOE estimate as a result of erroneous focal length estimate $\hat{f}$. Note that in plot (b), the estimated FOE appears quite close to the true FOE because the residual error map is plotted in angular units.

Overall, we found that the actual shift in the FOE estimate (in terms of angular deviation) for real images is not significant even for relatively large error in the focal length estimate, unless this shift is caused by the presence of the local minima. The experiments conducted demonstrate that, even with a relatively dense set of feature points, non-ideal effects such as non-uniform feature distribution and image noise, rather than calibration errors, could play a potentially more significant role in affecting the accuracy of FOE recovery.

5. Conclusions

Error analysis for SFM has always been plagued by the complexity of the problem. This complexity becomes even more daunting in the face of possible calibration errors. In this paper we have developed clear analytical expressions describing the error behavior of the egomotion estimates when the fixed intrinsic parameters are calibrated with error. The key results in this paper are independent of the algorithm used to perform egomotion estimation and calibration. They explain the various phenomenon often observed empirically in uncalibrated SFM but taken as true without theoretical explanation. We show that as a result of error in the estimate $\hat{f}$, the bas-relief valley is rotated in a direction that depends on the relationship between the translation and the rotation. Underestimating the focal length would have the effect of shortening the bas-relief valley and making it less well-defined in character. It also gives rise to a larger bias in the FOE estimate though with a smaller variance. On the other hand, over-estimating the focal length result in less change to the bas-relief valley and the FOE estimate would have smaller bias but larger variance. This large variance could indeed cause more error in the FOE estimate when the feature points are sparse and clustered. We also obtain an analytical bound that quantifies the effect of an erroneous focal length on the FOE estimate. For a typical figure of 10% error in the estimate $\hat{f}$ and given certain generic motion-scene conditions (such as rotation not too dominant), the bound obtained for the relative FOE shift might turn out to be acceptable. Furthermore, this bound is likely to be conservative as the actual shift obtained in simulation is consistently much smaller. Error in the principal point

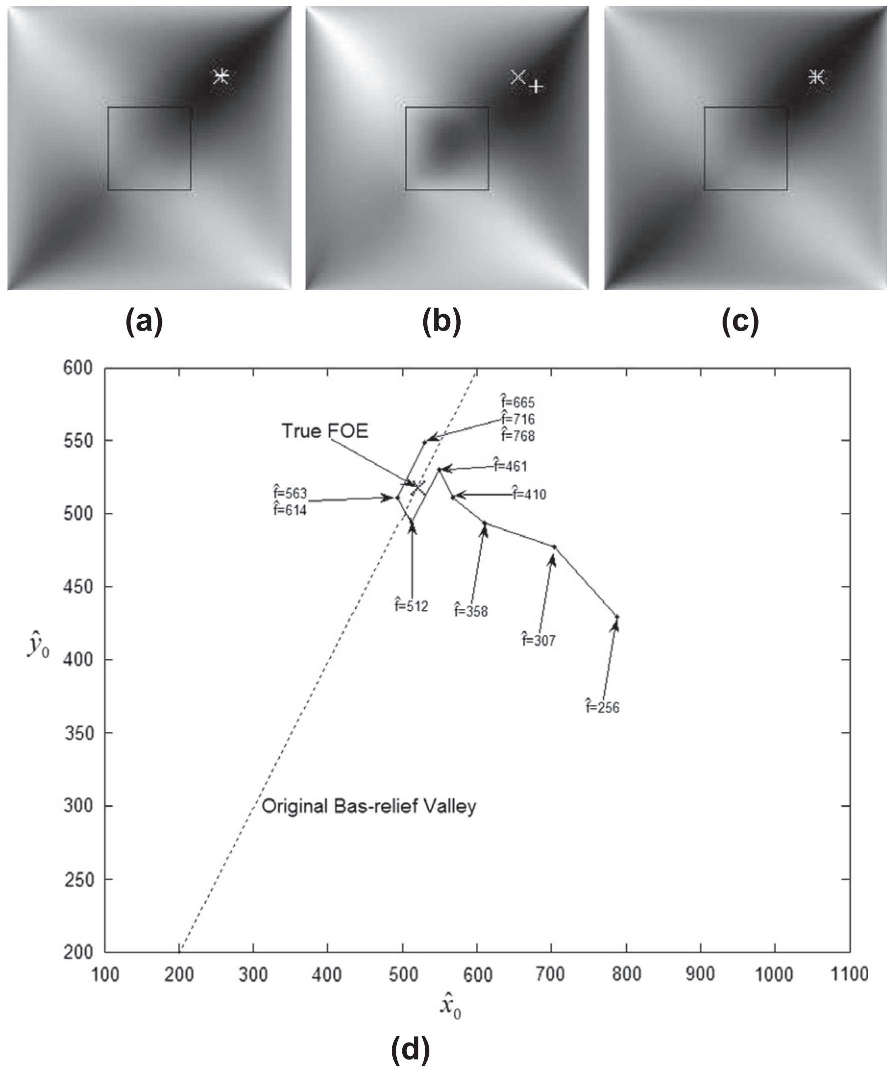

Fig. 15. The experimental conditions are the same as in Fig. 14, except that the number of feature points is 5000. The much denser feature set here results in much smaller errors in the FOE estimate compared to that of Fig. 14.

estimate is shown to result in a simple change to the error surface. The entire bas-relief valley is shifted by a constant amount such that it passes through the true principal point. Real-world effects such as image noise and non-uniform feature distribution are briefly investigated in the experimental section, with results showing that these non-ideal effects are likely to play a much more significant role than the errors in the calibration parameters.

The conclusion of this paper is that if the image quality is acceptable and the feature distribution is relatively dense and uniform, we might want to use a calibrated SFM algorithm even though the focal length estimate or the principal point estimate has small errors. The resultant small loss in accuracy might be acceptable compared to the uncertainty faced in estimating the focal length or principal point using a general uncalibrated SFM algorithm. Furthermore, if one chooses to perform self-calibration, one should err on the high side under such condition, as over-estimating the focal length results in a smaller bias in the FOE estimate. If, however, one has to deal with high image noise or sparse and clustered feature distribution, the perennial problems that plague SFM estimation even for the calibrated case would certainly be compounded by the calibration errors, posing grim problems for any general 2-frame SFM recovery algorithm. In particular, the presence of local minima under such noisy condition means that there is a distinct advantage in under-estimating the focal length. While it results in larger bias in the FOE estimate, the solution is nevertheless more stable with respect to the influence of local minima. These are the factors that one should consider under a particular operating condition and decide whether it is better to perform self-calibration or to use an existing calibrated setting despite small error in the parameters.

Appendix A

A.1. Procedure for computing J_R

The simulations carried out in this paper is based on computing the cost function in Eq. (5) or equivalently Eq. (6). In particular, we adopt the "epipolar reconstruction" scheme, that is, setting $\mathbf{n}_i$ in both these equations to be along the estimated epipolar direction (the results obtained are independent of the choice of $\mathbf{n}_i$ since the formation of the bas-relief valley is mainly dictated by the numerator terms). Given this scheme and for a particular FOE candidate $((\hat{x}_0), \hat{y}_0)$, J_R can be expressed as:

$$J_R = \sum \left(\frac{c_1 - \left(c_2\hat{\alpha} + c_3\hat{\beta} + c_4\hat{\gamma} \right)}{\eta} \right)^2 \tag{17}$$

where

$$c_1 = u(y - \hat{y}_0) - v(x - \hat{x}_0)$$
$$c_2 = \frac{xy}{\hat{f}}(y - \hat{y}_0) - \left(\frac{y^2}{\hat{f}} + \hat{f}\right)(x - \hat{x}_0)$$
$$c_3 = \frac{xy}{\hat{f}}(x - \hat{x}_0) - \left(\frac{x^2}{\hat{f}} + \hat{f}\right)(y - \hat{y}_0)$$
$$c_4 = x(x - \hat{x}_0) + y(y - \hat{y}_0)$$
$$\eta = \sqrt{(x - \hat{x}_0)^2 + (y - \hat{y}_0)^2}$$

and we minimize J_R over all points in the image to solve for the rotation variables $\hat{\alpha}, \hat{\beta}, \hat{\gamma}$. This is a typical linear least squares fitting problem, which we solved by the 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.

A.2. Effect of erroneous principal point

We use (x_s, y_s) to represent an image pixel location in an image coordinate system with its origin located at the lower left corner of the image. If the principal point of the camera is situated at (O_x, O_y) in this new coordinate system, then (x, y) and (x_s, y_s) are related by $(x, y) = (x_s - O_x, y_s - O_y)$. Given an error (O_{x_e}, O_{y_e}) in the principal point estimate, the corresponding error function J_R can be shown to be given by[2]

$$J_R = \sum \left(\frac{(x + O_{x_e} - \hat{x}_0, y + O_{y_e} - \hat{y}_0) \cdot \left(v_{rot_e} - \frac{y_{0e} + O_{y_e}}{Z}, \frac{x_{0e} + O_{x_e}}{Z} - u_{rot_e} \right)}{(x + O_{x_e} - \hat{x}_0, y + O_{y_e} - \hat{y}_0) \cdot \mathbf{n}} \right)^2$$

where u_{rot_e} and v_{rot_e} are given by:

$$u_{rot_e} = -\left(\beta f - \hat{\beta}\hat{f}\right) + \frac{\alpha}{f}xy - \frac{\hat{\alpha}}{\hat{f}}(x + O_{x_e})(y + O_{y_e}) - \frac{\beta}{f}x^2 + \frac{\hat{\beta}}{\hat{f}}(x + O_{x_e})^2 + \gamma y - \hat{\gamma}(y + O_{y_e})$$
$$v_{rot_e} = (\alpha f - \hat{\alpha}\hat{f}) + \frac{\alpha}{f}y^2 - \frac{\hat{\alpha}}{\hat{f}}(y + O_{y_e})^2 - \frac{\beta}{f}xy + \frac{\hat{\beta}}{\hat{f}}(x + O_{x_e})(y + O_{y_e}) - \gamma x + \hat{\gamma}(x + O_{x_e})$$

The corresponding terms in $\mathbf{t}_1$ and $\mathbf{t}_2$ are:

$$\mathbf{t}_{1,0} = \left(-\hat{x}_0 + O_{x_e}, -\hat{y}_0 + O_{y_e}\right)^T$$
$$\mathbf{t}_{1,1} = (x, y)^T$$
$$\mathbf{t}_{2,0} = ((\alpha f - \hat{\alpha}\hat{f}), (\beta f - \hat{\beta}\hat{f}))^T$$
$$\mathbf{t}_{2,1} = (-\gamma x + \hat{\gamma}(x + O_{x_e}), -\gamma y + \hat{\gamma}(y + O_{y_e}))^T$$
$$\mathbf{t}_{2,2} = \left(\frac{\alpha}{f}y^2 - \frac{\hat{\alpha}\hat{f}(y + O_{y_e})^2 - \frac{\beta}{f}xy + \frac{\hat{\beta}}{\hat{f}}(x + O_{x_e})(y + O_{y_e})}{,} - \frac{\alpha}{f}xy + \frac{\hat{\alpha}}{\hat{f}}(x + O_{x_e})(y + O_{y_e}) + \frac{\beta}{f}x^2 - \frac{\hat{\beta}}{\hat{f}}(x + O_{x_e})^2\right)^T$$
$$\mathbf{t}_{2,Z} = \left(-\frac{y_{0_e} + O_{y_e}}{Z}, \frac{x_{0_e} + O_{x_e}}{Z}\right)^T$$

To derive the conditions conducive for the formation of the bas-relief ambiguity, we apply the same condition that the constant-direction vectors $(\mathbf{t}_{2,0} + \mathbf{t}_{2,Z})$ and $\mathbf{t}_{1,0}$ should be perpendicular to each other. We obtain, analogous to Eq. (11), the following:

$$\frac{y_{0_e} + O_{y_e} - \alpha f Z + \hat{\alpha}\hat{f}Z}{x_{0_e} + O_{x_e} + \beta f Z - \hat{\beta}\hat{f}Z} = \frac{\hat{y}_0 - O_{y_e}}{\hat{x}_0 - O_{x_e}} \tag{18}$$

The corresponding condition for making $\|\mathbf{t}_{2,2}\|$ small gives rise to the following:

$$\frac{\alpha xy}{f} = \frac{\hat{\alpha}(x + O_{x_e})(y + O_{y_e})}{\hat{f}}$$
$$\frac{\beta xy}{f} = \frac{\hat{\beta}(x + O_{x_e})(y + O_{y_e})}{\hat{f}}$$
$$\frac{\alpha y^2}{f} = \frac{\hat{\alpha}(y + O_{y_e})^2}{\hat{f}}$$
$$\frac{\beta x^2}{f} = \frac{\hat{\beta}(x + O_{x_e})^2}{\hat{f}}$$

which are obviously not satisfiable at all points of the image. However, if we make the assumption that the second order effect $\|\mathbf{t}_{2,2}\|$ only comes into play at the peripheral image points where x and y are large and that the magnitude of the error (O_{x_e}, O_{y_e}) is small compared to x and y at these peripheral points, then the original constraint $\frac{\alpha}{f} = \frac{\hat{\alpha}}{\hat{f}}, \frac{\beta}{f} = \frac{\hat{\beta}}{\hat{f}}$ of Eq. (13) is still approximately true. Substituting this into Eq. (18), we obtain, after some manipulation, the following form:

$$\frac{y_{0e} - \alpha f Z\left(1 - \left(\frac{\hat{f}}{f}\right)^2\right)}{x_{0e} + \beta f Z\left(1 - \left(\frac{\hat{f}}{f}\right)^2\right)} = \frac{\hat{y}_0 - O_{y_e}}{\hat{x}_0 - O_{x_e}}$$

[2] Note that in deriving these equations and plotting the figures, the true and the estimated FOEs should be independent of the choice of the principal point, as the FOE actually indicates a direction in space—that of the 3D translation.

References

[1] S. Bougnoux, From projective to Euclidean space under any practical situation, a criticism of self-calibration, in: ICCV98, 1998, pp. 790–796.
[2] M.J. Brooks, W. Chojnacki, L. Baumela, Determining the egomotion of an uncalibrated camera from instantaneous optical flow, JOSA A 14 (10) (1997) 2670–2677.
[3] L.-F. Cheong, T. Xiang, Characterizing depth distortion under different generic motions, IJCV 44 (3) (2001) 199–217.
[4] L.-F. Cheong, X. Xiang, Error characteristics of SFM with erroneous focal length, in: ACCV, Hyderabad India, LNCS, vol. 3851, 2006, pp. 714–723.
[5] L.-F. Cheong, X. Xiang, What do we perceive from motion pictures? A computational account, J. Opt. Soc. Am. A 24 (2007) 1485–1500.
[6] A. Chiuso, R. Brockett, S. Soatto, Optimal structure from motion: local ambiguities and global estimates, IJCV 39 (3) (2000) 195–228.
[7] J.I. González, J.C. Gámez, C.G. Artal, A.M.N. Cabrera, Stability study of camera calibration methods, in: CI Workshop en Agentes Físicos, WAF, Spain, 2005.
[8] R.I. Hartley, C. Silpa-Anan, Reconstruction from two views using approximate calibration, in: ACCV, 2002, pp. 338–343.
[9] F. Kahl, B. Triggs, Critical motions in euclidean structure from motion, in: CVPR99, 1999, pp. II: 366–372.
[10] K. Kanatani, A. Nakatsuji, Y. Sugaya, Stabilizing the focal length computation for 3D reconstruction from two uncalibrated views, IJCV 66 (2) (2006) 109–122.
[11] A.B. Lee, J. Huang, Brown range image database, 2000. <http://www.dam.brown.edu/ptg/brid/index.html>.
[12] D.N. Lee, The optic flow field: the foundation of vision, Philos. Trans. R. Soc. Lond. B Biol. Sci. 290 (1980) 169–179.
[13] D. Liebowitz, A. Zisserman, Resolving ambiguities in auto-calibration, Roy. Soc. Lond. Philos. Trans. A A-356 (1998) 1193–1211.
[14] H.C. Longuet-Higgins, A computer algorithm for reconstructing a scene from two projections, Nature 293 (1981) 133–135.
[15] B.D. Lucas, Generalized image matching by the method of differences, PhD thesis, Carnegie-Mellon University, 1984.
[16] Y. Ma, J. Košecká, S. Sastry, Optimization criteria and geometric algorithms for motion and structure estimation, IJCV 44 (3) (2001) 219–249.
[17] R.C. Nelson, J.Y. Aloimonos, Obstacle avoidance using flow using flow field divergence', PAMI 11 (10) (1989) 1102–1106.
[18] G.N. Newsam, D.Q. Huynh, M.J. Brooks, H.-P. Pan, Recovering unknown focal lengths in self-calibration: an essentially linear algorithm and degenerate

configurations, in: Int. Arch. Photogram. Remote Sensing, vol. 31-B3-III, Vienna, Austria, 1996, pp. 575–580.
[19] J. Oliensis, A multi-frame structure-from-motion algorithm under perspective projection, IJCV 34 (2–3) (1999) 163–192.
[20] J. Oliensis, A critique of structure-from-motion algorithms, CVIU 80 (2) (2000) 172–214.
[21] P.F. Sturm, Critical motion sequences for the self-calibration of cameras and stereo systems with variable focal length, IVC 20 (5–6) (2002) 415–426.
[22] P.F. Sturm, Z.L. Cheng, P.C.Y. Chen, A.N. Poo, Focal length calibration from two views: method and analysis of singular cases, CVIU 99 (1) (2005) 58–95.
[23] M. Subbarao, Interpretation of Visual Motion: A Computational Study, Pitman, London, 1988.
[24] T. Svoboda, P. Sturm, A badly calibrated camera in ego-motion estimation—propagation of uncertainty, in: Proc. Int. Conf. on Computer Analysis of Images and Patterns, Kiel, Germany, 1997, pp. 183–190.
[25] T. Xiang, L.-F. Cheong, Understanding the behavior of SFM algorithms: a geometric approach, IJCV 51 (2) (2003) 111–137.
[26] Z.Y. Zhang, Q.T. Luong, O.D. Faugeras, Motion of an uncalibrated stereo rig: self-calibration and metric reconstruction, RA 12 (1) (1996) 103–113.
[27] M. Zucchelli, J. Košecká, Motion bias and structure distortion induced by intrinsic calibration errors, IVC 26 (5) (2008) 639–646.