Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE DIRECT®

Computer Vision and Image Understanding 93 (2004) 221–244

Computer Vision and Image Understanding

www.elsevier.com/locate/cviu

Depth distortion under calibration uncertainty

Loong-Fah Cheong* and Chin-Hwee Peh

Department of Electrical and Computer Engineering, National University of Singapore, Singapore 119260, Singapore

Received 1 March 2002; accepted 16 September 2003

Abstract

There have been relatively little works to shed light on the effects of errors in the intrinsic parameters on motion estimation and scene reconstruction. Given that the estimation of the extrinsic and intrinsic parameters apts to be imprecise, it is important to study the resulting distortion on the recovered structure. By making use of the iso-distortion framework, we explicitly characterize the geometry of the distorted space recovered from 3D motion with freely varying focal length. This characterization allows us: (1) to investigate the effectiveness of the visibility constraint in disambiguating errors in calibration parameters by studying the negative distortion regions and (2) to make explicit those ambiguous error situations under which the visibility constraint is not effective. An important finding is that under these ambiguous situations, the direction of heading can be accurately recovered and the structure recovered experienced a well-behaved distortion. The distortion is given by a relief transformation which preserves ordinal depth relations. Thus in the case where the change in focal length is not well estimated, structure information in the form of depth relief can be obtained. Experiments were presented to support the use of the visibility constraint in obtaining such partial motion and structure solutions.

Keywords: Structure from motion; Depth distortion; Space perception; Calibration parameters

1. Introduction

Many a time, in the solving of motion estimation problems, it is common to assume that the camera has been pre-calibrated off-line and that the intrinsic

* Corresponding author. Fax: +65-6779-1103.
E-mail address: eleclf@nus.edu.sg (L.-F. Cheong).

1077-3142/$ - see front matter
doi:10.1016/j.cviu.2003.09.003

parameters of the camera remain fixed throughout its operation. However, in actual camera operation, the focal length of a moving camera often changes. Some variations may be intentional. For instance, the use of a zoom lens in an active vision system. Other variations may be accidental, due to the camera undergoing small mechanical or thermal changes. In many situations, it is not practical to interrupt the operation of the camera to calibrate it with a calibration apparatus. These considerations establish the need to perform on-line calibration, also known as self-calibration [22,24].

Despite the various works focusing on the self-calibration problem, most implementations face difficulties in estimating the intrinsic parameters accurately. One way to circumvent this problem is to enforce special camera displacements to obtain better estimate of the intrinsic parameters [1,7,13,29,37]. For example, a pure rotating camera can lead to a more accurate calibration parameters estimation [13]. Conversely, the recovery of the intrinsic parameters can be rendered ambiguous when certain motion, commonly known as critical motion, is employed [18,31,32]. In particular, Hayman et al. [15] presented ill-conditioned configurations that resulted in near-ambiguities when rotating cameras undergo self-calibration. Instances whereby the Kruppa equations used to solve self-calibration become degenerate and ill-conditioned have also been identified [23].

While the mainstay of the research efforts in self-calibration adopts the discrete approach, recent works in [3,37] have formulated the problem in the continuous domain. Most of the schemes presented assume that the intrinsic parameters across the frames are constant and does not change over time. However, with the common and extensive use of zooming camera, this restrictive assumption can no longer holds true under many situations. A zooming camera changes its focal length and is likely to bring about a shift in the position of its principal point. A more general treatment of the problem, allowing for varying intrinsic parameters, is given in [2,3,12,16, 25,30,37]. Nevertheless, such formulation imposes new constraints or more frames are usually required.

Despite the difficulty in obtaining calibration parameters, many researchers argue that as far as scene reconstruction is concerned, several structures (Projective, Affine) can be obtained without the need to recover the intrinsic parameters completely or accurately [26,35]. Accurate intrinsic parameters are necessary for the attainment of Euclidean reconstruction. A highly debatable question is whether Euclidean reconstruction is a must. Many recent findings had demonstrated that only partial or qualitative scene reconstruction is required for many robotic tasks. The properties of these partial scene structures have been well studied in [8,19].

There have been relatively little works to shed light on the effects of errors in the intrinsic parameters on motion estimation and scene reconstruction. Florou and Mohr [10] used the statistical approach to study reconstruction errors with respect to calibration parameters. Svoboda and Sturm [34] studied how uncertainty in the calibration parameters gets propagated to the motion parameters. Viéville and Faugeras studied the partial observability of rotational motion, calibration, and depth map in [37]. Bougnoux [2] offered a critique of the self-calibration problem, finding that the estimation of various intrinsic parameters are unstable. However, it was observed, partly empirically, that despite uncertainty in the focal length estimation, the quality of the reconstruction does not seem to be affected. Certain geometrical properties such as parallelism seemed to be preserved.

Aside from these observations, there has not really been an in-depth geometrical characterization of the errors in the reconstructed depth given some errors in both the intrinsic and the extrinsic parameters. In a recent work, Cheong and Xiang [6] studied the aforementioned depth distortion under specific motion scene configuration whereby the translation is either in the forward or in the lateral direction. In this paper, we consider the more general situation of a camera moving arbitrarily in a 3D environment with the added possibility of zoom operation. Thus, the focal length of the camera can be freely varying across frames, resulting in a zoom field (considering infinitesimal motion) which is difficult to separate from that of a translation along the optical axis. This, together with the perennial problem of the coupling between translation and rotation, means that distortion in the recovered structure is likely to be present.

This paper attempts to make the geometry of this distortion explicit by using the iso-distortion framework introduced in [4]. The analysis is completely general and valid for any algorithm or any scene structure. As it is a purely geometrical analysis, it also does not contain the usual assumptions entailed in statistical error analyses [28,38]. The motivation for performing such distortion-oriented geometrical analysis is twofold. First, by gaining a deeper understanding of the geometry of depth distortion, we can have a better notion of what the proper space representation should be. Compared with the usual stratified viewpoint of space [8]—Projective, Affine, and Euclidean—such a distortion-oriented viewpoint represents an alternative look at the problem of space representation. Via this alternative look, we showed that the stratification should be extended, as, in general, we can only recover depth up to a transformation which is even more general than the projective transformation [5]. The other motivation, which is the main focus of this paper, is that such characterization of depth distortion allows us to understand how depth distortion in turn interacts with the motion (including zoom) estimation process, thereby revealing the additional constraint that can be imposed on motion estimation. So far, in the structure from motion literature, motion estimation and depth estimation are very much treated as two independent sub-problems. Indeed, we usually eliminate the depth component first so as to make the dimensionality of the motion estimation problem manageable. The result of such elimination is that we are often ignorant of how different types of scene structure may affect the motion estimation process. For instance, it is not clear how we can use the depth-is-positive constraint (also known as the visibility constraint) under different scene types to constrain the estimation of both the intrinsic and the extrinsic motion parameters. To address these problems, this paper investigates the effectiveness of the visibility constraint in disambiguating calibration errors and describes those ambiguous error situations under which the visibility constraint is not effective.

This paper is structured along the following lines. First comes some preliminaries regarding the iso-distortion framework in Section 2, followed by an extension of this framework to the self-calibration problem. Several major features of the resulting distortion are then made explicit. The main goals of Section 3 are: (1) to elaborate the relations between the depth distortion and the estimation of both the intrinsic and extrinsic parameters and (2) to study certain well-behaved depth distortion resulting from ambiguous solutions. Section 4 presents experiments to support the use of the visibility constraint in obtaining partial solutions to the estimation of both motion and structure. The paper ends with a summary of the work.

2. Iso-distortion framework

2.1. Optical flow equations with varying focal length

Let (x, y) be the image location resulting from the projection of a point in the 3D world onto the 2D image plane by a real camera. Due to the imperfect imaging process and depending on the actual operation (e.g., zooming), the coordinate system of a real camera may deviate slightly from the ideal. We use (x_s, y_s) to represent an image pixel location in the latter 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, the relationship between the two coordinate systems can then be represented by $(x, y) = (x_s - O_x, y_s - O_y)$. In this paper, we shall assume the common situation that the skew angle is equal to 90° (i.e., the pixel are rectangular) and that this condition does not change over a long period of time. The effect of radial distortion is also assumed to be negligible.

When the camera moves rigidly with respect to its 3D world with a translation (U, V, W) and a rotation (α, β, γ), together with a zooming operation that results in a change in the focal length and a shift in the principal point, the resulting optical flow $u = (u_x, u_y)$ at a local point in the image plane can be represented as follows under the two different coordinate systems:

$$\begin{aligned} u_x &= \frac{W}{Z}(x - x_0) + \frac{xy}{f}\alpha - f\left(1 + \frac{x^2}{f^2}\right)\beta + \gamma y + \frac{\dot{f}}{f}x + \dot{O}_x \\ &= \frac{W}{Z}((x_s - O_x) - x_0) - f\beta + \gamma(y_s - O_y) + \frac{\dot{f}}{f}(x_s - O_x) + \dot{O}_x + \mathcal{O}_u^2, \end{aligned} \tag{1}$$

$$\begin{aligned} u_y &= \frac{W}{Z}(y - y_0) - \frac{xy}{f}\beta + f\left(1 + \frac{y^2}{f^2}\right)\alpha - \gamma x + \frac{\dot{f}}{f}y + \dot{O}_y \\ &= \frac{W}{Z}((y_s - O_y) - y_0) + f\alpha - \gamma(x_s - O_x) + \frac{\dot{f}}{f}(y_s - O_y) + \dot{O}_y + \mathcal{O}_v^2, \end{aligned} \tag{2}$$

where

f is the focal length of the camera with its rate of change with respect to time defined by $\dot{f}$;

Z is the depth of scene point;

$(x_0, y_0) = (f\frac{U}{W}, f\frac{V}{W})$ is the Focus of Expansion (FOE) of the flow field;

$(\dot{O}_x, \dot{O}_y)$ is the shift in the principal point location;

$$(\mathcal{O}_u^2, \mathcal{O}_v^2) = \left(\frac{(x_s - O_x)(y_s - O_y)}{f}\alpha - \frac{(x_s - O_x)^2}{f}\beta, -\frac{(x_s - O_x)(y_s - O_y)}{f}\beta + \frac{(y_s - O_y)^2}{f}\alpha \right)$$

contains the second-order terms in (x_s, y_s).

2.2. Space distortion arising from calibration errors

In a recent work [4], the geometric laws under which the recovered scene is distorted due to some errors in the viewing geometry is represented by a distortion transformation. It was called the iso-distortion framework whereby distortion in the perceived space can be visualized by families of iso-distortion lines. In the present study, this framework has been extended to characterize the types of distortion experienced by a visual system where a change in the focal length and principal point location may result in further difficulties and errors in the estimation of its intrinsic parameters.

From the two motion Eqs. (1) and (2), one can recover the relative depth of a scene point using several possible schemes. For instance, in the normal flow approach [9,17], one can choose to reconstruct depth along the normal flow direction, given by (n_x, n_y). In general, (n_x, n_y) can be any other direction, in which case Z (i.e., the scaled depth) can be obtained from the flow u_n projected along that direction

$$u_n = u \cdot (n_x, n_y)$$
$$u_n = \left(\frac{W}{Z}(x - x_0, y - y_0) + u_r + u_f \right) \cdot (n_x, n_y),$$

and expressed as

$$Z = \frac{(x - x_0, y - y_0) \cdot (n_x, n_y)}{(u_n - (u_r + u_f) \cdot (n_x, n_y))} = \frac{(x_s - O_x - x_0, y_s - O_y - y_0) \cdot (n_x, n_y)}{(u_n - (u_r + u_f) \cdot (n_x, n_y))}, \tag{3}$$

where u_r is the rotational flow and u_f is the zoom flow caused by a change in the focal length.

If there are some errors in the estimation of the intrinsic and/or the extrinsic parameters, this will in turn cause errors in the estimation of the scaled depth, and thus a distorted version of space will be computed. We denote the estimated parameters with the hat symbol ($\hat{}$). Hence

$$\hat{Z} = \frac{(x - \widehat{x_0}, y - \widehat{y_0}) \cdot (n_x, n_y)}{\left(\widehat{u_n} - (\widehat{u_r} + \widehat{u_f}) \cdot (n_x, n_y)\right)} = \frac{\left(x_s - \widehat{O_x} - \widehat{x_0}, y_s - \widehat{O_y} - \widehat{y_0}\right) \cdot (n_x, n_y)}{\left(\widehat{u_n} - (\widehat{u_r} + \widehat{u_f}) \cdot (n_x, n_y)\right)}, \tag{4}$$

where $\widehat{u_n} = u_n + (N_x, N_y) \cdot (n_x, n_y)$ and (N_x, N_y) is the noise in the optic flow along the x- and y-direction.

If we represent errors in the estimated parameters with the subscript e (where error of any estimate p is defined as $p_e = p - \hat{p}$), upon substituting Eq. (3) into Eq. (4), the estimated relative depth $\hat{Z}$ may be expressed in terms of the actual depth Z as follows:

$$\hat{Z} = Z\left(\frac{\left(x_s - \widehat{O_x} - \widehat{x_0}\right)n_x + \left(y_s - \widehat{O_y} - \widehat{y_0}\right)n_y}{(x_s - O_x - x_0, y_s - O_y - y_0)\cdot(n_x, n_y) + u_{re}\cdot(n_x, n_y)Z + u_{fe}\cdot(n_x, n_y)Z + N.(n_x, n_y)Z}\right), \tag{5}$$

where

$$\begin{aligned}
u_{re} &= u_r - \hat{u}_r \\
&= \Big(-(f\beta - \hat{f}\hat{\beta}) + \gamma(y_s - O_y) - \hat{\gamma}\left(y_s - \widehat{O_y}\right) + \mathcal{O}^2_{ue}, \\
&\qquad (f\alpha - \hat{f}\hat{\alpha}) - \gamma(x_s - O_x) + \hat{\gamma}\left(x_s - \widehat{O_x}\right) + \mathcal{O}^2_{ve}\Big) \\
&= \left(-\widehat{\beta_f} + \gamma_e y - \hat{\gamma}O_{ye} + \mathcal{O}^2_{ue}, \widehat{\alpha_f} - \gamma_e x + \hat{\gamma}O_{xe} + \mathcal{O}^2_{ve}\right); \\
\widehat{\beta_f} &= f\beta - \hat{f}\hat{\beta}; \\
\widehat{\alpha_f} &= f\alpha - \hat{f}\hat{\alpha}; \\
u_{fe} &= u_f - \hat{u}_f \\
&= \Bigg(\frac{\dot{f}}{f}(x_s - O_x) - \left(\widehat{\frac{\dot{f}}{f}}\right)\left(x_s - \hat{O}_x\right) + \dot{O}_x - \hat{\dot{O}}_x, \\
&\qquad \frac{\dot{f}}{f}(y_s - O_y) - \left(\widehat{\frac{\dot{f}}{f}}\right)\left(y_s - \hat{O}_y\right) + \dot{O}_y - \hat{\dot{O}}_y\Bigg) \\
&= \left(\delta_e x - \delta O_{xe} + \zeta_{xe}, \delta_e y - \delta O_{ye} + \zeta_{ye}\right); \\
\delta &= \frac{\dot{f}}{f}; \\
\delta_e &= \frac{\dot{f}}{f} - \left(\widehat{\frac{\dot{f}}{f}}\right); \\
\left(\mathcal{O}^2_{ue}, \mathcal{O}^2_{ve}\right) &= \left(\mathcal{O}^2_u - \widehat{\mathcal{O}^2_u}, \mathcal{O}^2_v - \widehat{\mathcal{O}^2_v}\right); \\
(O_{xe}, O_{ye}) &= \left(O_x - \widehat{O_x}, O_y - \widehat{O_y}\right); \text{ and} \\
(\zeta_{xe}, \zeta_{ye}) &= \left(\dot{O}_x - \hat{\dot{O}}_x, \dot{O}_y - \hat{\dot{O}}_y\right).
\end{aligned}$$

Eq. (5) shows that errors in the motion estimates distort the recovered relative depth by a factor D, given by the terms in the bracket:

$$D = \frac{\left(x_s - \widehat{O_x} - \widehat{x_0}\right)n_x + \left(y_s - \widehat{O_y} - \widehat{y_0}\right)n_y}{(x_s - O_x - x_0, y_s - O_y - y_0)\cdot(n_x,n_y) + u_{re}\cdot(n_x,n_y)Z + u_{fe}\cdot(n_x,n_y)Z + N\cdot(n_x,n_y)Z}$$
$$= \frac{(x-\widehat{x_0})n_x + (y-\widehat{y_0})n_y}{(x'-x_0, y'-y_0)\cdot(n_x,n_y) + u_{re}\cdot(n_x,n_y)Z + u_{fe}\cdot(n_x,n_y)Z + N\cdot(n_x,n_y)Z}, \quad (6)$$

where $x' = x - O_{xe}$ and $y' = y - O_{ye}$.

Eq. (6) describes, for any fixed direction (n_x, n_y) and any fixed distortion factor D, a surface $f(x, y, Z) = 0$ in xyZ-space, which has been called the iso-distortion surface. For specific values of the parameters $x_0, y_0, \hat{x}_0, \hat{y}_0, \widehat{\alpha_f}, \widehat{\beta_f}, \gamma_e, \delta, \delta_e, O_{xe}, O_{ye}, \dot{O}_{xe}, \dot{O}_{ye}$ and (n_x, n_y), this iso-distortion surface has the obvious property that points lying on it are distorted in depth by the same multiplicative factor D. The distortion of the estimated space can be studied by looking at these iso-distortion surfaces. In this paper, we choose to reconstruct depth along the estimated epipolar direction. Such a choice is reasonable because the estimated epipolar direction contains the strongest translational flow. Henceforth the direction

$$(n_x, n_y) = \frac{(x-\widehat{x_0}, y-\widehat{y_0})}{\sqrt{(x-\widehat{x_0})^2 + (y-\widehat{y_0})^2}}$$

will be chosen; and Eq. (6) becomes:

$$D = \left((x-\widehat{x_0})^2 + (y-\widehat{y_0})^2\right)/((x'-x_0, y'-y_0)\cdot(x-\widehat{x_0}, y-\widehat{y_0}) + u_{re}\cdot(x-\widehat{x_0}, y-\widehat{y_0})Z + u_{fe}\cdot(x-\widehat{x_0}, y-\widehat{y_0})Z + N\cdot(x-\widehat{x_0}, y-\widehat{y_0})Z). \quad (7)$$

Now, if we assume that the camera has a small field of view so that u_{re} becomes $(-\widehat{\beta_f} + \gamma y, \widehat{\alpha_f} - \gamma x)$, we have

$$D = \left((x-\widehat{x_0})^2 + (y-\widehat{y_0})^2\right)/((x'-x_0, y'-y_0)\cdot(x-\widehat{x_0}, y-\widehat{y_0}) + (\Gamma_e, \Lambda_e)\cdot(x-\widehat{x_0}, y-\widehat{y_0})Z + (\delta_e x, \delta_e y)\cdot(x-\widehat{x_0}, y-\widehat{y_0})Z), \quad (8)$$

where $\Gamma_e = -\widehat{\beta_f} + \gamma_e\widehat{y_0} - \hat{\gamma}O_{ye} - \delta O_{xe} + \zeta_{xe} + N_x$ and $\Lambda_e = \widehat{\alpha_f} - \gamma_e\widehat{x_0} + \hat{\gamma}O_{xe} - \delta O_{ye} + \zeta_{ye} + N_y$.

Eq. (8) shows that, under any general motion types, a complicated distortion characteristic may arise. One way to simplify the analysis is to consider only specific motion types such as lateral and in-depth motion as in [6]. Here, we study the configurations under any general motion types.

The complexity of Eq. (8) can be better understood with a graphical representation of the equation. In particular, we are interested in deriving iso-distortion surfaces (i.e., surfaces that give rise to constant distortion factor D) spanning the xyZ-space. The simplest form of Eq. (8) reveals that the $D = 0$ surface exists only when both conditions $x = \widehat{x_0}$ and $y = \widehat{y_0}$ are simultaneously satisfied. This constraint is represented by a line, parallel to the Z-axis, piercing through the xyZ-space at $x = \widehat{x_0}$ and $y = \widehat{y_0}$. Before describing the distortion surface for any general value of D, let us first look at the formation of the $D = \pm\infty$ surface in the xyZ-space. This surface not only defines the distortion characteristics of the whole perceived space but also delineates the region where the visibility constraint will be violated.

By setting the denominator of Eq. (8) to zero, we get

$$Z = \frac{-(x - O_{xe} - x_0)(x - \widehat{x_0}) - (y - O_{ye} - y_0)(y - \widehat{y_0})}{\Gamma_e(x - \widehat{x_0}) + \Lambda_e(y - \widehat{y_0}) + \delta_e x(x - \widehat{x_0}) + \delta_e y(y - \widehat{y_0})}. \tag{9}$$

The intersection of the surface given by Eq. (9) with any plane $Z = k$ (where k is a constant) corresponds to a circle. Hence, for a range of Z values, we may generate a 3D volume enclosed by the $D = \pm\infty$ surface, whose intersection with the frontal parallel plane is given by a circle with the following parameters:

$$\begin{aligned} &\text{Radius, } R \\ &\quad = \frac{1}{2(1+\delta_e Z)}\sqrt{(x_0 - (1+\delta_e Z)\widehat{x_0} + O_{xe} - \Gamma_e Z)^2 + \left(y_0 - (1+\delta_e Z)\widehat{y_0} + O_{ye} - \Lambda_e Z\right)^2}, \\ &\text{Center, } (x, y) \\ &\quad = \left(\frac{x_0 + (1+\delta_e Z)\widehat{x_0} + O_{xe} - \Gamma_e Z}{2(1+\delta_e Z)}, \frac{y_0 + (1+\delta_e Z)\widehat{y_0} + O_{ye} - \Lambda_e Z}{2(1+\delta_e Z)}\right). \end{aligned} \tag{10}$$

Note that for $Z = -1/\delta_e$, the radius of the circle becomes infinitely large.

Figs. 1 and 2 give examples of the $D = \pm\infty$ iso-distortion surface in the xyZ-space under specific error configurations. Fig. 1 shows the existence of a discontinuity in the $D = \pm\infty$ surface at the plane $Z = -1/\delta_e$. The volume enclosed by the two surfaces on either side of the $Z = -1/\delta_e$ discontinuity formed two elongated 'cone-like' structures with one of them orientating in the positive Z direction and the other in the reverse direction. When δ_e is positive, the plane that separates the two cones lies in the negative Z region. In this case, the volume enclosed by the $D = \pm\infty$ surface in front of the image plane forms a single continuous cone structure (Fig. 2).

From here, we may extend our analysis to derive the iso-distortion surfaces for any real values of $D \neq 0, \pm\infty$. The characteristics of the distortion surfaces can be

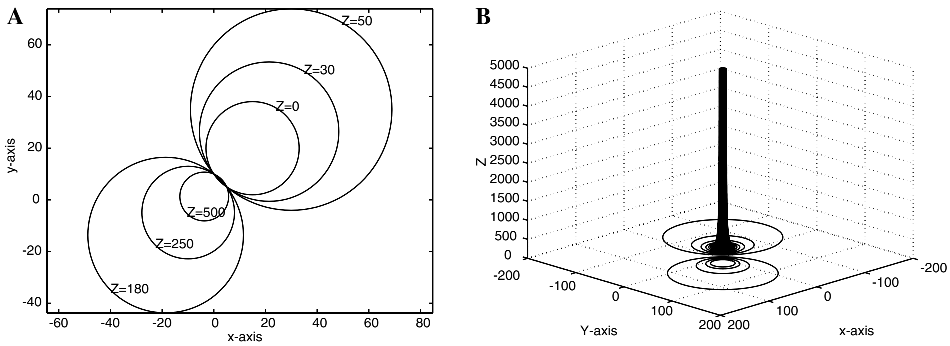

Fig. 1. $D = \pm\infty$ iso-distortion surfaces in the xyZ-space for $\delta_e = -0.01$ ($x_0 = 30, y_0 = 30, \widehat{x_0} = 0, \widehat{y_0} = 10, O_{xe} = O_{ye} = 0, \Gamma_e = 0.001, \Lambda_e = 0.001$): (A) 2D and (B) 3D view (□, true FOE and △, estimated FOE).

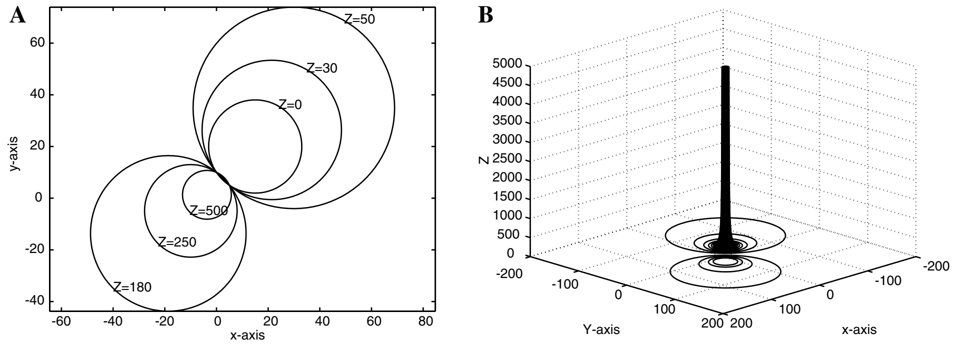

Fig. 2. $D = \pm\infty$ iso-distortion surfaces in the xyZ-space for $\delta_e = 0.01$ ($x_0 = 30, y_0 = 30, \widehat{x_0} = 0, \widehat{y_0} = 10, O_{xe} = O_{ye} = 0, \Gamma_e = 0.001, \Lambda_e = 0.001$): (A) 2D and (B) 3D view ($\square$, true FOE and $\triangle$, estimated FOE).

better understood from their level curves with depth Z as the 'height'. It can be shown that the level curves for any constant distortion factor D correspond to circles in the xyZ-space with the following parameters:

$$\text{Radius}, R = \frac{D}{2(D + D\delta_e Z - 1)} \sqrt{(x_0 - (1+\delta_e Z)\widehat{x_0} + O_{xe} - \Gamma_e Z)^2 + \left(y_0 - (1+\delta_e Z)\widehat{y_0} + O_{ye} - \Lambda_e Z\right)^2},$$

$$\text{Center}, (x, y) = \left(\frac{D(x_0 + (1+\delta_e Z)\widehat{x_0} + O_{xe} - \Gamma_e Z) - 2\widehat{x_0}}{2D(1+\delta_e Z) - 2}, \frac{D\left(y_0 + (1+\delta_e Z)\widehat{y_0} + O_{ye} - \Lambda_e Z\right) - 2\widehat{y_0}}{2D(1+\delta_e Z) - 2}\right).$$

For various D values, we obtained circles (i.e., level curves) of different radii, centered at different location. Figs. 3 and 4 correspond to two particular cases of the level curves viewed when slicing the xyZ volume with $Z = 0$ and $Z = 200$ planes for $\delta_e < 0$ and $\delta_e > 0$, respectively.

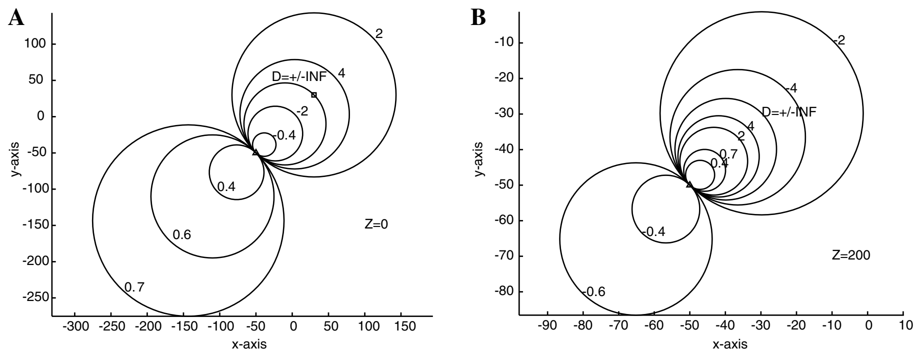

Fig. 3. Iso-distortion surfaces for $\delta_e = -0.01$ at (A) $Z = 0$ and (B) $Z = 200$ with different D: ($x_0 = 30, y_0 = 30, \widehat{x_0} = -50, \widehat{y_0} = -50, O_{xe} = O_{ye} = 0, \Gamma_e = 0.001, \Lambda_e = 0.001$).

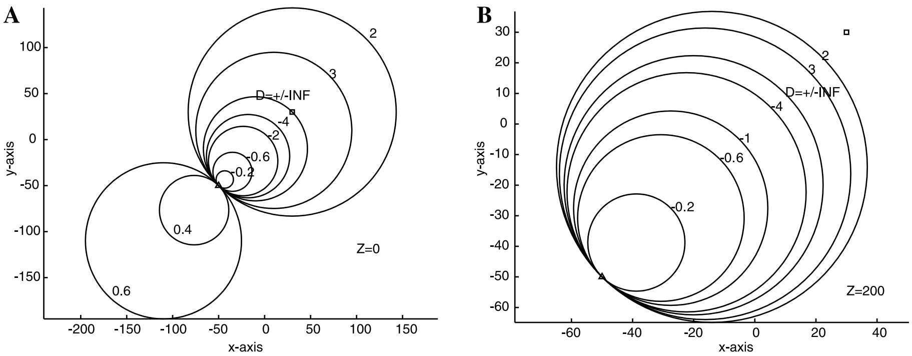

Fig. 4. Iso-distortion surfaces for $\delta_e = 0.01$ at (A) $Z = 0$ and (B) $Z = 200$ with different D: ($x_0 = 30, y_0 = 30, \widehat{x_0} = -50, \widehat{y_0} = -50, O_{xe} = O_{ye} = 0, \Gamma_e = 0.001, \Lambda_e = 0.001$).

It is shown in Fig. 4 and can be proven in general that for $\delta_e > 0$, the negative $D < 0$ surfaces are always enclosed by the $D = \pm\infty$ surface for all $Z > 0$. That is, the negative distortion region forms a cone in front of the image plane, defined by the $D = \pm\infty$ surface. On the other hand, for $\delta_e < 0$, since the asymptotic plane lies in the $Z > 0$ region, the negative depth surfaces may or may not be fully enclosed by the $D = \pm\infty$ surface, depending on the side of the asymptotic plane being considered. For example, in Fig. 3, the negative depth surfaces are enclosed by the $D = \pm\infty$ surface in the region $0 < Z < |1/\delta_e|$. In the region $Z > |1/\delta_e|$, however, it is the positive depth surfaces that are being enclosed by the $D = \pm\infty$ surface.

Much of the information that Eq. (8) contains can thus be visualized by considering a family of iso-distortion surfaces on a three-dimensional xyZ-space. Each family is defined by 11 parameters: $x0$, $y0$ and the nine error terms x_{0e}, y_{0e}, α_e, β_e, δ_e, O_{xe}, O_{ye}, ζ_{xe}, and ζ_{ye}. Within each family, a particular D value defines an iso-distortion surface. In the next subsection, we shall determine some salient geometrical properties of the iso-distortion surfaces.

2.3. Salient properties

Several salient features can be identified from the iso-distortion surfaces:

1. The $D = 0$ line (a degenerate surface) runs parallel to the Z-axis at $(x = \widehat{x_0}, y = \widehat{y_0})$. It is contained by the intersections of all $D \neq 0$ surfaces (see Figs. 3 and 4) and hence, the distortion factor on this special surface is actually undefined. Any adjustment made to the estimated FOE will move this line perpendicularly in the xyZ-space.
2. The $D = \pm\infty$ surface is discontinuous at the plane $Z = -1/\delta_e$. If the errors O_{xe} and O_{ye} are zero, the true FOE location is always found on the intersection of this surface with the $Z = 0$ plane. Otherwise, $(x_0 + O_{xe}, y_0 + O_{ye})$ is found.
3. As Z tends to infinity, the circles representing the level curves of the surfaces for all D where $D \neq 0$ decrease in size and approach a circle with radius

$$\sqrt{\frac{1}{4}\left(\frac{\Gamma_e}{\delta_e}+\widehat{x_0}\right)^2+\frac{1}{4}\left(\frac{\Lambda_e}{\delta_e}+\widehat{y_0}\right)^2}$$

and at center

$$\left(\frac{\delta_e\widehat{x_0}-\Gamma_e}{2\delta_e},\frac{\delta_e\widehat{y_0}-\Lambda_e}{2\delta_e}\right).$$

A decreasing value in δ_e shrinks the circles and shifts the circles away from the optical axis.

4. The horizontal asymptote for each surface where $D \neq 0$ is the plane $Z = (1 - D)/D\delta_e$. Hence, each surface has a different horizontal asymptotic plane which is dependent on the values of D. However, the horizontal asymptote for the surface $D = 1$ is always the $Z = 0$ plane, independent of all other parameters. A diminishing value in δ_e simultaneously moves all the horizontal asymptotes away from the $Z = 0$ plane.
5. The noise terms in the optical flow field can be found in Γ_e and Λ_e (refer to Eq. (8)). Hence, in a noisy situation, the 'cone-like' structure becomes very fuzzy since the radii and centers of the level curves are altered by various degree depending on the noise level at each optical flow vector.

2.4. Effects of large field of view

If we consider errors in the second-order flow (as would be present in a large field of view setting), the denominator of Eq. (8) (or the distortion function) becomes a complicated polynomial of third order. This means that the negative volume will not be entirely enclosed by the level curves as before, especially in the region whereby the third-order influence is substantial. The conditions for the third-order terms to be substantial are when the magnitudes of α_e, β_e, and Z are large and when the focal length is small. Fig. 5 shows an example of the simulated surface for $D = \pm\infty$.

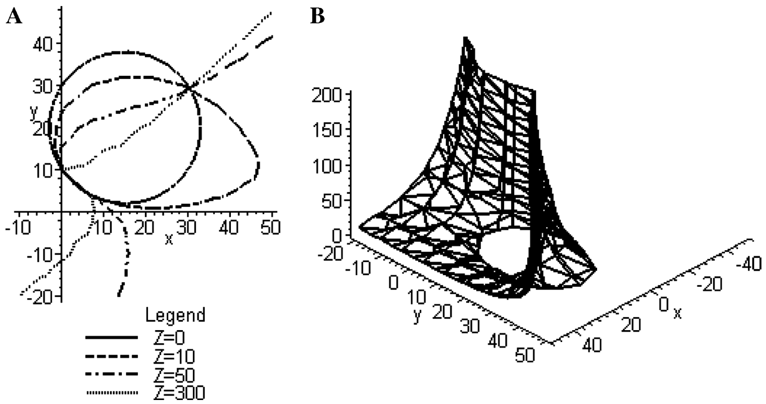

Fig. 5. $D = \pm\infty$ surface with second-order effect for $\delta_e > 0$ ($x_0 = 30, y_0 = 30, \widehat{x_0} = 0, \widehat{y_0} = 10, O_{xe} = O_{ye} = 0, \Gamma_e = 0.001, \Lambda_e = 0.001$). Left: 2D and right: 3D.

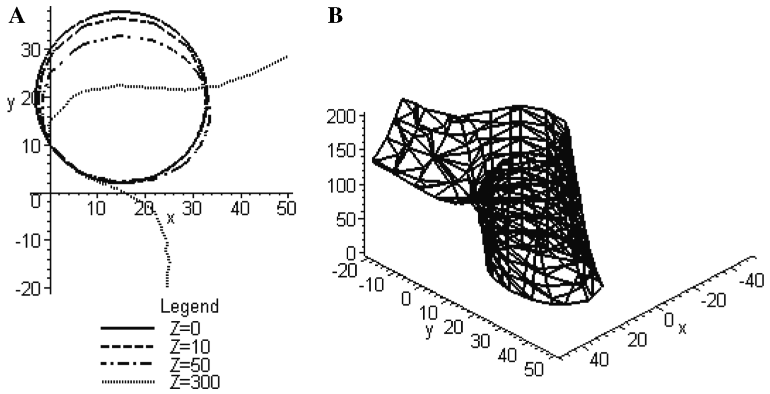

Fig. 6. $D = \pm\infty$ surface with second-order effect for large focal length. Left: 2D and right: 3D.

As can be seen, the qualitative property of the level curves changes when Z is large. Specifically, the curves open up and the negative area in the cross-section becomes unbounded. Fig. 6 shows the $D = \pm\infty$ surface for a larger focal length (i.e., smaller field of view). In this case, the opening up of the level curves only begins at a much larger Z.

The opening of the level curves represents the important role played by wide angle view in motion estimation. In order to achieve zero negative volume, the denominator of the distortion function needs to be positive. In this case, since it is a cubic curve, we can always find real root for some x and y. Hence, the denominator of the distortion function (i.e., Eq. (8)) can be positive or negative for a fixed Z. This makes it hard to satisfy the conditions with no negative depth in the presence of errors. The likelihood of getting ambiguous solutions is thus lessened.

3. What can the distortion contours tell us?

3.1. The visibility constraint

Direct motion algorithms [9,17] often attempt to find the solution by minimizing the number of negative depth found. This is known as the visibility constraint. In particular, Longuet-Higgins [21] showed that any spurious solutions arising from a moving plane could be ruled out with the constraint. However, its usage in the estimation of calibration parameters is relatively unexplored. We would now like to examine this constraint in the light of the negative distortion volume. The geometry of the negative distortion volume allows us to examine these questions: does the veridical solution have the minimum number of negative depth? Are there combination of estimation errors such that the visibility constraint is not sufficient to discriminate them from the true solution? Do these ambiguous solutions exhibit any peculiar properties in terms of their recovered structure or their motion estimates?

3.2. Constraints on motion errors

As observed in Figs. 3 and 4, the intersections of the iso-distortion surfaces with any Z-plane comprise of a family of circles belonging to different D values. Irrespective of the actual scene structure, there will be configurations whereby a large number of negative depth estimates are obtained, thereby furnishing the possibility of using the visibility constraint to rule out that particular solution.

We consider three cases according to different values of δ_e, and depict these cases in Fig. 7. In contrast to Figs. 3 and 4, Fig. 7 is obtained by slicing the iso-distortion volume with the $y = 0$ plane. Plotted in this way, how the region of negative depth (all the shaded region) varies as a function of depth is better illustrated. It can be seen that the negative region is either outside or inside the "cone" formed by the $D = \pm\infty$ surfaces (from this perspective, the cone is manifested as two vertical lines: the vertical asymptotic branch of the $D = \pm\infty$ surface and the $D = 0$ surface).

- When $\delta_e > 0$, most of the negative distortion region lie behind the image plane (see Fig. 7A). What remains in front of the image plane is a band of negative distortion volume, bounded by the $D = \pm\infty$ surface. This is true for all positive values of δ_e. Clearly, the negative distortion volume can be minimized by letting $(\widehat{x_0}, \widehat{y_0})$ approach (x_0, y_0) so that the cone vanishes.
- When $\delta_e < 0$ and $|\delta_e|$ is small such that $Z_{\max} < |1/\delta_e|$, we also obtain a small volume of negative distortion region (see darker shaded region in Fig. 7B). As discussed previously, $-1/\delta_e$ is the horizontal asymptotic plane of the $D = \pm\infty$ surface. Similar to the case of $\delta_e > 0$, the number of negative depth can be minimized by letting $(\widehat{x_0}, \widehat{y_0})$ approach (x_0, y_0) so that the cone vanishes.
- When $\delta_e < 0$ and $|\delta_e|$ is large (i.e., the asymptotic plane is close to the $Z = 0$ plane with $Z_{\max} > 1/\delta_e$), we would then obtain the negative distortion region as shown in Fig. 7C. Minimizing the negative depth volume does not yield the veridical solution. In this case, we would need to both minimize and maximize the negative depth volume in the range $Z < 1/|\delta_e|$ and $Z > 1/|\delta_e|$, respectively, to obtain the veridical solution (i.e., setting the radii of circles forming the $D = \pm\infty$ surface to zero). Finding this distance to allow both maximization and minimization of negative volume would be difficult in practice. The negative depth volume would

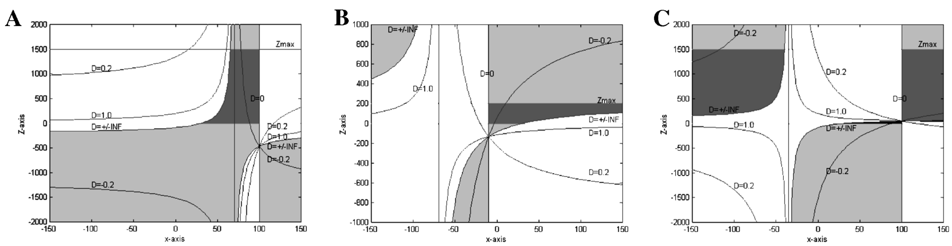

Fig. 7. 2D representations of iso-distortion volume sliced at $y = 10$ ($x_0 = 30, y_0 = 30, \widehat{y_0} = 10, O_{xe} = O_{ye} = 0, \Gamma_e = 0.001, \Lambda_e = 0.001$): (A) $\delta_e = 0.005$, $\widehat{x_0} = 100$; (B) $\delta_e = -0.005$, $\widehat{x_0} = -10$; and (C)$\delta_e = -0.01$, $\widehat{x_0} = 100$ (shaded region represents negative distortion region and darker shaded region represents scene in view that falls under the negative distortion region).

also reach a constant value once $(\widehat{x_0}, \widehat{y_0})$ lies outside the image plane so that the right vertical contour of the cone is out of view.

From the above, it follows that if the true FOE is inside the image plane, the condition on the zoom error conducive for ambiguity is such that:

$$\delta_e > 0 \quad \text{or} \quad \frac{1}{\delta_e} < -Z_{\max}. \tag{11}$$

If $\delta_e < 0$ and $|\delta_e|$ is large, large negative distortion region always results. Conversely, if the true FOE is out of the image plane, then all values of δ_e can lead to ambiguity. For example, if $\delta_e > 0$ and both (x_0, y_0) and $(\widehat{x_0}, \widehat{y_0})$ were out of the image plane (and on the same side), then even if they do not meet, the negative distortion band would be outside the field of view. Thus, this solution will not yield any negative depth estimates and would be totally ambiguous too. If $\delta_e < 0$ and both (x_0, y_0) and $(\widehat{x_0}, \widehat{y_0})$ were out of the image plane (and on different side), we would also obtain the configuration whereby the negative band is always outside the field of view.

Given condition (11), we shall now derive further conditions on the motion errors such that the band of negative volume will be minimized (i.e., ambiguity is maximized). To derive these combinations, we first arbitrarily fix the error δ_e given by condition (11). The constraint on the other parameters $\hat{\Gamma}, \hat{\Lambda}, \widehat{x_0}$, and $\widehat{y_0}$ that will yield minimum negative distortion region depends on whether an algorithm solves for these parameters separately or simultaneously:

1. If $\hat{\Gamma}$ and $\hat{\Lambda}$ are solved first and the estimates contain errors Γ_e and Λ_e, then the $\widehat{x_0}$ and $\widehat{y_0}$ that minimize the negative depth volume, given fixed Γ_e and Λ_e, are

$$\begin{aligned}\widehat{x_0} &= \frac{(x_0 + O_{xe} + (\Gamma_e/\delta_e))\ln[(1+\delta_e Z_{\max})/(1+\delta_e Z_{\min})]}{\delta_e(Z_{\max} - Z_{\min})} - \frac{\Gamma_e}{\delta_e},\\ \widehat{y_0} &= \frac{(y_0 + O_{ye} + (\Lambda_e/\delta_e))\ln[(1+\delta_e Z_{\max})/(1+\delta_e Z_{\min})]}{\delta_e(Z_{\max} - Z_{\min})} - \frac{\Lambda_e}{\delta_e},\end{aligned} \tag{12}$$

   where we have assumed that depths in the scene are uniformly distributed between $Z_{\min}$ and $Z_{\max}$. A check with the second derivative test reveals that this critical point is a minimum point for $Z_{\max} > Z_{\min}$ which is always true. Furthermore, in order for the above two equations to be valid, the terms inside the *ln* operator must be positive. It requires that $Z_{\min} > -1/\delta_e$ for $\delta_e > 0$ and $Z_{\max} < 1/|\delta_e|$ for $\delta_e < 0$, respectively, which are always satisfied given Eq. (11).

2. If all of the parameters, namely $\widehat{x_0}, \widehat{y_0}, \hat{\Gamma}$, and $\hat{\Lambda}$ are solved together, then the solution that minimizes the negative depth volume is given by

$$\begin{aligned}\hat{x}_0 &= -\frac{\Gamma_e}{\delta_e} = x_0 + O_{xe},\\ \hat{y}_0 &= -\frac{\Lambda_e}{\delta_e} = y_0 + O_{ye}.\end{aligned} \tag{13}$$

   This can be obtained by setting the radii of the circles forming the $D = \pm\infty$ surface to be zero for all $Z \neq 1/\delta_e$. In this case, the negative depth volume in front of the image plane vanishes. The distortion surfaces then become planes with

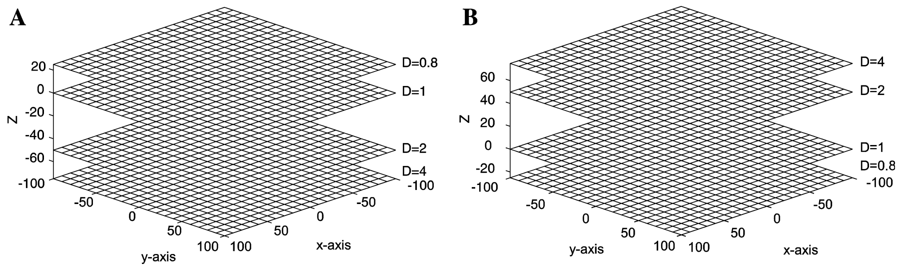

Fig. 8. Iso-distortion surfaces when $\hat{x}_0 = -\Gamma_e/\delta_e = x_0$ and $\hat{y}_0 = -\Lambda_e/\delta_e = y_0$ ($x_0 = 30, y_0 = 30, O_{xe} = 0, O_{ye} = 0, \Gamma_e = 0.3, \Lambda_e = 0.3$): when (A) $\delta_e = 0.01$ and (B) $\delta_e = -0.01$.

positive D. Depending on the sign of δ_e, we obtain either $D < 1$ or $D > 1$ distortion surfaces lying in front of the image plane (see Fig. 8). Also, unless O_{xe} and O_{ye} are both zero, the estimated FOE will never be the veridical one.

In the case whereby the negative depth volume in front of the image plane vanishes, the distortion factor D reduces itself to

$$D = \frac{1}{1 + \delta_e Z}. \tag{14}$$

This corresponds to a relief transformation $1/(a + bZ)$, where $a = 1$ and $b = \delta_e$. This relief transformation preserves the ordering of points; its general properties were recently discussed and analyzed by [11,20]. For instance, it is well known that depth relief can be reliably recovered from shading cue [20]. In the case of motion cues, it is believed and psychophysically demonstrated that human subject can recover the tilt—the relief structure of the scene—better than the slant of a planar scene. Here, we obtain even stronger results: that even with errors in the 3D motion estimates, the ordinal structure can be correctly recovered as long as the visibility constraint can be effectively applied.

It is well to note at this point that whilst we have derived the likely conditions for solutions to be ambiguous, we are not proposing any new algorithm for motion estimation. Rather, we are arguing that the visibility constraint can be a possible supplementary method to prune candidates obtained by other motion estimation algorithms. That is, from a solution set obtained by the latter, we choose those candidates that minimize the negative volume as the final solution. In addition, our analysis also elucidates the types of motion errors still likely to be present in the final solution even with the application of the visibility constraint. Such a characterization of the errors is important in view of the imperfect nature of most structure from motion algorithms and the attendant need to understand the limitations of such algorithms.

3.3. Summary

The preceding analysis shows that the use of the visibility constraint does not lift the ambiguities that exist among various kinds of motions. However, it does restrict the solution set so that those yielding the minimum negative depth estimates possess certain nice properties. In particular, when the constraint in Eq. (11) is satisfied, and both the rotation and translation parameters are simultaneously estimated, then the direction of heading is correctly estimated (up to an offset term (O_{xe}, O_{ye})). Furthermore, as can be seen from Fig. 8, the iso-distortion surfaces become parallel planes lying perpendicular to the optical axis, which would result in well-behaved distortion. As a result of this distortion, the reconstructed scene may appear visually perfect even though the depths have been squashed or stretched to various degrees. It is of interest to compare this result with that demonstrated by Bougnoux [2]: that the uncertainty on the focal length estimation leads to a Euclidean calibration up to a quasi-anisotropic homothety, which in turn yields visually good-looking reconstruction.

4. Experiments

This section presents the experiments carried out to support the theoretical findings established in the preceding section. Specifically, we demonstrate the possibility of estimating the heading direction of the camera correctly based on minimizing the number of negative depth estimates. The distortion effects due to erroneous motion estimates on simple surfaces were also tested. Both synthetic and real images were used.

4.1. Synthetic images with slanted planes

A set of noise-free synthetic images with dimension 240 pixels by 320 pixels were generated. The focal length of the projection was fixed at 600 pixels. This gave a viewing angle of near 30°. The 3D scene contained three slanted planes orientated with different slant angles. The slant profiles of the three planes were as shown in Fig. 9. The true FOE was located at $(30, 30)$ of the image plane and the 3D rotational parameters (α, β, γ) were $(0.00025, 0.0006, 0)$. There was no change in the focal length (i.e., $\dot{f}/f = 0$). The computed optical flow magnitude ranges between 0.000904 and 2.42017.

Recall that the condition on δ_e leading to ambiguity is very loose; thus many values of δ_e would yield minimum negative volume. To report on some specific numerical values, we first arbitrarily fixed the error δ_e to be some positive number. In this experiment, we fixed it to be 0.0001. We then solved for the rotational parameter $(\hat{f}\hat{\alpha}, \hat{f}\hat{\beta})$ and the FOE $(\widehat{x_0}, \widehat{y_0})$ in the following manner: For each hypothesized $(\widehat{x_0}, \widehat{y_0})$, we selected the best $(\hat{f}\hat{\alpha}, \hat{f}\hat{\beta})$ candidate such that the minimum number of negative depth estimates was obtained. The search for (x_0, y_0) ranges from $(-160, -140)$ to $(160, 140)$ in step of 1 pixel along each direction. This represents a total of 76,800 hypothesized FOEs. The search range for $(f\alpha, f\beta)$ is between $(0.348, 0.134)$ and $(0.372, 0.166)$ in step of 0.001 for each parameter. The hypothesized FOE that gave the global minimum negative depth was chosen as the FOE estimate.

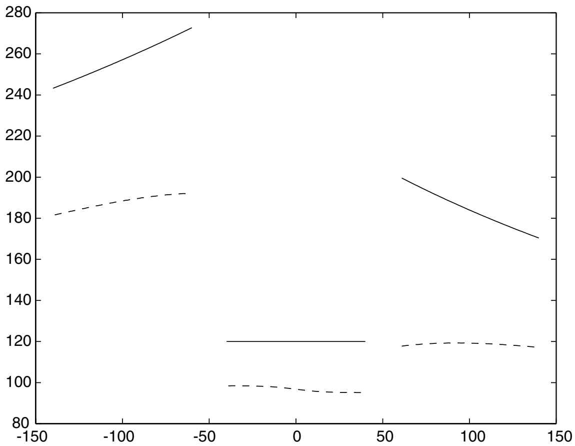

Fig. 9. Slant profiles for the three synthetic planes in x–Z-plane: solid lines, original profiles and broken lines, reconstructed profiles.

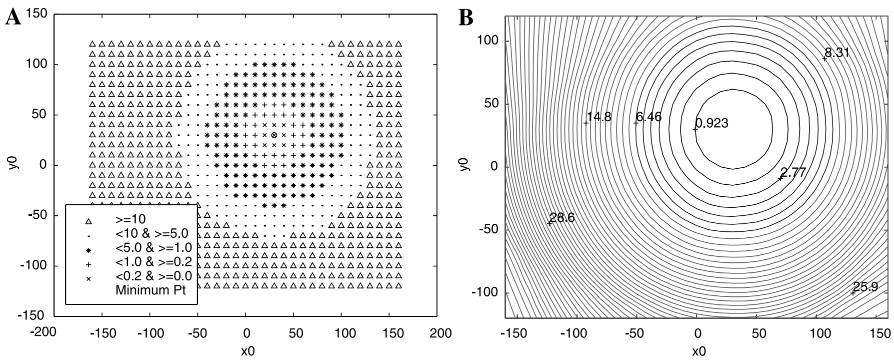

Fig. 10. Minimum negative depth distribution with coarse sampling of FOE: (A) scatter plot and (B) contour plot in plane scene.

Fig. 10 shows the location of the estimated FOE using the global minimum negative depth criteria, as well as the distribution of the minimum negative depth with the hypothesized FOEs. In this case, we have successfully obtained the global minimum negative depth position at $(30, 30)$. Using the erroneous motion estimates $(\widehat{x_0}, \widehat{y_0})$, $(\hat{f}\hat{\alpha}, \hat{f}\hat{\beta})$, and δ_e that resulted in the least amount of negative depth estimates, we attempted to reconstruct the synthetic planes. Fig. 9 shows the plan view of the three synthetic planes, together with their reconstructed versions. It can be seen that the relief of the plane remained unchanged after the transformation, i.e., the ordinal depth were preserved. Note that the metric aspect of the plane orientations (their slants), however, was altered. This change can be related to the calibration uncertainties via the complex rational function given by $1/(1 + \delta Z)$.

4.2. Image sequences

We conducted similar analysis on several image sequences, both synthetic (but realistic) and real. Optical flow fields were computed from these sequences using the Lucas–Kanade algorithm. These sequences, together with their results, are shown in Figs. 11–14, respectively. Fig. 15 shows the reconstructed depth map for the four image sequences. The intrinsic and extrinsic motion parameters for the sequences are tabulated in Table 1, while the search ranges for $(\widehat{x_0}, \widehat{y_0})$ and $(f\alpha, f\beta)$ are shown in Table 2. The value of δ_e has been fixed at 0.0001 and the FOE is hypothesized at a regular interval of 3 pixels to reduce computational overload. The estimated FOE for the four sequences were found to be at $(0, -6)$, $(-8, -13)$, $(-8, 124)$, and $(59, 61)$, respectively.

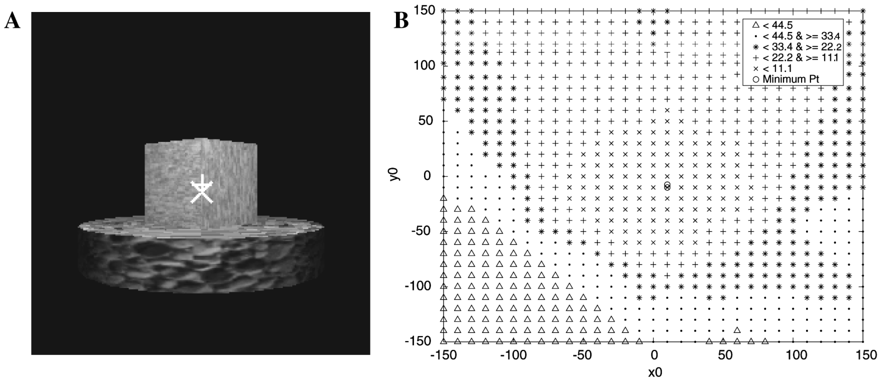

Fig. 11. Image Sequences I. Left: image ('+', true FOE and 'X', estimated FOE) and right: scatter plot of minimum negative depth distribution.

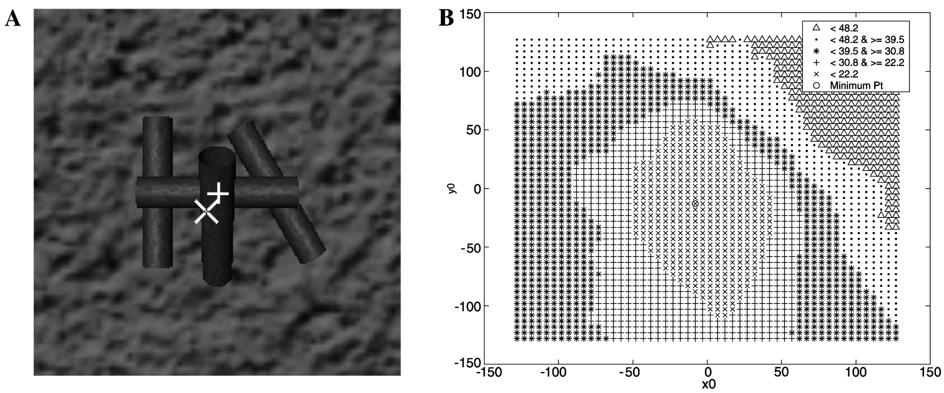

Fig. 12. Image Sequences II. Left: image ('+', true FOE and 'X', estimated FOE) and right: scatter plot of minimum negative depth distribution.

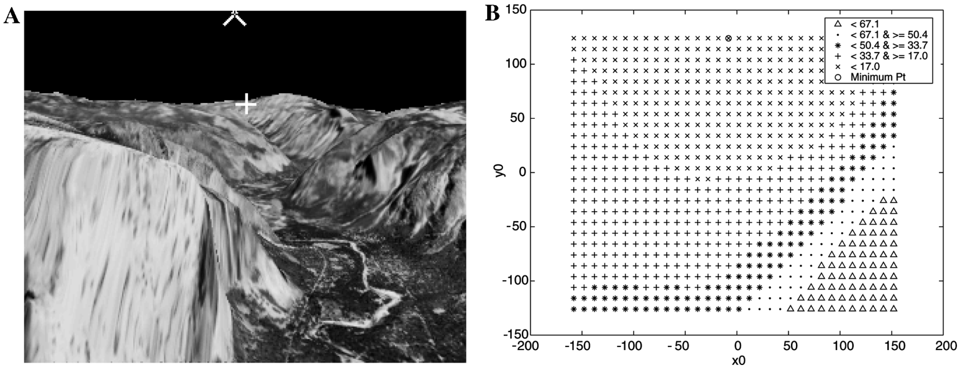

Fig. 13. Image Sequences III. Left: image ('+', true FOE and 'X', estimated FOE) and right: scatter plot of minimum negative depth distribution.

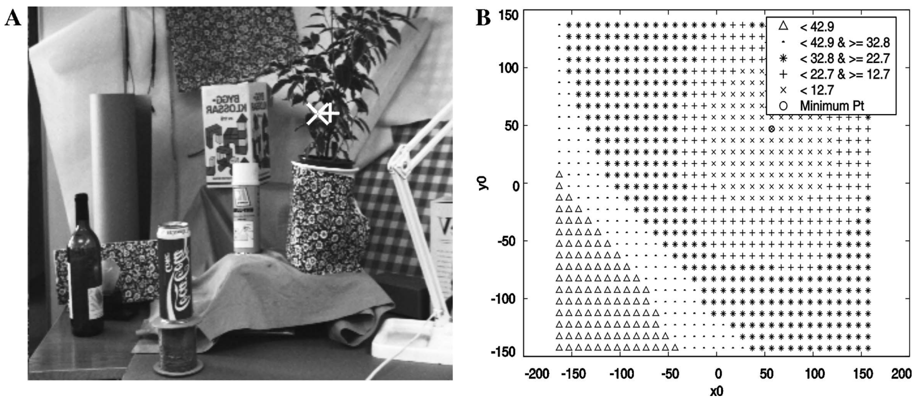

Fig. 14. Image Sequences IV. Left: image ('+', true FOE and 'X', estimated FOE) and right: scatter plot of minimum negative depth distribution.

4.3. Discussions

The results obtained seem to corroborate the various predictions made in this paper. In particular, while the use of visibility constraint cannot be used to effect a full recovery of all the parameters, minimizing the number of negative depth estimates do result in certain nice properties of the solutions if certain constraints discussed in previous section are met. It seems that at least in the case where the only unknown intrinsic parameter is the zoom flow and that the true FOE is inside the image plane, structure information in the form of depth relief can be obtained from the motion cue. The reconstructed depths did look visually alright due to the preservation of the depth relief in the synthetic image experiment (see Fig. 9). However, the reconstructed depth maps from real image sequences in Fig. 15 did not appear as good since the FOEs were not accurately determined in most of these cases.

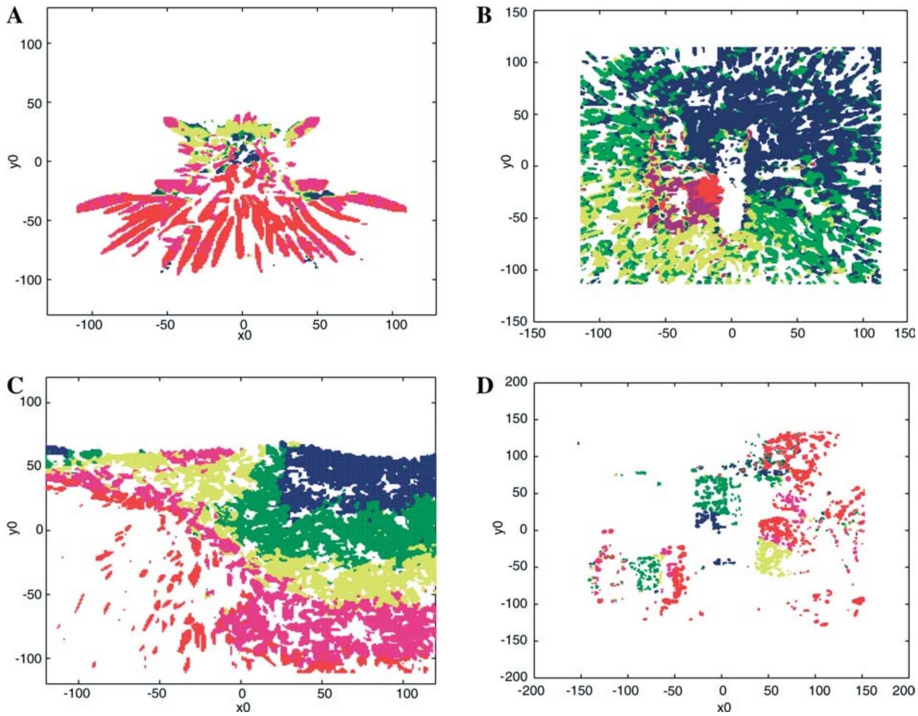

Fig. 15. Reconstructed depth map. Top left: Image Sequence I; top right: Image Sequence II; bottom left: Image Sequence III; and bottom right: Image Sequence IV. (Color representation from near to far: red, magenta, yellow, green and blue, respectively.)

Table 1
Intrinsic and extrinsic parameters

Image Sequence	(x_c, y_c)	f	(x_0, y_0)	(α, β, γ)	$\dot{f}$
I	$(0, 0)$	309	$(0, 0)$	$(0, 0, 0)$	0
II	$(0, 0)$	309	$(0, 0)$	$(0, 0, 0)$	0
III	$(0, 0)$	337.5	$(0, 59.5)$	$(0.0002319, 0.001625, -0.0002341)$	0
IV	$(3, -9)$	620	$(65, 73)$	$(-0.00025, -0.00013, 0)$	0

Table 2
Boundaries and step-sizes of searched parameters (subscripts i, f, and s denote initial, final, and step-size, respectively)

Seq.	$\widehat{x_0}_{\mathrm{i}}$	$\widehat{x_0}_{\mathrm{f}}$	$\widehat{x_0}_{\mathrm{s}}$	$\widehat{y_0}_{\mathrm{i}}$	$\widehat{y_0}_{\mathrm{f}}$	$\widehat{y_0}_{\mathrm{s}}$	$\hat{f}\hat{\alpha}_{\mathrm{i}}$	$\hat{f}\hat{\alpha}_{\mathrm{f}}$	$\hat{f}\hat{\alpha}_{\mathrm{s}}$	$\hat{f}\hat{\beta}_{\mathrm{i}}$	$\hat{f}\hat{\beta}_{\mathrm{f}}$	$\hat{f}\hat{\beta}_{\mathrm{s}}$
I	−128	128	3	−128	128	3	−0.00128	0.00128	0.001	−0.00128	0.00128	0.001
II	−128	128	3	−128	128	3	−0.00128	0.00128	0.001	−0.00128	0.00128	0.001
III	−158	158	3	−126	126	3	0.0657	0.0909	0.001	0.5326	0.5642	0.001
IV	−163	163	3	−143	143	3	−0.02205	0.00655	0.001	0.0318	0.0008	0.001

The reliability in the depth reconstruction was greatly hampered by the inherent noise in the flow vectors. In our theoretical study, we show that when $\delta_e > 0$ and if the FOE were not correctly estimated, a cone of negative volume (bounded by the true and estimated FOEs on the image plane) is obtained. Under such configuration, scene points belonging to the same depth do not necessarily get distorted by the same factor. For example, in Fig. 7A, for any constant Z, the distortion factor is different on the left and right of the shaded region. Ordinal depth is only preserved in localized regions that are away from the negative depth band where the contours lie flat. This characteristic was unanimously observed in our sequences where the FOEs were not veridically determined. For example, in Image Sequence II, the lower-left region appears to be nearer than the upper-right region in its reconstruction despite that they are of rough equivalent distance from the observation point. However, within each localized region, the ordinal depth relationship was still well preserved except for the area surrounding the estimated FOE. In Image Sequence III, despite the large error found in the FOE estimation, the ordinal depth was still well preserved since the depth values were largely computed in region where the distortion contours were flat (i.e., the negative depth region enclosed by the estimated and true FOE is close to the image border). Similarly in Image Sequence IV, the ordinal depth between the three objects (i.e., coke can, bottle, and table cloth) in the bottom-left region of the image is preserved. Again, this region is away from the FOEs. Our results suggest that, while accurate 3D scene reconstruction in the global sense can be difficult, qualitative recovery for localized image region is reliable. Overall, we found that the reconstructed depth maps for all of the sequences were rather flat; clear distinctions of object distances showed up only when these were with large separating distance between them.

Since the depth values had been computed from optical flow values, the two important criterion for the visibility constraint to work well were adequate number of feature points and their variations in scene depth. It holds that while the underlying negative distortion region may have increased in size, there may not be any increase in the number of negative depth estimates, due to a lack of scene point residing in the negative distortion regions. Evidently, under such circumstances, the number of negative depth estimates may not exhibit a monotonic increase as the error in the FOE increases. The effect is especially dependent on the features in the surrounding regions of the true FOE. Theoretically, we have established that if $\delta_e > 0$, (x_0, y_0) and $(\widehat{x_0}, \widehat{y_0})$ both lie on the intersection of $D = \pm\infty$ surface with the image plane. In other words, the two points lie on the circumference of a circle on the image plane. It can be seen from Eq. (10) that as $(\widehat{x_0}, \widehat{y_0})$ approaches (x_0, y_0), the radius of this circle decreases. Together with other conditions stated in Eq. (11), the entire encompassing negative volume thins out, finally vanishing when the two points coincide. Nevertheless, if the region around (x_0, y_0) is featureless or the noise is such that no negative depth is computed for this region, we would also obtain several likely possible solutions.

The dependency of solution on scene structure was evidently found in Sequences I and II. Although both sequences featured a forward moving camera along the optical axis (i.e., $(x_0, y_0) = (0, 0)$), we did not obtain the same FOE estimates (assuming effect of noise to be minimal). A close observation revealed that the estimates were slightly biased towards nearby region of the true FOE where feature points were sparse. This tendency was also found in Sequence III. In this sequence, the true FOE was just located near to the boundary of the featureless sky. When searching for the FOE in the featureless region, the entire negative volume bounded by (x_0, y_0) and $(\widehat{x_0}, \widehat{y_0})$ might fall into this region. Since no negative depth value was computed in this region, the amount of negative depth could be minimal as well. In order to substantiate our claim, we limited our search range for $\widehat{y_0}$ to be from -126 to 57 inclusive. The global minimum negative volume solution was found to be at $(7, 57)$, which was indeed close to our true solution.

Besides scene structure, the presence of noise is yet another problem that plagues most image sequences. The effect of any noise N at a particular image pixel is to alter the terms Γ_e and Λ_e in the numerators of Eq. (13). Thus, to this particular image point, its effective FOE estimates has shifted and part of the problem lies in that this shift has different effects on the various solution candidates. For the case of those solutions where the negative distortion region in front of the image plane would have vanished under noiseless conditions, this noise-induced shift away from $x = x_0$ and $y = y_0$ may result in that particular depth estimate becoming negative again (depending on the sign and magnitude of that N). For the case of other solutions, this shift may have the contrary effect of moving that point out of the negative distortion region. It becomes plausible that the "desired" solutions (i.e., those satisfying (11)) may not have the minimum number of negative depth estimates. Thus, the overall effect of noise is to reduce the effectiveness of the visibility constraint in getting the "desired" solutions. In the presence of random noise, the net effect of N may get canceled out. However, the analysis of the equation can be rendered more complicated with noise of anisotropic nature.

5. Conclusions

This paper represents a first look at the distortion in the perceived space resulting from errors in the estimates of calibration parameters. The geometry of the negative distortion region allows us to answer questions such as whether the visibility constraint is adequate for resolving ambiguity. Specifically, we show that, under small field of view condition, despite the presence of zoom error, the heading direction can be recovered with visibility constraint. It is also found that while Euclidean reconstruction is difficult, the resulting distortion in the structure satisfies the relief transformation, which means that ordinal depth is preserved. The above results were obtained based on the assumptions that the skew angle remained constant and radial distortion could be ignored. Whilst the skew angle can be quite stable, we acknowledge that the assumption on radial distortion may not be true over a long period of time. The latter can be solved by rectifying the radial distortion through a corrective mapping method as suggested by [14]. Although this rectification may change the effective focal length of the image, this does not affect our analysis since it allows for uncertainty in the focal length estimate.

In this paper, we had successfully qualified the effects of zoom estimation error on structure and heading direction recovery where the visibility constraint has been taken into account. However, the overall effects on the recovery of motion parameters in any general motion estimation algorithm is still not adequately explained. For instance, how do such intrinsic errors affect the residual error surface of a motion estimation algorithm? Does it result in additional or changed local minima condition on the residual error surface? This shall form the basis of our future work.

To close this paper, the remark should be added that there are many potential applications of the results of our research to areas like multimedia video indexing, searching, and browsing, where it is common practice to use zoom lenses. It is desirable to incorporate partial scene understanding capabilities under freely varying focal length, yet without having to go through elaborate egomotion estimation to obtain the scene information. The conclusion of this paper is that while it is very difficult to extract metric scene descriptions from video input, qualitative representations based on ordinal representation constitute a viable avenue.

References

[1] M. Armstrong, A. Zissermann, P. Beardsley, Euclidean structure from uncalibrated images, in: Proc. 5th British Machine Vision Conference, UK, September, 1994, pp. 508–518.

[2] S. Bougnoux, From Projective to Euclidean Space under any practical situation, a criticism of self-calibration, in: Proc. Int. Conf. on Computer Vision '98, 1998, pp. 790–796.

[3] M.J. Brooks, W. Chojnacki, Determining the egomotion of an uncalibrated camera from instantaneous optical flow, J. Opt. Soc. Amer. A 14 (10) (1997) 2670–2677.

[4] L.-F. Cheong, C. Fermüller, Y. Aloimonos, Effects of errors in the viewing geometry on shape estimation, Comput. Vis. Image Understand. 71 (3) (1998) 356–372.

[5] L.-F. Cheong, K. Ng, Geometry of distorted visual space and cremona transformation, Int. J. Comput. Vision 32 (2) (1999) 195–212.

[6] L.-F. Cheong, T. Xiang, Characterizing depth distortion under different generic motions, Int. J. Comput. Vision 44 (3) (2001) 199–217.

[7] L. Dron, Dynamic camera self-calibration from controlled motion sequences, in: Proc. IEEE Conf. on Computer Vision and Pattern Recognition, New York, NY, 1993, pp. 501–506.

[8] O. Faugeras, Stratification of three-dimensional vision: projective, affine, and metric representations, J. Opt. Soc. Amer. A 12 (3) (1995) 465–484.

[9] C. Fermüller, Passive navigation as a pattern recognition problem, Int. J. Comput. Vision 14 (1995) 147–158.

[10] G. Florou, R. Mohr, What accuracy for 3D measurement with cameras? in: Proc. Int. Conf. on Pattern Recognition, Los Alamitos, CA, 1996, pp. 354–358.

[11] J. Gårding, J. Porrill, J.E.W. Mayhew, J.P. Frisby, Stereopsis, vertical disparity and relief transformations, Vision Res. 35 (5) (1995) 703–722.

[12] R.I. Hartley, Calibration of cameras using the essential matrix, in: Proc. ARPA Image Understanding Workshop, Morgan Kaufmann Publishers Inc, Los Altos, CA, 1992, pp. 911–915.

[13] R.I. Hartley, Self-calibration of stationary cameras, Int. J. Comput. Vision 22 (1) (1997) 5–23.

[14] R. Hartley, A. Zisserman, Multiple View Geometry in Computer Vision, Cambridge University Press, UK, 2000, Chapter 6, pp. 166–183.

[15] E. Hayman, L. Agapito, I.D. Reid, D.W. Murray, The role of self-calibration in euclidean reconstruction from two rotating and zooming cameras, in: Proc. 6th Eur. Conf. on Computer Vision II, Dublin, Ireland, 2000, pp. 477–492.

[16] A. Heyden, K. Åström, Euclidean reconstruction from image sequences with varying and unknown focal length and principal points, in: Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 1997, pp. 438–443.

[17] B.K.P. Horn, E.J. Weldon Jr., Direct method for recovering motion, Int. J. Comput. Vision 2 (1988) 51–76.
[18] F. Kahl, B. Triggs, K. Åström, Critical motions for auto-calibration when some intrinsic parameters can vary, J. Math. Imag. Vision 13 (2000) 131–146.
[19] J.J Koenderink, A.J. van Doorn, Affine structure from motion, J. Opt. Soc. Amer. A 8 (2) (1991) 377–385.
[20] J.J. Koenderink, A.J. van Doorn, Relief: pictorial and otherwise, Image Vision Comput. 13 (5) (1995) 321–334.
[21] H.C. Longuet-Higgins, The visual ambiguity of a moving plane, in: Proc. of Royal Society of London Series B 223, (1984) pp. 165–175.
[22] Q.-T. Luong, O. Faugeras, Self-calibration of a moving camera from point correspondences and fundamental matrices, Int. J. Comput. Vision 22 (3) (1997) 261–289.
[23] Y. Ma, R. Vidal, J. Košecká, S. Sastry, Kruppa equation revisited: its renormalization and degeneracy, in: Proc. Eur. Conf. on Computer Vision II, Dublin, Ireland, 2000, pp. 561–577.
[24] S.J. Maybank, O.D. Faugeras, A theory of self-calibration of a moving camera, Int. J. Comput. Vision 8 (2) (1992) 123–151.
[25] M. Pollefeys, L. Van Gool, A. Oosterlinck, Euclidean reconstruction from image sequences with variable focal length, in: B. Buxton, R. Cipolla (Eds.), Eur. Conf. on Computer Vision '96, volume 1064 of Lecture notes in Computer Science, Springer-Verlag, 1996, pp. 31–44.
[26] L. Quan, T. Kanade, Affine structure from line correspondences with uncalibrated affine cameras, IEEE Trans. Pattern Analysis and Machine Intelligence 19 (8) (1997) 834–845.
[27] M. Spetsakis, Models of statistical visual motion estimation, Comput. Vision, Graphics, Image Process. 60 (1994) 300–312.
[28] G.P. Stein, Accurate internal camera calibration using rotation with analysis of sources of error, in: Proc. Int. Conf. on Computer Vision '95, Boston, MA, 1995, pp. 230–236.
[29] P. Sturm, Self-calibration of a moving zoom lens camera by pre-calibration, Image Vision Comput. 15 (8) (1997) 583–589.
[30] P. Sturm, Critical motion sequences for monocular self-calibration and uncalibrated euclidean reconstruction, in: Proc. Conf. on Computer Vision and Pattern Recognition, Puerto Rico, USA, June, 1997, pp. 1100–1105.
[31] P. Sturm, Critical motion sequences for the self-calibration of cameras and stereo systems with variable focal length, in: Proc. British Machine Vision Conference, Nottingham, UK, September, 1999, pp. 63–72.
[32] 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.
[33] B. Triggs, Factorization methods for projective structure and motion, in: Proc. IEEE Conf. on Computer Vision and Pattern Recognition, San Francisco, CA, June, 1996, pp. 845–851.
[34] T. Viéville, O. Faugeras, The first order expansion of motion equations in the uncalibrated case, Comput. Vision, Graphics, Image Process.: Image Understand. 64 (1) (1996) 128–146.
[35] G.-S.J. Young, R. Chellappa, Statistical analysis of inherent ambiguities in recovering 3-d motion from a noisy flow field, IEEE Trans. Pattern Analysis and Machine Intelligence 14 (1992) 995–1013.