

# Characterizing Depth Distortion under Different Generic Motions

LOONG-FAH CHEONG\*and TAO XIANG†

Electrical and Computer Engineering Department, National University of Singapore

10 Kent Ridge Crescent, Singapore 119260

## Abstract

Given that errors in the estimates for the intrinsic and extrinsic camera parameters are inevitable, it is important to understand the behaviour of the resultant distortion in depth recovered under different motion-scene configurations. The main interest in this study is to look for generic motion type that can render depth recovery more robust and reliable. To this end, lateral and forward motions are compared both under calibrated and uncalibrated scenarios. For lateral motion, we found that although Euclidean reconstruction is difficult, ordinal depth information is obtainable; while for forward motion, depth information (even partial one) is difficult to recover. In the uncalibrated case, with fixed intrinsic parameters, the preceding statements still hold. However, if intrinsic parameter variations are allowed, then for lateral motion, depth relief can only be preserved locally. In general, lateral motion yields a distortion relationship that belongs to the projective transformation of a very simple type, while the distortion transformations for general motions including forward motion belong to the Cremona transformation. As an aside, we also provide an analysis of the distortion in the depth recovered using the least square procedure as compared to the epipolar reconstruction approach.

**Keywords:** Structure from motion, Visual perception, Uncalibrated motion analysis, Depth distortion, Shape representation

---

\*Corresponding author. Tel: 65-874-2290; Fax: 65-779-1103; Email: eleclf@nus.edu.sg

†Email: engp8806@nus.edu.sg

# 1 Introduction

The estimation of the 3-D motion and structure is notorious for its noise sensitivity; a small amount of error in the image measurements can lead to very different solutions. Structure from motion (SFM) algorithms proposed in the past decade faced this problem to varying extent. This has led to many error analysis (Adiv 1989; Daniilidis and Spetsakis 1996; Dutta and Snyder 1990; Thomas et al. 1993; Weng and Huang 1991; Young 1992) in the past. Recently, there have been a number of papers further investigating the error behaviour of SFM algorithms, specifically its local minima and ambiguities (Oliensis 1999; Soatto and Brockett 1998). The view has also been expressed (Oliensis 2000) that since current SFM algorithms perform well only in restricted domains, and different types of algorithms do well on quite different types of sequences, it was important to evaluate the limits of applicability of these algorithms. That is, each algorithm should be evaluated specifically against likely problem conditions. If such understanding could be achieved, it then becomes possible to fuse the results of several algorithms, and might even be the best strategy.

The main concern of these previous approaches seems to be on the reliability of the motion estimates. The corresponding problem of the reliability of the depth estimates has been dealt with to a lesser extent. While some of the works (Weng and Huang 1991; Szeliski and Kang 1997; Grossmann and Victor 2000) predicted the sensitivity of the depth estimates to small amounts of image noise, the situation where the errors in the depth estimates arise from the erroneous 3-D motion parameters has not been dealt with, except in the case of critical surface pairs (Horn 1987; Negahdaripour 1989). In the case of uncalibrated motion, the projective transformation is used to characterize the effect of unknown intrinsic parameters on the recovered depth. Again, the general behaviour of the distorted depths arising from errors in these camera parameters are not dealt with, except for the case of special motions termed as critical motions (Kahl 1999; Sturm 1997). The need to characterize such depth distortion arising from errors in the motion estimates prompted the work of Cheong et al. (Cheong et al. 1998), which gave an account of the systematic nature of the errors in the depth estimates via the so-called iso-distortion framework. It showed that under this case, the transformation from physical to perceptual space is more complicated than that of the projective transformation, and belongs to the family of Cremona transformations.

The results of (Cheong et al. 1998) not only corroborate the view that recovery of metrical depth information is in general very difficult, but also show that even recovery of partial depth

such as ordinal depth information might not be possible under all situations. The latter result lends support to Oliensis’ view (Oliensis 2000) that SFM algorithms perform well only in restricted domains—in this case it is possible to perform partial depth recovery under lateral motion—and that it is important to evaluate the limits of applicability of these algorithms. We adopted this viewpoint, and this paper attempts to characterize the reliability of depth recovery under different motion-scene configurations. In particular, we made use of the iso-distortion framework put forth in (Cheong et al. 1998) to investigate motion types that allow robust recovery of depth information.

The current work differs from the works of (Oliensis 1999; Soatto and Brockett 1998) since it deals with the reliability of the depth estimates rather than that of the motion estimates. In general, the reliability of a reconstructed scene might have quite a different behaviour from that of the motion estimates. For instance, if the motion contains dominant lateral translation, it might be very difficult to lift the ambiguity between translation and rotation. However, in spite of such ambiguity, certain aspect of depth information seems recoverable with robustness. Indeed, in the biological world, lateral motions are often executed to judge distance and relative ordering. On the other hand, psychophysical experiments (Ullman 1979) reported that under pure forward translation, human subjects were unable to recover structure unless favorable conditions such as large field of view exist. Thus it seems that not all motions are equal in terms of robust depth recovery and that there exists certain dichotomy between forward and lateral translation. In the case of uncalibrated motion, in spite of uncertainty in the focal length which further compounds the recovery of motion parameters, certain qualitative aspect of the recovered depth such as parallelism seemed not to be affected (Bougnoux 1998; Cheong and Peh 2000). Thus it is important to treat the question of the reliability of depth reconstruction in its own right. In addition, this would allow us to address the fundamental issue: Do we need to perform calibration of intrinsic parameters in order to recover certain aspects of depth robustly?

If we understand the reliability of the depth estimates under different motion-scene configurations, we can design good motion strategy to reveal reliable depth information. The idea of executing intelligent controlled movements so as to accomplish tasks robustly is of course the central tenet of the active vision paradigm. While there have been many motion-based works under this paradigm, we find that most of them dealt with problems whose purpose is to perform robust navigation. For instance, Santos-Victor et al. (Santos-Victor et al. 1993) and Coombs and Roberts (Coombs and Roberts 1993) present methods to steer a camera between two walls, and to veer

around obstacles, both methods being based on simple analysis of the optical flow without going through depth recovery. Much less analysis has been conducted on how to execute movements so as to recover interesting structure information (besides that used for avoiding obstacles in navigation). While Chaumette et al. (Chaumette et al. 1994) dealt with the optimal estimation of 3-D structures using visual servoing, the errors they dealt with concerned only discretization and measurement errors, and the analysis was applied only to specific shape primitives such as spheres and cylinders. In this paper, we address depth recovery under the case where the 3-D motion parameters themselves are estimated with some errors and where the scene in view is arbitrary in shape.

In the face of such errors, what motion strategy should we adopt to recover robust depth information? If self motions can be controlled perfectly or if there are no other constraints, then of course any pure translation would be a good motion strategy. However in the case where rotation often accompanies translation, such as in the case of the human visual system, the rotation inevitably confounds the recovery of the translation. The question becomes: What *kind* of translation is the best strategy, given the mechanical constraints of the visual systems? Or consider the case where the camera is equipped with a zoom lens and the focal length can be freely varying across frames, or in the human visual system where the intrinsic parameters vary from fovea to the periphery, and also with time and external factors. While the estimation of the intrinsic parameters from the direct observation of the environment is a solvable problem, most approaches face numerical difficulties in estimating these parameters accurately. Given some errors in these intrinsic parameters, how would this affect our choice of motion strategy? Elucidating these questions forms the subject of this paper. In view of the various results that seem to imply that not all motions are equal in terms of depth recovery, we seek to use the iso-distortion framework to analyze the nature of the depths recovered under two major types of motions—lateral and forward motion. In this paper, we use three parameters to describe the intrinsic optical parameters, namely, the focal length of the optical sensor and the principal point position. We consider the case where these parameters are unknown but fixed and the case where they are varying. Lastly, in the appendix, we consider the effects that different schemes of recovering depth (in particular, the least square procedure and the epipolar reconstruction approach) would have on the geometries in the attendant depth distortions.

The organization of this paper is as follows. First, we briefly review in Section 2 the iso-distortion framework requisite for subsequent analysis in this paper. Then, in Section 3, we consider various aspects of depth recovery under generic types of calibrated motion. This is followed by

similar analyses for the case of uncalibrated motion in Section 4. In particular, we first look at the perceived space obtained with inaccurate estimates of the fixed focal length  $f$  and the fixed principal point  $(O_x, O_y)$ . Then we allow for a dynamically changing focal length which results in a zoom field and a changing principal point in the motion recovery process. In Section 5, we conduct experiments to verify the various theoretical predictions. The paper ends with a conclusion of the work and an appendix comparing the different distortion geometries resulting from different methods of depth recovery.

## 2 Iso-distortion Framework

In (Cheong et al. 1998), the geometric laws under which the recovered scene is distorted due to some errors in the estimated motion parameters is represented by a distortion transformation. The distortion in the perceived space can then be visualized by looking at the locus of constant distortion. This approach was termed the iso-distortion framework.

We adopt the standard perspective image formation model. A camera is moving rigidly with respect to a coordinate system  $OXYZ$  fixed to its nodal point  $O$  with a translation  $(U, V, W)$  and a rotation  $(\alpha, \beta, \gamma)$ ; the image plane is located at a focal length  $f$  pixels from  $O$  along the  $Z$ -axis; a point  $P$  at  $(X, Y, Z)$  in the world produces an image point  $p$  at  $(x, y)$  on the image plane where  $(x, y)$  is given by  $(\frac{fX}{Z}, \frac{fY}{Z})$ . The resulting optical flow  $(u, v)$  at an image location  $(x, y)$  can then be expressed with the following well-known equations (Longuet-Higgins 1981):

$$\begin{aligned}
 u &= u_{trans} + u_{rot} \\
 &= (x - x_0) \frac{W}{Z} + \frac{\alpha xy}{f} - \beta \left( \frac{x^2}{f} + f \right) + \gamma y \\
 v &= v_{trans} + v_{rot} \\
 &= (y - y_0) \frac{W}{Z} + \alpha \left( \frac{y^2}{f} + f \right) - \frac{\beta xy}{f} - \gamma x
 \end{aligned} \tag{1}$$

where  $(x_0, y_0) = (f \frac{U}{W}, f \frac{V}{W})$  is the focus of expansion (FOE),  $Z$  is the depth of a scene point,  $u_{trans}, v_{trans}$  are the horizontal and vertical components of the flow due to translation, and  $u_{rot}, v_{rot}$  the horizontal and vertical components of the flow due to rotation, respectively.

Since the depth can only be derived up to a scale factor, we can set  $W = 1$  without loss of generality. Then the scaled depth of a scene point recovered can be written as

$$Z = \frac{(x - x_0, y - y_0) \cdot (n_x, n_y)}{(u - u_{rot}, v - v_{rot}) \cdot (n_x, n_y)} \quad (2)$$

where  $(n_x, n_y)$  is a unit vector which specifies a direction.

If there are some errors in the estimation of the extrinsic parameters, this will in turn cause errors in the estimation of the scaled depth, and thus a distorted version of the space will be computed. Denoting the estimated parameters with the hat symbol ( $\hat{\cdot}$ ) and errors in the estimated parameters with the subscript  $e$  (where error of any estimate  $r$  is defined as  $r_e = r - \hat{r}$ ), the estimated depth  $\hat{Z}$  can be readily shown to be related to the actual depth  $Z$  as follows:

$$\hat{Z} = Z \left( \frac{(x - \hat{x}_0, y - \hat{y}_0) \cdot (n_x, n_y)}{(x - x_0, y - y_0) \cdot (n_x, n_y) + (u_{rot_e}, v_{rot_e}) \cdot (n_x, n_y) Z + (u_e, v_e) \cdot (n_x, n_y) Z} \right) \quad (3)$$

where  $(u_e, v_e)$  is a noise term representing error in the estimate for the optical flow. In the forthcoming analysis we do not attempt to model the statistics of the noise and we will therefore ignore the noise term, that is,  $(\hat{u}, \hat{v}) = (u, v)$ .

From (3) we can see that  $\hat{Z}$  is obtained from  $Z$  through multiplication by a factor given by the terms inside the bracket, which we denote by  $D$  and call the distortion factor. The expression for  $D$  contains the term  $(n_x, n_y)$  which in this paper does not necessarily refer to the image intensity gradient direction. Its value depends on the scheme we use to recover depth. For instance, the normal flow approach (Fermüller 1995) recovers depth along the normal direction in which case  $(n_x, n_y)$  is the gradient direction. In the optical-flow based approach, however, a possible scheme is to recover depth along the estimated epipolar direction, based on the intuition that the epipolar direction contains the strongest translational flow. It means that we first project optical flow along the direction emanating from the estimated FOE and then recover depth along that direction, i.e.  $(n_x, n_y) = \frac{(x - \hat{x}_0, y - \hat{y}_0)}{\sqrt{(x - \hat{x}_0)^2 + (y - \hat{y}_0)^2}}$ , or in the case of  $\hat{W} = 0$  where the estimated FOE is at infinity,  $(n_x, n_y) = -\frac{(\hat{U}, \hat{V})}{\sqrt{\hat{U}^2 + \hat{V}^2}}$ . In the forthcoming analysis, we will study the properties of the recovered depth based on the epipolar reconstruction approach. Another important alternative of recovering depth, which we do no more than performing a brief analysis in this paper, is the linear least square

reconstruction approach where  $(n_x, n_y) = \frac{(u-u_{\hat{rot}}, v-v_{\hat{rot}})}{\sqrt{(u-u_{\hat{rot}})^2+(v-v_{\hat{rot}})^2}}$ . See the appendix for the derivation of this expression, as well as the statistical and geometrical reasons for not choosing this scheme of reconstructing depth as the main focus of the present study.

Upon substituting the corresponding value of  $(n_x, n_y)$  for the case of epipolar reconstruction approach, we obtain the following expression for the distortion factor:

$$D = \frac{(x - \hat{x}_0)^2 + (y - \hat{y}_0)^2}{(x - x_0, y - y_0) \cdot (x - \hat{x}_0, y - \hat{y}_0) + (u_{rot_e}, v_{rot_e}) \cdot (x - \hat{x}_0, y - \hat{y}_0) Z} \quad (4)$$

For specific values of the parameters  $x_0, y_0, \hat{x}_0, \hat{y}_0, \alpha_e, \beta_e,$  and  $\gamma_e$ , and for any fixed distortion factor  $D$ , equation (4) describes a surface  $f(x, y, Z) = 0$  in the  $xyZ$ -space, which we call an iso-distortion surface. This iso-distortion surface has the obvious property that points lying on it are distorted in depth by the same multiplicative factor  $D$ . The systematic nature of the distortion can then be made clear by looking at the organization of these iso-distortion surfaces. Sometimes to facilitate the pictorial description of these surfaces, we slice them with planes parallel to either the  $xZ$ -plane or the  $xy$ -plane. We call the curves thus obtained on the planar slice the iso-distortion contours. Some examples are plotted in Figures 1(a) and 1(b); these plots show that the distortion in general is far more complicated than usually expected.

**Figure 1 here**

Algebraically, it was shown from (Cheong and Ng 1999) that the transformation from physical to perceptual space belongs to the family of Cremona transformations. We recapitulate some notations that will be useful for this paper. We denote the homogeneous co-ordinates of a point  $P^3$  by  $[\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \mathcal{W}]$ , which is related to the non-homogeneous co-ordinates  $(X, Y, Z)$  by  $(X, Y, Z) = \left[ \frac{\mathcal{X}}{\mathcal{W}}, \frac{\mathcal{Y}}{\mathcal{W}}, \frac{\mathcal{Z}}{\mathcal{W}}, 1 \right]$ . Denoting the homogeneous co-ordinates of the estimated position  $\hat{P}^3$  by  $[\hat{\mathcal{X}}, \hat{\mathcal{Y}}, \hat{\mathcal{Z}}, \hat{\mathcal{W}}]$ , we look for a distortion transformation  $\phi : P^3 \rightarrow \hat{P}^3$ . Note that to obtain the estimated  $\hat{X}$ , we use the back-projection given by  $\hat{X} = \frac{x\hat{Z}}{f} = D \frac{xZ}{f} = DX$ ; similarly,  $\hat{Y} = DY$ . The image  $[\hat{\mathcal{X}}, \hat{\mathcal{Y}}, \hat{\mathcal{Z}}, \hat{\mathcal{W}}]$  of a point  $[\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \mathcal{W}]$  can then be expressed as follows:

$$[\hat{\mathcal{X}}, \hat{\mathcal{Y}}, \hat{\mathcal{Z}}, \hat{\mathcal{W}}] = [\phi_1, \phi_2, \phi_3, \phi_4]$$

Similarly, the inverse transformation  $\phi^{-1} : \hat{P}^3 \rightarrow P^3$  can be expressed as:

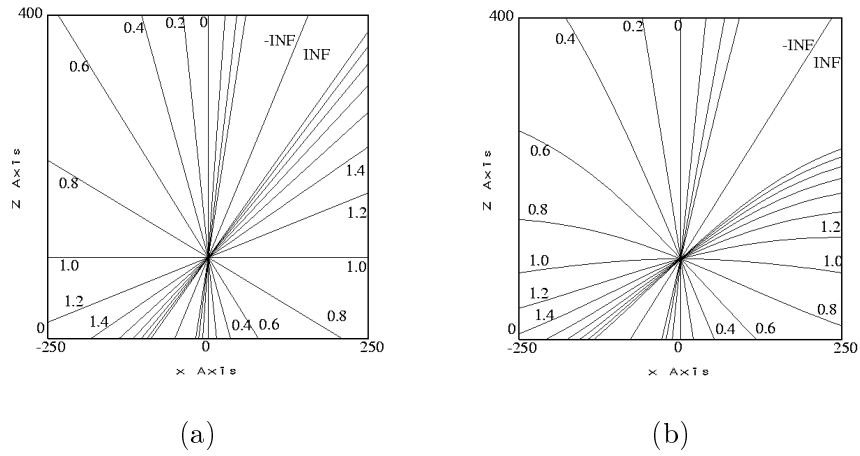


Figure 1: Families of iso-distortion contours in the  $xZ$ -plane, parameterized by  $x_0, y_0, \hat{x}_0, \hat{y}_0, y, \alpha_e, \beta_e$  and  $\gamma_e$ . The number beside each contour denotes the distortion factor  $D$  of that contour. INF denotes  $\infty$ . 1(b) illustrates the effects of including second order terms on the iso-distortion contours of 1(a), with  $\text{FOV}=50^\circ$ .

$$[\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \mathcal{W}] = [\phi_1^{-1}, \phi_2^{-1}, \phi_3^{-1}, \phi_4^{-1}]$$

where the quantities  $\phi_i$  are homogeneous polynomials in  $[\mathcal{X}, \mathcal{Y}, \mathcal{Z}, \mathcal{W}]$  and  $\phi_i^{-1}$  are homogeneous polynomials in  $[\hat{\mathcal{X}}, \hat{\mathcal{Y}}, \hat{\mathcal{Z}}, \hat{\mathcal{W}}]$ . In general, these homogeneous polynomials are of degree greater than one. The resulting transformation  $\phi$  is a Cremona transformation; such transformation is bijective almost everywhere except on the set of fundamental elements where all the  $\phi_i$ 's vanish, under which the correspondence is one-to-many. However, under some special cases, the transformation may reduce to that of a projective transformation, in which case the homogeneous polynomials  $\phi_i$  and  $\phi_i^{-1}$  are of degree one.

Even from such brief geometric and algebraic analyses, it is clear that in general it is very difficult to recover metric depth accurately. What is less clear is the feasibility of recovering some of the less metrical depth representations often argued by researchers. For instance, the ordinal representation of depth constitutes one such reduced representation of depth. In many cases, knowing that ordinal depth is preserved is enough for us to carry out some visual tasks. Unfortunately, the distortion equation in the most general case (as illustrated in Figure 1) shows that it may not be possible to recover even ordinal relationships under all situations. Nevertheless, in the ensuing sections, we shall show that there exists generic motions that allow robust recovery of partial depth information. In particular, we shall show that when translation is coupled with rotation, with known or unknown intrinsic parameters, lateral motion is better than forward motion in terms of yielding ordinal depth information and other aspects of depth recovery.

Before embarking on such analysis, we would like to make a few reasonable assumptions. Since these generic types of motions are likely to be purposely executed for depth recovery, we expect that the agent executing such motion is at least aware that such generic type of motion is being executed. That is,

- When lateral motion is executed,  $\hat{W} = W = 0$ .
- When forward motion is executed,  $\hat{U} = U = \hat{V} = V = 0$

Furthermore, we make an assumption that will allow us to better grasp the geometrical organization of the iso-distortion surfaces: within a limited field of view, quadratic terms in the image

co-ordinates are small relative to linear and constant terms. This is typically the case when the field of view is small or when the visual system focuses its attention on the foveal region. Furthermore, we assume that the contribution of  $\gamma_e$  is small, so that  $(u_{\text{rot}_e}, v_{\text{rot}_e})$  becomes  $(-\beta_e f, \alpha_e f)$ . In typical visual systems, rotation about the optical axis is usually not executed unless as a result of perturbation. In any case, given their typical magnitudes, these terms do not qualitatively affect the organization of the iso-distortion surfaces (see Figure 1(b)).

### 3 Depth Recovery under Calibrated Motion

#### 3.1 Lateral motion

We derive the distorted depth under lateral motion following the same procedure as in section 2, except that we express the translational parameters in equations (3) in terms of  $U$  and  $V$  to handle the case of FOE at infinity:

$$\hat{Z} = Z \left( \frac{(\hat{U}, \hat{V}) \cdot (n_x, n_y)}{(U, V) \cdot (n_x, n_y) + Z(\beta_e, -\alpha_e) \cdot (n_x, n_y)} \right) \quad (5)$$

For the epipolar reconstruction scheme of recovering depth, since the estimated FOE lies in the infinity, all  $(n_x, n_y)$  will be in the same direction given by  $-\frac{(\hat{U}, \hat{V})}{\sqrt{\hat{U}^2 + \hat{V}^2}}$ . For notational convenience, we can set this  $(n_x, n_y)$  to be  $(1, 0)$  via a rotation of the  $x$ - and  $y$ -axes without loss of generality (it can be easily shown that even without such a change in the coordinate system, the distortion expression obtained has identical form). After this simplification, we obtain the distortion factor as follows:

$$D = \frac{\hat{U}}{U + Z\beta_e} \quad (6)$$

where  $U, \hat{U}$  and  $\beta_e$  are understood to be the corresponding quantities in the rotated coordinate system. Thus the equation of the iso-distortion surface is:

$$Z = \frac{1}{D} \frac{\hat{U}}{\beta_e} - \frac{U}{\beta_e}$$

which represents plane parallel to the image plane.

## Figure 2 here

Figure 2 depicts how the perceived space is distorted. It shows that there exists a  $D = 1$  iso-distortion surface which divides the whole space into two parts: one in which the space is expanded ( $D > 1$ ) and the other in which the space is compressed ( $D < 1$ ). Its equation is given by  $Z = \frac{\hat{U}-U}{\beta_e}$ . Whether the  $D$  values increase or decrease with  $Z$  depends on the sign of  $\beta_e$ . However, we shall show later that in both cases, we are able to recover the ordinal depth, provided that we take proper care of the sign. An estimated depth will be negative if it falls between the region bounded by the  $D = 0$  and the  $D = -\infty$  surfaces. In this case, the  $D = 0$  surface is always located at infinity as its equation is given by  $Z = \pm\infty$ , and the  $D = \pm\infty$  surface is located at  $Z = -\frac{U}{\beta_e}$ .

### 3.2 Forward motion

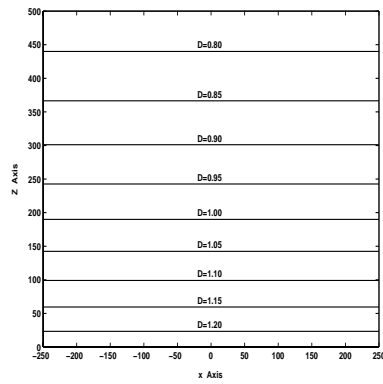
For the case of forward motion, we again make use of the assumptions stated at the end of last section. Conducting “epipolar reconstruction”, the direction  $(n_x, n_y)$  can be expressed as  $\frac{(x,y)}{\sqrt{x^2+y^2}}$ . Substituting into equation (3), we obtain:

$$\hat{Z} = Z \left( \frac{x^2 + y^2}{x^2 + y^2 + Z(-\beta_e f, \alpha_e f) \cdot (x, y)} \right) \quad (7)$$

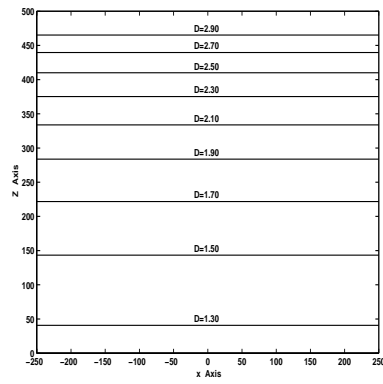
or expressing  $\hat{Z}$  in terms of  $DZ$ , the above equation can be expressed as:

$$x^2 + y^2 + \left( \frac{DZf}{D-1} \right) (-\beta_e x + \alpha_e y) = 0 \quad (8)$$

For a particular value of  $D$ , the corresponding iso-distortion surface is a cone. The  $D = \pm\infty$  surface is of special interest as all other region in space where  $D$  is negative is encompassed by the cone formed by this  $D = \pm\infty$  surface. This negative volume is illustrated schematically in Figure 3(a). If we slice these cones with planes parallel to the image plane, we obtain a family of circles, each with center at  $\left( \frac{DZf\beta_e}{2(D-1)}, -\frac{DZf\alpha_e}{2(D-1)} \right)$  and radius as  $\frac{1}{2} \left| \frac{D}{D-1} \right| Zf \sqrt{\beta_e^2 + \alpha_e^2}$ . It can also be shown readily that all  $D$  surfaces intersect on a common line, which is the  $Z$ -axis (see Figure 3(a)). In other word, on this line, the distortion factor is undefined, or equivalently, the  $Z$ -axis is the fundamental element of the distortion transformation.



(a)



(b)

Figure 2: Families of iso-distortion contours for lateral motion in calibrated case. The parameters are:  $U = 0.81$ ,  $\hat{U} = 1.0$ , and  $\beta_e = 0.001$  for (a) and  $\beta_e = -0.001$  for (b).

### Figure 3 here

If we further intersect these cones with planes parallel to the  $xZ$ -plane, we obtain the iso-distortion contours as shown in Figures 3(b), (c) and (d), respectively for the cases of  $y = 0$ ,  $y = 50$ , and  $y = -50$ . Several salient features can be identified from the plot:

- 1) As far as depth reconstruction is concerned, the key observation here is that, unlike the case of the lateral motion, the value of the distortion factor depends on the image co-ordinate position, thus giving rise to difficulty in depth reconstruction.
- 2) In terms of depth recovery, what the fundamental element  $Z$ -axis means is that around this region, the distortion factor changes value rapidly in a small neighborhood (Figure 3(b)), resulting in poor depth estimates that are not even locally smooth. As we move away from the fundamental element (e.g. Figures 3(c) and (d)), the distortion contours no longer intersect together. Nevertheless, as  $Z \rightarrow \infty$ , the contours approach asymptotically towards the line  $x = \frac{\alpha_e}{\beta_e}y$ , again resulting in rapidly changing distortion values (on the line itself,  $D$  happens to have the value of one in this case). The size of the region where distortion changes rapidly depends on the magnitude of  $\alpha_e, \beta_e$ . In the limiting case where  $\alpha_e, \beta_e$  approach zero, this region shrinks to the asymptotic line itself.
- 3) For forward motion, the  $D = 1$  surface always coincides with the plane  $Z = 0$ . This means that metrical values of depths in the near ground can be judged relatively accurately, which is not necessarily the case for lateral motion.

### 3.3 Ordinal depth

Looking at the specific case of the lateral motion, the distortion factor expressed in equation (6) has the form  $\frac{1}{a+bZ}$ , where  $a = \frac{U}{V}$  and  $b = \frac{\beta_e}{U}$  are constants for all the scene points. Such distortion has the effect of generating a relief transformation which has some nice properties (Koenderink and Van Doorn 1995). In particular, consider two points in space with depths  $Z_1 > Z_2$ . It can be shown that, given the following conditions:

$$(a + bZ_1)(a + bZ_2) > 0 \text{ if } a > 0$$

$$(a + bZ_1)(a + bZ_2) < 0 \text{ if } a < 0$$



the transformation preserves the depth order of the two points, that is,  $\hat{Z}_1 > \hat{Z}_2$ . Since  $a = \frac{U}{\hat{U}}$  here, the condition  $a > 0$  means that  $U$  and  $\hat{U}$  have the same sign. This condition can easily be met by most visual systems; thus we can just focus on the first condition. The requirement  $(a + bZ_1)(a + bZ_2) > 0$  simply means that the two estimated depths have the same sign. However, even if the two estimated depths have different signs, we know that if  $a > 0$ , it is the greater depth whose estimate will have a negative sign, which means that we can always restore the correct depth order under all conditions.

If we take into account the full rotational error flow ( $\gamma_e$  and the second order terms), then the  $b$  term in the transformation  $\frac{1}{a+bZ}$  is no longer constant. What this means is that global ordinality is no longer preserved; we can only obtain ordinality within a neighborhood where  $b$  can be approximately treated as constant (the size of this neighborhood depends on the size of the motion errors, the respective image co-ordinates, and the depth differences). This means that even with lateral motion, global ordinal depth information may not be obtainable over a large field of view, or if motion perturbation results in unaccounted-for rotation about the  $Z$ -axis.

For the case of forward motion, even local ordinal depth information is difficult to obtain, as the distortion factor changes value significantly in a local region. Regions near the fundamental element of the distortion transformation (Figures 3(a) and (b)) or near the asymptotic lines illustrated in Figures 3(c) and (d) are particularly susceptible to depth reversal. The size of the neighborhood in which we can determine ordinal relationship is in general small and depends on several factors. If we again consider two points in space with depths  $Z_1 > Z_2$ , we found that given the following condition, the depth order will be preserved:

$$\frac{Z_1 - Z_2 + (b_2 - b_1)Z_1Z_2}{(1 + b_1Z_1)(1 + b_2Z_2)} > 0$$

where  $b_1 = \left(\frac{(-\beta_e f, \alpha_e f) \cdot (x_1, y_1)}{x_1^2 + y_1^2}\right)$  and  $b_2 = \left(\frac{(-\beta_e f, \alpha_e f) \cdot (x_2, y_2)}{x_2^2 + y_2^2}\right)$ . From this expression, we can only say that, in general, the more we approach towards the image periphery, the size of this ordinal neighborhood increases (this is also reflected in Figure 3, where the iso-distortion contours become more parallel to the image plane near the periphery, i.e. less dependent on the image co-ordinates). However, this increase in size in the image space probably does not signify much in the 3-D space due to the large perspective effect at the periphery of the image plane.

### 3.4 Distortion Transformation

Under lateral motion, we obtain from equation (6) linear expressions in  $\phi_i$ 's; thus the distortion transformation can be expressed with the following matrix:

$$\begin{bmatrix} \hat{\mathcal{X}} \\ \hat{\mathcal{Y}} \\ \hat{\mathcal{Z}} \\ \hat{\mathcal{W}} \end{bmatrix} = \begin{bmatrix} \phi_1 \\ \phi_2 \\ \phi_3 \\ \phi_4 \end{bmatrix} = \begin{bmatrix} \hat{U} & 0 & 0 & 0 \\ 0 & \hat{U} & 0 & 0 \\ 0 & 0 & \hat{U} & 0 \\ 0 & 0 & \beta_e & U \end{bmatrix} \begin{bmatrix} \mathcal{X} \\ \mathcal{Y} \\ \mathcal{Z} \\ \mathcal{W} \end{bmatrix} \quad (9)$$

It is obvious that the inverse transformation can be expressed as a matrix with similar form.

As can be seen, the original complex Cremona transformation has now reduced to an invertible projective transformation. Furthermore most of the elements of the matrix representing the transformation are zero, which means that it is really a “well-behaved” kind of projective transformation. In particular, the tilt of a surface, which represents the ordinal aspect of depth information, is preserved although the slant is not. Furthermore, looking at the matrix, if the term  $\beta_e$  in the last row approaches zero, the transformation will tend to preserve the plane at infinity (i.e. it is an affine transformation), in which case all the first order and second order shapes are preserved. In general, the nice properties of approaching such an affine transformation close enough may indeed be sufficient for most vision systems.

For the case of forward motion, we obtain from equation (7) homogeneous polynomials  $\phi_i$  of degree three, given by:

$$\begin{aligned} \hat{\mathcal{X}} &= \phi_1 = (\mathcal{X}^2 + \mathcal{Y}^2) \mathcal{X} \\ \hat{\mathcal{Y}} &= \phi_2 = (\mathcal{X}^2 + \mathcal{Y}^2) \mathcal{Y} \\ \hat{\mathcal{Z}} &= \phi_3 = (\mathcal{X}^2 + \mathcal{Y}^2) \mathcal{Z} \\ \hat{\mathcal{W}} &= \phi_4 = (\mathcal{X}^2 + \mathcal{Y}^2) \mathcal{W} + (-\beta_e \mathcal{X} + \alpha_e \mathcal{Y}) \mathcal{Z}^2 \end{aligned}$$

Under this transformation, we can only say that a general element is distorted into an element of the same nature: a point remains as a point, a surface remains as a surface, and a curve remains as a curve. By general element, we mean that the element does not contain any fundamental elements (in this case the  $Z$ -axis). If an element is not general, then a point may blow up into a plane, or a plane may reduce to a line under such a transformation.

## 4 Depth Recovery under Uncalibrated Motion

### 4.1 Intrinsic Parameters Unknown but Fixed

The true optical flow can be expressed in the following form to take into account intrinsic parameters:

$$\begin{aligned} u &= \frac{1}{Z} ((x_s - O_x)W - fU) - \beta f + \gamma(y_s - O_y) + O_u^2(x_s, y_s) \\ v &= \frac{1}{Z} ((y_s - O_y)W - fV) + \alpha f - \gamma(x_s - O_x) + O_v^2(x_s, y_s) \end{aligned}$$

where  $(x_s, y_s)$  represents the image pixel location in a new co-ordinate system with origin located at the lower left corner of the image,  $(O_x, O_y)$  is the location of the principal point in the new co-ordinate system, and  $O_u^2(x_s, y_s), O_v^2(x_s, y_s)$  represent second order terms in  $(x_s, y_s)$ . Note that  $(x_s, y_s)$  is related to  $(x, y)$  by  $(x, y) = (x_s - O_x, y_s - O_y)$ .

In the uncalibrated case, both the focal length and the location of the principal point are unknown or estimated with error. We denote the estimated focal length and the estimated principal point as  $\hat{f}$  and  $(\hat{O}_x, \hat{O}_y)$  respectively.

If the influence of the second order terms and  $\gamma_e$  is ignored, we can rewrite the iso-distortion factors for the case of lateral motion as:

$$D = \frac{\hat{f}\hat{U}}{fU + \widehat{\beta}_f Z} \quad (10)$$

where  $U, \hat{U}$  and  $\beta_e$  are again quantities in the rotated coordinate system. For the case of forward motion, we have:

$$D = \frac{x^2 + y^2}{(x'x + y'y) + \left(-\widehat{\beta}_f x + \widehat{\alpha}_f y\right) Z}$$

where  $\left(\widehat{\beta}_f, \widehat{\alpha}_f\right) = \left(\beta f - \hat{\beta}\hat{f}, \alpha f - \hat{\alpha}\hat{f}\right)$ , and  $(x', y') = (x - O_{x_e}, y - O_{y_e})$ .

It can be seen from equation (10) that the error in the principal point estimate has no impact on the depth reconstruction for the case of lateral motion at all, while error in the focal length estimate

alters some constant parameters in the expression for the iso-distortion factor without changing its form. The upshot is that all the previous results regarding ordinal depth is still applicable and that the distortion transformation is still a projective transformation with equation (9) revised as:

$$\begin{bmatrix} \hat{\mathcal{X}} \\ \hat{\mathcal{Y}} \\ \hat{\mathcal{Z}} \\ \hat{\mathcal{W}} \end{bmatrix} = \begin{bmatrix} \hat{f}\hat{U} & 0 & 0 & 0 \\ 0 & \hat{f}\hat{U} & 0 & 0 \\ 0 & 0 & \hat{f}\hat{U} & 0 \\ 0 & 0 & \widehat{\beta}_f & fU \end{bmatrix} \begin{bmatrix} \mathcal{X} \\ \mathcal{Y} \\ \mathcal{Z} \\ \mathcal{W} \end{bmatrix}$$

For the case of forward motion, we again resort to the iso-distortion plot to visualize the distortion. Figure 4 shows that not much difference from Figure 3 can be found. Each iso-distortion surface is still a cone:

$$x^2 + y^2 + \frac{D}{D-1} \left( (-Z \widehat{\beta}_f - O_{xe})x + (Z \widehat{\alpha}_f - O_{ye})y \right) = 0 \quad (11)$$

If we slice these cones with planes parallel to the image plane, we obtain a family of circles, each with center at  $\left( \frac{D(O_{xe} + Z \widehat{\beta}_f)}{2(D-1)}, \frac{D(O_{ye} - Z \widehat{\alpha}_f)}{2(D-1)} \right)$  and radius equal to  $\frac{1}{2} \left| \frac{D}{D-1} \right| \sqrt{(O_{xe} + Z \widehat{\beta}_f)^2 + (O_{ye} - Z \widehat{\alpha}_f)^2}$ . Slicing these cones with planes parallel to the  $xZ$ -plane yields the iso-distortion contours as illustrated in Figure 4(b). The distortion transformation remains a Cremona one; we will not show the expressions for its homogeneous polynomials  $\phi_i$  and  $\phi_i^{-1}$  here.

**Figure 4 here**

As a whole, we can say that fixed uncalibrated intrinsic parameters may change the distortion factor equations, but they do not alter the essential properties of the distortion for both the cases of lateral motion and forward motion. For the case of lateral motion, depth order is still preserved, given the same conditions stated in the calibrated case. For the case of forward motion, as can be seen from Figure 4, there are regions which exhibit large distortion variation as before and are therefore likely to undergo depth reversal. The  $D = 1$  surface has also shifted away from the  $Z = 0$  plane due to the presence of the term  $(O_{xe}, O_{ye})$ .

## 4.2 Intrinsic Parameters Unknown and Varying

Varying the focal length will result in a zoom field (considering infinitesimal motion) which is hard to separate from a field of translation along the optical axis. Usually a changing focal length will

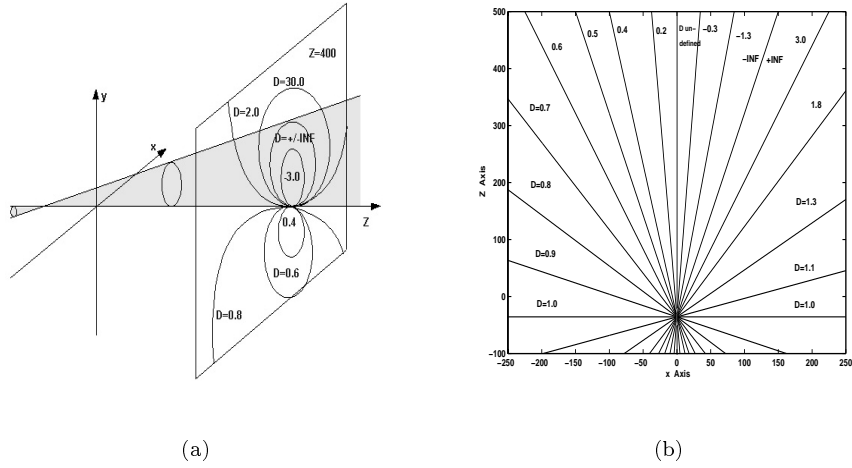


Figure 4: Families of iso-distortion contours for forward motion in uncalibrated case with fixed intrinsic parameters. (a) Schematic representation of the iso-distortion surfaces in the  $xyZ$ -space, as in Figure 3(a). (b) Iso-distortion contours obtained by slicing the iso-distortion surfaces with the  $y = 0$  plane.  $(O_{x_e}, O_{y_e}) = (10, -10)$ ,  $\beta = 0$ ,  $\hat{\beta} = -0.001$ ,  $\alpha = 0$ ,  $\hat{\alpha} = -0.001$ ,  $f = 309.0$ , and  $\hat{f} = 280.0$ .

also be accompanied by a change in the principal point. We will see that in such cases, depth ordinality in the global sense is lost even for the case of lateral motion.

With focal length and principal point variation, the resulting optical flow  $(u, v)$  can be expressed as:

$$\begin{aligned} u &= \frac{1}{Z} ((x_s - O_x)W - fU) - \beta f + \gamma(y_s - O_y) + \dot{O}_x + \frac{\dot{f}}{f}(x_s - O_x) + O_u^2(x_s, y_s) \\ v &= \frac{1}{Z} ((y_s - O_y)W - fV) + \alpha f - \gamma(x_s - O_x) + \dot{O}_y + \frac{\dot{f}}{f}(y_s - O_y) + O_v^2(x_s, y_s) \end{aligned}$$

where  $(\dot{O}_x, \dot{O}_y)$  is the rate of change of the principal point and  $\dot{f}$  is the rate of change of the focal length.

We assume that the principal point and focal length and their corresponding change rates are all estimated with errors. We make use of the following notations:  $(\zeta_{u_e}, \zeta_{v_e}) = (\dot{O}_x - \hat{\dot{O}}_x, \dot{O}_y - \hat{\dot{O}}_y)$  and  $\sigma_e = \frac{\dot{f}}{f} - \left(\frac{\hat{\dot{f}}}{f}\right)$ .

#### 4.2.1 Lateral motion

Again, by an appropriate rotation of the  $xy$ -coordinate system, we can express the iso-distortion factor as follows:

$$D = \frac{\hat{f}\hat{U}}{fU + \left(\widehat{\beta}_f - \zeta_{u_e} + \frac{\dot{f}}{f}O_{xe}\right)Z - \sigma_e x Z} \quad (12)$$

from which the following can be derived:

$$Z = \frac{\hat{f}\hat{U} - DfU}{D\left(\widehat{\beta}_f - \zeta_{u_e} + \frac{\dot{f}}{f}O_{xe} - \sigma_e x\right)}$$

**Figure 5 here**

The presence of an error in the zoom estimate  $\sigma_e$  could significantly change the topological distribution of the iso-distortion surfaces. As illustrated in Figure 5, the iso-distortion contours become reciprocal curves with a common vertical asymptote given by  $x = \frac{\widehat{\beta}_f - \zeta_{u_e} + \frac{\dot{f}}{f}O_{xe}}{\sigma_e}$ . As can be

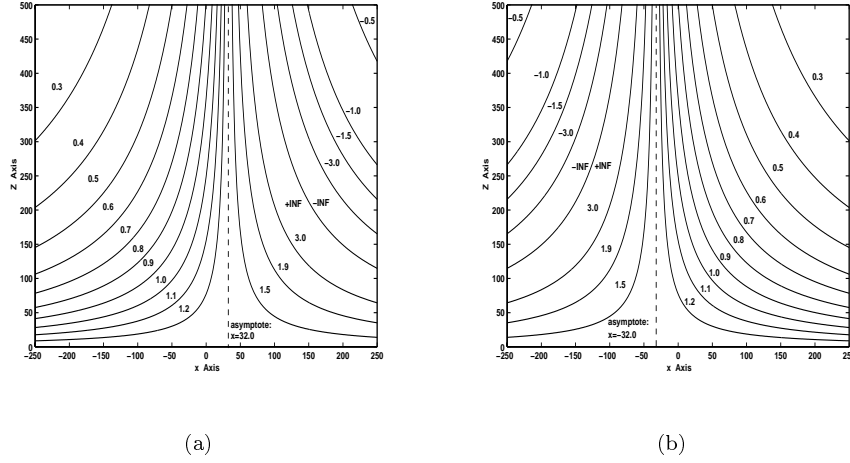


Figure 5: Families of iso-distortion contours for lateral motion with varying intrinsic parameters. The parameters are:  $U = 0.81$ ,  $\hat{U} = 1.0$ ,  $\beta = 0$ ,  $\hat{\beta} = -0.001$ ,  $f = 309.0$ ,  $\hat{f} = 330.0$ ,  $O_{xe} = 10.0$ ,  $\zeta_{ue} = 0.01$ , and  $\dot{f} = 0$ . The two figures illustrate the effects of the sign of  $\sigma_e$ :  $\sigma_e = 0.01$  for (a) and  $\sigma_e = -0.01$  for (b). On the asymptote, the value of  $D$  is  $\frac{\hat{U}}{\hat{f}} = 1.23$ .

seen, the position of this asymptote depends very much on the zoom error term  $\sigma_e$ . This means that any flow due to zoom motion must be estimated accurately for meaningful depth information to be derived from lateral motion. When  $\sigma_e$  approaches zero, the vertical asymptote approaches infinity, and the contours will tend to flatness again.

Algebraically, the distortion transformation can be written as:

$$\begin{bmatrix} \hat{x} \\ \hat{y} \\ \hat{z} \\ \hat{w} \end{bmatrix} = \begin{bmatrix} \hat{f}\hat{U} & 0 & 0 & 0 \\ 0 & \hat{f}\hat{U} & 0 & 0 \\ 0 & 0 & \hat{f}\hat{U} & 0 \\ -f\sigma_e & 0 & \widehat{\beta}_f - \zeta_{u_e} + \frac{\hat{f}}{f}O_{xe} & fU \end{bmatrix} \begin{bmatrix} \mathcal{X} \\ \mathcal{Y} \\ \mathcal{Z} \\ \mathcal{W} \end{bmatrix}$$

It is still a projective transformation with a simple form. However the presence of the  $\sigma_e$  term in the last row of the matrix means that the distortion transformation is no longer a relief transformation (see also equation (12)). Ordinality is only preserved if the image distance between two points satisfies certain inequality, which we derive as follows. Consider two points with true depth  $Z_1$  and  $Z_2$  respectively and that  $Z_1 > Z_2$ . To determine the geometry of this local neighborhood within which ordinal depth is preserved, we need to determine the sign of  $\hat{Z}_1 - \hat{Z}_2$ . However, directly deciding the sign of  $\hat{Z}_1 - \hat{Z}_2$  is difficult; instead we consider  $\frac{1}{\hat{Z}_1} - \frac{1}{\hat{Z}_2}$ . If  $\hat{Z}_1$  and  $\hat{Z}_2$  have same sign, then if  $\frac{1}{\hat{Z}_1} - \frac{1}{\hat{Z}_2} < 0$ , we can say that the depth order in the perceived space is preserved.

$$\frac{1}{\hat{Z}_1} - \frac{1}{\hat{Z}_2} = \frac{f\frac{U}{\hat{U}}(Z_2 - Z_1) + \frac{\sigma_e}{\hat{U}}(x_2 - x_1)Z_1Z_2}{fZ_1Z_2}$$

Since the denominator on the right hand side of the above equation is positive, the condition needed for preservation of depth ordinality can be expressed as:

$$\frac{\sigma_e}{\hat{U}}(x_2 - x_1) < -f\frac{U}{\hat{U}}\frac{(Z_2 - Z_1)}{Z_1Z_2}$$

If we further make the reasonable assumption that  $\hat{U}$  and  $U$  have same sign, we have:

$$x_2 - x_1 < -f\frac{U}{\sigma_e}\frac{(Z_2 - Z_1)}{Z_1Z_2} \quad \text{if } \frac{\sigma_e}{\hat{U}} > 0 \quad (13)$$

$$x_2 - x_1 > -f\frac{U}{\sigma_e}\frac{(Z_2 - Z_1)}{Z_1Z_2} \quad \text{if } \frac{\sigma_e}{\hat{U}} < 0 \quad (14)$$

In either cases, given two fixed depths whose estimates have same sign, the geometry of the local neighborhood on the image plane is that of a half-plane. Furthermore, it is noted that errors in the principal point and the rotational parameter estimates do not influence the properties of the neighborhood. The simple geometry of this local neighborhood means that if bounds can be given to the various terms found in the inequalities, the region of the neighborhood can be approximated.

#### 4.2.2 Forward motion

The expression for the iso-distortion factor is complex:

$$D = \frac{x^2 + y^2}{(x'x + y'y) + \left( -\widehat{\beta}_f + \zeta_{u_e} - \frac{\dot{f}}{f}O_{x_e}, \widehat{\alpha}_f + \zeta_{v_e} - \frac{\dot{f}}{f}O_{y_e} \right) \cdot (x, y)Z + \sigma_e(x^2 + y^2)Z} \quad (15)$$

Each iso-distortion surface has the following expression:

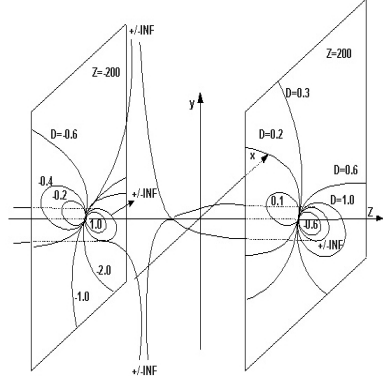
$$x^2 + y^2 + \frac{D}{D + D\sigma_e Z - 1} ((C_1 Z - O_{x_e})x + (C_2 Z - O_{y_e})y) = 0 \quad (16)$$

where  $(C_1, C_2) = \left( -\widehat{\beta}_f + \zeta_{u_e} - \frac{\dot{f}}{f}O_{x_e}, \widehat{\alpha}_f + \zeta_{v_e} - \frac{\dot{f}}{f}O_{y_e} \right)$ .

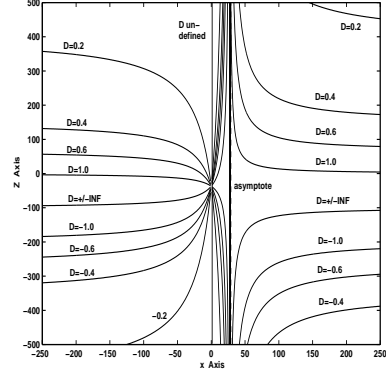
**Figure 6 here**

In the  $xyZ$ -space, each iso-distortion surface is no longer a cone, but a third-order surface, due to the  $\sigma_e$  term. Slicing these iso-distortion surfaces with a frontal-parallel plane would still yield circles, all with one end anchored at the  $Z$ -axis. Each circle has its center at  $\left( -\frac{1}{2} \frac{D(C_1 Z - O_{x_e})}{D + D\sigma_e Z - 1}, -\frac{1}{2} \frac{D(C_2 Z - O_{y_e})}{D + D\sigma_e Z - 1} \right)$  and radius as  $\frac{1}{2} \left| \frac{D}{D + D\sigma_e Z - 1} \right| \sqrt{(C_1 Z - O_{x_e})^2 + (C_2 Z - O_{y_e})^2}$ . As  $Z \rightarrow \infty$ , the circle radius becomes constant, that is, in the  $xyZ$ -space, the iso-distortion surface forms a cylinder. Slicing these iso-distortion surfaces with the  $xZ$ -plane yields the iso-distortion contour plot shown in Figure 6(b)

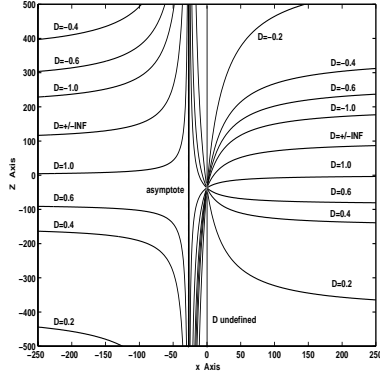
In contrast to the previous cases, each iso-distortion surface now also approaches asymptotically towards the frontal parallel plane given by  $Z = \frac{1-D}{D\sigma_e}$  (see the iso-distortion contour plots in Figure 6). The asymptotic plane of the  $D = \pm\infty$  surface is determined by  $\sigma_e$  only as the plane is given by  $Z = \frac{1}{\sigma_e}$ . Its position has particular significance as it determines the distribution of the positive and negative distortion regions. If the sign of  $\sigma_e$  is positive (Figure 6(b)), then the space in front of the image plane mostly experiences a positive distortion factor, except for the small negative distortion



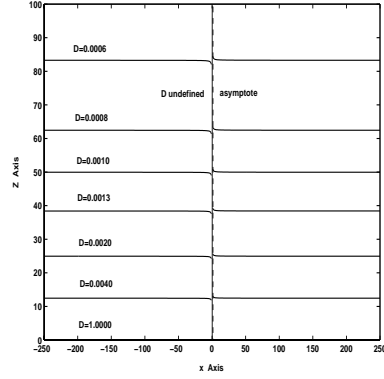
(a)



(b)



(c)



(d)

Figure 6: Distribution of iso-distortion surfaces and contours for forward motion with varying intrinsic parameters. (a) Schematic representation of the iso-distortion surfaces in the  $xyZ$ -space. The region where  $D$  is negative is not shaded due to the complexity of the region. Figure (b) depicts the iso-distortion contours obtained by slicing the iso-distortion surfaces with the  $y = 0$  plane. Figures (c) and (d) illustrate the effects of various values of  $\sigma_e$  on the iso-distortion contours. In (a) and (b),  $\left(\frac{\hat{f}}{f}\right) = -0.01$ . (c) Reverse the sign of  $\sigma_e$ :  $\left(\frac{\hat{f}}{f}\right) = 0.01$ . (d) Large  $\sigma_e$ :  $\left(\frac{\hat{f}}{f}\right) = -0.01$ . Other parameters are the same for all the plot:  $\beta = 0$ ,  $\alpha = 0$ ,  $\hat{\alpha} = -0.001$ ,  $\hat{\beta} = -0.001$ ,  $\frac{\hat{f}}{f} = 0$ ,  $f = 309.0$ ,  $\hat{f} = 280.0$ ,  $(O_{x_e}, O_{y_e}) = (10.0, -10.0)$  and  $(\zeta_{u_e}, \zeta_{v_e}) = (0.01, -0.01)$  for (a) and (b);  $\left(\frac{\hat{f}}{f}\right) = -20.0$  for (c) and (d);  $y = 0$  for (a), (b) and (c) and  $y = 50$  for (d).

region enclosed by the  $D = \pm\infty$  surface. Conversely, if the sign of  $\sigma_e$  is negative (Figure 6(c)), the distortion configuration flips about the  $Z = 0$  plane, and most of the region in front of the image plane would have negative distortion factor. In other words, if we impose the “depth is positive” constraint in our motion estimation algorithm, the sign of the error for the zoom estimate is more likely to be positive than negative.

Figure 6(d) shows that the larger  $\sigma_e$  is, the flatter the iso-distortion contours will be. That is, the distortion factor varies less with the image co-ordinate position. Indeed by letting the  $\sigma_e$  term in equation (15) approach infinity, we obtain  $D = \frac{1}{1 + \sigma_e Z}$  which is a relief transformation. This might seem to suggest that ironically, large error in estimating zoom field parameter will result in more “well-behaved” recovered depths, in the sense of preserving its ordinality. However these recovered depths are very much compressed and 3-D information is to a great extent lost.

## 5 Experiments

This section presents the experiments carried out to support the theoretical findings established in the preceding sections. Specifically, we want to demonstrate that lateral motion is more amenable to preserving ordinal depth and yields a less distorted scene reconstruction.

We used SOFA image sequences<sup>1</sup> for conducting experiments. SOFA is a package of 9 synthetic sequences designed for testing research works in motion analysis. It includes full ground truth on motion and camera parameters. Sequence 1 and 5 (henceforth abbreviated as SOFA1 and SOFA5) were chosen for our experiments, the former depicting a lateral motion and the latter a forward motion. Both of them have an image dimension of  $256 \times 256$  pixels, a focal length of 309 pixels and a field of view of approximately  $45^\circ$ . Camera focal length and principal point were fixed for the whole sequence. The optical flow was obtained using Lucas and Kanade’s method (Lucas 1984), with a temporal window of 15 frames. Depth was recovered for frame 9.

The 3-D scene for SOFA1 consisted of a cube resting on a cylinder (Figure 7(a)). The camera trajectory was a circular route on a plane perpendicular to the world  $Y$ -axis, with constant translational parameter  $(U, V, W) = (0.8137, 0.5812, 0)$  and constant rotational parameters  $(\alpha, \beta, \gamma) = (-0.0203, 0.0284, 0)$ . The resulting motion is a constant lateral translation with quite

---

<sup>1</sup>courtesy of the Computer Vision Group, Heriot-Watt University (<http://www.cee.hw.ac.uk/mtc/sofa>)

significant rotational components. If the observer or the system is aware that a lateral motion is being executed, then  $\hat{W} = 0$ , and the following equation will be used to recover depth:

$$\hat{Z} = \frac{-\hat{f}(\hat{U}, \hat{V}) \cdot (n_x, n_y)}{(u, v) \cdot (n_x, n_y) - (u_{rot}, v_{rot}) \cdot (n_x, n_y) - \left(\frac{\hat{f}}{f}x, \frac{\hat{f}}{f}y\right) \cdot (n_x, n_y)}$$

ignoring terms involving the principal point (since they have little effect on depth reconstruction, we ignore them in the experiments). The erroneous motion estimates were arbitrarily fixed at  $(\hat{U}, \hat{V}) = (1, 0)$  and  $(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) = (-0.0213, 0.0274, 0.001)$ . We further chose the scheme where the depth was recovered along the estimated epipolar directions. Since we estimated  $(\hat{U}, \hat{V})$  to be  $(1, 0)$ , it means that  $(n_x, n_y)$  would be fixed in the horizontal direction  $(1, 0)$ .

### Figure 7 here

We first simulated the case where there are no errors in the intrinsic parameters. Using the erroneous motion estimates, we performed depth reconstruction. The results were illustrated in Figures 7(b) and (c). In Figure 7(b), the recovered depths were depicted using a color coding scheme; cool colors such as deep blue meant that the object points were close to the observer, while warm colors such as red represented points that were far away from the observer. The mapping between the colors and the depth ranges was performed individually for each experiment so as to render the plots readable. In Figure 7(c), the reconstructed 3-D depths were displayed using a 3-D plot viewed from the side. As these figures showed, the reconstruction is good despite significant errors in both the estimates for the translational and the rotational parameters. Depth orders were preserved for most of the feature points, except for those which were probably affected by noise. Figure 7(c) also showed that although all the recovered depths were under-estimated, they tend to be under-estimated more at the points which have larger physical depths. This phenomena can be explained by the iso-distortion contours in Figure 2, where it was shown that the value of the iso-distortion factor decreases with depth.

Next we simulated the case where the intrinsic parameters are fixed and estimated with errors. Adding an error to the focal length ( $\hat{f} = 330.0$ ) yielded Figure 7(d), depicting the case of fixed intrinsic parameters estimated with errors. No significant difference can be observed from that of Figure 7(b); this corroborates our theoretical prediction that an erroneous focal length will not influence the distortion properties for the case of lateral motion. Finally, we simulated the case

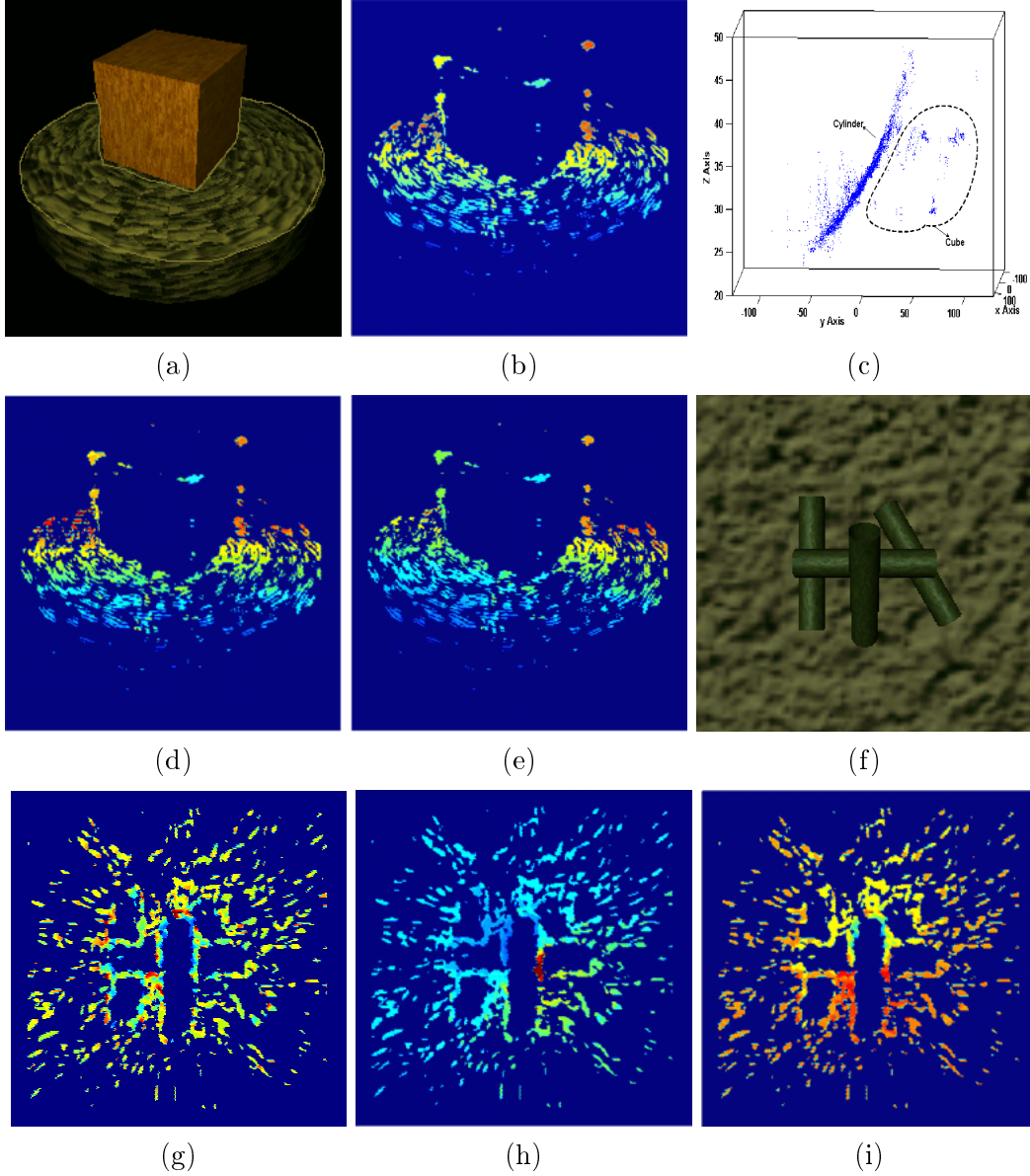


Figure 7: Motion sequences and depth reconstructions for SOFA1 ((a) to (e)) and SOFA5 ((f) to (i)). (a) SOFA1 frame 9. (b), (c) Reconstruction with true focal length and zoom parameter (i.e. no zoom). (d) Reconstruction with erroneous focal length  $\hat{f} = 330.0$  and true zoom parameter. (e) Reconstruction with true focal length and an erroneous zoom field of  $\hat{\dot{z}} = -0.01$ . Other parameters were the same throughout (b) to (e):  $(\hat{U}, \hat{V}, \hat{W}) = (1, 0, 0)$  and  $(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) = (-0.0213, 0.0274, 0.001)$ . (f) SOFA5 frame 9. (g) Reconstruction with accurate motion parameters. (h) Reconstruction with errors only in the extrinsic parameters. (i) Reconstruction with erroneous focal length  $\hat{f} = 280.0$  and a large erroneous zoom field of  $\hat{\dot{z}} = -20.0$ . (h) and (i) have the same translational and rotational parameter estimates:  $(\hat{U}, \hat{V}, \hat{W}) = (0, 0, 1)$  and  $(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) = (-0.001, -0.001, 0.01)$ .

where the intrinsic parameters are varying and estimated with errors. Figure 7(e) were obtained when we assumed an erroneous zoom field with  $\hat{f} = -0.01$  (or  $\sigma_e = 0.01$ ). This was the situation analyzed in Section 4.2.1, whereby we concluded that if  $\frac{\sigma_e}{f} > 0$ , ordinal depth will be preserved only if the image distance between the two feature points satisfies the inequality (13). Figure 7(e) showed clearly that in a local region ordinal depth was preserved, but for points which were far from each other, such as the points which were on the top face of the cylinder and on either sides of the cube, some depth orders were reversed. We also found that the recovered depths on the left side of the image were under-estimated more than those on the right side. This is consistent with the iso-distortion contours depicted in Figure 5.

Since no ground truth for the depth orders is available for the SOFA sequences, it is difficult to obtain comprehensive numerical results on the correctness of the recovered depth orders. Nevertheless, we observed that for the points on the top face of the cylinder in SOFA1 (which are delineated in Figure 7(a) and accounts for 4328 of the 5092 feature points), the true depth orders are known—the larger the  $y$  value is, the larger the  $Z$  value. For this particular region, the rates of correct ordinal depth recovery were calculated and tabulated in Table 1.

Table 1: Rates of correct ordinal depth recovery for points on the cylinder’s top face (SOFA1)

Figures 7(b) and (c)	Figure 7(d)	Figure 7(e)
92.42%	92.50%	82.28%

The SOFA5 sequence was used to perform experiments to verify predictions in the case of forward motion. The 3-D scene for SOFA5 comprised of a pile of 4 cylinders stacking upon each other and in front of a frontal-parallel background (Figure 7(f)). The camera trajectory for SOFA5 was parallel to the world  $Z$ -axis and the corresponding translational and rotational parameters were  $(U, V, W) = (0, 0, 1)$  and  $(\alpha, \beta, \gamma) = (0, 0, 0)$  respectively. We assumed that the observer or the system is aware that a forward motion is being executed, i.e.  $\hat{U} = \hat{V} = 0$ . Performing epipolar reconstruction, the equation for calculating depth for each feature point would be (again ignoring terms involving the principal point):

$$\hat{Z} = \frac{x^2 + y^2}{(u, v) \cdot (x, y) - (u_{\hat{rot}}, v_{\hat{rot}}) \cdot (x, y) - \frac{\hat{f}}{f} (x^2 + y^2)}$$

In Figures 7 (g), (h) and (i), the recovered depths were again rendered using the color coding

scheme. Figure 7 (*g*) depicted the case of no errors in the extrinsic and the intrinsic parameters. As could be seen, even for the case of no errors in the motion parameters, the ordinality of the depths recovered was error-prone, possibly as a result of the noise in optical flow computation. Figure 7 (*h*) depicted the case of erroneous extrinsic parameters. The depths recovered in the lower right part of the image were expanded while those on the upper left part were compressed. Such distortion was consistent with the iso-distortion contours depicted in Figure 3. Figure 7(*i*) was obtained when we estimated an arbitrarily large zoom field which in actual fact was zero. This corresponds to the case where some intrinsic parameters’ variations are not estimated correctly. The resultant iso-distortion contours should approximate that of Figure 6(*d*), although we did not introduce any error in the principal point estimate in our experiments. As predicted, the depths recovered seemed “better” than that of Figure 7 (*h*) in the sense that there was no systematic bias in the global arrangement of the depths. However, the recovered depths were compressed within a very small range, which meant that most of the depth details were lost. Thus it still cannot be deemed a good reconstruction.

Since no ground truth on the depth is available, we conducted the following check on the depth orders recovered. We singled out those points on the cylinders as foreground points and the points on the frontal-parallel plane as background points. We then compared the recovered depth of each foreground point with that of each background point. The former should be smaller than the latter if the depth order is preserved. Although such a test is not a comprehensive one, it gives a fair indication on the rates of correct ordinal depth recovery under various error scenarios. Using 1509 foreground points and 5925 background points, we obtained the following results. With true 3D motion (Figure 7(*g*)), 75.14% of such foreground-background depth orders were preserved. Most of these errors were randomly distributed in spatial location, as could be seen from Figure 7(*g*), and could be attributed to image noise. This result also corroborates our hypothesis that forward motion is not well suited for depth recovery, the distortion configuration of the perceived space being very sensitive to noise perturbations. For comparison, in the case of SOFA1 (lateral motion configuration) under no errors in the motion parameters, the rate of correct ordinal depth recovery for points on the cylinders top face was a much better figure of 92.89%<sup>2</sup>. For the case of forward motion with errors in the extrinsic parameters (Figure 7(*h*)), the rate of correct ordinal depth

---

<sup>2</sup>No ground truth was available for comparing the relative accuracy of the optical flow fields computed for SOFA1 and SOFA5, but from a visual inspection of the flow fields computed, they appeared to be of similar quality and thus should not be a significant factor contributing to the different rates of correct ordinal depth recovery.

recovery dropped to an almost chance level of 56.60%. Adding an error to the focal length estimate ( $\hat{f} = 280$ ) yielded a similar figure of 56.91%. With a small zoom error of 0.01, we obtained 56.67%. These two scenarios were not illustrated in Figure 7. Finally, for the case of large zoom error of 20 (Figure 7(i)), we obtained a poor result of 50.15%, confirming our earlier prediction that under this scenario, ordinal depth recovery is in practice very difficult.

## 6 Conclusions and Future Directions

This paper presented an investigation on the reliability of depth recovery given some errors in the estimates for the intrinsic and extrinsic parameters. Specifically, we sought for some generic motion types that rendered depth recovery more robust and reliable. Lateral and forward motions were compared both under calibrated and uncalibrated scenarios. For lateral movement, we found that although Euclidean reconstruction is difficult, the resulting distortion in the structure possesses many nice properties. For the case of calibrated motion, the distortion preserves the depth relief, which means that ordinal depth is preserved. In the uncalibrated case, if the intrinsic parameters are fixed, the situation remains largely the same. If focal length variation is also allowed and the resulting zoom field is not estimated correctly, the ordinal depth can be preserved only locally. In all these three cases, the transformations from the physical space to the perceived space are projective transformations and have very simple forms. For forward movement, whether calibrated or uncalibrated, depth information (even partial one) is hard to recover, except for those points close to the observer. Again, if the intrinsic parameters are fixed but estimated with some errors, things change very little from that of the calibrated case. If intrinsic parameters' variations are allowed, a large error in the zoom estimate has in theory the potential of improving ordinal depth recovery, although the resultant drastic loss of 3-D depth information would mean that this recovery is in practice very much suspect. The transformation relating the physical space to the perceived space is a Cremona transformation of degree three. Experiments conducted seemed to support the preceding theoretical predictions.

The conclusion is that under lateral movement, while it might be very difficult to resolve the ambiguity between translation and rotation, ordinal depth can be recovered with robustness. Conversely, it seems to explain the psychophysical phenomenon that under pure forward translation, human subjects were unable to recover structure unless favorable conditions such as large field of

view exist. In the case of uncalibrated motion, in spite of uncertainty in the focal length, the qualitative aspect of the recovered depth indeed is not affected, regardless of whether it is a lateral or a forward motion. Thus as far as depth recovery is concerned, if the intrinsic parameters are fixed, then calibration of these parameters are not the determining factors for accurate ordinal depth recovery. However, if the intrinsic parameters are allowed to vary, then it is important to estimate the zoom field correctly, although it is safe to forgo the variation of the principal point for depth reconstruction purpose. Otherwise, even in the case of lateral translation, global ordinality of depth will be lost.

This work represents the first step towards achieving partial scene understanding. More work needs to be done to build up such understanding under different motion-scene configurations. These capabilities, together with the ability to characterize the robustness of the partial depth information, would be important for many applications such as robotics, active vision and multimedia video indexing.

### Appendix: Using the Linear Least Square Scheme to Recover Depth

For any SFM algorithm based on optical flow as input, after obtaining the 3-D motion parameters, the next task would be to recover depth for each scene point according to equation (2). For notational convenience, we re-write them as follows:

$$\hat{Z} = \frac{\hat{\mathbf{e}} \cdot \mathbf{n}}{\hat{\mathbf{r}} \cdot \mathbf{n}} \quad (17)$$

where  $\hat{\mathbf{e}} = (x - \hat{x}_0, y - \hat{y}_0)$ ,  $\hat{\mathbf{r}} = (u - u_{\hat{r}ot}, v - v_{\hat{r}ot})$  and  $\mathbf{n} = (n_x, n_y)$ . If there is no error in the optical flow and the 3-D motion estimates, the choice of  $\mathbf{n}$  is immaterial, for any direction can give rise to correct depth recovery. However, when the 3-D motion estimates contain errors,  $\hat{\mathbf{e}}$  and  $\hat{\mathbf{r}}$  would not be parallel, and choosing different  $\mathbf{n}$  will yield different depth recovery, which leads to different depth recovery schemes. In particular, the standard linear least square estimate  $\hat{Z}_{LLSR}$  is given by  $\hat{Z}$  which minimizes the “estimated measurement error”  $\|\hat{\mathbf{e}} - \hat{Z}\hat{\mathbf{r}}\|$ , from which the following is obtained:

$$\hat{Z}_{LLSR} = \frac{\hat{\mathbf{e}} \cdot \hat{\mathbf{r}}}{\hat{\mathbf{r}} \cdot \hat{\mathbf{r}}}$$

In other words,  $\mathbf{n}$  is given by  $\frac{\hat{\mathbf{r}}}{\|\hat{\mathbf{r}}\|}$ , instead of  $\frac{\hat{\mathbf{e}}}{\|\hat{\mathbf{e}}\|}$  in the case of epipolar reconstruction.

Statistically, while the least square scheme is optimal given the necessary conditions, it must be noted that this approach has a serious qualification. In this problem, not only the observation

term  $\hat{\mathbf{e}}$  contains error, the measurement matrix also contains errors since the entries of the matrix are themselves estimates. Such errors, depending on their magnitudes, could be malign and affect the validity of the least square procedure.

There are also geometrical reason for not choosing to study the linear least square scheme, as we shall see in the following brief study of the distortion properties under the linear least square scheme.

In the lateral motion case, the iso-distortion factor for the linear least square reconstruction scheme can be expressed as:

$$D = \frac{(\hat{U}, \hat{V}) \cdot (U + Z\beta_e, V - Z\alpha_e)}{(U + Z\beta_e)^2 + (V - Z\alpha_e)^2} \quad (18)$$

where we have made the same assumptions as in Section 2. Figure 8(a) shows that the iso-distortion surfaces are also planes parallel to the image plane, identical to the case of epipolar reconstruction scheme. However, equation (18) shows that the distortion transformation does not belong to the class of relief transformation; thus there is no guarantee that the depth orders will be preserved.

**Figure 8 here**

In the forward motion case, the equation for the iso-distortion factor is:

$$D = \frac{(x, y) \cdot (x - \beta_e f Z, y + \alpha_e f Z)}{(x - \beta_e f Z)^2 + (y + \alpha_e f Z)^2}$$

Figures 8(b) and (c) show the iso-distortion contours for the forward motion case. Different  $y$  values will yield different distributions of iso-distortion contours. These contours are more or less similar to the contours we obtained using the epipolar reconstruction scheme. For instance, the  $D = 1$  contour still lies on the  $x$ -axis, and depth orders will in general not be preserved.

In conclusion, it is evident that different  $(n_x, n_y)$  would result in different distortion geometries. We are not trying to argue for any particular approach of recovering depth, but whichever scheme is adopted, it is the attendant distortion geometry that we are interested in. From this brief analysis, it seems that, geometrically, the “epipolar reconstruction approach” has certain favourable properties; it leads to the ordinal depth being preserved in the case of lateral motion, which is not so for the least square approach.

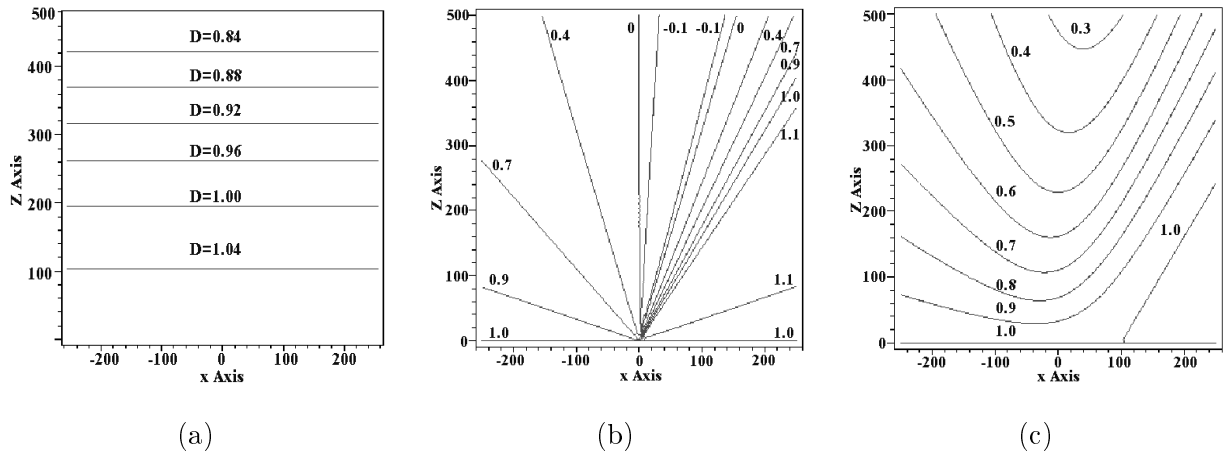


Figure 8: Families of iso-distortion contours using the linear least square scheme for depth reconstruction. (a) Lateral motion with  $U = 0.81$ ,  $\hat{U} = 1.0$ ,  $V = 0.58$  and  $\hat{V} = 0.4$ ; (b) forward motion case with  $y = 0$ ; (c): forward motion case with  $y = 100$ . All the other parameters are identical:  $\alpha_e = 0.001$ ,  $\beta_e = 0.001$  and  $f = 309.0$ .

## References

- Adiv, G. 1989. Inherent ambiguities in recovering 3-D motion and structure from a noisy flow field. *IEEE Trans. PAMI* **11**: 477–489.
- Bougnoux, S. 1998. From Projective to Euclidean Space under any practical situation, a criticism of self-calibration. In *Proc. Sixth Int. Conf. on Computer Vision* , pp. 790-796.
- Chaumette, F., Boukir, S., Bouthemy, P. and Juvin, D. 1994. Optimal estimation of 3D structures using visual servoing. In *Proc. IEEE Conf. Computer Vision and Pattern Recognition* , pp. 347–354.
- Cheong, L-F., Fermüller, C. and Aloimonos, Y. 1998. Effects of errors in the viewing geometry on shape estimation. *Computer Vision and Image Understanding*, 71(3):356–372.
- Cheong, L-F. and Ng, K. 1999. Geometry of Distorted Visual Space and Cremona Transformation. *International Journal of Computer Vision*, 32(2):195–212.
- Cheong, L-F. and Peh, C.H. 2000. Characterizing Depth Distortion due to Calibration Uncertainty. Accepted by *Sixth European Conf. on Computer Vision*, Dublin, Ireland.
- Coombs, D. and Roberts. K. 1993. Centering behaviour using peripheral vision. In *Proc. Conf. Computer Vision and Pattern Recognition*. pp 440-445.
- Daniilidis, K. and Spetsakis, M.E. 1995. Understanding Noise Sensitivity in Structure from Motion. In *Visual Navigation: From Biological Systems to Unmanned Ground Vehicles*, Y. Aloimonos (Ed.), Lawrence Erlbaum Assoc., Pub.
- Darrell, T. and Pentland, A. 1991. Robust estimation of a multi-layered motion representation. In *Proc. IEEE Workshop on Visual Motion*, pp. 173–178.
- Dutta, R. and Snyder, M.A. 1990. Robustness of correspondence-based structure from motion. In *Proc. Int'l Conf. on Computer Vision*, 106–110.
- Fermüller, C. Passive navigation as a pattern recognition Problem. *Int'l Journal of Computer Vision*, 14:147–158.
- Grossmann, E. and Victor, J.S. 2000. Uncertainty analysis of 3D reconstruction from uncalibrated views. *Image Vision Computing*, 18(9): 686–696.

- Horn, B.K.P. 1987. Motion fields are hardly ever ambiguous. *Int'l Journal of Computer Vision*, 1:259–274, 1987.
- Kahl, F. 1995. Critical motions and ambiguous Euclidean reconstructions in auto-calibration. In *Proc. Int. Conf. on Computer Vision*, pp. 469–475.
- Koenderink, J.J. and van Doorn, A.J. 1995. Relief: Pictorial and Otherwise *Image and Vision Computing*, 13(5):321–334.
- Longuet-Higgins, H.C. A computer algorithm for reconstruction of a scene from two projections. *Nature* 293, 133-135.
- Lucas, B.D. 1984. Generalized Image Matching by the Method of Differences. PhD Dissertation, Department of Computer Science, Carnegie-Mellon University.
- Negahdaripour, S. Critical surface pairs and triplets. *Int'l Journal of Computer Vision*, 3:293–312.
- Oliensis, J. 1999. A new structure-from-motion ambiguity. In *Proc. Conf. Computer Vision and Pattern Recognition*, pages 185–191.
- Oliensis, J. 2000. A critique of structure-from-motion algorithms. NECI Technical Report, April 1997. Updated February 2000. <http://www.neci.nj.nec.com/homepages/oliensis/Critique.html>.
- Santos-Victor, J., Sandini, G, Curotto, F. and Garibaldi. S. 1993. Divergent stereo for robot navigation: Learning from bees. *Proc. Conf. Computer Vision and Pattern Recognition*, New York, pp. 434–439.
- Soatto, S. and Brockett, R. 1998. Optimal structure from motion: local ambiguities and global estimates. In *Proc. Conf. Computer Vision and Pattern Recognition*, pp. 282–288.
- Sturm, P. 1997. Critical motion sequences for monocular self-calibration and uncalibrated Euclidean reconstruction. In *Proc. Conf. Computer Vision and Pattern Recognition*, pp. 1100–1105.
- Szeliski, R. and Kang, S.B. 1997. Shape ambiguities in structure from motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(5): 506–512.
- Thomas, J.I., Hanson, A. and Oliensis, J. 1993. Understanding noise: The critical role of motion error in scene reconstruction. In *Proc. DARPA Image Understanding Workshop*, pp. 691–695.

- Todd, J.T. and Reichel, F.D. 1989. Ordinal structure in the visual perception and cognition of smoothly curved surfaces. *Psychological Review*, 96(4):643–657.
- Ullman, S. 1979. *The Interpretation of Visual Motion*, MIT Press, Cambridge and London.
- Weng, J., Huang, T.S and Ahuja, N. 1991. *Motion and Structure from Image Sequences*, Springer-Verlag.
- Young, G.S. and Chellapa, R. 1992. Statistical analysis of inherent ambiguities in recovering 3-D motion from a noisy flow field. *IEEE Trans. PAMI*, 14:995–1013.