COMPUTER VISION AND IMAGE UNDERSTANDING
Vol. 71, No. 3, September, pp. 356–372, 1998
ARTICLE NO. IV970649

Effects of Errors in the Viewing Geometry on Shape Estimation*

LoongFah Cheong, Cornelia Fermüller, and Yiannis Aloimonos

Computer Vision Laboratory, Center for Automation Research, University of Maryland, College Park, Maryland 20742-3275

Received August 15, 1996; accepted July 3, 1997

A sequence of images acquired by a moving sensor contains information about the three-dimensional motion of the sensor and the shape of the imaged scene. Interesting research during the past few years has attempted to characterize the errors that arise in computing 3D motion (egomotion estimation) as well as the errors that result in the estimation of the scene's structure (structure from motion). Previous research is characterized by the use of optic flow or correspondence of features in the analysis as well as by the employment of particular algorithms and models of the scene in recovering expressions for the resulting errors. This paper presents a geometric framework that characterizes the relationship between 3D motion and shape in the presence of errors. We examine how the three-dimensional space recovered by a moving monocular observer, whose 3D motion is estimated with some error, is distorted. We characterize the space of distortions by its level sets, that is, we characterize the systematic distortion via a family of iso-distortion surfaces, which describes the locus over which the depths of points in the scene in view are distorted by the same multiplicative factor. The framework introduced in this way has a number of applications: Since the visible surfaces have positive depth (visibility constraint), by analyzing the geometry of the regions where the distortion factor is negative, that is, where the visibility constraint is violated, we make explicit situations which are likely to give rise to ambiguities in motion estimation, independent of the algorithm used. We provide a uniqueness analysis for 3D motion analysis from normal flow. We study the constraints on egomotion, object motion, and depth for an independently moving object to be detectable by a moving observer, and we offer a quantitative account of the precision needed in an inertial sensor for accurate estimation of 3D motion.

1. INTRODUCTION

Motion perception is one of the most important visual faculties of biological systems. It is concerned with the inference of our movement through the world and the detection of other moving bodies. It also allows us to infer the structure of the world around us. The apparent ease with which we perform these tasks belies the underlying nontrivial computational issues, as evidenced by the lack of robust, real-time, and accurate algorithms for estimating 3D motion and shape that exist today.

The main difficulty faced by such algorithms is the ill-conditioned nature of the problem. While many solutions to the structure from motion problem have been proposed, either using feature correspondences or flow fields, the resulting algorithms do not work well for real, complex scenes, and most of them degrade ungracefully as the quality of the data deteriorates. Many error analyses have been carried out in the past for the general case of 3D motion [1, 5, 7, 22, 26, 27] as well as for the special case of stereo [20]; a recent illuminating and critical survey is presented in [6]. They attempt to model either the errors in the motion estimates or those in the depth estimates, but due to the large number of unknowns in the problem, most of them deal with restricted conditions such as planarity of the scene [1, 5] or nonbiasedness of the estimators [5, 27]. Notably absent in all of these analyses is an account of the systematic nature of the errors in the depth estimates that result from errors in the motion estimates. In other words, the highly correlated nature of the depth errors at different spatial locations is not reflected adequately in these analyses. While [22] attempts to capture such systematic relationships using an error covariance matrix, the representation used there does not lend to a clear understanding of the relationship involved. Due to the lack of such an analysis, the relationship between the distorted surfaces and the true surfaces is not well understood, except in the case of critical surface pairs [14, 17–19], where the relationship is explicitly stated.

Given that human estimation of 3D motion is likely to be imprecise, the understanding of this distortion relationship is important in explaining various geometrical properties of perceived visual space, especially with regard to its non-veridical aspects. More importantly, the distortion relationship can also be used for studying the invariant aspects of perceived visual space. For instance, we would like to know if the ordinal relationship between the depths of different points [10, 16, 23] is preserved. As far as motion estimation is concerned, what is of great computational interest is those regions in space where the distortions are such that the perceived depths become negative. At these points, the visibility constraint is violated—any image

* The support of the Office of Naval Research under Grant N00014-96-1-0587 is gratefully acknowledged. The first author was also supported in part by the Tan Kah Khee Postgraduate Scholarships. Special thanks to Sara Larson for her editorial and graphics assistance.

1077-3142/98 $25.00

point, being visible, cannot lie behind the camera. Since the visibility constraint is essential because it is the only approach to motion analysis not relying on assumptions and heuristics, it is important to answer questions such as the sufficiency of the visibility constraint (an issue currently vaguely understood) and the stability of algorithms that employ the visibility constraint. Besides the negative distortion of space, we are also interested in a comparison of the distortion at different scene locations. As 3D scenes are usually piecewise smooth, we would like to know how the distortion of space affects the computation of shape in order to employ smoothness constraints in the development of processes for 3D motion estimation and segmentation.

In this paper, we propose a new framework, the iso-distortion surfaces, to capture the distortion relationship and address a number of issues related to 3D motion estimation. The main goal of the paper is to provide an intuitive picture of the inherent relationship between shape and motion as it arises from perceptual estimation processes. For this purpose, we discuss the distortion of space by presenting, in a successive way, its properties, and we attempt to illuminate its importance by discussing a number of computational issues arising in visual navigation problems. In accompanying papers, the same framework was used to address in a more elaborate way specific issues. In particular, in [13] it was used to explain the psychophysics of the distortion of visual space experienced by human observers from stereo or motion. In [4], among others, algorithms were developed for independent motion detection by a moving observer that exploit the relationships among the distortion spaces in the time domain; and in [11], on the basis of the iso-distortion framework, an error analysis of the structure from motion problem was carried out.

The organization of the paper is as follows: Section 2 develops the equations leading to the structure of the iso-distortion surfaces and describes some of the geometric properties of the distortion space. Section 3 is devoted to the study of the ambiguities that could arise in estimating 3D motion from normal flow by exploiting the structure of the distortion space and, in particular, its negative distortion subset. In this section we also study (a) the relationship between the errors of the various parameters of the 3D motion that are most likely to give rise to ambiguities regarding 3D motion estimation; and (b) the disambiguating power—with regard to obtaining 3D motion solutions—of a surface patch containing differently oriented features, i.e., what we can achieve from a local analysis. This second result provides a quantitative account of how likely it is for a moving observer to detect independent motion as a function of the 3D motion error (difference between egomotion and object motion) and the depth of the scene in view. Section 4, using the insight gained in the previous sections, studies whether one can obtain a unique solution for 3D motion using normal flow. The analysis also shows that motion fields are never ambiguous, a result conjectured in [14] and recently proved in [3]. Considering a limited field of view, however, ambiguities may arise. An analysis of the positions of the image areas giving rise to ambiguities is provided. Constraints are developed on the field of view for accurate 3D motion estimation. Section 5, using the theoretical results obtained in the previous sections, gives a quantitative analysis of the precision needed by an inertial sensor for accurate 3D motion estimation (the inertial sensor estimates rotation) and provides experiments. Section 6 summarizes the work.

2. THE ISO-DISTORTION FRAMEWORK

To characterize the distortion of depth due to erroneous motion estimates, we consider those points in 3D space whose estimated depth $\hat{Z}$ would be distorted by the same multiplicative factor D:

$$\hat{Z} = DZ.$$

The locus of such points constitutes a surface, which we call an iso-distortion surface. To facilitate the pictorial description of these surfaces in the following section, we slice them with planes parallel to the x-Z plane. We call the curves thus obtained iso-distortion contours.

2.1. Technical Prerequisites

We adopt the standard model for image formation, as illustrated in Fig. 1, with (U, V, W) and (α, β, γ) representing respectively the translation and the rotation of the observer in the coordinate system $OXYZ$. As a consequence of the well-known scale ambiguity, only the focus of expansion or FOE $(x0,\ y0)$,

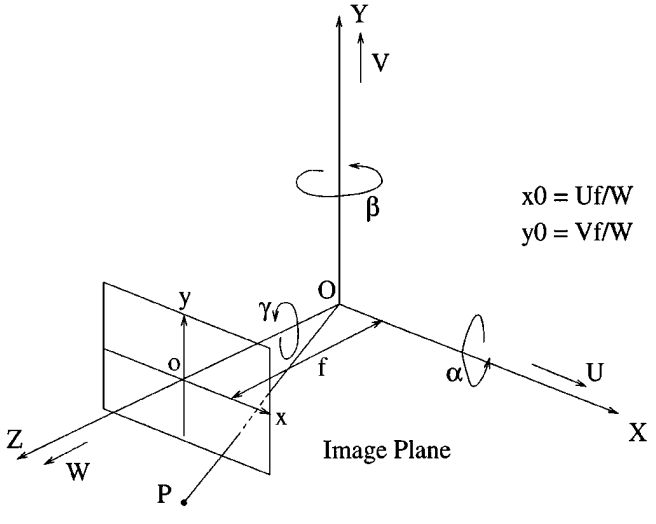


FIG. 1. The image formation model. $OXYZ$ is a coordinate system fixed to the camera. O is the optical center and the positive Z-axis is the direction of view. The image plane is located at a focal length f pixels from O along the Z-axis. A point P at (X, Y, Z) in the world produces an image point p at (x, y) on the image plane where (x, y) is given by $(\frac{fX}{Z}, \frac{fY}{Z})$. The instantaneous motion of the camera is given by the translational vector (U, V, W) and the rotational vector (α, β, γ).

given by $(\frac{fU}{W}, \frac{fV}{W})$, the rotational parameters (α, β, γ), and the scaled depth $\frac{Z}{W}$ are obtainable from flow information. Without loss of generality, we can set $W = 1$; henceforth Z shall represent the scaled depth, unless explicitly noted otherwise. In the derivation that follows, we assume that these five motion parameters have been estimated, though probably with some errors. Our focus is on the depth estimation stage, where the goal is to describe the distortion in depth as a function of the motion errors. We rearrange the familiar flow equation so that the quantity of interest, the depth Z, is on the left hand side:

$$Z = \frac{(x - x0, y - y0) \cdot (n_x, n_y)}{u_n - \boldsymbol{u_r} \cdot (n_x, n_y)}. \tag{1}$$

Here (n_x, n_y) denotes the direction of the vector along which normal flow is measured; depending on circumstances, we may also write it as $(\cos\theta, \sin\theta)$. u_n is the magnitude of the optic flow projected in direction (n_x, n_y), and $\boldsymbol{u_r}$ is the rotational flow vector, given by $(\alpha\frac{xy}{f} - \beta(\frac{x^2}{f} + f) + \gamma y, \alpha(\frac{y^2}{f} + f) - \beta\frac{xy}{f} - \gamma x)$. To refer to the various estimates, we use the hat sign $\hat{}$ to represent estimated quantities, and subscript e to represent errors. We also allow for a noise term N in the estimate for normal flow u_n:

$$\begin{aligned}
(\widehat{x0}\,, \widehat{y0}) &= (x0 - x0_e, y0 - y0_e) \\
(\hat{\alpha}, \hat{\beta}, \hat{\gamma}) &= (\alpha - \alpha_e, \beta - \beta_e, \gamma - \gamma_e) \\
\widehat{\boldsymbol{u_r}} &= \boldsymbol{u_r} - \boldsymbol{u_{r_e}} \\
\widehat{u_n} &= u_n + N.
\end{aligned}$$

Bringing the various estimated quantities into (1), the computed depth becomes

$$\hat{Z} = \frac{(x - \widehat{x0}\,, y - \widehat{y0}) \cdot (n_x, n_y)}{\widehat{u_n} - \widehat{\boldsymbol{u_r}} \cdot (n_x, n_y)}. \tag{2}$$

Writing $\widehat{u_n}$ as $(\frac{1}{Z}(x - x0, y - y0) + \boldsymbol{u_r}) \cdot (n_x, n_y) + N$, we obtain

$$\hat{Z} = Z\left(\frac{(x - \widehat{x0})n_x + (y - \widehat{y0})n_y}{(x - x0, y - y0) \cdot (n_x, n_y) + \boldsymbol{u_{r_e}} \cdot (n_x, n_y)Z + NZ}\right). \tag{3}$$

From (3), we can see that $\hat{Z}$ is distorted by a multiplicative factor given by the term inside the brackets, which we denote as D, the distortion factor. Further denoting the components of $\boldsymbol{u_{r_e}}$ as $(u^x_{r_e}, u^y_{r_e})$, we can write D as

$$D = \frac{(x - \widehat{x0})n_x + (y - \widehat{y0})n_y}{\big(x - x0 + u^x_{r_e}Z\big)n_x + \big(y - y0 + u^y_{r_e}Z\big)n_y + NZ}. \tag{4}$$

Ignoring the noise term for the moment, we see that for any fixed (n_x, n_y) and fixed distortion factor D, the above equation is of the form $Z = f(x, y)$ and therefore defines a surface in the 3D space, which we term the iso-distortion surface. Henceforth, when we talk about a family of iso-distortion surfaces, it is always with respect to a certain direction (n_x, n_y) defined at every image point. It is important to realize, on the basis of the preceding analysis, that the distortion of depth is different for different directions on the image plane where flow is estimated! This simply means that if one estimates depth from optic flow in the presence of errors, the results may be very different depending on whether the horizontal or vertical component (or any other component) is used!

In order to obtain the iso-distortion surfaces in 3D space (i.e., XYZ space) instead of visual space (i.e., xyZ space) we substitute $x = f\frac{X}{Z}$ and $y = f\frac{Y}{Z}$ in (4). This gives

$$\begin{aligned}
&D((\alpha_e XY - \beta_e(X^2 + Z^2) + \gamma_e YZ)\, n_x \\
&\quad + (\alpha_e(Y^2 + Z^2) - \beta_e XY - \gamma_e XZ)\, n_y) + \frac{NZ^2}{f} \\
&\quad - \left(X - \frac{\widehat{x0}\;Z}{f} - D\left(X - \frac{x0\;Z}{f}\right)\right) n_x \\
&\quad - \left(Y - \frac{\widehat{y0}\;Z}{f} - D\left(Y - \frac{y0\;Z}{f}\right)\right) n_y = 0.
\end{aligned} \tag{5}$$

Equation (5) describes the iso-distortion surfaces as quadratic surfaces, in the general case hyperboloids. An illustration is given in Fig. 2. However, for most of the analysis conducted in this paper, we study the surfaces (and contours) in visual space.

To throw more light on the nature of these surfaces, we first consider their intersections with planes parallel to the x-Z plane, and look at the resultant iso-distortion contours. After understanding the geometrical properties of the iso-distortion contours, we then proceed to describe the iso-distortion surfaces in 3D, and how they are related to each other.

2.2. Iso-distortion Contours

In what follows, we first perform a simplification—to be removed later—which, though not theoretically necessary, will allow us to better grasp the geometrical organization of the iso-distortion contours. The simplification basically amounts to ignoring some terms in $(u^x_{r_e}, u^y_{r_e})$ that result in secondary effects. In particular, we assume that the system's field of view (FOV) is not large and that the contribution of γ_e (i.e., $\gamma_e y$, $-\gamma_e x$) is small compared to that of α_e and β_e, so that $(u^x_{r_e}, u^y_{r_e})$ becomes $(-\beta_e f, \alpha_e f)$, denoted henceforth as $(-\beta_f, \alpha_f)$. If we now fix (n_x, n_y) to be in the horizontal direction, and ignore the noise term N for the moment, we can rewrite (4) as follows, assuming D to be non-zero,

$$Z = \frac{D - 1}{D}\frac{x}{\beta_f} - \frac{1}{D}\left(\frac{x0_e}{\beta_f} + \frac{(D - 1)x0}{\beta_f}\right) \tag{6}$$

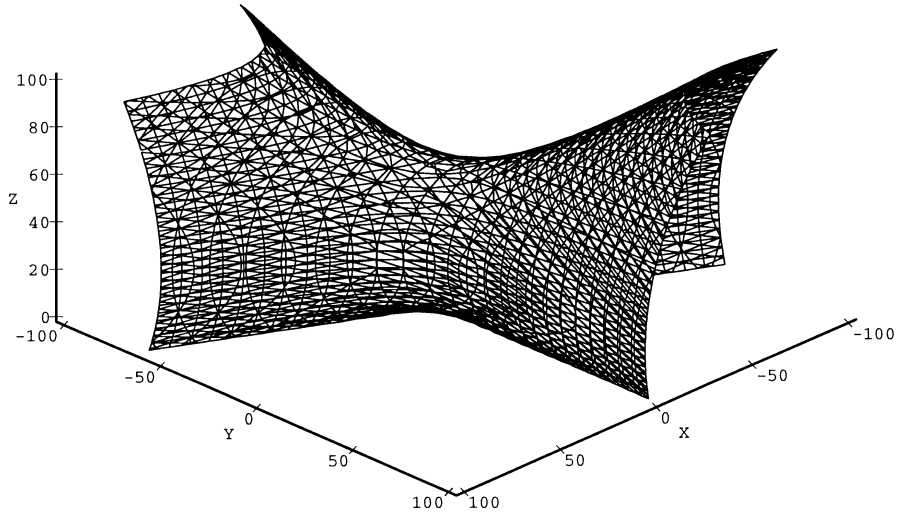


FIG. 2. Iso-distortion surface in XYZ-space. Only the part in front of the image plane is shown.

which describes the iso-distortion surfaces as a set of planes perpendicular to the x-Z plane. Much of the information that (6) contains can thus be visualized by considering a family of iso-distortion contours on a two-dimensional x-Z plane. Each family is defined by three parameters: $x0$ and the two error terms $x0_e$ and β_f. Within each family, a particular D defines an iso-distortion contour. Figure 3 shows several families of iso-distortion contours; Figs. 3c and 3d correspond to the special cases of $\beta_f = 0$ and $x0_e = 0$, respectively. In the next section, we shall determine the salient geometrical properties of such simplified iso-distortion contours, after which we will re-incorporate the terms that we have ignored.

2.3. Geometrical Properties of Iso-distortion Contour Plots

2.3.1. Negative Distortion Regions

The depth estimate $\hat{Z}$ becomes negative when the distortion factor D is negative. On the distortion plots, this negative region lies between the $D = 0$ line and the $D = -\infty$ line, represented as -INF on the plot. The size of this region determines the efficacy of the visibility constraint. As remarked earlier, (6) is valid only if D is non-zero. Taking this into account, we derive the equations for both the $D = 0$ and the $D = -\infty$ contours as

$$\begin{aligned} D = 0 &: x = \widehat{x0} \\ D = -\infty &: Z = \frac{1}{\beta_f}(x - x0) \qquad \text{if } \beta_f \neq 0 \\ &: x = x0 \qquad\qquad\quad \text{otherwise.} \end{aligned}$$

The $D = 0$ contour is a vertical line, whereas the $D = -\infty$ line has slope given by $\frac{1}{\beta_f}$, and x-intercept given by $x0$. Therefore, for the case where $\beta_f = 0$, the negative distortion region is a vertical band defined by $x = x0$ and $x = \widehat{x0}$ (see Fig. 3c). As β_f deviates from zero, its effect is to rotate the $D = -\infty$ contour while pivoting about the x-intercept $x0$. The size of the negative distortion region changes accordingly.

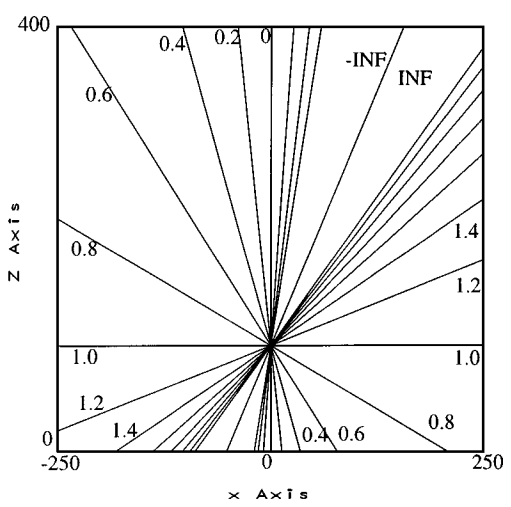
2.3.2. The $D = 1$ Contours Are Horizontal

This is the contour on which the depth estimates are not distorted. The equation of the contour is given by $Z = -\frac{x0_e}{\beta_f}$. Together with the $D = 0$ vertical line, it divides the x-Z plane into four regions. Ignoring the negative distortion regions, we can roughly characterize two of them as having the effect of overestimating depth, and two of them as underestimating depth. If the location of $\widehat{x0}$ is well outside the image, so that the $D = 0$ contour is not in view, then the x-Z plane is divided by the $D = 1$ contour into two half-planes. Depths on one side of the contour are subjected to either contraction only or expansion only (and vice versa for the other side), depending on the errors in the motion estimates. The point where the $D = 0$ and the $D = 1$ contours meet also defines the common intersection point of all the distortion contours. At this point, the distortion factor is undefined.

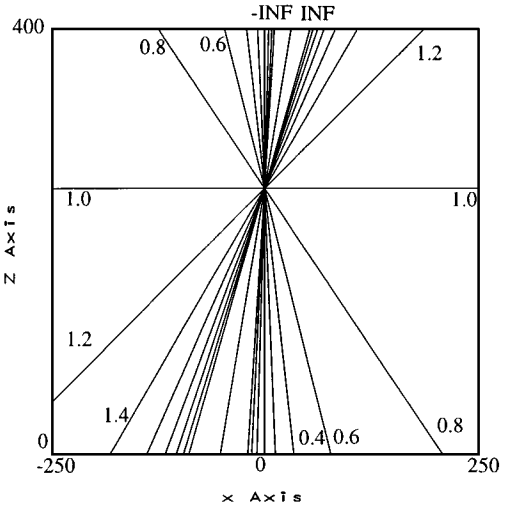
2.3.3. Flows in Other Directions

The previous analysis has considered the case of horizontal flow. Now, we extend the analysis to any arbitrary direction (n_x, n_y). With our simplification, this is particularly easy to handle. By substituting

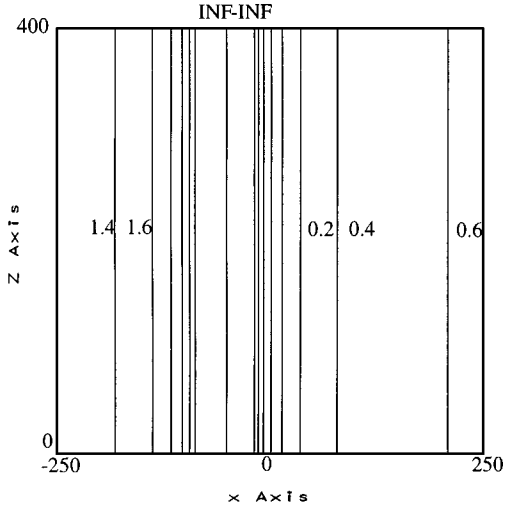
$$\begin{aligned} x0' &= (x0, y0) \cdot (n_x, n_y) \\ x0'_e &= (x0_e, y0_e) \cdot (n_x, n_y) \\ x' &= (x, y) \cdot (n_x, n_y) \\ \beta'_f &= (\beta_f, -\alpha_f) \cdot (n_x, n_y) \end{aligned}$$



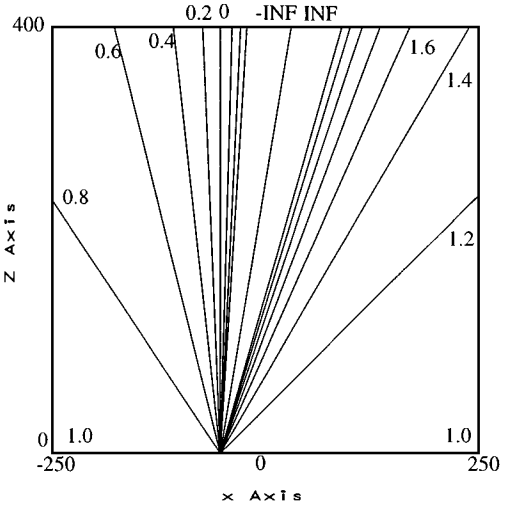
(a) $x0 = -50, x0_e = -50, \beta_f = 0.5$



(b) $x0 = -50, x0_e = -50, \beta_f = 0.2$



(c) $x0 = -50, x0_e = -50, \beta_f = 0.0$



(d) $x0 = -50, x0_e = 0, \beta_f = 0.2$

FIG. 3. Families of iso-distortion contours parameterized by $x0$, $x0_e$, and β_f. The number beside each contour denotes the distortion factor D of that contour. INF denotes ∞. The contours are spaced $D = 0.2$ apart, except at the region where the magnitude of D becomes very large. The $-$INF and INF contours coincide.

we get for any direction

$$Z = \frac{D-1}{D}\frac{x'}{\beta'_f} - \frac{1}{D}\left(\frac{x0'_e}{\beta'_f} + \frac{(D-1)x0'}{\beta'_f}\right)$$

which has the same form as (6). Therefore, each direction exhibits its own iso-distortion pattern, characterized by its $x0'$, $x0'_e$, and β'_f.

2.4. Effects of FOV, γ_e, and Noise

The results of the previous section were derived under the assumptions that the FOV is not large and γ_e is small. We now consider the effects of these assumptions on the iso-distortion contours.

- *FOV.* Figure 4 presents two iso-distortion plots, characterized by the same triplet ($x0$, $x0_e$, β_e) that defines Fig. 3a, but considering the rotational terms quadratic in the image coordinates. Figure 4a has a FOV of 50°, whereas Fig. 4b has a FOV of 70°. It can be observed that the iso-distortion contours become curved at the periphery of the image, notably in Fig. 4b. The above qualifications notwithstanding, the topology of the contours remains the same, so that many of the remarks made in Subsections 2.3.1 and 2.3.2 are applicable with few modifications.

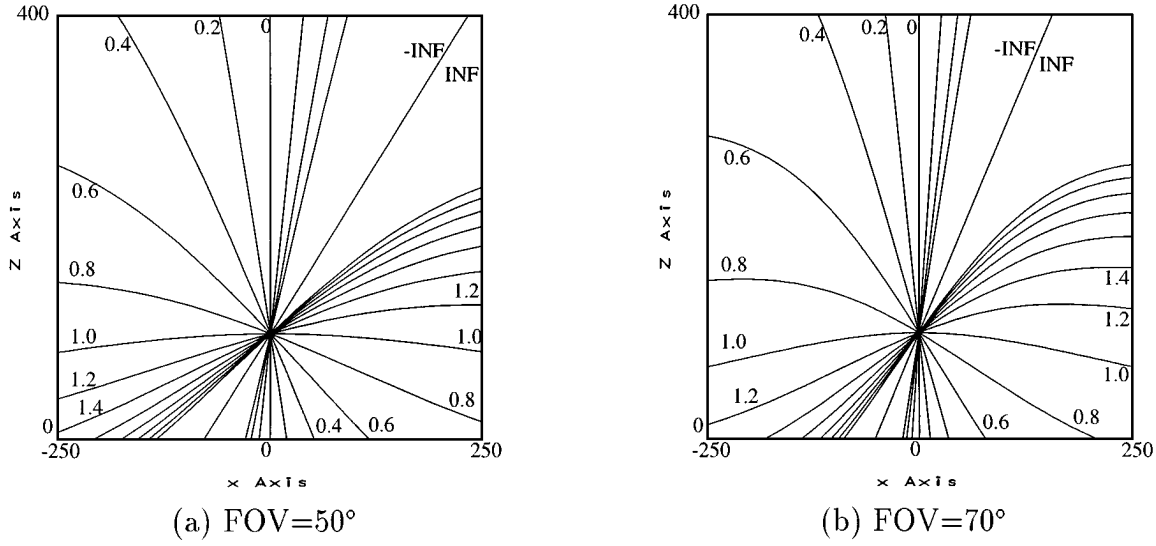


FIG. 4. Effects of wide field of view on the iso-distortion contours of Fig. 1a. $x0 = -50$, $x0_e = -50$, and $\beta_e \approx 0.001$. The different FOVs of (a) and (b) are achieved by using different focal lengths: 500 for FOV $= 50^\circ$ and 350 for FOV $= 70^\circ$.

• *Noise.* The effect of noise is to change the β_f term in (6) to $\beta_f - N$; thus we obtain

$$Z = \frac{D-1}{D}\frac{x}{\beta_f - N} - \frac{1}{D}\left(\frac{x0_e}{\beta_f - N} + \frac{(D-1)x0}{\beta_f - N}\right). \quad (7)$$

Therefore, flows with different noise have different iso-distortion contours. The corresponding effects on the shape of the negative distortion region will be investigated in Subsection 3.2, where we consider the noise sensitivity of motion estimators that are based on the visibility constraint.

• γ_e. In the horizontal direction, the contribution of γ_e is the term $\gamma_e y$. Since y is fixed on any particular horizontal plane where we view the iso-distortion contours, the effect of γ_e is to increase or decrease the rotational component, by a constant term, thus changing β_f to $\beta_f - \gamma_e y$. In other words, the iso-distortion contours in different horizontal planes are governed by different rotational components, but within a particular horizontal plane, the rotational term is fixed. The effect is better visualized in 3D space, which we turn to in the next subsection.

2.5. Iso-distortion Surfaces

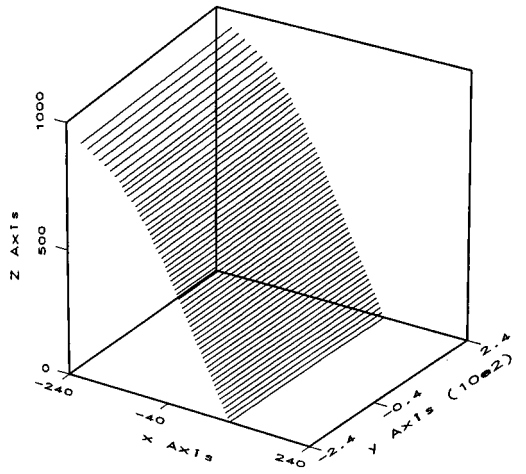
Figures 5a and 5b illustrate two iso-distortion surfaces, corresponding to $D = 0.4$ and $D = 1.8$, respectively, and with respect to the gradient $(n_x, n_y) = (1, 0)$. The same triplet $(x0, x0_e, \beta_e)$ as in Fig. 4b was used, and α_e and γ_e were set to zero. The effects of α_e and γ_e were respectively introduced in Figs. 5c and 5d, and the $D = 1.8$ iso-distortion surface was shown. Comparing Fig. 5c to 5b, we can see that the effect of the $\alpha_e \frac{xy}{f}$ term on the distortion surface is quite small, whereas comparing Fig. 5d to 5b, we can see that the effect of $\gamma_e y$ is quite considerable. For Fig. 5d, the effects of a changing rotational component are quite evident, namely, an ever-changing slope and a changing intercept in different x-Z planes.

3. WHAT CAN THE ISO-DISTORTION FRAMEWORK TELL US?

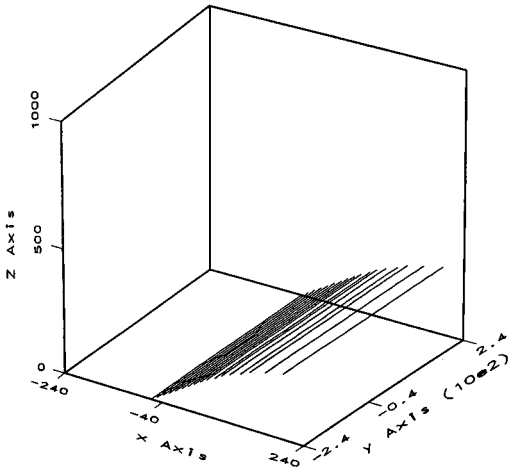
Using the framework, we now delve into questions concerning computational issues related to 3D motion estimation using normal flow (the movement of the image along a direction normal to the intensity gradient) [8, 9, 15]. The basic tenet of the normal flow approach (also known as the direct approach) is that local measurement of movement in the image is not sufficient to determine the movement of the corresponding point in space, and therefore only normal flow should be used, from which motion is inferred directly without computing correspondence or optic flow. These approaches rely on the visibility constraint, that is, the positivity-of-depth constraint, to find the location of the FOE in the image plane. Therefore it is important to understand the properties of the visibility constraint. With regard to the iso-distortion surfaces, the region where the visibility constraint is violated corresponds to the region where the distortion factor is negative. Therefore geometrical knowledge of this negative distortion region can be used as a basis on which computational issues can be studied. For much of the discussion, it suffices to use families of iso-distortion contours as representations of the iso-distortion surfaces. Furthermore, when generality is not lost, the simplified model is used; that is, the second-order effects in the rotational error flows are ignored.

3.1. The Localization of the FOE

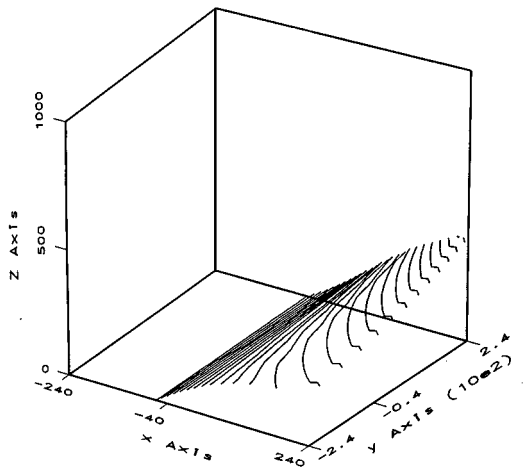
In [2, 15, 21], the problem of determining the FOE given a known rotation is examined. Suppose the known rotation is subtracted from the normal flow; the remaining normal flow corresponds to that arising from a purely translational flow field. By exploiting the visibility constraint, the problem of FOE estimation can be converted to a constrained optimization problem.



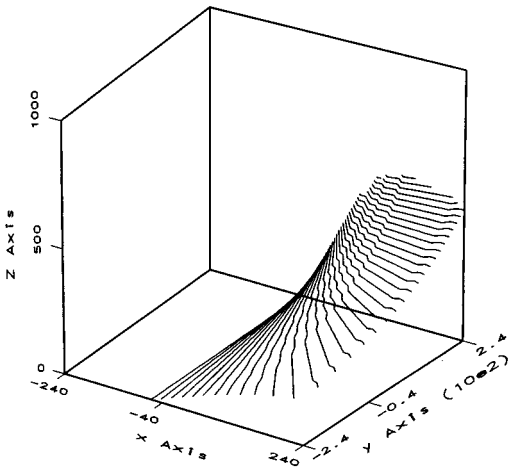
(a) $D = 0.4,\ (\alpha_e, \beta_e, \gamma_e) = (0.0, 0.001, 0.0)$



(b) $D = 1.8,\ (\alpha_e, \beta_e, \gamma_e) = (0.0, 0.001, 0.0)$



(c) $D = 1.8,\ (\alpha_e, \beta_e, \gamma_e) = (0.001, 0.001, 0.0)$



(d) $D = 1.8,\ (\alpha_e, \beta_e, \gamma_e) = (0.0, 0.001, 0.001)$

FIG. 5. Iso-distortion surfaces for gradients in the x direction. $x0 = -50$; $x0_e = -50$; FOV $= 70°$. Note that the x-Z planes as presented in the plots become vertical rather than horizontal.

The basic underlying idea is illustrated in Fig. 6. It is clear that if $\boldsymbol{u_n}$ contains only translational flow, the FOE must lie in the shaded half plane defined by the line ε. Thus every point in that half plane receives a vote for being the FOE. The best solution for the FOE corresponds to the location with the highest number of votes.

We next investigate the behavior of such algorithms for the estimation of translation in the presence of small rotational motion (that is, if either the rotation is estimated with some error, or a flow field considered to be only due to translation also contains small rotational components). Figure 7 depicts the negative distortion areas of several FOE candidates, each with different amounts of error $x0_e$. The negative distortion region is bounded by two contours: the $D = -\infty$ and the $D = 0$ contours, whose

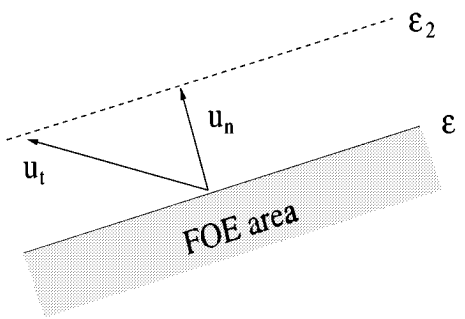


FIG. 6. The translational flow vector $\boldsymbol{u_t}$ has its tip anywhere along the line ε_2. The focus of expansion lies on the (shaded) half plane defined by the line ε that does not contain possible vectors $\boldsymbol{u_t}$.

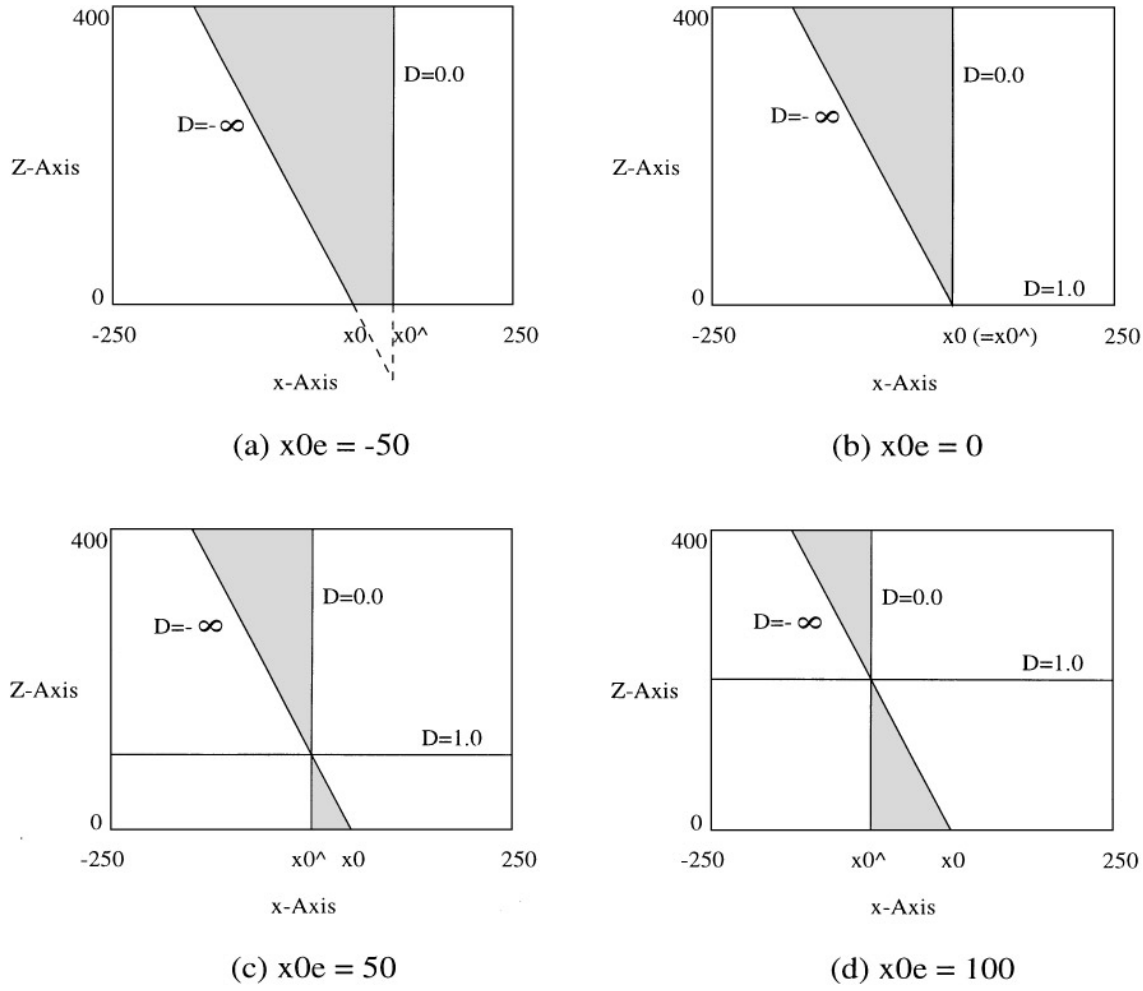


FIG. 7. The shaded area represents the negative distortion band. Errors in the FOE estimates are expected since (b), the correct candidate, does not have the smallest negative distortion area. $\beta_f = -0.5$, $x0 = 50$, $\widehat{x0} = x0 - x0_e$.

equations are given by $Z = \frac{1}{\beta_f}(x - x0)$ and $x = \widehat{x0}$, respectively. The unaccounted-for rotational flow in all four cases is given by the same β_f and thus the $D = -\infty$ contour is the same in all cases. The equation for the $D = 0$ contour is different though. The resultant negative distortion regions, represented by the shaded areas in Fig. 7, clearly indicate that the correct candidate (Fig. 7b) does not have the smallest negative distortion region; therefore usually the correct $x0$ will not be estimated. Furthermore, it is also evident from the figure that the area of negative depth volumes is anisotropic about the true FOE; in this particular case the estimation for $x0$ is skewed towards the side with positive $x0_e$. A complete analysis must take into account the normal flow measurements obtained from all gradient directions. As we shall see in Subsection 3.3.2, if the scene in view is approximately fronto-parallel, then the preceding observation still holds even if we utilize normal flow measurements from all gradient directions.

By examining the geometry of the negative distortion areas, one can partially compensate for the estimation error if one has some knowledge about how the near and far scene distances are approximately distributed in the image. This is accomplished by noting the fact that the $D = \pm\infty$ contours (they are coincident) pass through $x0$ on the iso-distortion plots; thus by observing how the very large positive $\hat{Z}$'s and the very large negative $\hat{Z}$'s are juxtaposed with respect to each other in the near ground, one can at least obtain the direction of the FOE error. There remains, however, the difficulty connected with noise, which we deal with in the next section.

3.2. Effect of Noise

We shall show in this section that even if a flow field contains no rotation, the presence of noise considerably reduces the robustness of the voting method. This may be expected, but what is not so obvious is that the negative distortion regions of some incorrect FOE candidates may not increase in size at all (probabilistically speaking). For illustration purposes, suppose the noise N has the following probability mass

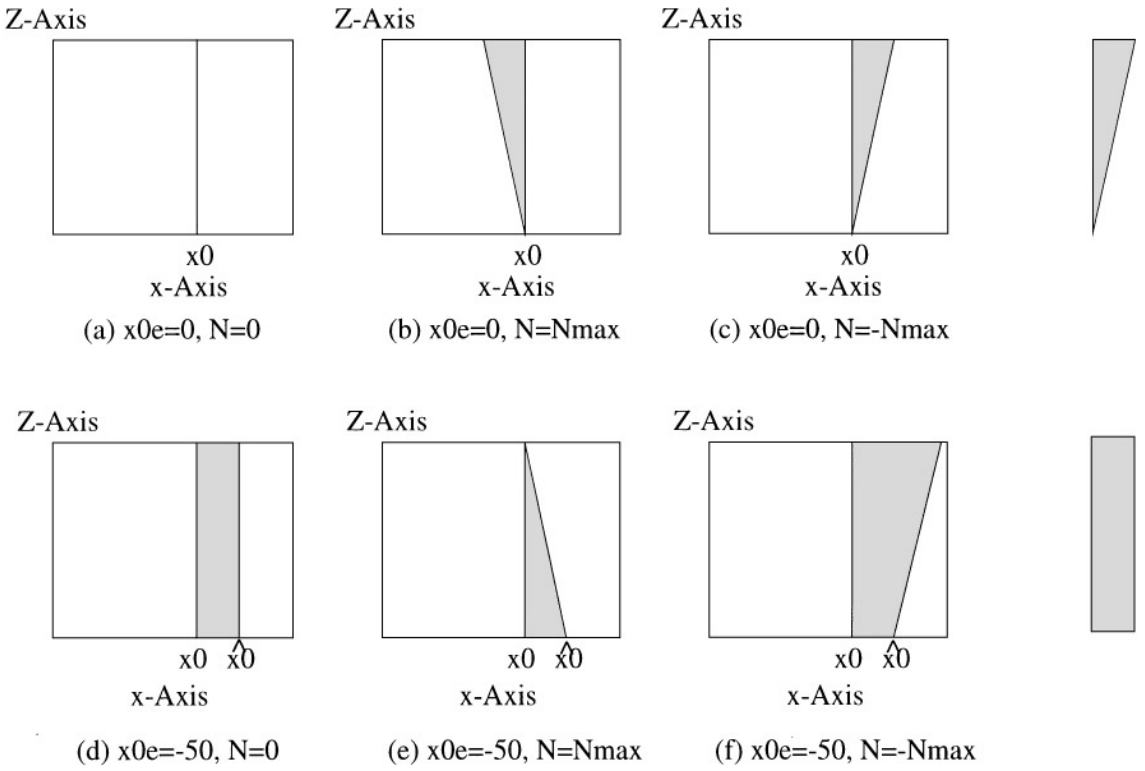


FIG. 8. Negative distortion regions of two FOE estimates under varying conditions of noise. The upper row represents the negative distortion regions of the correct candidates. The shaded figures at the end of each row represent the expected sizes (not the shape) of the negative distortion regions of the respective candidates given the probability mass function $P(N)$.

function:

$$P(N = N_{\max}) = \frac{1}{2}$$
$$P(N = -N_{\max}) = \frac{1}{2}.$$

Figure 8 illustrates the negative distortion regions of two FOE candidates under various noise conditions and zero rotational error, with the upper row corresponding to that of the correct FOE estimate. Since the rotational error is zero, the slope of the $D = -\infty$ contour is simply given by $\frac{1}{-N}$. The expected area of the negative distortion region is calculated using the given probability mass function of the noise N and its size is shown at the end of each row. By comparing these figures, we see that the effect of noise, while introducing negative depth estimates for the correct FOE candidate, does not result in an increase in the size of the negative distortion region of the incorrect FOE candidate.

The form of the noise distribution is not a matter of primary importance here; our aim is to demonstrate influences on the areas of negative depth as they arise from the distribution of the scene features and the statistical nature of the noise. A further point can be made from Fig. 8: For the correct FOE candidate the positivity of distant points is more susceptible to noise. This is related to the observation that the localization of the FOE improves with the magnitude of the translational flow.

3.3. Situations Giving Rise to Ambiguities

The purpose of this section is to make explicit those situations under which the visibility constraint may not be sufficient to discriminate between alternative motion solutions. The factors that give rise to ambiguity are examined from two perspectives: first, the types of motion errors; second, the influence of the location of a surface patch.

3.3.1. Types of Motion Errors

We first establish a condition on the motion errors that most likely leads to ambiguities. It follows from the observation that if the iso-distortion diagram is such that the point at which its $D = 0$ contour and its $D = -\infty$ contour intersect lies behind the image plane, then irrespective of the actual scene structure, negative depth estimates will always be obtained, since for some image point P (see Fig. 9a), the space in front of the image region is entirely spanned by the negative distortion area. However, if the motion errors are such that this intersection point lies in front of the image plane for all gradient directions, the likelihood of ambiguity increases considerably. Therefore the conditions on

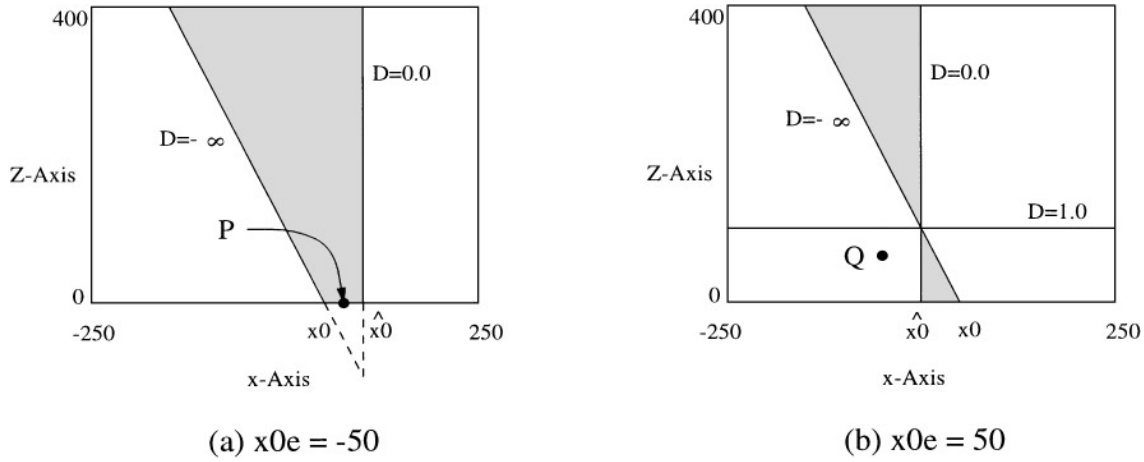


FIG. 9. The iso-distortion diagram on the left will always result in negative depth estimates, regardless of the actual depth distribution. This is so because the whole region in space between $x = x0$ and $x = \widehat{x0}$ is spanned by the negative distortion region.

the motion errors for ambiguity can be stated as, first, $\gamma_e = 0$; and second, the Z-coordinate (the absolute depth) of the point at which the $D = 0$ contour and its $D = -\infty$ contour intersect should be such that

$$-W\,\frac{(x0_e, y0_e)\cdot(\cos\theta, \sin\theta)}{(\beta_f, -\alpha_f)\cdot(\cos\theta, \sin\theta)} > 0 \qquad \forall\theta.$$

This condition can be satisfied only if $W(x0_e, y0_e) = -\lambda(\beta_f, -\alpha_f)$ for some positive λ. This constitutes our condition on the type of motion error that most likely gives rise to ambiguity.

Given this condition, if the scene in view is such that it avoids the negative distortion region for all image points (for instance, point Q in Fig. 9b avoids the negative distortion area), then ambiguities arise. The constraints on the regions within which point Q can reside so as to yield ambiguities can be obtained directly from Fig. 9b by using the equations for the $D = 0$ and the $D = -\infty$ contours.

$$\mathrm{sgn}(\beta'_f)Z < \frac{\mathrm{sgn}(\beta'_f)}{\beta'_f}((x - x0)\cos\theta + (y - y0)\sin\theta)$$
$$\text{if } (x - \widehat{x0})\cos\theta + (y - \widehat{y0})\sin\theta > 0 \quad (8)$$

$$\mathrm{sgn}(\beta'_f)Z > \frac{\mathrm{sgn}(\beta'_f)}{\beta'_f}((x - x0)\cos\theta + (y - y0)\sin\theta)$$
$$\text{if } (x - \widehat{x0})\cos\theta + (y - \widehat{y0})\sin\theta < 0, \quad (9)$$

where we have used β'_f to denote $(\beta_f\cos\theta - \alpha_f\sin\theta)$ and $\mathrm{sgn}(\cdot)$ to denote the sign function. As long as the scene in view satisfies these constraints, no negative depth estimates will arise. In actual fact, the scene in view is not composed of spatially unrelated elements where Z can vary freely to meet the constraints. A scene patch, with gradients in different directions, and therefore experiencing different constraints on depth, may not be capable of meeting all of them owing to its smoothness property. The likelihood of such a scene patch giving rise to negative depth estimates depends on a number of factors, which we investigate in the next section.

3.3.2. Using a Patch

Consider a smooth surface patch. We would like to know the likelihood of the resultant surface estimate yielding some negative depth values if the feature gradients are distributed over a sufficiently wide range. This will indicate regions on the image plane in which the aggregate of normal flow field measurements is most effective in disambiguating alternative motion solutions, and it also provides a quantitative assessment of how likely it is for a moving observer to be able to detect independent motion.

For the purpose of quantifying this effectiveness, we adopt the following measure: Given some errors in the motion estimates and a fixed small region in space, what is the range of directions in which the gradient must lie so as to yield negative depth estimates? The larger this θ range is, the more effective the patch.

The disambiguating power of a scene patch depends on a number of factors such as the image location of the patch, the actual depth of the patch, and the range of gradient directions on the patch. For instance, with regard to gradient directions, we note that they are most effective around the direction at which the bound on Z (expressions (8) and (9)) changes from an upper bound to a lower bound. If the gradients of the patch happen to lie on either side of this direction, the depth of the patch will be constrained to lie in the range given by the upper and the lower bound. Referring to (8) and (9), the direction is given by $(x - \widehat{x0})\cos\theta + (y - \widehat{y0})\sin\theta = 0$. Unfortunately, depths are usually not computed for such directions, since being perpendicular to lines emanating from the estimated FOE, these gradients are deemed to carry no depth information. In spite of this fact, if the range around the direction within which negative depth estimates arise is large, then the patch becomes potentially useful in disambiguating motion solutions.

In the following analysis we drop, as before, the quadratic terms in the rotational error. The analysis is further aided by rotating the x-y coordinate system so that in the new coordinate system, there is no error in the y-component of the rotational estimate, i.e., $\alpha_e = 0$. The distortion factor can now be written as

$$D = \frac{x - \widehat{x0} + (y - \widehat{y0})\tan\theta}{x - x0 - (\beta_f - \gamma_e y)Z + (y - y0 - \gamma_e x Z)\tan\theta}.$$

We denote the numerator and denominator by $h(\theta)$ and $k(\theta)$, and the angles θ at which they become zero as θ_h and θ_k, respectively. The distortion factor D becomes negative when $h(\theta)$ and $k(\theta)$ have different signs. We consider θ between $-\frac{\pi}{2}$ and $\frac{\pi}{2}$. The size of the θ range within which $h(\theta)$ and $k(\theta)$ have different signs can be expressed as

$$R = \begin{cases} |\theta_h - \theta_k| & \text{if } \dfrac{\partial h(\theta)}{\partial \tan\theta} \cdot \dfrac{\partial k(\theta)}{\partial \tan\theta} > 0 \\ \pi - |\theta_h - \theta_k| & \text{otherwise,} \end{cases} \tag{10}$$

where R is the size of the interval. To compute R, for convenience of notation we let $t = \tan(\theta_h - \theta_k)$. Then the angle $\tan^{-1} t$, ranging from $-\pi$ to π, is given by

$$\tan^{-1} t = \tan^{-1}\frac{-(x - \widehat{x0})}{y - \widehat{y0}} - \tan^{-1}\frac{-(x - x0 - (\beta_f - \gamma_e y)Z)}{y - y0 - \gamma_e x Z}.$$

Taking the tangent of both sides, we obtain

$$\begin{aligned} t((y - \widehat{y0})(y - y0 - \gamma_e x Z) + (x - \widehat{x0})(x - x0 \\ - (\beta_f - \gamma_e y)Z)) - (x - x0 - (\beta_f - \gamma_e y)Z)(y - \widehat{y0}) \\ + (x - \widehat{x0})(y - y0 - \gamma_e x Z) = 0. \end{aligned} \tag{11}$$

Introducing the following coordinate translation,

$$\begin{aligned} x_{\text{old}} &= x_{\text{new}} + x0 - \frac{x0_e}{2} = x_{\text{new}} + \widehat{x0} + \frac{x0_e}{2} \\ y_{\text{old}} &= y_{\text{new}} + y0 - \frac{y0_e}{2} = y_{\text{new}} + \widehat{y0} + \frac{y0_e}{2} \end{aligned} \tag{12}$$

and doing away with the subscripts in x and y, we finally obtain the equation

$$\begin{aligned} (t - \gamma_e Z)x^2 + (t - \gamma_e Z)y^2 - \left(t\beta_f Z + y0_e + \frac{t\gamma_e Z y0_e}{2}\right. \\ \left. + \frac{\gamma_e Z x0_e}{2}\right)x + \left(\beta_f Z + x0_e + \frac{t\gamma_e Z x0_e}{2} - \frac{\gamma_e Z y0_e}{2}\right)y \\ + \left(\frac{\beta_f Z y0_e}{2} - \frac{t\beta_f Z x0_e}{2} - \frac{t x0_e^2}{4} - \frac{t y0_e^2}{4}\right) = 0 \end{aligned} \tag{13}$$

which can be shown to give equations of circles for different values of t. All the circles pass through the two points $(\widehat{x0}, \widehat{y0})$ and $((x0 + (\beta_f - \gamma_e y0)Z)/(1 + \gamma_e^2 Z^2),\ (y0 + \gamma_e Z(x0 + \beta_f Z))/(1 + \gamma_e^2 Z^2))$. In particular, the $t = 0$ locus is a straight line defined by the above two points.

To obtain R, these loci of t must be modified according to (10):

$$R = \begin{cases} |\tan^{-1} t| & \text{if } \dfrac{\partial h(\theta)}{\partial \tan\theta} \cdot \dfrac{\partial k(\theta)}{\partial \tan\theta} > 0 \\ \pi - |\tan^{-1} t| & \text{otherwise.} \end{cases}$$

Figure 10 illustrates the case $\gamma_e = 0$. Figure 10a shows the loci of $\tan^{-1} t$, whereas Fig. 10b shows the actual range R. The centers of all the circles lie on a straight line perpendicular to

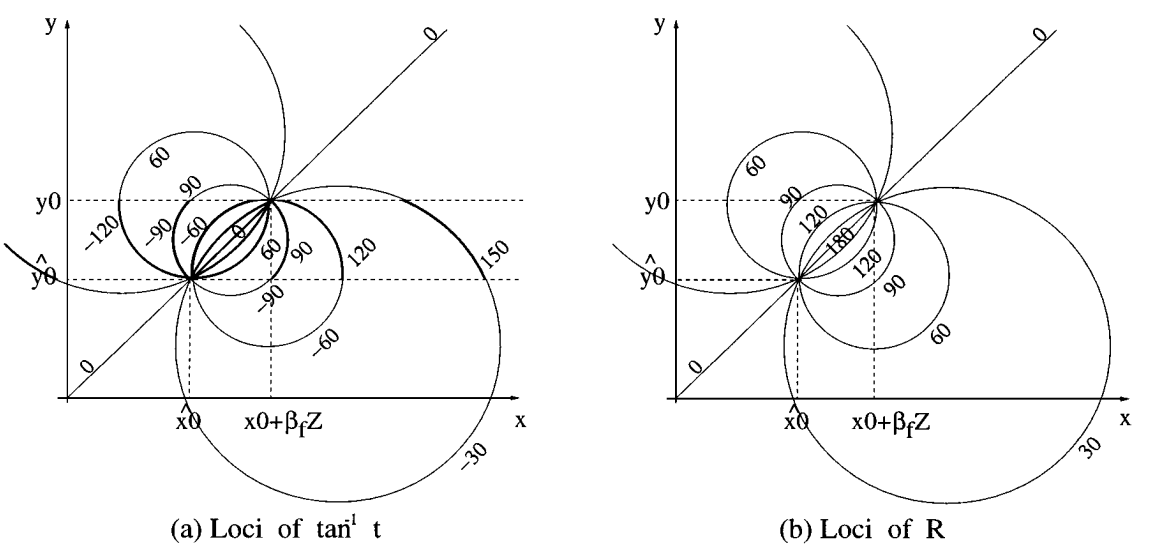


FIG. 10. (a) The circles describe loci of t on the image plane ($\gamma_e = 0$). (b) The values of R are indicated by the numbers on the circles.

the $t=0$ locus.[1] Several points can be made. First, there exists a band in the image plane where R is large. The direction of the band depends on the relative magnitudes of two terms: $y0_e$ and $x0_e+\beta_f Z$. Second, for a correct FOE candidate ($x0=\widehat{x0}$), the radii of the circles may be larger than for an incorrect FOE candidate. Therefore the remarks made in Subsection 3.1 still apply even when scene patches with differently oriented gradients are utilized in the motion estimation process. In particular, when the motion errors and the depth of the patch are such that $\widehat{x0}=x0+\beta_f Z$, $\widehat{y0}=y0-\alpha_f Z$, and $\gamma_e=0$, the circles collapse. The interval R becomes null everywhere, unless one takes into account the quadratic terms in the rotational flow. The above conditions are readily satisfied by a fronto-parallel scene and a camera with a small field of view. Therefore such a scenario is likely to yield multiple solutions.

4. UNIQUENESS ANALYSIS

In the previous sections, we examined the problem of ambiguity from various aspects. The aim is to obtain an intuitive picture of various conditions that give rise to multiple solutions. In this section, we shall formulate mathematically the conditions under which the solution is unique. Specifically, we ask the question: If we are allowed to arbitrarily choose a gradient distribution, can we find a gradient field such that the actual 3D motion and the estimated motion must give rise to different normal flow fields, whatever the corresponding scenes in view are (observing the positivity of depth constraint, of course)? If the answer is affirmative, then the actual 3D motion has a unique solution. We first develop the constraints which guarantee a motion estimate to be disambiguated from the correct solution and show that for an infinite image plane we always have a unique solution; then we investigate the geometry of the constraints for a limited field of view.

4.1. Uniqueness Analysis for an Unlimited Field of View

To simplify the algebra, we rotate the x-y coordinate system so that in the new coordinate system, $y0_e$ is zero. Henceforth, $x0_e$, α_e, and β_e are with respect to the new coordinate system. For convenience of presentation, we introduce the following notation for the various terms found in the expression for D (4):

$$\begin{aligned} r(\theta) &= \left(\beta_e\left(\frac{x^2}{f}+f\right)-\alpha_e\frac{xy}{f}-\gamma_e y\right) \\ &\quad +\left(-\alpha_e\left(\frac{y^2}{f}+f\right)+\beta_e\frac{xy}{f}+\gamma_e x\right)\tan\theta \\ &= -u^x_{r_e}-u^y_{r_e}\tan\theta \end{aligned}$$

$$p(\theta)=x-x0+(y-y0)\tan\theta$$

$$q(\theta)=x-\widehat{x0}+(y-y0)\tan\theta.$$

Furthermore, denote the angles where $p(\theta)$, $q(\theta)$, and $r(\theta)$ become zero as θ_p, θ_q, and θ_r, respectively, that is,

$$\theta_p=\tan^{-1}\frac{-(x-x0)}{y-y0}$$

$$\theta_q=\tan^{-1}\frac{-(x-\widehat{x0})}{y-y0}$$

$$\theta_r=\tan^{-1}\frac{u^x_{r_e}}{-u^y_{r_e}}=\tan^{-1}\frac{-(\beta_e(x^2/f+f)-\alpha_e(xy/f)-\gamma_e y)}{(-\alpha_e(y^2/f+f)+\beta_e(xy/f)+\gamma_e x)}.$$

From this, it is immediately clear that $p(\theta_q)=-x0_e$ and $q(\theta_p)=x0_e$. Two different motions must be disambiguated, if for every point (x,y) there exists a gradient direction θ, such that $D<0$ for all positive values Z. Disregarding noise, the expression (4) for $D<0$ becomes

$$\frac{q(\theta)}{p(\theta)-r(\theta)Z}<0. \tag{14}$$

Thus, we have to find a gradient direction θ such that

$$\mathrm{sgn}(r(\theta))=\mathrm{sgn}(q(\theta))=-\mathrm{sgn}(p(\theta)). \tag{15}$$

Since $\partial p/\partial\theta=\partial q/\partial\theta$, $p(\theta)$ and $q(\theta)$ have different signs only when θ is between θ_p and θ_q. This satisfies one equality in (15), assuming $x0_e\neq 0$. With regard to the relationship of $r(\theta)$ with $p(\theta)$ and $q(\theta)$, it is sufficient to consider the relationships at the extremal angles θ_p and θ_q. In a neighborhood around θ_p, we obtain values $p(\theta)<0$ and values $p(\theta)>0$. Thus, if $r(\theta_p)q(\theta_p)>0$, then in some region around θ_p, (15) must hold. If the constraint $r(\theta_p)q(\theta_p)>0$ does not hold, it is equally admissible to have $r(\theta_q)p(\theta_q)<0$ at θ_q. To summarize, we need either $r(\theta_p)q(\theta_p)>0$ or $r(\theta_q)p(\theta_q)<0$.

Substituting the expressions for $r(\theta)$, θ_p, and θ_q, $p(\theta_q)=-x0_e$ and $q(\theta_p)$, we obtain our requirements as either of

$$(-x0_e)\left(-u^x_{r_e}-u^y_{r_e}\frac{-(x-x0)}{y-y0}\right)<0 \tag{16}$$

$$(-x0_e)\left(-u^x_{r_e}-u^y_{r_e}\frac{-(x-\widehat{x0})}{y-y0}\right)<0. \tag{17}$$

Multiplying out the terms in both (16) and (17), the requirements become either (18) or (19),

$$-x0_e(y-y0)\,(Ax^2+Bxy+Cy^2+Dx+Ey+F)>0 \tag{18}$$

$$-x0_e(y-y0)\,(\hat{A}x^2+\hat{B}xy+\hat{C}y^2+\hat{D}x+\hat{E}y+\hat{F})>0, \tag{19}$$

[1] The radius $r(t)$ and center $(a(t),b(t))$ of a circle describing locus t are given in the shifted coordinate system (12) by $r(t)=(1/2t)\sqrt{(1+t^2)(y0_e^2+(x0_e+\beta_f Z)^2)}$; $(a(t),b(t))=(\beta_f Z/2+y0_e/2t,\,-\beta_f Z/2t-x0_e/2t)$. The centers of all the circles lie on a straight line, parametrically described by $(a(t),b(t))$, or equivalently, $y=-((x0_e+\beta_f Z)/y0_e)(x-\beta_f Z/2)$, which is perpendicular to the line defined by the locus $t=0$.

where

$$A = \gamma_e + \frac{\beta_e y0}{f}$$
$$B = -\left(\frac{\beta_e x0}{f} + \frac{\alpha_e y0}{f}\right)$$
$$C = \gamma_e + \frac{\alpha_e x0}{f}$$
$$D = -(\alpha_e f + \gamma_e x0)$$
$$E = -(\beta_e f + \gamma_e y0)$$
$$F = \alpha_e f x0 + \beta_e f y0$$

and $\hat{A}, \hat{B}, \hat{C}, \hat{D}, \hat{E}, \hat{F}$ are similar to A, B, C, D, E, F, respectively, except that $x0$ is replaced by $\widehat{x0}$.

Relationship with motion flow. If two motion solutions are to be disambiguated, one of the two inequalities (18) and (19) must be satisfied. Conversely, for ambiguities to exist, both of the following must be true:

$$x0_e(y - y0)(Ax^2 + Bxy + Cy^2 + Dx + Ey + F) > 0 \quad (20)$$

$$x0_e(y - y0)(\hat{A}x^2 + \hat{B}xy + \hat{C}y^2 + \hat{D}x + \hat{E}y + \hat{F}) > 0. \quad (21)$$

The same equations expressed in a spherical coordinate system have been obtained in [3] for the cases where optic flow values can give rise to ambiguous solutions subject to the visibility constraint. As can easily be understood, requiring D to be positive in all gradient directions yields the same constraint as that obtained from requiring the same motion flow subject to the visibility constraint. It automatically follows from the studies conducted in [14, 19] that if we do not consider the positivity of depth, the scene in view must be a critical surface in order for ambiguity to arise. On the other hand, in [3], it was shown that for a half-sphere or equivalently an infinitely large image plane, the regions where (18) and (19) hold are always non-empty. Therefore the conclusion is that if we consider the constraint of positive depth, the full motion field on a half-sphere uniquely constrains the 3D motion independently of the scene in view.

Of course for a practical imaging system, we do not have an infinitely large image plane. In the following we study the geometry of the areas defined by (20) and (21) in order to investigate the potential confusion that exists between different motions for image planes of limited size.

4.2. Practical Implications

Geometry of the areas yielding negative depth. For convenience of notation, we use $f(x, y)$ to denote $Ax^2 + Bxy + Cy^2 + Dx + Ey + F$, $\hat{f}(x, y)$ to denote $\hat{A}x^2 + \hat{B}xy + \hat{C}y^2 + \hat{D}x + \hat{E}y + \hat{F}$, and $g(y)$ to denote $-x0_e(y - y0)$. To obtain the geometry of the regions described by (18) and (19), we first make the following observation: the equation $f(x, y) = 0$ describes the locus on the image plane where the translational flow vectors defined by $(x0, y0)$ are parallel to the rotational flow vectors defined by $(\alpha_e, \beta_e, \gamma_e)$, as can be readily verified by $f(x, y) = (x - x0, y - y0) \cdot (-u^y_{r_e}, u^x_{r_e})$. In [12], this locus is termed the zero iso-motion contour because the resultant flow vector could potentially be zero due to cancellation between the translational and the rotational flow vectors. The locus describes conics on the image plane, passing through the FOE $(x0, y0)$ and the point $(\frac{\alpha_e f}{\gamma_e}, \frac{\beta_e f}{\gamma_e})$, which is where the axis of rotation as defined by $(\alpha_e, \beta_e, \gamma_e)$ pierces the image plane. It separates the image plane into regions where $f(x, y)$ is positive or negative. Similar arguments apply for $\hat{f}(x, y)$, except that the FOE is moved to $(\widehat{x0}, \widehat{y0})$.

Similarly, the equation $g(y) = 0$ separates the image plane into two half-planes where $g(y)$ is positive or negative. It passes through the FOEs of both the real and the estimated motion (the coordinate system is chosen such that $y0 = \widehat{y0}$). Thus the line defines the locus of points where the real translational flow is parallel to the estimated translational flow.

Some examples of regions defined by the two conics $f(x, y)$, $\hat{f}(x, y)$ and the line $g(y)$ are shown in Fig. 11. The regions where (18) or (19) holds, that is, $g(y) f(x, y) > 0$ or $g(y) \hat{f}(x, y) > 0$, are represented by all the shaded areas. Depending on the values of $(x0, y0)$, $(\widehat{x0}, \widehat{y0})$ and $(\alpha_e, \beta_e, \gamma_e)$, the curves $f(x, y) = 0$ and $\hat{f}(x, y) = 0$ could be ellipses, hyperbolas, parabolas, or one of the degenerate forms. Figure 11 illustrates the cases where the conics are ellipses and hyperbolas. The limiting case where

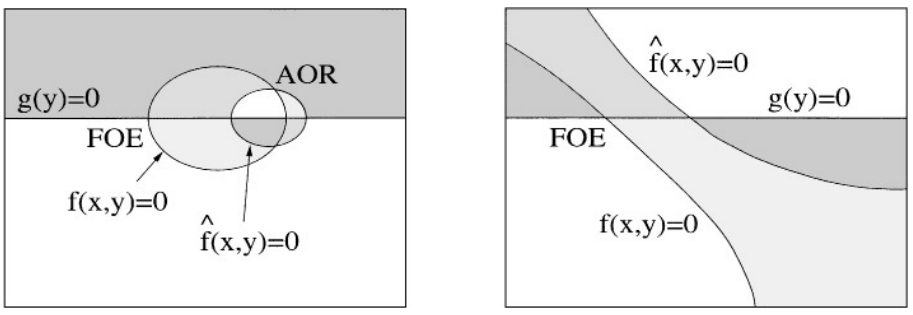


FIG. 11. Areas where motion fields due to actual and erroneous estimated motion must differ: Areas where $g(y) f(x, y) > 0$ are shown in light gray, areas where $g(y) \hat{f}(x, y) > 0$ are shown in medium gray, and areas where both $g(y) f(x, y) > 0$ and $g(y) \hat{f}(x, y) > 0$ are shown in dark gray.

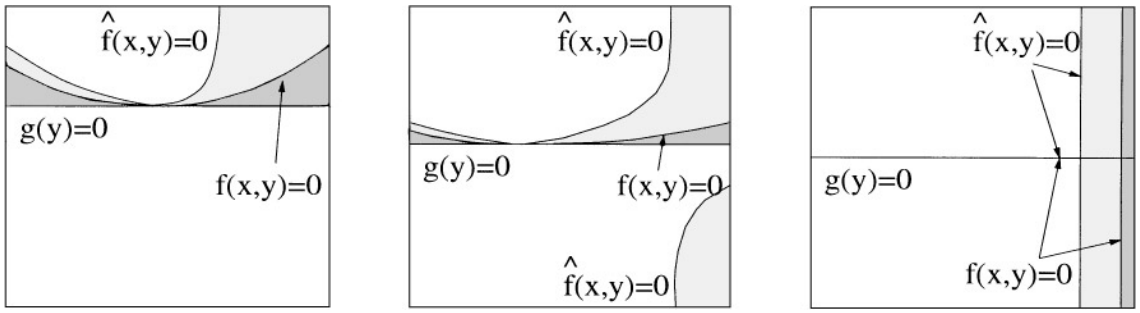


FIG. 12. Diagram illustrating the areas where the the motion fields arising from the erroneous motion estimates must differ from the true motion field, for the case where $W(x0_e, y_e) = -\lambda(\beta_e, -\alpha_e)$ for some positive λ. The diagrams from left to right illustrate the change in these areas for increasingly small $y0$. As can be seen, as $y0$ tends to zero, both $f(x, y) = 0$ and $\hat{f}(x, y) = 0$ degenerate into two pairs of intersecting lines.

an infinitely large image plane is needed for uniqueness of the solution corresponds to the case where $f(x, y)$ and $\hat{f}(x, y)$ degenerate into two intersecting lines, as shown next.

FOV for accurate FOE estimation. Consider the case investigated earlier, namely, $W(x0_e, y0_e) = -\lambda(\beta_e, -\alpha_e)$ for some positive λ or, in the rotated coordinate system, $y0_e = \alpha_e = \gamma_e = 0$ and $\frac{-Wx0_e}{\beta_e} > 0$. Then it can be shown that $f(x, y) = 0$ and $\hat{f}(x, y) = 0$ become respectively

$$\frac{\beta_e y0}{f}x^2 - \frac{\beta_e x0}{f}xy - \beta_e f(y - y0) = 0$$

$$\frac{\beta_e y0}{f}x^2 - \frac{\beta_e \widehat{x0}}{f}xy - \beta_e f(y - y0) = 0.$$

If $y0$ is non-zero, these are equations describing hyperbolas. As $y0$ becomes zero, the loci of both $f(x, y) = 0$ and $\hat{f}(x, y) = 0$ degenerate into two intersecting straight lines:

$$x = \frac{f^2}{x0}; \qquad y = y0 \quad \text{for } f(x, y) = 0 \tag{22}$$

$$x = \frac{f^2}{\widehat{x0}}; \qquad y = y0 \quad \text{for } \hat{f}(x, y) = 0. \tag{23}$$

Figure 12 shows the evolution of these curves as $y0$ approaches zero. As can be seen, as $y0$ approaches zero, the area on the image plane where the two motion solutions can be theoretically disambiguated (independently of scene depth) becomes smaller. The errors in the FOE $(x0_e, y0_e)$ are most likely to be parallel to $(x0, y0)$, so that along the direction where the projection of $(x0_e, y0_e)$ is zero, the projection of $(x0, y0)$ is also zero. Consider the case where $x0$ is also zero. Then one of the lines resulting from $f(x, y) = 0$ lies at infinity, whereas that resulting from $\hat{f}(x, y)$, if it were to approach the true solution ($\widehat{x0} \to 0$), would also lie well outside the image plane. For instance, even allowing a 5° error in $\widehat{x0}$, we still require a field of view of 170° to have the line $x = f^2/\widehat{x0}$ within the image. In other words, on the basis of equal motion flow, together with the visibility constraint, if the FOE is at (0, 0) and the depths of the scene are appropriately chosen, then all $\widehat{x0}$ estimates with less than 5° error could yield motion fields which are not distinguishable from the correct solution in a system with a field of view of 170°.

5. INERTIAL SENSOR FOR FOE ESTIMATION

5.1. Analysis

Viéville and Faugeras [25] describe how inertial and visual cues can be combined. Angular velocity information is provided by low-cost gyrometers, with a precision of 0.04°/s.[2] As analyzed in Subsection 3.1, errors in the rotational estimates lead to errors in the FOE estimates. Thus we would like to know the degree of precision required of an inertial sensor, so that the position of the FOE can be determined within some small bound σ.

For the present purpose, we content ourselves with a system which does not have a large FOV and we assume $\gamma_e = 0$. Furthermore, we do not consider the finite size of the image plane, which might introduce bias due to the position of the FOE in the image, and we assume that the gradient distribution is uniform, that is, the image gradients are uniformly distributed in every direction and at every depth within the depth range of a given scene. Now, given any rotational error (α_e, β_e) due to inaccuracies in the inertial sensor, we define a new coordinate system such that the new x-axis is along the direction given by $(\beta_e, -\alpha_e)$. Henceforth, $x0_e$, $y0_e$, α_e, β_e are taken with respect to the new coordinate system, in other words, α_e is zero.

The following analysis is based on the fact that along any gradient direction $\boldsymbol{n}$, the size of the negative distortion region always decreases at first as the FOE estimate begins to deviate from the true FOE (see Fig. 7). It reaches a minimum when the error $(x0_e, y0_e) \cdot \boldsymbol{n}$ is such that $Z = -\frac{(x0_e, y0_e) \cdot \boldsymbol{n}}{(\beta_e f, 0) \cdot \boldsymbol{n}}$, that is, the $D = 1$ contour, is at the middle of the depth range in the scene (see Fig. 7d). This depth is referred to as Z_{mid}. Beyond this error,

[2] The human inertial sensor is comparable in resolution to such low-cost sensors, the main difference being that the human angular sensor is an angular accelerometer [24].

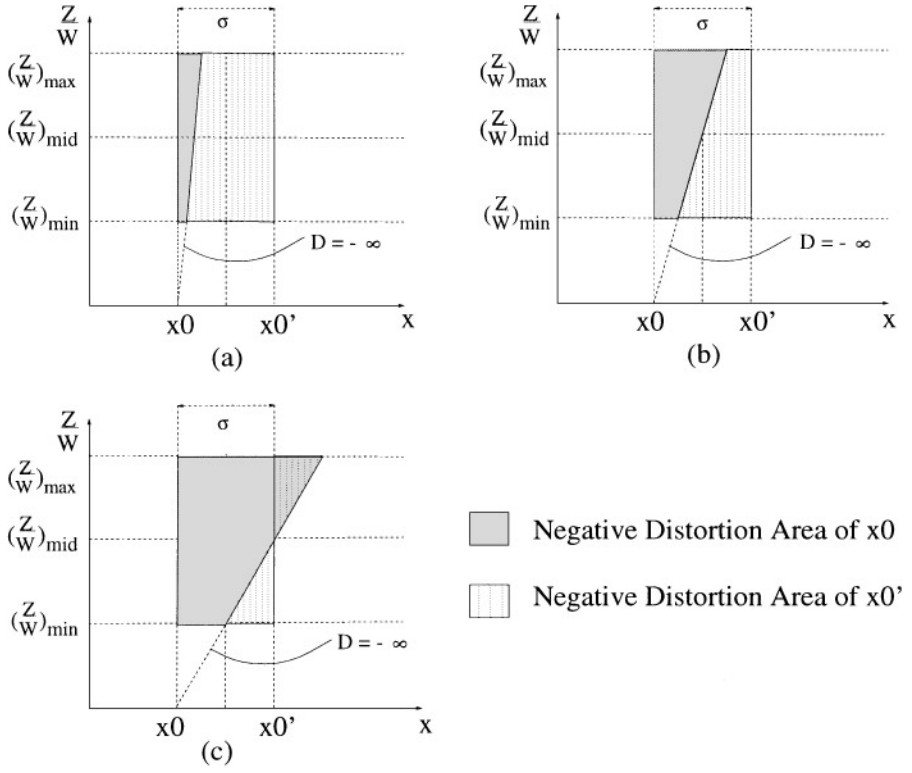


FIG. 13. The negative distortion areas of $x0$, the true FOE (represented by shaded regions), and its neighbor $x0'$ σ units away (represented by dotted regions), under varying values of $\beta_e f$. $(\frac{Z}{W})_{\min}$, $(\frac{Z}{W})_{\max}$, and $(\frac{Z}{W})_{\text{mid}}$ represent the minimum, the maximum, and the average scaled depth in the scene respectively. Our criterion (see the text) requires that the negative distortion area of $x0$ should be smaller than that of $x0'$.

the negative area increases monotonically with the size of the FOE error.

Given this relationship, what is the FOE error $(x0_e, y0_e)$ that minimizes the negative areas in all gradient directions? Since $\alpha_e = 0$, if we also set $y0_e = 0$ and $x0_e = -\beta_e f(\frac{Z}{W})_{\text{mid}}$, then the negative distortion areas in every gradient direction will be minimized, since the depth Z given by $Z = -\frac{(x0_e,0)\cdot \boldsymbol{n}}{(\beta_e f,0)\cdot \boldsymbol{n}}$ occurs at $(\frac{Z}{W})_{\text{mid}}$ for every $\boldsymbol{n}$. Thus, the overall negative distortion area is minimized. To have the desired FOE accuracy, it therefore suffices to ensure that the FOE error $x0_e$ given above is well within the required error bound σ (see Fig. 13). In other words,

$$\left|\beta_e f \left(\frac{Z}{W}\right)_{\text{mid}}\right| < \sigma.$$

Thus the desired degree of precision needed in an inertial sensor can be written as

$$|\beta_e| < \left|\frac{\sigma}{f(Z/W)_{\text{mid}}}\right|. \tag{24}$$

If angular velocity information is provided by a low-cost gyrometer with a typical precision of 0.04°/s, then (24) imposes a constraint on the type of scene in which such a system can operate. Substituting the values $\beta_e = 0.04$°/s (≈0.0007 rad/s), and $\sigma = 1°$ of visual arc, say, we obtain $(\frac{Z}{W})_{\text{mid}} < 25\ \text{s}^{-1}$. A robot moving in an indoor environment, with a typical walking speed of 1 to 2 m/s, would amply satisfy these criteria. Thus it would seem that such a low-cost inertial sensor measures up well to the task of accurate FOE estimation (here, within 1° of visual field) in an indoor environment.

5.2. Experiments

First, we used synthetic images to determine the extent to which a low-cost inertial sensor is effective for FOE estimation. The conditions of the experiments were as follows:

(1) The observer's focus of expansion was located at $x0 = 0$, $y0 = 0$. The field of view was 36°.

(2) The world features were placed randomly within the following depth ranges for each experiment: $1 < Z < 20$, $1 < Z < 50$, $10 < Z < 100$. The forward speed W was 1.8 m/s in all three cases.

(3) Noise in the normal flow was introduced as Gaussian distributed perturbations of the magnitude. The errors ranged from 0% to 10%, and they were not correlated either spatially or temporally.

(4) The estimates of α, β, and γ had errors up to 0.04°/s.

(5) Four FOE candidates were tested, with $y0_e = 0$ and $x0_e$ ranging from $-1°$ to 2° of visual field.

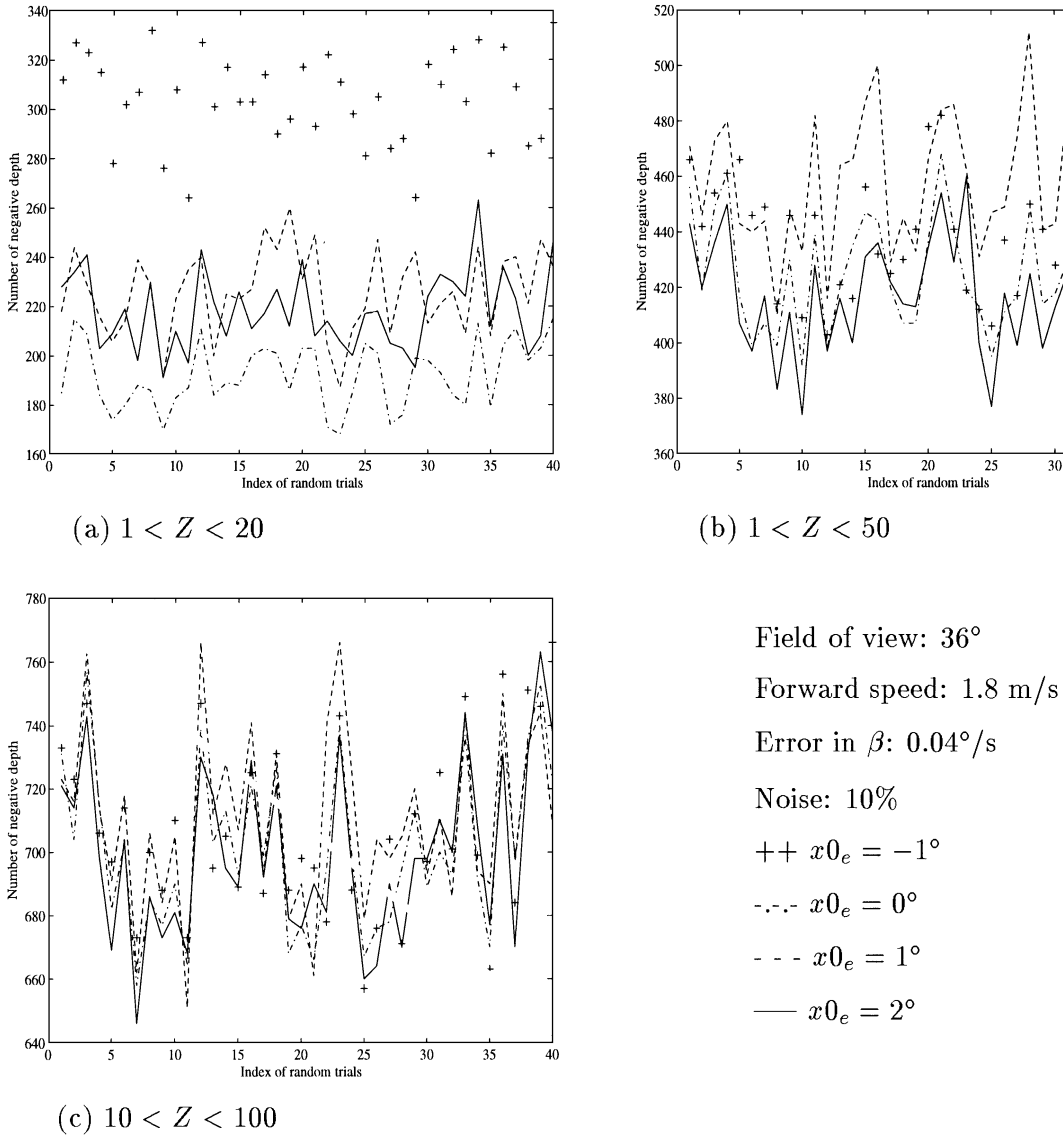


FIG. 14. Number of estimated negative depth values over 40 trials for different depth ranges.

Numbers of negative depths for the four FOE candidates were recorded over 40 random trials, and the results were represented as four different curves in each of the three graphs in Fig. 14. As can be seen by comparing Figs. 14a–14c the effect of increasing depth is to gradually degrade the efficacy of the inertial sensor estimates. Only in Fig. 14a (indoor scene) is the correct FOE candidate well separated from the other candidates. For images with greater scene distances, the local nature of the visibility constraint makes the test sensitive to the distribution of features and the effect of noise. Thus, the results support our criteria in Subsection 5.1, namely, $(\frac{Z}{W})_{\text{mid}}$ must be less than 25 s^{-1}.

These results are further confirmed by experiments performed on real images (Fig. 15a) of an indoor scene with objects about 4–5 m away. Here the different scaled depth ranges were obtained by changing the forward speed W. Estimates of the rotational velocities were again obtained from an inertial sensor with an accuracy of 0.04°/s. Figures 15b–15d show the numbers of negative depths obtained, in terms of level curves, as the FOE estimates move away from the true FOE location, each figure representing a varying value of $(\frac{Z}{W})_{\text{mid}}$. The "sinks" of the level curves represent the FOE candidates that yield the minimum number of depth values, while the true FOE is represented by the cross. It can be seen that the errors become significant when $(\frac{Z}{W})_{\text{mid}}$ increases.

6. CONCLUSIONS

The extraction of 3D motion and shape from a sequence of images represents an important problem in computational vision that has attracted much attention. When an estimate of 3D

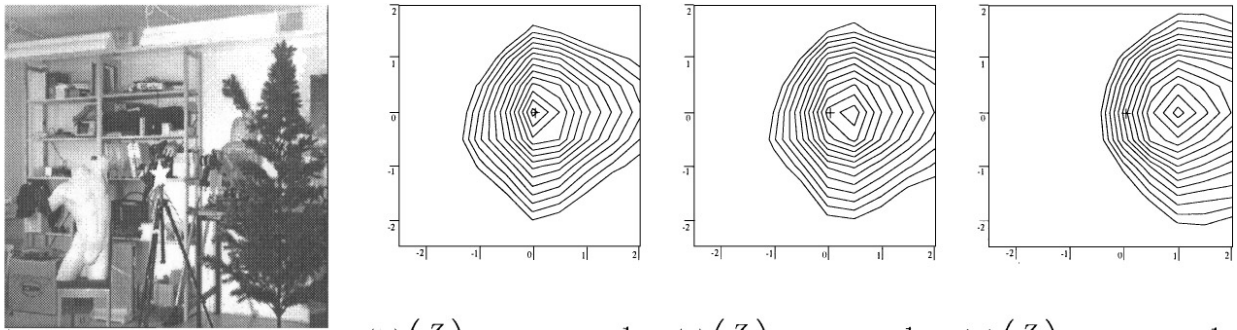


FIG. 15. The contours in (b), (c), and (d) show the variation in the number of negative depths as the FOE estimates move away from the true FOE (indicated by the cross). The best FOE estimate is denoted by the "bottom" of the contours (minimum number of negative depths). The axes of these contour plots represent the error of the FOE in degrees; they are not plotted at the same scale as the image in (a). FOV $= 30^\circ$; $\beta_e = 0.04^\circ/\text{s}$.

motion is available, it can be used with image motion measurements to estimate the structure (relative depth) of the scene in view. In this paper we have shown that when an error exists in the 3D motion estimate, the computed structure of the scene is distorted, and we have introduced the framework of iso-distortion to characterize the estimated distorted space through its level sets. We then employed this framework to present a number of geometric arguments regarding the inherent ambiguity in image sequences as far as 3D motion estimation is concerned.

REFERENCES

1. G. Adiv, Inherent ambiguities in recovering 3-D motion and structure from a noisy flow field, *IEEE Trans. Pattern Anal. Mach. Intell.* **11**, 1989, 477–489.
2. Y. Aloimonos, E. Rivlin, and L. Huang, Designing vision systems: Purposive navigation, in *Active Perception* (Y. Aloimonos, Ed.), Advances in Computer Vision, pp. 47–102, Erlbaum, Hillsdale, NJ, 1993.
3. T. Brodsky, C. Fermüller, and Y. Aloimonos, Directions of motion fields are hardly ever ambiguous, *Int. J. Comput. Vision* **26**(1), 1998, 5–24.
4. L. Cheong, *The Geometry of the Interaction between 3D Shape and Motion*, Ph.D. thesis, Department of Computer Science, University of Maryland, 1996.
5. K. Daniilidis and H.-H. Nagel, The coupling of rotation and translation in motion estimation of planar surfaces, in *Proc. IEEE Conference on Computer Vision and Pattern Recognition, 1993*, pp. 188–193.
6. K. Daniilidis and M. E. Spetsakis, Understanding noise sensitivity in structure from motion, in *Visual Navigation: From Biological Systems to Unmanned Ground Vehicles* (Y. Aloimonos, Ed.), Chap. 4, Erlbaum, Hillsdale, NJ, 1997.
7. R. Dutta and M. Snyder, Robustness of correspondence-based structure from motion, in *Proc. International Conference on Computer Vision, 1990*, pp. 106–110.
8. C. Fermüller, Passive navigation as a pattern recognition problem, *Int. J. Comput. Vision* **14**, 1995, 147–158.
9. C. Fermüller and Y. Aloimonos, Direct perception of three-dimensional motion from patterns of visual motion, *Science* **270**, 1995, 1973–1976.
10. C. Fermüller and Y. Aloimonos, Representations for active vision, in *Proc. International Joint Conference on Artificial Intelligence*, August 1995.
11. C. Fermüller and Y. Aloimonos, Ambiguity in structure from motion: sphere versus plane, *Int. J. Comput. Vision*, in press.
12. C. Fermüller and Y. Aloimonos, On the geometry of visual correspondence, *Int. J. Comput. Vision* **21**(3), 1997, 223–247.
13. C. Fermüller, L. Cheong, and Y. Aloimonos, Visual space distortion, *Biol. Cybernetics* **77**, 1997, 323–337.
14. B. K. P. Horn, Motion fields are hardly ever ambiguous, *Int. J. Comput. Vision* **1**, 1987, 259–274.
15. B. K. P. Horn and E. J. Weldon, Jr., Direct method for recovering motion, *Int. J. Comput. Vision* **2**, 1988, 51–76.
16. J. J. Koenderink and A. J. van Doorn, Relief: Pictorial and otherwise. *Image Vision Comput.* **13**, 1995, 321–334.
17. S. J. Maybank, *Theory of Reconstruction from Image Motion*, Springer-Verlag, Berlin, 1993.
18. S. Negahdaripour, Ambiguities of a motion field, in *Proc. International Conference on Computer Vision, 1987*, pp. 607–612.
19. S. Negahdaripour, Critical surface pairs and triplets, *Int. J. Comput. Vision* **3**, 1989, 293–312.
20. H. Sahabi and A. Basu, Analysis of error in depth perception with vergence and spatially varying sensing, *Comput. Vision Image Understanding* **63**(3), 1996, 447–461.
21. D. Sinclair, A. Blake, and D. Murray, Robust estimation of egomotion from normal flow, *Int. J. Comput. Vision* **13**, 1994, 57–69.
22. J. I. Thomas, A. Hanson, and J. Oliensis, Understanding noise: The critical role of motion error in scene reconstruction, in *Proc. DARPA Image Understanding Workshop, 1993*, pp. 691–695.
23. J. Todd and F. D. Reichel, Ordinal structure in the visual perception and cognition of smoothly curved surfaces, *Psychological Rev.* **96**, 1989, 643–657.
24. T. Viéville and O. D. Faugeras, Computation of inertial information on a robot, in *Fifth International Symposium on Robotics Research* (H. Miura and S. Arimoto, Eds.), pp. 57–65, MIT Press, Cambridge, MA, 1989.
25. T. Viéville and O. D. Faugeras, Cooperation of the inertial and visual systems, in *Traditional and Non-traditional Robotic Sensors* (T. Henderson, Ed.), Springer-Verlag, Berlin, 1989.
26. J. Weng, T. S. Huang, and N. Ahuja, *Motion and Structure from Image Sequences*, Springer-Verlag, Berlin, 1991.
27. G. S. Young and R. Chellappa, Statistical analysis of inherent ambiguities in recovering 3-D motion from a noisy flow field, *IEEE Trans. Pattern Anal. Mach. Intell.* **14**, 1992, 995–1013.