
International Journal of Computer Vision, Vol. no. 28, Issue No. 2, 137–154 (1998)

Ambiguity in Structure from Motion: Sphere versus Plane*

CORNELIA FERMÜLLER AND YIANNIS ALOIMONOS
Computer Vision Laboratory, Center for Automation Research, Institute for Advanced Computer Studies and Computer Science Department, University of Maryland, College Park, MD 20742-3275

fer@cfar.umd.edu

yiannis@cfar.umd.edu

Received October 30, 1996; Revised April 22, 1997; Accepted May 15, 1997

Abstract. If 3D rigid motion can be correctly estimated from image sequences, the structure of the scene can be correctly derived using the equations for image formation. However, an error in the estimation of 3D motion will result in the computation of a distorted version of the scene structure. Of computational interest are these regions in space where the distortions are such that the depths become negative, because in order for the scene to be visible it has to lie in front of the image, and thus the corresponding depth estimates have to be positive. The stability analysis for the structure from motion problem presented in this paper investigates the optimal relationship between the errors in the estimated translational and rotational parameters of a rigid motion that results in the estimation of a minimum number of negative depth values. The input used is the value of the flow along some direction, which is more general than optic flow or correspondence. For a planar retina it is shown that the optimal configuration is achieved when the projections of the translational and rotational errors on the image plane are perpendicular. Furthermore, the projection of the actual and the estimated translation lie on a line through the center. For a spherical retina, given a rotational error, the optimal translation is the correct one; given a translational error, the optimal rotational error depends both in direction and value on the actual and estimated translation as well as the scene in view. The proofs, besides illuminating the confounding of translation and rotation in structure from motion, have an important application to ecological optics. The same analysis provides a computational explanation of why it is easier to estimate self-motion in the case of a spherical retina and why shape can be estimated easily in the case of a planar retina, thus suggesting that nature's design of compound eyes (or panoramic vision) for flying systems and camera-type eyes for primates (and other systems that perform manipulation) is optimal.

Keywords: error analysis, normal flow, negative depth, spherical and planar imaging surfaces

1. Introduction

The general problem of structure from motion is defined as follows: given a number of views of a scene, to recover the rigid transformations between the views and the structure (shape) of the scene in view. In the field of computational vision a lot of effort has been devoted to this problem because it lies at the heart of several applications in pose estimation, recognition, calibration, and navigation (Faugeras, 1992; Horn, 1986). For reasons related to the tractability of the exposition and without loss of generality, we consider here the case of differential motion for a camera moving in a static environment with the goal of recovering the camera's 3D rigid motion and the structure of the scene (Bruss and Horn, 1983; Nelson and Aloimonos, 1988; Ullman, 1979). The problem has been traditionally treated

*The support of the Office of Naval Research under Grant N00014-96-1-0587 is gratefully acknowledged.

in a two-step approach. The first step attempts to establish the correspondence between successive image frames, i.e., to identify in consecutive images features that are the projections of the same feature in the 3D scene. Such correspondence is expressed through displacement vectors or optic flow—an approximation of the motion field which represents the projection of the velocity field of scene points on the image. The second step attempts to interpret this correspondence or flow field and recover 3D motion and structure.

During the Eighties, questions related to the uniqueness of solutions were answered for both the discrete case of point matches (Longuet-Higgins, 1981; Tsai and Huang, 1984), and the differential case Longuet-Higgins, 1980; Waxman and Wohn, 1985) of optical flow, and closed form solutions were studied. Similar problems were solved in the photogrammetric literature (Slama, Theurer and Henriksen, 1980). The algorithms developed during this phase of research were based on two frames (or views) and the use of point features. A bit later, in (Spetsakis and Aloimonos, 1990), algorithms were introduced for the case of three (or multiple) frames with the formulation of the trilinear constraints and these constraints were generalized in (Faugeras and Mourrain, 1995) using geometric algebra. Also, at the same time, algorithms appeared that made use of line correspondences (Spetsakis and Aloimonos, 1990), as well as algorithms that used both point and line correspondences. In the Nineties, these results were generalized to the case of uncalibrated cameras, a situation in which only projective (or, under some assumptions, affine) structure can be recovered (Åström, 1996; Faugeras, 1992b; Hartley, 1994; Koenderink and van Doorn, 1991).

The promise of the uniqueness studies gave rise to an exciting quest for practical and robust algorithms for recovering 3D structure and motion from image sequences, but this was soon to be frustrated by high sensitivity to noise in the input used (optic flow or correspondence). While many solutions have been proposed, they become problematic in the case of realistic scenes and most of them degrade ungracefully as the quality of the input deteriorates. This has motivated research on the stability of the problem; (Daniilidis and Spetsakis, 1996) contains an excellent survey of existing error analyses. We will discuss the most important results in Section 3 in more technical detail after some mathematical prerequisites are given in Section 2. In summary, it can be concluded that the majority of the existing analyses attempt to model the errors in either the 3D motion estimates or the depth estimates, and due to the large number of unknowns in the problem, they deal with restricted conditions such as planarity of the scene in view or non-biasedness of the estimators. Notably absent in published efforts is an account of the systematic nature of the errors in the depth estimates due to errors in the 3D motion estimates. Put in different terms, there exists an interplay between 3D motion and depth. In existing approaches, however, the highly correlated nature of the depth errors at different image locations, due to 3D motion errors, is not reflected adequately. Furthermore, all analyses are based on the two-step approach, analyzing the estimation of 3D motion from noise-contaminated optic flow or correspondence. However, as has been shown in previous work, the estimation of 3D motion does not necessarily require the prior computation of exact correspondence (Fermüller and Aloimonos, 1995; Fermüller and Aloimonos, 1986). Flow measurements, or even their signs, along some direction in the image, such as—for example—the one provided by the spatial gradient, are sufficient for recovering 3D motion (Brodsky, 1998). Such measurements can be computed by even the simplest systems—biological or artificial—using, for example, Reichardt detectors or equivalent energy models (Poggio and Reichardt, 1973; Reichardt, 1961; Reichardt, 1987; van Santen and Sperling, 1984).

In this paper an approach that is independent of any algorithm or estimator is taken. Due to the geometry of image formation any spatiotemporal representation in the image is due to the 3D motion and the structure of the scene in view. If the 3D motion can be estimated correctly, the structure can be derived correctly using the equations of image formation. However, an error in the estimation of the 3D motion will result in the computation of a distorted version of the actual scene structure. Of computational interest are those regions in space where the distortions are such that the depths become negative. Not considering any scene interpretation the only fact we know about the scene is that for it to be visible it has to lie in front of the image and thus the corresponding depth estimates have to be positive. Therefore the number of image points whose corresponding scene points would yield negative values due to erroneous 3D motion estimation should be kept as small as possible. This is the computational principle behind the error analysis presented in this paper. We are interested in the relationship of the errors in the translational and rotational components of the motion

for the purpose of understanding its impact on the design of optimal 3D motion estimation algorithms. In particular, assuming there is an error in the estimation of the rotational (translational) motion components, we ask what the error in the translational (rotational) components is that leads to a minimization of the negative depth values computed, and how structure from motion algorithms are affected by these error configurations. The analysis is carried out for a complete field of view as perceived by an imaging sphere, and for a restricted field of view on a constrained image plane.

2. Overview and Problem Statement

2.1. Prerequisites

We consider an observer moving rigidly with translation $\mathbf{t} = (U, V, W)$ and rotation $\boldsymbol{\omega} = (\alpha, \beta, \gamma)$ in a stationary environment. Thus each scene point $\mathbf{R} = (X, Y, Z)$ measured with respect to a coordinate system $OXYZ$ fixed to the camera's nodal point O has a velocity $\dot{\mathbf{R}} = -\mathbf{t} - \boldsymbol{\omega} \times \mathbf{R}$ relative to the camera. The image formation is based on perspective projection.

If the image is formed on a plane orthogonal to the Z axis at distance f from the nodal point (see Fig. 1) the image points $\mathbf{r} = (x, y, f)$ are related to the scene points $\mathbf{R}$ through equation

$$\mathbf{r} = \frac{f}{\mathbf{R} \cdot \mathbf{z}_0} \mathbf{R}$$

with $\mathbf{z}_0$ a unit vector in the direction of the Z axis and "$\cdot$" denoting the inner product. Thus, the 2D image velocity amounts to

$$\dot{\mathbf{r}} = \frac{\mathbf{v}_{\mathrm{tr}}(\mathbf{r})}{Z} + \mathbf{v}_{\mathrm{rot}}(\mathbf{r}) = -\frac{1}{Z}(\mathbf{z}_0 \times (\mathbf{t} \times \mathbf{r})) + \frac{1}{f}(\mathbf{z}_0 \times (\mathbf{r} \times (\boldsymbol{\omega} \times \mathbf{r}))) \tag{1}$$

where $\frac{\mathbf{v}_{\mathrm{tr}}(\mathbf{r})}{Z}$ and $\mathbf{v}_{\mathrm{rot}}(\mathbf{r})$ are the translational and rotational flow components respectively and $Z = \mathbf{R} \cdot \mathbf{z}_0$.

Similarly, if the image is formed on a sphere of radius f (i.e., $\mathbf{r} \cdot \mathbf{r} = f^2$) (see Fig. 2), the image $\mathbf{r} = (x, y, z)$ of any point $\mathbf{R}$ is

$$\mathbf{r} = \frac{\mathbf{R} f}{|\mathbf{R}|}$$

with $|\mathbf{R}|$ being the norm of $\mathbf{R}$ and denoting the range; thus the 2D image motion is

$$\dot{\mathbf{r}} = \frac{\mathbf{v}_{\mathrm{tr}}(\mathbf{r})}{|\mathbf{R}|} + \mathbf{v}_{\mathrm{rot}}(\mathbf{r}) = -\frac{1}{|\mathbf{R}| f}(\mathbf{r} \times (\mathbf{t} \times \mathbf{r})) - \boldsymbol{\omega} \times \mathbf{r} \tag{2}$$

The component of the flow u_n along any direction $\mathbf{n}$ is therefore

$$\begin{aligned} u_n &= \dot{\mathbf{r}} \cdot \mathbf{n} = \tfrac{\mathbf{v}_{\mathrm{tr}}}{Z} \cdot \mathbf{n} + \mathbf{v}_{\mathrm{rot}} \cdot \mathbf{n} \\ \text{or} \quad u_n &= \dot{\mathbf{r}} \cdot \mathbf{n} = \tfrac{\mathbf{v}_{\mathrm{tr}}}{|\mathbf{R}|} \cdot \mathbf{n} + \mathbf{v}_{\mathrm{rot}} \cdot \mathbf{n} \end{aligned} \tag{3}$$

As can be seen from (1) and (2), the effects of translation and scene structure cannot be disentangled and thus we can only obtain the direction of translation $\mathbf{t}/|\mathbf{t}|$ and the depth (range) of the scene up to a scaling factor, that is $\frac{Z}{|\mathbf{t}|}\left(\frac{|\mathbf{R}|}{|\mathbf{t}|}\right)$. For the sake of simplicity, we will assume $\mathbf{t}$ to be of length 1 and we will no longer mention the scaling in the computation of structure.

2.2. Previous work

It is in general a very hard task to develop analytical results about the stability or error sensitivity of structure from motion. This is due to the nonlinearities and the large number of parameters that are involved. As a result a fair number of observations and intuitive arguments have been developed by a multitude of authors over the years. Most important, a small number of studies have given rise to three crisp results regarding

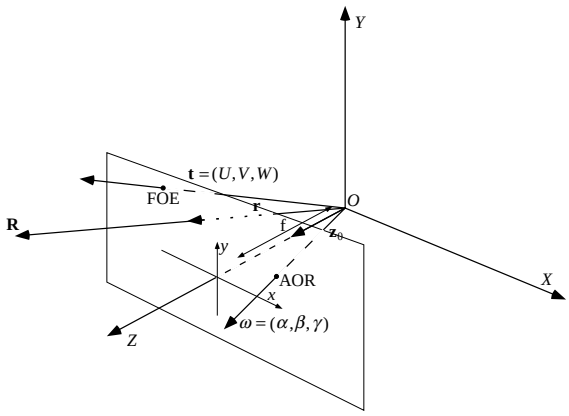


Fig. 1. Image formation under perspective projection on a planar retina: The instantaneous rigid motion is described through a translation $\mathbf{t} = (U, V, W)$ and a rotation $\boldsymbol{\omega} = (\alpha, \beta, \gamma)$. The focus of expansion (FOE), given by $(\frac{U}{W} f, \frac{V}{W} f)$, denotes the direction of translation, and the AOR (axis of rotation point), given by $(\frac{\alpha}{\gamma} f, \frac{\beta}{\gamma} f)$, denotes the intersection of the rotation axis and the image.

noise sensitivity in structure from motion (Daniilidis and Spetsakis, 1996).These are:

(a) A translation can be easily confounded with a rotation in the case of a small field of view under the assumption of lateral motion and insufficient variation of depth (Adiv, 1989; Daniilidis, 1992).More precisely, translation along the x axis can be confused with rotation around the y axis and translation along the y axis with rotation around the x axis. Evidence for this result can be obtained intuitively from the flow equation (1). As can be seen, if the scene in view is a plane, then the flow becomes a polynomial in the retinal coordinates x, y with the terms $(U + \beta, V - \alpha)$ representing the zero-order terms. A proof of this fact using techniques from estimation theory has been presented in (Daniilidis, 1992)for the case of unbiased estimators.

(b) Usually 3D motion estimation is addressed by evaluating the deviation from the epipolar constraint. Some error metric has to be developed to capture the deviation and its minimization provides a solution for 3D motion and subsequently for structure. If this metric is not appropriately normalized, in the case of a small field of view the translation estimate is biased toward the viewing direction. This can be seen directly from the epipolar constraint. By projecting both sides of (1) onto $\mathbf{z}_0 \times (\mathbf{z}_0 \times (\mathbf{t} \times \mathbf{r}))$ and setting $f = 1$, we obtain the epipolar constraint in its instantaneous form as $(\mathbf{t} \times \mathbf{r}) \cdot (\dot{\mathbf{r}} - \boldsymbol{\omega} \times \mathbf{r}) = 0$.
A solution coming from the minimization of $\sum_i ((\mathbf{t} \times \mathbf{r}_i) \cdot (\dot{\mathbf{r}}_i - \boldsymbol{\omega} \times \mathbf{r}_i))^2$ is bound to be biased, because the crossproduct $(\mathbf{t} \times \mathbf{r}_i)$ introduces the sine of the angle between $\mathbf{t}$ and $\mathbf{r}_i$. Thus the minimization prefers vectors $\mathbf{t}$ that are close to the center of gravity of the points $\mathbf{r}_i$ (Spetsakis, 1994; Spetsakis and Aloimonos, 1988).Techniques from statistics such as maximum likelihood estimation (Spetsakis, 1994) or Rayleigh optimization (Spetsakis and Aloimonos, 1988)can be used to deal with the bias, but they have their own problems.

(c) The third result is due to Maybank (Maybank, 1986; Maybank, 1987; Maybank, 1993) and also Jepson (Jepson and Heeger, 1990)who showed that in the case of a small field of view, a translation $\mathbf{t}$ far away from the image points where measurements are taken, and an irregular surface, the cost function $\sum_i ((\mathbf{t} \times \mathbf{r}_i) \cdot (\dot{\mathbf{r}}_i - \boldsymbol{\omega} \times \mathbf{r}_i))^2$, resulting from the epipolar constraint, takes its minima along a line in the space of translation directions which passes through the true translation and the viewing direction. This means that the tilt of the direction of $\mathbf{t}$ can be estimated more reliably than its slant.

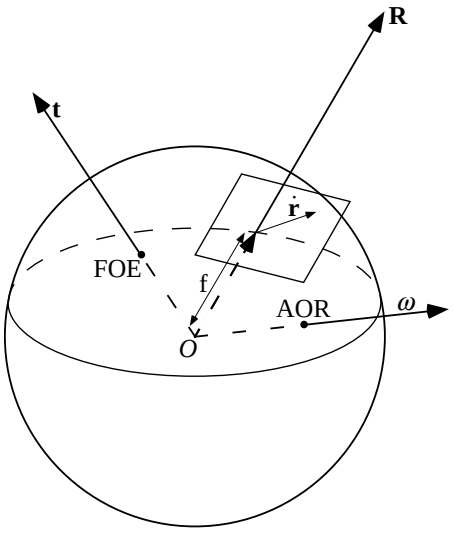


Fig. 2. Image formation under perspective projection on a spherical retina.

Insight into the instability of the motion estimation problem is also gained from uniqueness studies. Multiple 3D-motion and scene configurations giving rise to the same flow field (Horn, 1987; Maybank,1993) or point correspondences (Longuet-Higgins, 1981; Tsai and Huang, 1984) have been investigated and it has been found that the scene in view has to correspond to certain quadrics, the so-called ambiguity-critical surfaces. The study of directions of motion fields conducted in (Brodsky et al, 1998)for an unbounded image plane ambiguity can only occur for motion configurations constrained by $(\mathbf{t} \times \hat{\mathbf{t}}) \cdot (\boldsymbol{\omega} \times \hat{\boldsymbol{\omega}}) = 0$. In the context of error sensitivity a perturbation analysis has been performed in (Horn, 1990) showing that the epipolar constraint is not affected by first-order deformations of the motion parameters if the points lie on a quadric surface with certain properties. The relationship between these instability-critical surfaces and the ambiguity-critical surfaces has been established in (Daniilidis, 1992; Hofmann, 1953).

Next, we study the relationship between errors in the estimation of the 3D motion and errors in the estimation

of the depth of the scene. This relationship is the basis for our subsequent error analysis.

2.3. Distorted space

Based on an exact computation of the motion parameters the depth (range) can be derived from (3). Let us assume, however, there is an error in the estimation of the five motion parameters, that is the two parameters of the direction of translation and the three parameters of rotation. As a consequence there will also be errors in the estimation of depth (range) and thus a distorted version of the space will be computed. A convenient way to describe the distortion of space is to sketch it through surfaces in space which are distorted by the same multiplicative factor, the iso-distortion surfaces (Cheong, Fermüller and Aloimonos, 1998).

In the following, in order to distinguish between the various estimates, we use letters with hat signs to represent the estimated quantities $(\hat{\mathbf{t}}, \hat{\boldsymbol{\omega}}, |\hat{\mathbf{R}}|, \hat{Z}, \hat{\mathbf{v}}_{\rm tr}, \hat{\mathbf{v}}_{\rm rot})$ and unmarked letters to represent the actual quantities $(\mathbf{t}, \boldsymbol{\omega}, |\mathbf{R}|, Z, \mathbf{v}_{\rm tr}, \mathbf{v}_{\rm rot})$. The subscript "$\epsilon$" is used to denote errors, where we define $\boldsymbol{\omega} - \hat{\boldsymbol{\omega}} = \boldsymbol{\omega}_\epsilon$ and $\mathbf{v}_{\rm rot} - \hat{\mathbf{v}}_{\rm rot} = \mathbf{v}_{{\rm rot}_\epsilon}$.

The estimated depth or range can be obtained from (3) as

$$\hat{Z} \text{ (or } |\hat{\mathbf{R}}|) \;=\; \frac{\hat{\mathbf{v}_{\rm tr}} \cdot \mathbf{n}}{\dot{\mathbf{r}} \cdot \mathbf{n} - \hat{\mathbf{v}}_{\rm rot} \cdot \mathbf{n}}$$

and we have on the image plane

$$\hat{Z} = Z \left(\frac{-f \left(\mathbf{z}_0 \times \left(\hat{\mathbf{t}} \times \mathbf{r}\right)\right) \cdot \mathbf{n}}{\begin{array}{c} -f \left(\mathbf{z}_0 \times (\mathbf{t} \times \mathbf{r})\right) \cdot \mathbf{n} \\ +\; Z \left(\mathbf{z}_0 \times (\mathbf{r} \times (\boldsymbol{\omega}_\epsilon \times \mathbf{r}))\right) \cdot \mathbf{n} \end{array}} \right) \tag{4}$$

and on the image sphere

$$|\hat{\mathbf{R}}| = |\mathbf{R}| \cdot \left(\frac{\left(\mathbf{r} \times \left(\hat{\mathbf{t}} \times \mathbf{r}\right)\right) \cdot \mathbf{n}}{\left(\mathbf{r} \times (\mathbf{r} \times \mathbf{t})\right) \cdot \mathbf{n} + f|\mathbf{R}| \left(\boldsymbol{\omega}_\epsilon \times \mathbf{r}\right) \cdot \mathbf{n}} \right) \tag{5}$$

From (4) it can be seen that $\hat{Z}$ can be expressed as a multiple of Z, where the multiplicative factor, which we denote by D, the distortion factor, is given by the term inside the brackets. Thus the distortion factor is

$$D = \frac{-f \left(\mathbf{z}_0 \times \left(\hat{\mathbf{t}} \times \mathbf{r}\right)\right) \cdot \mathbf{n}}{\begin{array}{c} -f \left(\mathbf{z}_0 \times (\mathbf{t} \times \mathbf{r})\right) \cdot \mathbf{n} \\ +\; Z \left(\mathbf{z}_0 \times (\mathbf{r} \times (\boldsymbol{\omega}_\epsilon \times \mathbf{r}))\right) \cdot \mathbf{n} \end{array}} \tag{6}$$

Similarly we can interpret the estimated range in (5) as a multiple of the actual range with distortion D, where

$$D = \frac{\left(\mathbf{r} \times \left(\hat{\mathbf{t}} \times \mathbf{r}\right)\right) \cdot \mathbf{n}}{\left(\mathbf{r} \times (\mathbf{t} \times \mathbf{r})\right) \cdot \mathbf{n} + f|\mathbf{R}| \left(\boldsymbol{\omega}_\epsilon \times \mathbf{r}\right) \cdot \mathbf{n}} \tag{7}$$

Equations (6) and (7) describe, for any fixed direction $\mathbf{n}$ and any distortion factor D, a surface in space. Any such surface is to be understood as the locus of points in space which are distorted in depth (range) by the same factor D, if the corresponding image measurements are in direction $\mathbf{n}$.

It should be emphasized that the distortion of depth also depends on the direction $\mathbf{n}$ of the flow measurement (hereafter called the flow direction) used as basis for the computations and therefore is different for different directions of flow. This means simply that if one estimates depth from optical flow in the presence of errors, the results can be very different, depending on whether the horizontal, vertical, or any other component is used. Depending on the direction, any value between $+\infty$ and $-\infty$ can be obtained!

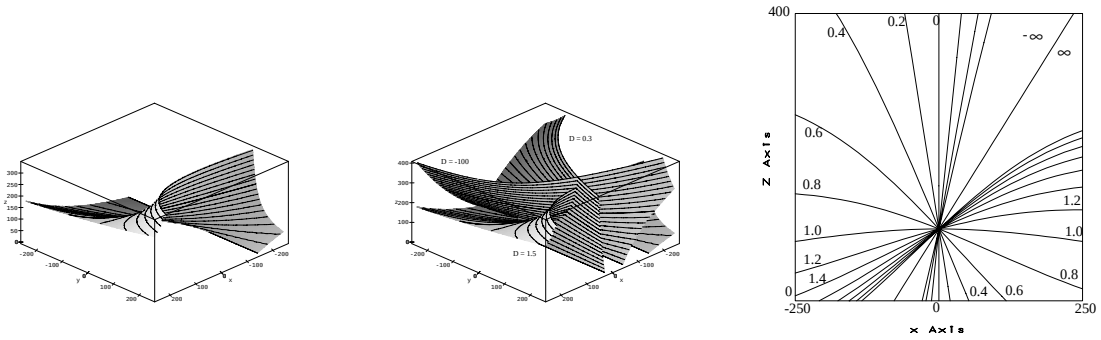
In the analysis in this paper, we are not interested in actual 3D space, but we consider the surfaces in visual space, that is, the space perceived under perspective projection where the dimensions parallel to the image are measured according to the size with which they appear on the image.

Fig. 3a gives an example of an iso-distortion surface, and Fig. 3b illustrates a family of iso-distortion surfaces corresponding to the same gradient direction but different distortion factors D. The same family is intersected with the xZ plane in Fig. 3c. In the plane the intersections give rise to a family of contours.

As can be seen the iso-distortion surfaces of a family intersect in a curve, and they change continuously as we vary D. Thus all the space between the 0 distortion surface and the $-\infty$ distortion surface (which is also the $+\infty$ distortion surface) is distorted by a negative distortion factor.

2.4. Description of results

In the forthcoming sections we employ a geometric statistical model to represent the negative depth values. We assume that the scene in view lies within a certain depth (range) interval between a minimum value and a maximum value. The flow representation vectors in the image are in different directions, and we assume some distribution for their directions. Our focus is on the



(a) (b) (c)

Fig. 3. (a) Iso-distortion surface in xyZ space. The parameters are: $(x_0, y_0) = (\frac{U}{W}f, \frac{V}{W}f) = (-50, -25)$, $(\hat{x}_0, \hat{y}_0) = (\frac{\hat{U}}{\hat{W}}f, \frac{\hat{V}}{\hat{W}}f) = (0, -20)$, $\boldsymbol{\omega}_\epsilon = (\alpha_\epsilon, \beta_\epsilon, \gamma_\epsilon) = (-0.005, 0.001, 0.003)$, $D = 1.5$, $\mathbf{n} = (1, 0)$, $f = 500$ (corresponding to a field of view of $50°$). (b) Family of iso-distortion surfaces for the same motion parameters ($\mathbf{n} = (1, 0)$). (c) Corresponding iso-distortion contours in the xZ plane.

points in space which for a 3D motion estimate yield negative depth (range) estimates.

For every direction $\mathbf{n}$ the points in space with negative depth estimates cover the space between the 0 and $-\infty$ distortion surface within the range covered by the scene. Thus for every direction we obtain a 3D subspace, covering a certain volume. The sum of all volumes for all directions, normalized by the flow distributions considered, represents a measure of the likelihood that negative depth values occur. We call it the "negative depth volume" or "negative range volume." The idea behind our error analysis lies in the minimization of this negative depth (range) volume—that is, we are interested in the relationship between the translational and rotational motion errors that minimizes this volume.

In our analysis we do not want to make any particular scene-related assumptions favoring particular orientations or depth values. We wish to treat all depth values and flow directions as having equal importance. To be more precise, we assume that the flow directions are uniformly distributed in every direction and at every depth (range) between a minimum value $Z_{\min}(|\mathbf{R}_{\min}|)$ and a maximum value $Z_{\max}(|\mathbf{R}_{\max}|)$. We do not wish to assume any particular distribution for the noise in the flow measurements. Therefore, we do not consider any noise in the measurements. The underlying idea is that noise which is of no particular bias does not change the qualitative characteristics of the function resulting from the minimization of negative depth values. Thus, one can view our analysis as a geometric investigation of the inherent confounding of translation and rotation, which is the reason behind the instability in structure from motion.

In summary, as an answer to the question about the coupling of motion errors, the following results are obtained:

(a) If we take the whole sphere as the imaging surface and we assume an error in the estimation of rotation, then the direction of translation that minimizes the negative depth volume is the correct direction of translation.
The practical implication of this result is that 3D motion estimation is most easily accomplished for a complete field of view, as provided by an imaging sphere. A working system (biological or artificial) is usually equipped with an inertial sensor which provides rotational information, though probably with some error. On the basis of this information, the best one can do to estimate the remaining translation is to assume that the flow field obtained by subtracting the estimated rotation is purely translational and apply a simple algorithm designed for only translation (Aloimonos and Duric, 1994; Horn and Weldon, Jr., 1988; Negahdaripour, 1986; Sinclair, Blake and Murray, 1994).
Such algorithms, if based only on the constraint that the depth is positive, are formulated basically as constrained minimization problems. The underlying idea is illustrated in Fig. 4. Assuming the observer is approaching the scene, the exact 2D motion vector at every point is away from the

FOE (the point where the translation axis pierces the image). Therefore, knowing the projection $\mathbf{u}_n$ of the flow vector on some direction $\mathbf{n}$, we know the FOE is to be found in the shaded half-plane defined by line ϵ. Thus the estimation of the translational direction can be implemented by simply voting for a half-plane at every point. The best solution corresponds to the location with the highest number of votes.

Estimation of purely translational motion is much simpler than estimation of complete 3D rigid motion, which requires techniques that decouple the translation from the rotation in some way, and if designed on the basis of the constraint of positive depth, require voting in higher dimensions (Fermüller and Aloimonos, 1995; Fermüller and Aloimonos, 1995; Fermüller and Aloimonos, 1997).

As demonstrated in the forthcoming analysis, however, a simple algorithm designed for translation only will find the correct solution. Thus insects with spherical eyes, such as bees and flies, have a big advantage in the task of 3D motion estimation.

(b) On the other hand, if we assume a certain error in the estimation of translation on a spherical image, the vector of the rotational error $\boldsymbol{\omega}_\epsilon$ lies in different directions on the sphere, whose exact value not only depends on the actual and estimated translation, but also on the range of the scene in view. However, if an estimate of translation is available, the estimation of rotation (which does not depend on depth) on the basis of negative depth, is algorithmically not meaningful. Thus, for the purpose of analyzing 3D motion estimation algorithms this error configuration is of no interest and will not be further investigated. We will, however, investigate it for the planar retina, because in this case it provides insights about global optimization algorithms (simultaneous estimation of all 3D motion parameters) as well as shape estimation.

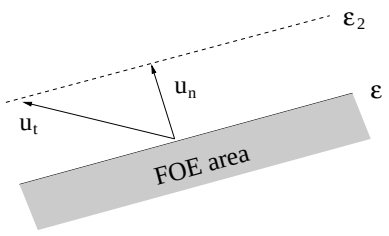


Fig. 4. The translational flow vector $_t$ has its tip anywhere along the line ε_2. The focus of expansion lies on the (shaded) half plane defined by the line ε that does not contain possible vectors $_t$.

(c) Considering as imaging surface a plane of limited extent, we find that the translational and rotational errors are perpendicular to each other. Using the notation $\frac{Uf}{W} - \frac{\hat{U}f}{\hat{W}} = x_{0_\epsilon}$ and $\frac{Vf}{W} - \frac{\hat{V}f}{\hat{W}} = y_{0_\epsilon}$, this means that $\frac{x_{0_\epsilon}}{y_{0_\epsilon}} = -\frac{\beta_\epsilon}{\alpha_\epsilon}$. If we fix the rotational error $(\alpha_\epsilon, \beta_\epsilon, \gamma_\epsilon)$, this provides us with a constraint on the direction of the translational error.

(d) If we fix the translational error $(x_{0_\epsilon}, y_{0_\epsilon})$ we obtain the same constraint, and in addition we find that $\gamma_\epsilon = 0$. Furthermore, if we fix only the amount of translational error $\sqrt{{x_{0_\epsilon}}^2 + {y_{0_\epsilon}}^2}$, we find that the actual and estimated FOE lie on a line passing through the image center, that is, $\frac{x_0}{y_0} = \frac{\hat{x_0}}{\hat{y_0}}$.

The results developed in this paper have a clear relationship with those of existing error analyses as described in Section 2.2, with the exception of the bias of translation towards the viewing direction, since this result has been obtained on the basis of particular minimization functions.

Regarding the confusion between translation and rotation, it has been experimentally observed and proven for simple scene structures, restricted fields of view, and certain estimation techniques using particular statistical estimators, that the translation along the x axis is coupled with rotation around the y axis and that the translation along the y axis is coupled with rotation around the x axis. Our results are in accordance with these findings. In addition, we show that for the minimization studied here, the confusions between x-translation and y-rotation and y-translation and x-rotation are not decoupled, and that the rotation around the Z axis is estimated most easily.

The results regarding the preference for estimation of translations whose projections lie on a line through the center and the actual projection of translation are consistent with those found in (Jepson and Heeger, 1990; Maybank, 1986). Here it is shown that in general for a limited field of view, there exists such a bias which previously has been proven for translation directions far away from the image points where measurements are taken on the basis of a cost function resulting from the epipolar constraint.

The importance of the results obtained for the plane also lies in their consequences for shape estimation. They can be translated into the statement

that planar retinas are advantageous in the computation of shape. As will be shown in Section 5, if $\frac{x_{0_\epsilon}}{y_{0_\epsilon}} = -\frac{\beta_\epsilon}{\alpha_\epsilon} = \frac{x_0}{y_0}$, and $\gamma_\epsilon = 0$, a depth map can be derived such that all scene points of the same depth are distorted by the same factor. Thus the level contours of the depth map are the correct ones!

3. Analysis on the Sphere

The following analysis considers a fixed rotational error. We need a parameterization for expressing all possible orientations $\mathbf{n}$ tangent to the sphere at every point. One way to achieve this that is convenient for our problem is through the selection of unit vectors of an arbitrary plane. Given a unit vector $\mathbf{s}$, at each point $\mathbf{r}$ of the sphere, the vector $\frac{\mathbf{s}\times\mathbf{r}}{|\mathbf{s}\times\mathbf{r}|}$ defines a direction at the tangent plane. As $\mathbf{s}$ varies along half a great circle, $\frac{\mathbf{s}\times\mathbf{r}}{|\mathbf{s}\times\mathbf{r}|}$ takes on every possible orientation in the tangent plane at every point $\mathbf{r}$ with the exception of the set of points $\mathbf{r}$ lying on the great circle of $\mathbf{s}$, which is of measure zero. To facilitate the analysis, we choose $\mathbf{s}$ perpendicular to $\boldsymbol{\omega}_\epsilon$.

As shown in Fig. 5, let $\boldsymbol{\omega}_\epsilon$ be parallel to the x axis and let $\mathbf{s}$ be the set of all the unit vectors in the yz plane with $\mathbf{s} = (0, \sin\chi, \cos\chi)$ and χ in the interval $[0 \ldots \pi]$. The flow directions $\mathbf{n}$ at every point are defined as $\mathbf{n} = \frac{\mathbf{s}\times\mathbf{r}}{|\mathbf{s}\times\mathbf{r}|}$. This parameterization, however, does not treat all orientations equally (as $\mathbf{s}$ varies along a great circle with constant speed, $\mathbf{s}\times\mathbf{r}$ accelerates and decelerates). In order to obtain a uniform distribution we must perform some normalization. Luckily, however, this normalization does not complicate matters in the following proof because, due to symmetry, its behavior with regard to monotonicity is the same as the one of the volumes of negative range for the functions considered.

We assume a uniform distribution for the directions $\mathbf{n}$. Thus, in order to obtain the negative range volume V_n, we have to integrate the individual volumes in each direction over all directions. If $\psi \in [0, \pi]$ provides a uniform parameterization for $\mathbf{n}$, as given in the appendix, $V(\psi)$ is the volume for a single direction $\mathbf{n}(\psi)$, and χ is the parameterization for $\mathbf{n}$ as defined above, the following transformation applies:

$$V_n = \int_0^\pi V(\psi)\, d\psi = \int_{g^{-1}(0)}^{g^{-1}(\pi)} V\left(g(\chi)\right) \left|\frac{\partial g(\chi)}{\partial \chi}\right| d\chi$$

where $\psi = g(\chi)$. For this parameterization the normalization term is

$$\left|\frac{\partial g(\chi)}{\partial \chi}\right| = \left|\frac{\sin\varphi_y}{\cos(\varphi_y)^2\cos(\chi-\varphi_x)^2 - 1}\right| \quad (8)$$

where φ_y is the angle between vector $\mathbf{r}$ and the yz plane, and φ_x is the angle between the projection of $\mathbf{r}$ on the yz plane and the positive y-coordinate axis. A derivation is given in the appendix.

Our focus is on the points in space with estimated negative range values $|\hat{\mathbf{R}}|$. Since $\mathbf{n} = \frac{\mathbf{s}\times\mathbf{r}}{|\mathbf{s}\times\mathbf{r}|}$ and $\mathbf{s}\cdot\boldsymbol{\omega}_\epsilon = 0$, we obtain from (5), by setting $f = 1$,

$$\left|\hat{\mathbf{R}}\right| = |\mathbf{R}|\,\frac{\left(\hat{\mathbf{t}}\times\mathbf{s}\right)\cdot\mathbf{r}}{(\mathbf{t}\times\mathbf{s})\cdot\mathbf{r} - |\mathbf{R}|\,(\boldsymbol{\omega}_\epsilon\cdot\mathbf{r})\,(\mathbf{s}\cdot\mathbf{r})} < 0 \quad (9)$$

From this inequality the following constraint on $|\mathbf{R}|$ can be derived:

$$\begin{aligned}&\mathrm{sgn}\left(\left(\hat{\mathbf{t}}\times\mathbf{s}\right)\cdot\mathbf{r}\right)\\ &\quad = -\mathrm{sgn}\left((\mathbf{t}\times\mathbf{s})\cdot\mathbf{r} - |\mathbf{R}|\,(\boldsymbol{\omega}_\epsilon\cdot\mathbf{r})\,(\mathbf{s}\cdot\mathbf{r})\right) \quad (10)\end{aligned}$$

At any point $\mathbf{r}$ in the image this constraint is either satisfied for all values $|\mathbf{R}|$, or it is satisfied for an interval of values $|\mathbf{R}|$ bounded from either above or below, or it is not satisfied for any value at all. Thus, inequality (9) provides a classification for the points on the sphere, and we obtain four different kinds of areas (types I–IV). The locations of these areas are defined by the signs of

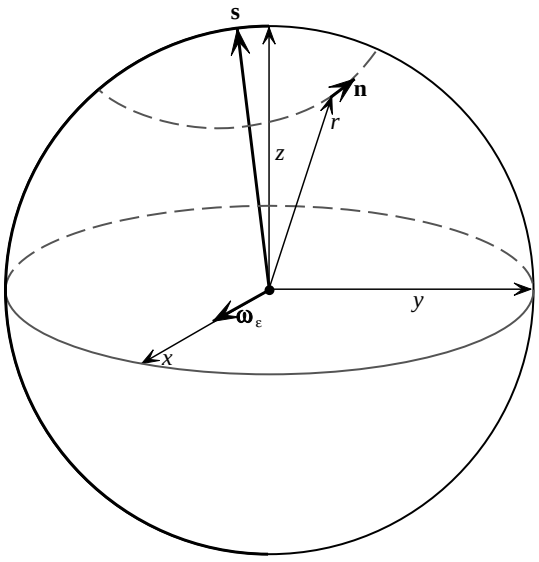


Fig. 5. Parameterization used in the analysis: $\boldsymbol{\omega}_\epsilon = \lambda(1,0,0)$, $\mathbf{s} = (0, \sin\chi, \cos\chi)$ with $\chi \in [0 \ldots \pi]$, $\mathbf{n} = \frac{\mathbf{s}\times\mathbf{r}}{|\mathbf{s}\times\mathbf{r}|}$.

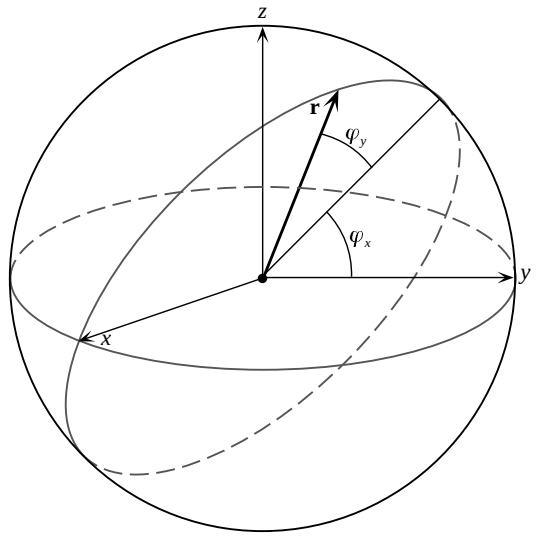


Fig. 6. Parameterization of $\mathbf{r}$: φ_y is the angle between $\mathbf{r}$ and the yz plane; φ_x is the angle between the projection $\mathbf{r}$ on the yz plane and the positive y-coordinate axis.

Table 1.

area	location	constraint on $\lvert\mathbf{R}\rvert$
I	$\begin{aligned}&\operatorname{sgn}((\mathbf{t}\times\mathbf{s})\cdot\mathbf{r})\\ =\ &\operatorname{sgn}((\hat{\mathbf{t}}\times\mathbf{s})\cdot\mathbf{r})\\ =\ &\operatorname{sgn}((\mathbf{r}\cdot\boldsymbol{\omega}_\epsilon)(\mathbf{r}\cdot\mathbf{s}))\end{aligned}$	$\lvert\mathbf{R}\rvert > \dfrac{(\mathbf{t}\times\mathbf{s})\cdot\mathbf{r}}{(\mathbf{r}\cdot\boldsymbol{\omega}_\epsilon)(\mathbf{r}\cdot\mathbf{s})}$
II	$\begin{aligned}&-\operatorname{sgn}((\mathbf{t}\times\mathbf{s})\cdot\mathbf{r})\\ =\ &\operatorname{sgn}((\hat{\mathbf{t}}\times\mathbf{s})\cdot\mathbf{r})\\ =\ &\operatorname{sgn}((\mathbf{r}\cdot\boldsymbol{\omega}_\epsilon)(\mathbf{r}\cdot\mathbf{s}))\end{aligned}$	all $\lvert\mathbf{R}\rvert$
III	$\begin{aligned}&\operatorname{sgn}((\mathbf{t}\times\mathbf{s})\cdot\mathbf{r})\\ =\ &-\operatorname{sgn}((\hat{\mathbf{t}}\times\mathbf{s})\cdot\mathbf{r})\\ =\ &\operatorname{sgn}((\mathbf{r}\cdot\boldsymbol{\omega}_\epsilon)(\mathbf{r}\cdot\mathbf{s}))\end{aligned}$	$\lvert\mathbf{R}\rvert < \dfrac{(\mathbf{t}\times\mathbf{s})\cdot\mathbf{r}}{(\mathbf{r}\cdot\boldsymbol{\omega}_\epsilon)(\mathbf{r}\cdot\mathbf{s})}$
IV	$\begin{aligned}&\operatorname{sgn}((\mathbf{t}\times\mathbf{s})\cdot\mathbf{r})\\ =\ &\operatorname{sgn}((\hat{\mathbf{t}}\times\mathbf{s})\cdot\mathbf{r})\\ =\ &-\operatorname{sgn}((\mathbf{r}\cdot\boldsymbol{\omega}_\epsilon)(\mathbf{r}\cdot\mathbf{s}))\end{aligned}$	none

the functions $(\hat{\mathbf{t}}\times\mathbf{s})\cdot\mathbf{r}$, $(\mathbf{t}\times\mathbf{s})\cdot\mathbf{r}$ and $(\boldsymbol{\omega}_\epsilon\cdot\mathbf{r})(\mathbf{s}\cdot\mathbf{r})$, as summarized in Table 1.

Thus for any direction $\mathbf{n}$ defined by a certain $\mathbf{s}$, we obtain a volume of negative range values consisting of the volumes above areas I, II, and III. An illustration for both hemispheres is given in Fig. 7. As can be seen, areas II and III cover the same amount of area, which has the size of the area between the two great circles $(\mathbf{t}\times\mathbf{s})\cdot\mathbf{r}=0$ and $(\hat{\mathbf{t}}\times\mathbf{s})\cdot\mathbf{r}=0$, and area I covers a hemisphere minus the area between $(\mathbf{t}\times\mathbf{s})\cdot\mathbf{r}=0$ and $(\hat{\mathbf{t}}\times\mathbf{s})\cdot\mathbf{r}=0$.

If the scene in view is unbounded, that is, $|\mathbf{R}|\in[0\ldots\infty]$, there is a range of values $|\mathbf{R}|$ above any point $\mathbf{r}$ in areas I and III which results in negative range estimates. If we consider a lower bound $|\mathbf{R}_{\min}|\neq 0$ and an upper bound $|\mathbf{R}_{\max}|\neq\infty$, we obtain two additional curves $C_{\min}$ and $C_{\max}$ with $C_{\min}=(\mathbf{t}\times\mathbf{s})\cdot\mathbf{r}-|\mathbf{R}_{\min}|\,(\boldsymbol{\omega}_\epsilon\cdot\mathbf{r})(\mathbf{s}\cdot\mathbf{r})=0$ and $C_{\max}=(\mathbf{t}\times\mathbf{s})\cdot\mathbf{r}-|\mathbf{R}_{\max}|\,(\boldsymbol{\omega}_\epsilon\cdot\mathbf{r})(\mathbf{s}\cdot\mathbf{r})=0$ as bounds for areas with negative range values (as shown in Fig. 7). As can be

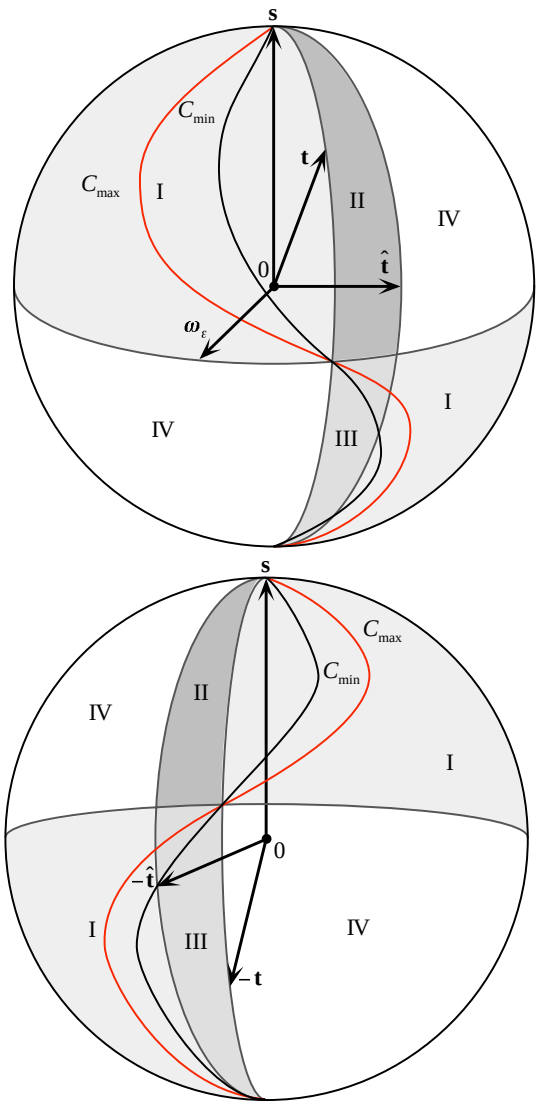


Fig. 7. Classification of image points according to constraints on $|\mathbf{R}|$. At $C_{\min}$ and $C_{\max}$, $|\mathbf{R}|$ is constrained to be greater (area I) or smaller (area III) than $|\mathbf{R}_{\min}|$ or $|\mathbf{R}_{\max}|$. The two hemispheres correspond to the front of the sphere and the back of the sphere, both as seen from the front of the sphere.

seen, the curves $C_{\min} = 0$, $C_{\max} = 0$, $(\mathbf{t} \times \mathbf{s}) \cdot \mathbf{r} = 0$ and $(\boldsymbol{\omega}_\epsilon \cdot \mathbf{r})(\mathbf{s} \cdot \mathbf{r}) = 0$ intersect at the same point.

In area I, we do not obtain any volume of negative range estimates for points $\mathbf{r}$ between the curves $(\boldsymbol{\omega}_\epsilon \cdot \mathbf{r})(\mathbf{s} \cdot \mathbf{r}) = 0$ and $C_{\max} = 0$; the volume for points $\mathbf{r}$ between $C_{\min} = 0$ and $C_{\max} = 0$ is bounded from below by $|\mathbf{R}| = \frac{(\mathbf{t} \times \mathbf{s}) \cdot \mathbf{r}}{(\boldsymbol{\omega}_\epsilon \cdot \mathbf{r})(\mathbf{s} \cdot \mathbf{r})}$ (and from above by $|\mathbf{R}_{\max}|$) and the volume for points $\mathbf{r}$ between $C_{\min} = 0$ and $(\mathbf{t} \times \mathbf{s}) \cdot \mathbf{r} = 0$ extends from $|\mathbf{R}_{\min}|$ to $|\mathbf{R}_{\max}|$. In area III we do not obtain any volume for points $\mathbf{r}$ between $(\mathbf{t} \times \mathbf{s}) \cdot \mathbf{r} = 0$ and $C_{\min} = 0$. The volume for points $\mathbf{r}$ between $C_{\min} = 0$ and $C_{\max} = 0$ is bounded from above by $|\mathbf{R}| = \frac{(\mathbf{t} \times \mathbf{s}) \cdot \mathbf{r}}{(\boldsymbol{\omega}_\epsilon \cdot \mathbf{r})(\mathbf{s} \cdot \mathbf{r})}$ (and from below by $|\mathbf{R}_{\min}|$) and the volume for points $\mathbf{r}$ between $C_{\max} = 0$ and $(\boldsymbol{\omega}_\epsilon \cdot \mathbf{r})(\mathbf{s} \cdot \mathbf{r}) = 0$ extends from $|\mathbf{R}_{\min}|$ to $|\mathbf{R}_{\max}|$.

We are given $\boldsymbol{\omega}_\epsilon$ and $\mathbf{t}$, and we are interested in $\hat{\mathbf{t}}$, which minimizes the negative range volume. For any $\mathbf{s}$ the corresponding negative range volume becomes smallest if $\hat{\mathbf{t}}$ is on the great circle of $\mathbf{t}$ and $\mathbf{s}$, that is, $(\mathbf{t} \times \mathbf{s}) \cdot \hat{\mathbf{t}} = 0$, as will be shown next.

Assume that the scene in view is unbounded, i.e., $|\mathbf{R}| \in [0, \ldots, \infty]$. Then $C_{\min} = (\mathbf{t} \times \mathbf{s}) \cdot \mathbf{r} = 0$ and $C_{\max} = (\boldsymbol{\omega}_\epsilon \cdot \mathbf{r})(\mathbf{s} \cdot \mathbf{r}) = 0$. Let us consider a $\hat{\mathbf{t}}$ such that $(\mathbf{t} \times \mathbf{s}) \cdot \hat{\mathbf{t}} \neq 0$ (i.e., $\hat{\mathbf{t}}$ does not lie on the great circle defined by $\mathbf{t}$ and $\mathbf{s}$) and let us change $\hat{\mathbf{t}}$ such that $(\mathbf{t} \times \mathbf{s}) \cdot \hat{\mathbf{t}} = 0$. As $\hat{\mathbf{t}}$ changes, the area of type II on the sphere becomes an area of type IV and the area of type III becomes an area of type I. Thus, the negative range volume obtained consists only of range values above areas of type I.

Let us use the following notation. A_{III-I} denotes the area which changes from type III to type I and V_{III} and $V_{I(III)}$ are the volumes before and after change. Similarly, A_{II-IV} denotes the area which changes from type II to type IV and V_{II} and V_{IV} are the corresponding volumes.

The change of $\hat{\mathbf{t}}$ does not have any effect on the volumes above the areas that did not change in type, as can be seen from the constraint on $|\mathbf{R}|$ in Table 1. However, the change of $\hat{\mathbf{t}}$ causes a change in the volume above the areas which changed in type. As can be seen from (8), the normalization term is the same for points $\mathbf{r}_1(\varphi_{x_1}, \varphi_{y_1})$ and $\mathbf{r}_2(\varphi_{x_2}, \varphi_{y_2})$ symmetric with respect to the great circle $\mathbf{s} \cdot \mathbf{r} = 0$, because $\varphi_{y_1} = \varphi_{y_2}$ and $\varphi_{x_1} + \varphi_{x_2} = 2k\pi$ with $k \in \mathbb{N}$, which means that we encounter the same normalization factors in areas A_{III-I} and A_{II-IV}. As $\hat{\mathbf{t}}$ changes, there is a decrease of the negative depth volume by $V_{II} + V_{III}$ and an increase by $V_{I(III)}$, but V_{II} is always greater than $V_{I(III)}$ and thus the overall volume decreases.

Next, assume the scene in view to be bounded. As can easily be understood, independent of the range of values in which the scene lies, the volume of negative range values for any $\mathbf{s}$ is smallest for $(\mathbf{t} \times \mathbf{s}) \cdot \hat{\mathbf{t}} = 0$. If we assume an upper bound $|\mathbf{R}_{\max}| \neq \infty$, or a lower bound $|\mathbf{R}_{\min}| \neq 0$, or both bounds on the scene in view, there exist points $\mathbf{r}$ in areas I and III above which there are no range values which contribute to the negative range volume. However as shown before, since the curves $C_{\min} = 0$, $C_{\max} = 0$, $(\boldsymbol{\omega}_\epsilon \cdot \mathbf{r})(\mathbf{s} \cdot \mathbf{r}) = 0$ and $(\mathbf{t} \times \mathbf{s}) \cdot \mathbf{r} = 0$ intersect at the same point, V_{II} must always be larger than $V_{I(III)}$.

For any $\mathbf{s}$ the smallest volume is obtained for $\mathbf{s}$, $\mathbf{t}$, and $\hat{\mathbf{t}}$ lying on a great circle. Therefore, in order to minimize the total negative range volume V_n, we must have $\mathbf{t} = \hat{\mathbf{t}}$.

Thus, in summary, we have shown that for any given rotational error $\boldsymbol{\omega}_\epsilon$ the negative range volume is smallest if the direction of the actual translation and the estimated translation coincide, that is, $\mathbf{t} = \hat{\mathbf{t}}$.

4. The Planar Case

Let us express (4) in the more common component notation: $\dot{\mathbf{r}} = (\dot{r}_1, \dot{r}_2, \dot{r}_3)$. $\dot{r}_3$ is zero. If we denote $\dot{r}_1$ by u and $\dot{r}_2$ by v and express the coordinates of the focus of expansion as $(x_0, y_0) = (\frac{Uf}{W}, \frac{Vf}{W})$ we obtain the well-known equations

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

$$v = \frac{v_{\text{tr}}}{Z} + v_{\text{rot}} = (y - y_0)\frac{W}{Z} + \alpha\left(\frac{y^2}{f} + f\right) - \beta\frac{xy}{f} - \gamma x \tag{11}$$

Since, due to the scaling ambiguity, only the direction of translation can possibly be obtained, we set $W = 1$

and $\hat{W} = 1$, and obtain from (7)

$$\hat{Z} = Z \left(\begin{array}{l} \dfrac{(x - \hat{x}_0)\, n_x + (y - \hat{y}_0)\, n_y}{(x - x_0)\, n_x + (y - y_0)\, n_y} \\ + \; Z\Big(\Big(\alpha_\epsilon \frac{xy}{f} - \beta_\epsilon \Big(\frac{x^2}{f} + f\Big) \\ + \; \gamma_\epsilon y\Big) n_x + \Big(\alpha_\epsilon \Big(\frac{y^2}{f} + f\Big) \\ - \; \beta_\epsilon \frac{xy}{f} - \gamma_\epsilon x\Big) n_y\Big) \end{array} \right) \tag{12}$$

where n_x and n_y denote the components of $\mathbf{n}$ in the x and y directions.

In the following analysis, we assume that the FOE and the estimated FOE are within the image. We do not consider the exact effects resulting from volumes of negative depth in different directions being outside the field of view. We also perform some simplification: For a limited field of view, the terms quadratic in the image coordinates, which appear in the rotational components, are small with respect to the linear and constant terms, and we therefore drop them.

The 0 distortion surface thus becomes

$$(x - \hat{x}_0)\, n_x + (y - \hat{y}_0)\, n_y = 0 \tag{13}$$

and the $-\infty$ distortion surface takes the form

$$\begin{aligned}(x - x_0)n_x + (y - y_0)n_y + Z((-\beta_\epsilon f + \gamma_\epsilon y)n_x \\ +(\alpha_\epsilon f - \gamma_\epsilon x)n_y) = 0\end{aligned} \tag{14}$$

The flow directions (n_x, n_y) can alternatively be written as $(\cos\psi, \sin\psi)$, with $\psi \in [0, \pi]$ denoting the angle between $[n_x, n_y]^T$ and the x axis.

To simplify the visualization of the volumes of negative depth in different directions, we perform the following coordinate transformation to align the flow direction with the x axis: for every ψ we rotate the coordinate system by angle ψ, to obtain the new coordinates

$$\begin{aligned}[x', y']^T = R[x, y]^T, [x_0', y_0']^T = R[x_0, y_0]^T \\ [\hat{x}_0', \hat{y}_0']^T = R[\hat{x}_0, \hat{y}_0]^T, [{\alpha_\epsilon}', {\beta_\epsilon}']^T = R[\alpha_\epsilon, \beta_\epsilon]^T\end{aligned}$$

where $R = \begin{bmatrix} \cos\psi & \sin\psi \\ -\sin\psi & \cos\psi \end{bmatrix}$.

Equations (13) and (14) thus become

$$\begin{aligned}(x' - \hat{x}_0') &= 0 \\ \text{and} \quad (x' - x_0') + Z\left(-{\beta_\epsilon}' f + \gamma_\epsilon y'\right) &= 0\end{aligned}$$

4.1. *Relationship between the translational and rotational errors*

In the following proof we first consider the case of $\gamma_\epsilon = 0$ and we then study the general case.

Part 1 ($\gamma_\epsilon = 0$). If $\gamma_\epsilon = 0$, the volume of negative depth values for every direction ψ lies between the surfaces

$$(x' - \hat{x}_0') = 0 \quad \text{and} \quad (x' - {x_0}') - {\beta_\epsilon}' f Z = 0$$

Equation $(x' - \hat{x}_0') = 0$ describes a plane parallel to the $y'Z$ plane at distance $\hat{x}_0'$ from the origin, and equation $(x' - {x_0}') - {\beta_\epsilon}' f Z = 0$ describes a plane parallel to the y' axis of slope $\frac{1}{{\beta_\epsilon}' f}$, which intersects the $x'y'$ plane in the x' coordinate ${x_0}'$. Thus we obtain a wedge-shaped volume parallel to the y' axis. Fig. 8 illustrates the volume through a slice parallel to the $x'Z$ plane and Fig. 9 gives an illustration of this volume in space.

The scene in view extends between the depth values $Z_{\min}$ and $Z_{\max}$. The $-\infty$ distortion surface intersects the planes $Z = Z_{\max}$ and $Z = Z_{\min}$ in the x' coordinates ${x_0}' + {\beta_\epsilon}' f Z_{\max}$ and ${x_0}' + {\beta_\epsilon}' f Z_{\min}$. We denote as A_ψ the area of the cross section parallel to the $x'Z$ plane through the negative depth volume in direction ψ.

If ${x_0}'$ lies between ${x_0}' + \beta_\epsilon' f Z_{\min}$ and ${x_0}' + \beta_\epsilon' f Z_{\max}$

$$\begin{aligned}A_\psi = \Big| {x_{0_\epsilon}}' (Z_{\max} + Z_{\min}) \\ + \frac{{\beta_\epsilon}' f}{2} \left({Z_{\max}}^2 + {Z_{\min}}^2\right) + \frac{{x_{0_\epsilon}}'^2}{\beta_\epsilon f} \Big|\end{aligned} \tag{15}$$

Let us fix β_ϵ'.

To obtain the minimum A_ψ we solve for ${x_{0_\epsilon}}'$ such that

$$\frac{\partial A_\psi}{\partial {x_{0_\epsilon}}'} = 0$$

and obtain

$${x_{0_\epsilon}}' = -\frac{{\beta_\epsilon}' f}{2} (Z_{\max} + Z_{\min}) \tag{16}$$

that is, the 0 distortion surface has to intersect the $-\infty$ distortion surface in the middle of the depth interval in the plane $Z = \frac{(Z_{\max}+Z_{\min})}{2}$.

Since $\frac{x'_{0_\epsilon}}{\beta'_\epsilon f}$ depends only on the depth interval and not on the direction ψ, the total negative depth volume is minimized if (16) holds for every direction and thus the volume is minimized for every direction. From the relations ${\beta_\epsilon}' = \cos\psi\beta_\epsilon - \sin\psi\alpha_\epsilon$ and ${x_{0_\epsilon}}' = \cos\psi x_{0_\epsilon} + \sin\psi y_{0_\epsilon}$, we obtain $\frac{x_{0_\epsilon}}{y_{0_\epsilon}} = -\frac{\beta_\epsilon}{\alpha_\epsilon}$. In other words, the rotational error $(\alpha_\epsilon, \beta_\epsilon)$ and the translational error $(x_{0_\epsilon}, y_{0_\epsilon})$ have to be perpendicular to each other.

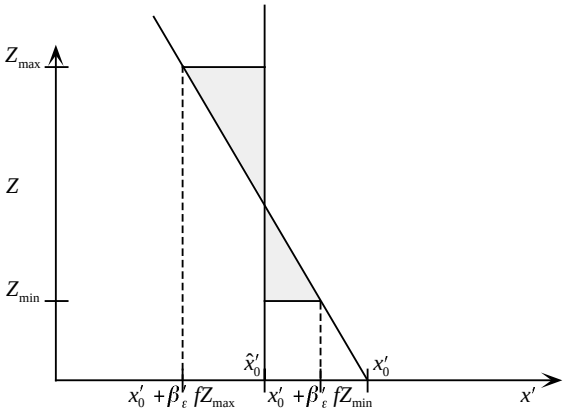


Fig. 8. Slice parallel to the $x'Z$ plane through the volume of negative estimated depth for a single direction.

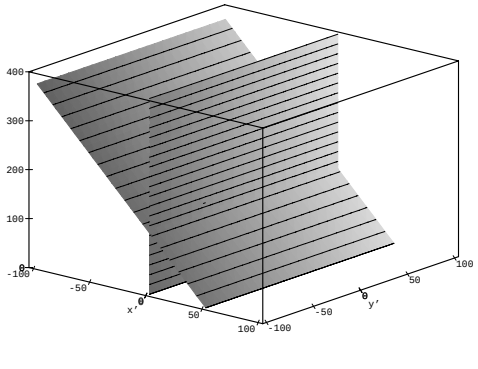


Fig. 9. $\gamma_\epsilon = 0$: The volume of negative depth values for a single direction between the 0 and $-\infty$ distortion surfaces.

If on the other hand we fix ${x_{0_\epsilon}}'$, we obtain by solving for ${x_{0_\epsilon}}'$ such that $\frac{\partial A_\psi}{\partial {\beta_\epsilon}'} = 0$:

$$ {x_{0_\epsilon}}' = -{\beta_\epsilon}' f \sqrt{\frac{\left({Z_{\max}}^2 + {Z_{\min}}^2\right)}{2}} \qquad (17) $$

that is, the 0 distortion surface and the $-\infty$ distortion surface intersect in the plane $Z = \sqrt{\frac{({Z_{\max}}^2+{Z_{\min}}^2)}{2}}$.

Again, if $\frac{x_{0_\epsilon}}{y_{0_\epsilon}} = -\frac{\beta_\epsilon}{\alpha_\epsilon}$, (17) holds for every direction, and thus the total negative depth volume is minimized.

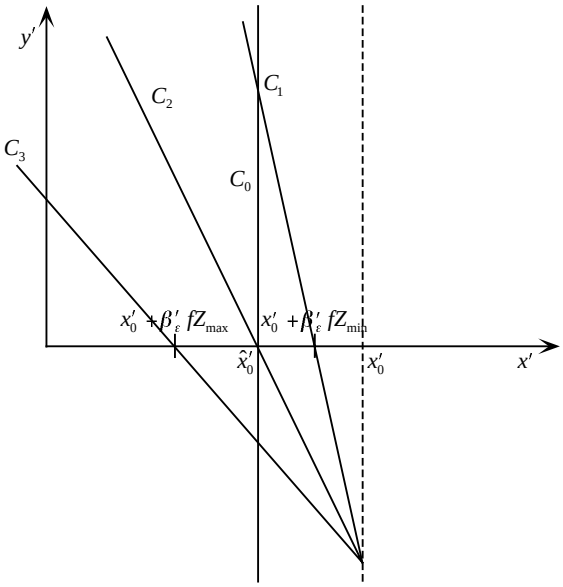


Fig. 10. Slices parallel to the $x'y'$ plane through the 0 distortion surface (C_0) and the $-\infty$ distortion surface at depth values $Z = Z_{\min}$ (C_1), $Z = -\frac{x'_{0_\epsilon}}{{\beta_\epsilon}' f}$ (C_2), and $Z = Z_{\max}$ (C_3).

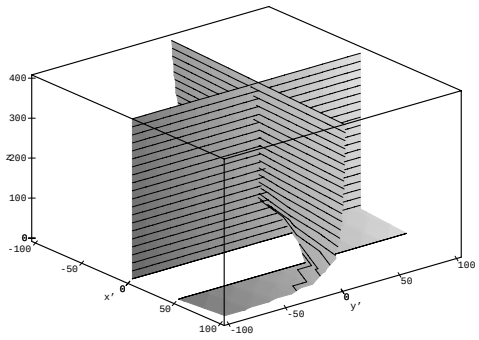


Fig. 11. $\gamma_\epsilon \neq 0$: volume of negative depth values between the 0 and $-\infty$ distortion surfaces.

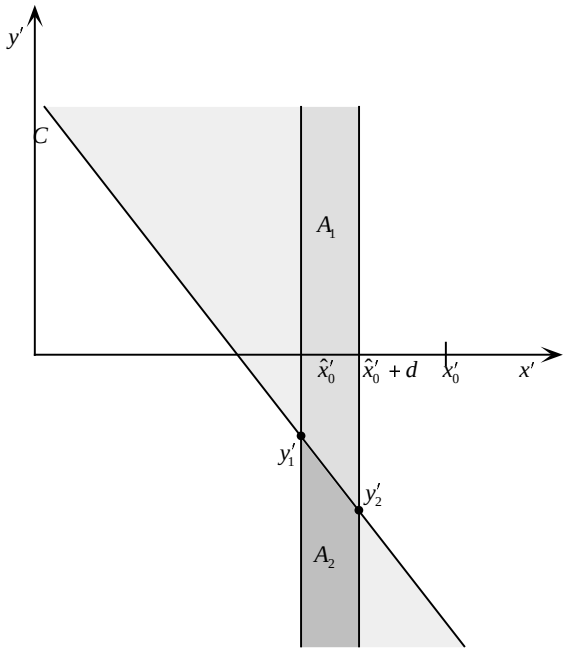


Fig. 12. A change of $\hat{x}'_0$ to $\hat{x}'_0 + d$ causes the area of negative depth values A_c to increase by area A_1 and to decrease by area A_2. This change amounts to $A_c = -(y'_1 + y'_2)d\,\mathrm{sgn}(\gamma_\epsilon)$.

Part 2 ($\gamma_\epsilon \neq 0$). If $\gamma_\epsilon \neq 0$, the $-\infty$ distortion surface becomes

$$(x' - {x_0}') + Z\left(-{\beta_\epsilon}' f + \gamma_\epsilon y'\right) = 0$$

This surface can be most easily understood by slicing it with planes parallel to the $x'y'$ plane. At every depth value Z, we obtain a line of slope $\frac{-1}{\gamma_\epsilon Z}$ which intersects the x' axis in $x' = {x_0}' + {\beta_\epsilon}' f Z$ (see Fig. 10). For any given Z the slopes of the lines in different directions are the same. An illustration of the volume of negative depth values is given in Fig. 11.

In part 1 of this analysis we found that for a given rotational error $(\alpha_\epsilon, \beta_\epsilon)$ and $\gamma_\epsilon = 0$, the smallest volume of negative depth values is obtained if $\hat{x}'_0 = {x_0}' + {\beta_\epsilon}' f \frac{(Z_{\max}+Z_{\min})}{2}$, and the intersection of the 0 and $-\infty$ distortion surfaces is at $Z = \frac{Z_{\min}+Z_{\max}}{2}$. In order to derive the position of $\hat{x}'_0$ that minimizes the negative depth values for the general case of $\gamma_\epsilon \neq 0$, we study the change of volume as $\hat{x}'_0$ changes from ${x_0}' + {\beta_\epsilon}' f \frac{(Z_{\max}+Z_{\min})}{2}$.

Referring to Fig. 12, it can be seen that for any depth value Z, a change in the position of $\hat{x}'_0$ to $\hat{x}'_0 + d$, assuming $Z \neq 0$, causes the corresponding area of negative depth values to change by A_c, where

$$A_c = -\left({y'}_1 + {y'}_2\right) d\,\mathrm{sgn}(\gamma_\epsilon)$$

and ${y'}_1$ and ${y'}_2$ denote the y' coordinates of the intersection point of the $-\infty$ distortion contour at depth Z with the 0 distortion contours $x' = \hat{x}'_0$ and $x' = \hat{x}'_0 + d$.

By intersecting the $-\infty$ distortion contour $x' - {x_0}' + Z(-{\beta_\epsilon}' f + \gamma_\epsilon y') = 0$ with the 0 distortion contours $x' - ({x_0}' + \frac{{\beta_\epsilon}' f}{2}(Z_{\min} + Z_{\max})) = 0$ and $x' - ({x_0}' + \frac{{\beta_\epsilon}' f}{2}(Z_{\min} + Z_{\max}) + d) = 0$, we obtain

$$\begin{aligned} y'_1 &= -\tfrac{{\beta_\epsilon}' f (Z_{\min}+Z_{\max})}{2Z\gamma_\epsilon} + \tfrac{{\beta_\epsilon}' f}{\gamma_\epsilon} \\ \text{and}\quad y'_2 &= -\tfrac{{\beta_\epsilon}' f}{2}\tfrac{(Z_{\min}+Z_{\max})}{Z\gamma_\epsilon} - \tfrac{d}{Z\gamma_\epsilon} + \tfrac{{\beta_\epsilon}' f}{\gamma_\epsilon} \end{aligned}$$

and therefore

$$A_c = -\mathrm{sgn}\left(\gamma_\epsilon\right) d \left(\frac{2{\beta_\epsilon}' f}{\gamma_\epsilon} - \left(\frac{{\beta_\epsilon}' f \left(Z_{\min} + Z_{\max}\right) + d}{Z\gamma_\epsilon} \right) \right)$$

The change V_c in negative depth volume for any direction is thus given by

$$V_c = \int_{Z_{\min}}^{Z_{\max}} A_c dZ$$

which amounts to

$$V_c = -\mathrm{sgn}\left(\gamma_\epsilon\right) d \left(\frac{2{\beta_\epsilon}' f}{\gamma_\epsilon} \left(Z_{\max} - Z_{\min}\right) - \left(\frac{{\beta_\epsilon}' f \left(Z_{\min} + Z_{\max}\right) + d}{\gamma_\epsilon} \right) \ln\left(\frac{Z_{\max}}{Z_{\min}} \right) \right)$$

It can be verified that in order for V_c to be negative, $\mathrm{sgn}({\beta_\epsilon}') = -\mathrm{sgn}(d)$. This means that $\hat{x_0}' + d$ lies between $\hat{x}'_0$ and ${x_0}'$ (as in Fig. 12).

We are interested in the d which minimizes V_c. By solving

$$\frac{\partial V_c}{\partial d} = 0$$

we obtain

$$d = {\beta_\epsilon}' f \left[\frac{\left(Z_{\max} - Z_{\min}\right)}{\ln\left(\frac{Z_{\max}}{Z_{\min}}\right)} - \frac{1}{2}\left(Z_{\min} + Z_{\max}\right) \right]$$

Thus

$$x'_{0_\epsilon} = -\beta'_\epsilon f \frac{(Z_{\max} - Z_{\min})}{\ln\left(\frac{Z_{\max}}{Z_{\min}}\right)} \tag{18}$$

Again, since $\frac{x'_{0_\epsilon}}{\beta'_\epsilon f}$ depends only on the depth interval, the total negative depth volume is obtained if the volume in every direction is minimized. Therefore we have the constraint independent of γ_ϵ.

$$\frac{x_{0_\epsilon}}{y_{0_\epsilon}} = -\frac{\beta_\epsilon}{\alpha_\epsilon}$$

For a given rotational error $(\alpha_\epsilon, \beta_\epsilon, \gamma_\epsilon)$ this constraint defines the direction of the FOE of the translational error on the image plane.

If on the other hand the translational error $(x_{0_\epsilon}, y_{0_\epsilon})$ is given, it can be seen using Fig. 10 that the area of negative depth values for every depth Z increases as γ_ϵ increases. Thus, for a given $(x_{0_\epsilon}, y_{0_\epsilon})$ the constraint $\frac{x_{0_\epsilon}}{y_{0_\epsilon}} = -\frac{\beta_\epsilon}{\alpha_\epsilon}$ defines the direction of the rotational error on the image. In addition, we must have $\gamma_\epsilon = 0$.

Some comment on the finiteness of the image is necessary here. The values A_c and V_c have been derived for an infinitely large image. If γ_ϵ is very small or some of the depth values Z in the interval $[Z_{\min}, Z_{\max}]$ are small, the coordinates of the intersections ${y'}_1$ and ${y'}_2$ do not lie within the image. The value of A_c can be at most the length of the image times d. Since the slope of the $-\infty$ distortion contour for a given Z is the same for all directions, this will have very little effect on the relationship between the direction of the translational and rotational motion errors. It has an effect, however, on the relative values of the motion errors. Only if the intersections are within the image, equation (18) describes adequately the value of x'_{0_ϵ} as a function of β'_ϵ and the interval of depth values of the scene in view.

4.2. Bias in the estimated direction of translation

We investigate here additional constraints on the size of the negative depth volumes as they arise for typical planar imaging systems. In particular, we consider the effects due to the finite size of the aperture and the effects due to fixation at a scene point.

Let us consider a circular aperture. We fix the amount of translational error, $({x_{0_\epsilon}}^2 + {y_{0_\epsilon}}^2)^{1/2}$, and study which is the direction of the translational error resulting in the smallest negative depth volume. Independent of the direction of translation, (17) describes the relationship of x_{0_ϵ} and β_ϵ for the smallest negative depth volume. Substituting (17) into (15), we obtain equation

$$A_\psi = \left| {x_{0_\epsilon}}' \left(\sqrt{2\left({Z_{\max}}^2 + {Z_{\min}}^2\right)} - (Z_{\max} + Z_{\min}) \right) \right|$$

which describes the cross sections through the negative depth volumes as a function of ${x_{0_\epsilon}}'$ and the depth interval. The negative depth volume for every direction ψ amounts to $A_\psi l_\psi$, where l_ψ denotes the average extent of the wedge-shaped negative depth volume in direction ψ. The total negative depth volume is minimized if $\int_0^\pi A_\psi l_\psi d\psi$ is minimized. Considering a circular aperture, this minimization is achieved if the largest A_ψ corresponds to the smallest extent l_ψ and the smallest A_ψ corresponds to the largest l_ψ. This happens when the actual and the estimated FOE lie on a line passing through the image center, that is, $\frac{x_0}{y_0} = \frac{\hat{x_0}}{\hat{y_0}}$, and the actual FOE is closer to the image center than the estimated FOE (see Fig. 13).

Next we investigate the negative depth volumes under fixation. We are interested in this configuration, because systems with planar eyes perform motion estimation usually during fixation. For example, typical eye movements in humans consist of a series of fixations separated by saccadic movements, and during the latter no processing of visual information takes place.

If the system is fixating, the flow at the center is zero, and thus from (11) we obtain

$$\begin{aligned} \frac{U}{Z_{\text{fix}}} &= -\beta, \quad \frac{V}{Z_{\text{fix}}} = \alpha \\ \text{and} \quad \frac{\hat{U}}{Z_{\text{fix}}} &= -\hat{\beta}, \quad \frac{\hat{V}}{Z_{\text{fix}}} = \hat{\alpha}, \end{aligned}$$

where Z_{fix} denotes the depth of the fixation point. Therefore $\frac{U}{V} = -\frac{\beta}{\alpha}$ and $\frac{\hat{U}}{\hat{V}} = -\frac{\hat{\beta}}{\hat{\alpha}}$, that is, the actual FOE and AOR (as well as the estimated FOE and AOR) lie on lines through the image center and perpendicular to each other. In addition, we obtain from the above equations $U_\epsilon = -\beta_\epsilon Z_{\text{fix}}$ and $V_\epsilon = \alpha_\epsilon Z_{\text{fix}}$, which are constraints on absolute velocity and depth measurements.

From the minimization of negative depth volumes (not considering fixation) we obtained that $x_{0_\epsilon} = -\beta_\epsilon f \sqrt{\frac{{Z_{\max}}^2 + {Z_{\min}}^2}{2}}$ and $y_{0_\epsilon} = \alpha_\epsilon f \sqrt{\frac{{Z_{\max}}^2 + {Z_{\min}}^2}{2}}$.

Let us use the notation $\frac{Z_{\text{ideal}}}{W} = \sqrt{\frac{{Z_{\max}}^2 + {Z_{\min}}^2}{2}}$. Combining the constraints from fixation and negative depth volume minimization we obtain

$$\frac{U}{W} - \frac{\hat{U}}{\hat{W}} = \frac{U_\epsilon Z_{\text{ideal}}}{Z_{\text{fix}} W} \quad \text{and} \quad \frac{V}{W} - \frac{\hat{V}}{\hat{W}} = \frac{V_\epsilon Z_{\text{ideal}}}{Z_{\text{fix}} W}$$

and therefore

$$\begin{aligned} \hat{W} &= \frac{W Z_{\text{fix}} \hat{U}}{U Z_{\text{fix}} - U_\epsilon Z_{\text{ideal}}} \\ \text{and} \quad \hat{W} &= \frac{W Z_{\text{fix}} \hat{V}}{V Z_{\text{fix}} - V_\epsilon Z_{\text{ideal}}} \end{aligned} \tag{19}$$

Both equations in (19) can be true only if either $Z_{\text{fix}} = Z_{\text{ideal}}$ or if $\frac{x_0}{y_0} = \frac{\hat{x_0}}{\hat{y_0}}$. This just means that comparing the case without fixation to the case of fixation, if the real and the estimated FOE are on a line through the center, the negative depth volumes are of the same amount. For any other configuration the negative depth volumes are larger for the case of fixation, as for these configurations the perpendicularity of the translational and rotational errors cannot be enforced. Thus fixation, too, increases the likelihood of a motion estimation in the form $\frac{x_0}{y_0} = \frac{\hat{x_0}}{\hat{y_0}}$. Furthermore, (19) defines $\hat{W}$ in terms of the actual translation, the estimated translation, and the depth of the fixation point. If the system has a good estimate of its forward translational velocity, this provides additional constraints on the motion errors, which in turn means that motion estimation is performed more accurately if the system fixates.

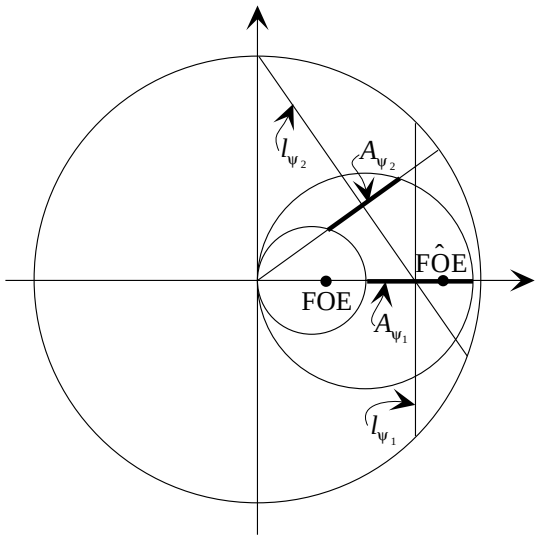


Fig. 13. Cross-sectional view of the wedge-shaped negative depth volumes in a circular aperture. The minimization of the negative depth volume for a given amount of translational error occurs when $\frac{x_0}{y_0} = \frac{\hat{x_0}}{\hat{y_0}}$. A_{ψ_i} and l_{ψ_i} denote the areas of the cross sections and average extents respectively, for two angles ψ_1 and ψ_2, and the circles at the boundaries of A_{ψ_i} are described as $x' = {x_0}' + \beta'_\epsilon f Z_{\min}$ and $x' = {x_0}' + \beta'_\epsilon f Z_{\max}$.

At this point, it is worth noting that results similar to those obtained here for the case of a planar eye are also achieved when one minimizes deviation from the epipolar constraint on the basis of optic flow, as opposed to minimizing negative depth on the basis of normal flow. In particular, if the depth of the scene in view is uniformly distributed, for a fixed translational error $(x_{0_\epsilon}, y_{0_\epsilon})$ the rotational error is of the form $-\frac{\beta_\epsilon}{\alpha_\epsilon} = \frac{x_{0_\epsilon}}{y_{0_\epsilon}}$, $\gamma_\epsilon = 0$. In the case of a global minimization, that is, a simultaneous estimation of rotation and translation, we have in addition $\frac{x_0}{y_0} = \frac{x_{0_\epsilon}}{y_{0_\epsilon}}$. In the case of a fixed rotational error the relationship is more intricate, depending on the exact value of the rotational error and the scene in view. The interested reader is referred to (Fermüller and Aloimonos, 1998).

5. Shape Estimation in the Presence of Distortion

The above results are of great importance for the analysis of shape estimation. An error of the form $\gamma_\epsilon = 0$, $\frac{x_{0_\epsilon}}{y_{0_\epsilon}} = -\frac{\beta_\epsilon}{\alpha_\epsilon} = \frac{x_0}{y_0}$ guarantees that shape maps of the scene are derived which are very well behaved.

If we use the estimated normal flow measurements to derive optical flow, we can obtain a global shape map which is affine in the inverse depth. The underlying idea is as follows: We use the 3D motion estimate obtained to detect discontinuities in the flow field; that is, on the basis of the 3D motion estimate, depth is derived from normal flow measurements (which are sufficiently different from the direction perpendicular to the estimated translational flow) only for the purpose of detecting depth discontinuities. Knowledge about the discontinuities allows for the derivation of better optical flow estimates. Then we compute the depth from the component of optical flow along the directions tangential to circles with center at 0. This allows us to cancel the terms linear and quadratic in the image coordinates. Thus, considering (n_x, n_y) in the direction $(-y, x)$ and considering $\gamma_\epsilon = 0$, we obtain from (12) the distortion factor as

$$D = \frac{\hat{x}_0 n_x + \hat{y}_0 n_y}{x_0 n_x + y_0 n_y + Zf\,(\beta_\epsilon n_x - \alpha_\epsilon n_y)} \tag{20}$$

If $\frac{x_{0_\epsilon}}{y_{0_\epsilon}} = -\frac{\beta_\epsilon}{\alpha_\epsilon} = \frac{x_0}{y_0}$ for any given Z, the numerator is a multiple of the denominator and thus the distortion factor is the same for every direction (n_x, n_y). This means that scene points of the same depth are distorted by the same factor and the computed depth map has

the same level contours as the actual depth map of the scene.

Also, the same distortion factor can be obtained approximately for the image near the fixation center, if depth is estimated directly from the normal flow measurements. As near the image center the image coordinates are very small, if we ignore them, D for any (n_x, n_y) takes the same form as in (20).

In the shape map obtained in this way, depending on the sign of the rotational error, there will either be an overestimation for the nearby scene and an underestimation for the far scene or vice versa. All the distortion, however, takes place only in the Z dimension. Thus the resulting depth function involves an affine transformation (that is, the inverse depth estimates can be expressed as $\frac{1}{Z} = \frac{a}{Z} + b$). The invariants of these shape maps have been studied in the work of Koenderink and van Doorn (Koenderink and van Doorn, 1994; Koenderink and van Doorn, 1995).

6. Conclusions

A stability analysis, which investigated the inherent ambiguities of structure from motion, has been presented. The analysis was based solely on the fact that the depth of the scene—in order for the scene to be visible—has to be positive. As input to the structure from motion process we considered the value of the flow at every point along some direction, a quantity more easily computable than optical flow or correspondence. Our stability analysis amounts to an understanding of the coupling of the translational and rotational error. Given an error in the translation (or the rotation), we asked: what is the value of the rotation (or the translation) that estimates the minimum number of negative depth values? We performed the analysis for both a spherical and a planar retina. For the case of a planar retina we found that the configuration of the rotational and translational errors resulting in minimum negative depth is the one in which the projections of the two error vectors on the image plane are perpendicular to each other, and that an estimated translation whose projection on the image passes through the image center and the projection of the actual translation is most likely. For the case of a spherical retina, we found that given a rotational error, the optimal translation is the correct one, while given an error in translation, the optimal rotation error has an intricate relationship to the translational error and the scene in view.

These results, besides their potential use in structure from motion algorithms, also represent a computational analysis assessing different eye construction in the natural world. The results on the sphere demonstrate that it is very easy for a system with panoramic vision to estimate its self-motion. Indeed, if the system possesses an inertial sensor providing its rotation with some error, we have shown that after derotation, a simple algorithm considering only translation based on normal flow will estimate the translation optimally. This suggests that spherical eye design is optimal for flying systems such as the compound eyes of insects and the panoramic vision of birds.

The analysis on the plane revealed that for an optimal configuration of errors, the estimated depth distorts only in the z direction, with the level contours of the depth function distorting by the same amount, thus making it feasible to extract meaningful shape representations. This suggests that the camera-type eyes of primates are possibly optimal for systems that need good shape computation capabilities.

Acknowledgements

Special thanks to Sara Larson for her editorial and graphics assistance.

Appendix

Re-parameterization of Flow Directions

Let us choose a uniformly distributed flow field direction $\mathbf{n}_1(\psi)$ as follows. The coordinates of $\mathbf{r} = [x, y, z]^T$ at every point on the unit sphere are obtained through a rotation of point $[0, 0, 1]^T$ by an angle φ_x around the x axis followed by a rotation of angle φ_y around the y axis. Thus the rotation matrix R is given by

$$R = \begin{bmatrix} \cos\varphi_y & 0 & \sin\varphi_y \\ -\sin\varphi_x \sin\varphi_y & \cos\varphi_x & \sin\varphi_x \cos\varphi_y \\ -\cos\varphi_x \sin\varphi_y & -\sin\varphi_x & \cos\varphi_x \cos\varphi_y \end{bmatrix}$$

and every point

$$\mathbf{r} = [\sin\varphi_y, \sin\varphi_x \cos\varphi_y, \cos\varphi_x \cos\varphi_y]^T.$$

Vectors $\mathbf{n}_1(\psi)$ are obtained through rotation of unit vector $[\sin\psi, \cos\psi, 0]^T$ at point $[0,0,1]^T$. Thus

$$\mathbf{n}_1(\psi) = [\cos\varphi_y \sin\psi, -\sin\varphi_x \sin\varphi_y \sin\psi + \cos\varphi_x \cos\psi, -\cos\varphi_x \sin\varphi_y \sin\psi - \sin\varphi_x \cos\psi]^T$$

On the other hand, the direction $\mathbf{n}_2(\chi)$ used in the analysis is chosen to be $\mathbf{n}_2(\chi) = \mathbf{r} \times \mathbf{s}(\chi)$ with $\mathbf{s} = [0, \sin\chi, \cos\chi]^T$.

Thus

$$\mathbf{n}_2(\chi) = [\cos\varphi_x \cos\varphi_y \sin\chi - \sin\varphi_x \cos\varphi_y \cos\chi, \sin\varphi_y \cos\chi, -\sin\varphi_y \sin\chi]^T.$$

In order for $\mathbf{n}_1(\psi)$ to be parallel to $\mathbf{n}_2(\chi)$ the following must hold:

$$(\mathbf{n}_1(\psi) \times \mathbf{n}_2(\chi)) \cdot \mathbf{r} = 0$$

Thus $\psi = g(\chi) = \arctan(\frac{\tan(\chi-\varphi_x)}{\sin\varphi_y})$ and the normalization factor $|\frac{\partial\psi}{\partial\chi}|$ is

$$\left|\frac{\partial\psi}{\partial\chi}\right| = \left|\frac{\sin(\varphi_y)}{\cos(\varphi_y)^2 \cos(\chi-\varphi_x)^2 - 1}\right|$$

For an illustration see Fig. 6.

References

Adiv, G. 1989. Inherent ambiguities in recovering 3-D motion and structure from a noisy flow field. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11:477–489.

Aloimonos, Y. and Duric, Z. 1994. Estimating the heading direction using normal flow. *International Journal of Computer Vision*, 13:33–56.

Åström, K. 1996. Invariancy Methods for Points, Curves and Surfaces in Computational Vision. PhD thesis, Department of Mathematics, Lund Institute of Technology, Lund, Sweden.

Brodsky, T., Fermüller, C. and Aloimonos, Y. 1998. Directions of motion fields are hardly ever ambiguous. *International Journal of Computer Vision*, 26:5–24.

Bruss, A. and Horn, B.K.P. 1983. Passive navigation. *Computer Vision, Graphics, and Image Processing*, 21:3–20.

Cheong, L., Fermüller, C. and Aloimonos, Y. 1998. Effects of errors in the viewing geometry on shape estimation. *Computer Vision and Image Understanding*, forthcoming. Earlier version available as Technical Report CAR-TR-773, June 1996.

Daniilidis, K. 1992. On the Error Sensitivity in the Recovery of Object Descriptions. PhD thesis, Department of Informatics, University of Karlsruhe, Germany. In German.

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

Faugeras, O.D. 1992. *Three-Dimensional Computer Vision*. MIT Press: Cambridge, MA.

Faugeras, O.D. 1992. What can be seen in three dimensions with an uncalibrated stereo rig? In *Proc. Second European Conference on Computer Vision*, Santa Margherita Ligure, Italy. Springer-Verlag: Berlin, pp. 563–578.

Faugeras, O.D. and Mourrain, B. 1995. On the geometry and algebra of the point and line correspondences between n images. In *Proc. 5th International Conference on Computer Vision*, Cambridge, MA, pp. 951–956.

Fermüller, C. and Aloimonos, Y. 1995. Direct perception of three-dimensional motion from patterns of visual motion. *Science*, 270:1973–1976.

Fermüller, C. and Aloimonos, Y. 1995. Qualitative egomotion. *International Journal of Computer Vision*, 15:7–29.

Fermüller, C. and Aloimonos, Y. 1997. On the geometry of visual correspondence. *International Journal of Computer Vision*, 21:223–247.

Fermüller, C. and Aloimonos, Y. 1998. What is computed by structure from motion algorithms? In *Proc. European Conference on Computer Vision*, Freiburg, Germany.

Fermüller, C., Cheong, L. and Aloimonos, Y. 1997. Visual space distortion. *Biological Cybernetics*, 77:323–337.

Hartley, R.I. 1994. Projective reconstruction and invariants from multiple images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16:1036–1041.

Hofmann, W. 1953. Das problem der "gefährlichen Flächen" in Theorie and Praxis. Deutsche Geodätische Kommission bei der Bayerischen Akademie der Wissenschaften, Reihe C, Heft 3, München.

Horn, B.K.P. 1986. *Robot Vision*. McGraw Hill: New York.

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

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

Horn, B.K.P. and Weldon, Jr., E.J. 1988. Direct methods for recovering motion. *International Journal of Computer Vision*, 2:51–76.

Jepson, A.D. and Heeger, D.J. 1990. Subspace methods for recovering rigid motion II: theory. Technical Report RBCV-TR-90-36, University of Toronto.

Koenderink, J.J. and van Doorn, A.J. 1991. Affine structure from motion. *Journal of the Optical Society of America*, 8:377–385.

Koenderink, J.J. and van Doorn, A.J. 1994. Two-plus-one-dimensional differential geometry. *Pattern Recognition Letters*, 15:439–443.

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

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

Longuet-Higgins, H.C. and Prazdny, K. 1980. The interpretation of a moving retinal image. *Proc. Royal Society, London B*, 208:385–397.

Maybank, S.J. 1986. Algorithm for analysing optical flow based on the least-squares method. *Image and Vision Computing*, 4:38–42.

Maybank, S.J. 1987. A Theoretical Study of Optical Flow. PhD thesis, University of London, England.

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

Negahdaripour, S. 1986. Direct Passive Navigation. PhD thesis, Massachusetts Institute of Technology, Cambridge, MA.

Nelson, R.C. and Aloimonos, J. 1988. Finding motion parameters from spherical flow fields (or the advantage of having eyes in the back of your head). *Biological Cybernetics*, 58:261–273.

Poggio, T. and Reichardt, W. 1973. Considerations on models of movement detection. *Kybernetik*, 13:223–227.

Reichardt, W. 1961. Autocorrelation, a principle for evaluation of sensory information by the central nervous system. In W. A. Rosenblith, editor, *Sensory Communication*. M.I.T. Press: Cambridge, MA, pp. 303–317.

Reichardt, W. 1987. Evaluation of optical motion information by movement detectors. *J. Comp. Physiol.*, 161:533–547.

Sinclair, D., Blake, A. and Murray, D. 1994. Robust estimation of egomotion from normal flow. *International Journal of Computer Vision*, 13:57–69.

Slama, C.C., Theurer, C. and Henriksen, S.W. 1980. *Manual of Photogrammetry*. American Society of Photogrammetry: Falls Church, VA.

Spetsakis, M.E. 1994. Models of statistical visual motion estimation. *Computer Vision, Graphics, and Image Processing*, 60:300–312.

Spetsakis, M.E. and Aloimonos, J. 1988. Optimal computing of structure from motion using point correspondence. In *Proc. Second International Conference on Computer Vision*, pp. 449–453.

Spetsakis, M.E. and Aloimonos, J. 1989. Optimal motion estimation. In *Proc. IEEE Workshop on Visual Motion*, pp. 229–237.

Spetsakis, M.E. and Aloimonos, J. 1990. Structure from motion using line correspondences. *International Journal of Computer Vision*, 4:171–183.

Spetsakis, M.E. and Aloimonos, J. 1990. A unified theory of structure from motion. In *Proc. DARPA Image Understanding Workshop*, pp. 271–283.

Tsai, R.Y. and Huang, T.S. 1984. Uniqueness and estimation of three-dimensional motion parameters of rigid objects with curved surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6:13–27.

Ullman. S. 1979. *The Interpretation of Visual Motion*. MIT Press: Cambridge, MA.

van Santen, J.P.H. and Sperling, G. 1984. Temporal covariance model of human motion perception. *Journal of the Optical Society of America A*, 1:451–473.

Waxman, A.M. and Wohn, K. 1985. Contour evolution, neighborhood deformation and global image flow: planar surfaces in motion. *International Journal of Robotics Research*, 4(3):95–108.