Observability of 3D Motion

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

Abstract

This paper examines the inherent difficulties in observing 3D rigid motion from image sequences. It does so without considering a particular estimator. Instead, it presents a statistical analysis of all the possible computational models which can be used for estimating 3D motion from an image sequence. These computational models are classified according to the mathematical constraints that they employ and the characteristics of the imaging sensor (restricted field of view and full field of view). Regarding the mathematical constraints, there exist two principles relating a sequence of images taken by a moving camera. One is the "epipolar constraint," applied to motion fields, and the other the "positive depth" constraint, applied to normal flow fields. 3D motion estimation amounts to optimizing these constraints over the image. A statistical modeling of these constraints leads to functions which are studied with regard to their topographic structure, specifically as regards the errors in the 3D motion parameters at the places representing the minima of the functions. For conventional video cameras possessing a restricted field of view, the analysis shows that for algorithms in both classes which estimate all motion parameters simultaneously, the obtained solution has an error such that the projections of the translational and rotational errors on the image plane are perpendicular to each other. Furthermore, the estimated projection of the translation on the image lies on a line through the origin and the projection of the real translation. The situation is different for a camera with a full (360 degree) field of view (achieved by a panoramic sensor or by a system of conventional cameras). In this case, at the locations of the minima of the above two functions, either the translational or the rotational error becomes zero, while in the case of a restricted field of view both errors are non-zero. Although some ambiguities still remain in the full field of view case, the implication is that visual navigation tasks, such as visual servoing, involving 3D motion estimation are easier to solve by employing panoramic vision. Finally, the analysis makes it possible to compare properties of algorithms that first estimate the translation and on the basis of the translational result estimate the rotation, algorithms that do the opposite, and algorithms that estimate all motion parameters simultaneously, thus providing a sound framework for the observability of 3D motion.

1 Introduction: Visual Servoing and Motion Estimation

A broad definition of visual servoing amounts to the control of motion on the basis of image analysis. Thus, a driver adjusting the turning of the wheel by monitoring some features in the scene, a system attempting to insert a peg through a hole, a controller at some control panel perceiving a number of screens and adjusting a number of levers, a navigating system trying to find its home, all perform visual servoing tasks. Such systems must respond "appropriately" to changes in their environment; thus, on a high level, they can be described as evolving or dynamical systems, which can be represented as a function from states and control signals to new states; both the states and the control variables may be functions of time. One approach to controlling such a system is to design an observer or state *estimator* to obtain an estimate of the state; for example, this estimator might implement a partial visual recovery process. This estimate is used by a controller or state *regulator* to compute a control signal to drive the dynamical system. Ideally, observation and control are separable in the sense that if we have an optimal controller and an optimal observer then the control system that results from coupling the two is guaranteed to be optimal [13]. Different agents, i.e., dynamical systems, have different capabilities and different amounts of memory, with simple reactive systems at one end of the spectrum and highly sophisticated and flexible systems, making use of scene descriptions and reasoning processes, at the other end.

In the first studies on visual control, efforts concentrated on the upper echelon of this spectrum, trying to equip systems with the capability of estimating accurate 3D motion and the shape of the environment. Assuming that this information could be acquired exactly, sensory feedback robotics was concerned with the planning and execution of the robot's activities. This was characterized as the "look and move" approach and the servoing approaches were "position-or-scene-based." The problem with such separation of perception from action was that both computational goals turned out to be very difficult. On the one hand, the reconstruction of 3D motion and shape is hard to achieve accurately and, in addition, many difficult calibration problems had to be addressed. On the other hand, spatial planning and motion control are very sensitive to errors in the description of the spatiotemporal environment. After this realization and with the emergence of active vision [3–6], attention turned to the lower part of the spectrum, minimizing visual processing and placing emphasis on the regulator, as opposed to the estimator. One of the main lessons learned from research on the static look and move control strategy was that there ought to be easier things to do with images than using them to compute 3D motion and shape. This led to formulations of visual servoing tasks which were such that the controller had to act on the image, i.e., to move the manipulator's joints in such a way that the scene ends up looking a particular way. In technical terms, the controller had to act in such a way that specific coordinate systems (hand, eye, scene, etc.) were put in a particular relationship with each other. Furthermore, this relationship could be realized by using directly available image measurements as feedback for the control loop. This approach is known as image-based robot servoing [16, 18, 21], and in recent years it has given very interesting research results [14, 23, 24, 28]. Most of these results deal with the computation of the image Jacobian (i.e., the differential relationship between the camera frame and the scene frame), along with the camera's intrinsic and extrinsic parameters.

Regardless of the philosophical approach one adopts in visual servoing (scene-based or image-based), the essential aspects of the problem amount to the recovery of the relationship between different coordinate systems (such as the ones between camera, gripper, robot, scene, object, etc.). This could be the relationship itself or a representation of the change of the relationship. Since a servoing system is in general moving (or observing moving parts), a fundamental problem is the recovery of 3D motion from an image sequence. Whether this recovery is explicit or implicit, without it the relationships between different coordinate frames cannot be maintained, and servoing tasks cannot be achieved. Thus, it is important to understand how accurately 3D motion can be observed, in order to understand how successfully servoing can be accomplished. Experience has shown that in practice 3D motion is very difficult to accurately observe, involving many ambiguities and sensitivities. Are there any regularities in which the ambiguity in 3D motion expresses itself? In trying to estimate 3D motion, do the introduced errors satisfy any constraints? As 3D rigid motion consists of the sum of a translation and a rotation, are there differences in the inherent ambiguities governing the recovery of the different components of 3D motion? In addition, we need to study this question independently of specific optimization algorithms. This is the problem studied in this paper.

2 The Approach

There is a veritable cornucopia of techniques for estimating 3D motion but our approach should be algorithm independent, in order for the results to be of general use. Equivalently, our approach should encompass all the possible computational models that can be used to estimate 3D motion from image sequences. To do so, we need to classify all possible approaches to 3D motion on the basis of the input used and the mathematical constraints that are employed.

As a system moves in some environment, every point in the scene has a velocity with regard to the system. The projection of these velocity vectors on the system's eye constitutes the so-called motion field. An estimate of the motion field, called optical flow, starts by first estimating the spatiotemporal derivatives of the image intensity function. These derivatives comprise the so-called normal flow which is the component of the flow along the local image gradient, i.e., normal to the local edge. A system could start 3D motion estimation using normal flow as input or it could first attempt to estimate the optical flow, though that is a very difficult problem, and subsequently 3D motion. This means that an analysis of the difficulty of 3D motion estimation must consider both inputs, optic flow fields and normal flow fields, and algorithms in the literature use one or the other input.

Regarding the mathematical constraints through which 3D motion is encoded in the input image motion, extensive work in this area has established that there exist only two such constraints if no information about the scene is available. The first is the "epipolar constraint" and the second, the "positive depth constraint." The epipolar constraint ensures that corresponding points in the sequence are the projection of the same point in the scene. The positive depth constraint requires that the depth of every scene point be positive, since the scene is in front of the camera. Since knowledge of normal flow does not imply knowledge of corresponding points in the sequence, the epipolar constraint cannot be used when normal flow is available.

The epipolar constraint has attracted most work. It can only be used when optic flow is available. Since many measurements are present, one develops a function that represents deviation from the epipolar constraint all over the image. A variety of approaches can be found in the literature using different metrics in representing epipolar deviation and using different techniques to seek the optimization of the resulting functions. Furthermore, there exist techniques that first estimate rotation and, on the basis of the result subsequently estimate translation [9, 36, 37]; techniques that do the opposite [1, 22, 26, 31, 33, 38, 40]; and techniques that estimate all motion parameters simultaneously [7, 8, 17, 35, 43]. The positive depth constraint, which has been used for normal flow fields, is relatively new and is employed in the so-called direct algorithms [8, 20, 27]. One has to search for the 3D motion that is consistent with the input and produces the minimum amount of negative depth. Put differently, in these approaches the function representing the amount of negative depth must be minimized. Finally, one may be able to use the "positive depth constraint" when optic flow is available, but there exist no algorithms yet in the literature implementing this principle.

Looking at nature we observe that there exists a great variety of eye designs in organisms with vision. An important characteristic is the field of view. There exist systems whose vision has a restricted field of view. This is achieved by the corneal eyes of land vertebrates. Human eyes and common video cameras fall in that category and they are geometrically modeled through central projection on a plane. We refer to these eyes as planar or camera-type eyes. There exist also systems whose vision has a full 360 degree field of view. This is achieved by the compound eyes of insects or by placing camera-type eyes on opposite sides of the head, as in birds and fish. Panoramic vision is adequately modeled geometrically by projecting on a sphere using the sphere's center as the center of projection. In studying the observability of 3D motion, we investigate both restricted field of view and panoramic vision. It turns out that they have different properties regarding 3D motion estimation.

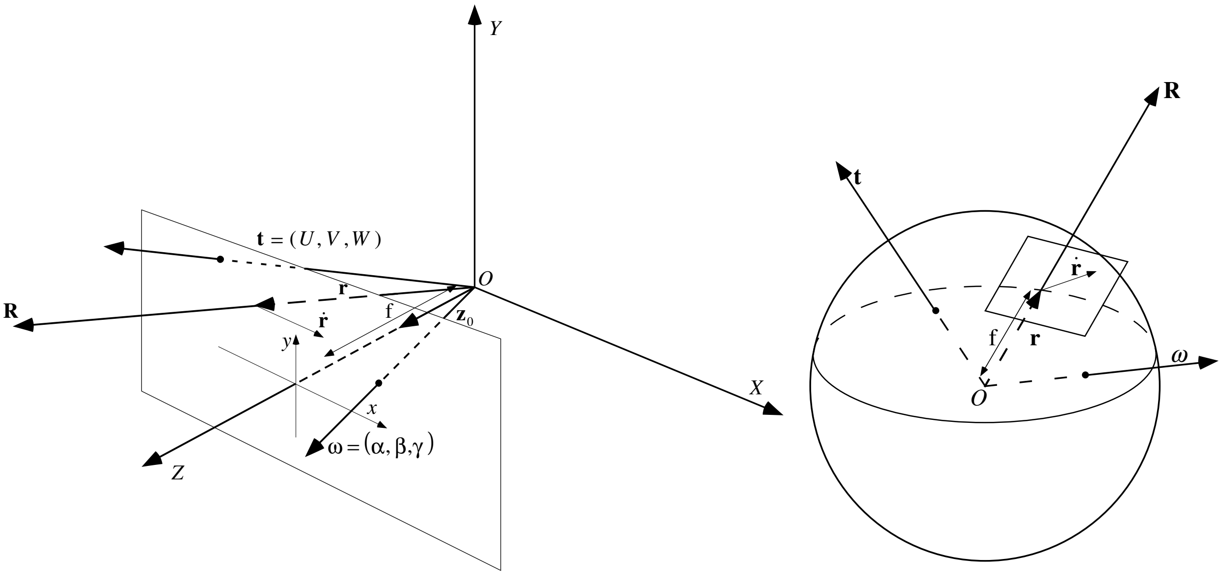
In our approach we employ a statistical model to represent the constraints to derive functions representing deviation from epipolar geometry or the amount of negative depth, for both a camera-type eye and a spherical eye. All possible, meaningful algorithms for 3D motion estimation can be understood from the minimization of these functions. Thus, we perform a topographic analysis of these functions and study their global and local minima. Specifically, we are interested in the relationships between the errors in 3D motion at the points representing the minima of these surfaces. The idea behind this is that in practical situations any estimation procedure is hampered by errors and usually local minima of the functions to be minimized are found as solutions.

3 Problem Statement

3.1 Prerequisites

We use the standard conventions for expressing image motion measurements for a monocular observer moving rigidly in a static environment. We describe the motion of the observer by its translational velocity $\mathbf{t} = (U, V, W)$ and its rotational velocity $\boldsymbol{\omega} = (\alpha, \beta, \gamma)$ with respect to a coordinate system $OXYZ$ fixed to the nodal point of the camera. Each scene point P with coordinates $\mathbf{R} = (X, Y, Z)$ has a velocity $\dot{\mathbf{R}} = -\mathbf{t} - \boldsymbol{\omega} \times \mathbf{R}$ relative to the camera. Projecting $\dot{\mathbf{R}}$ onto a retina of a given shape gives the image motion field. If the image is formed on a plane (Figure 1a) orthogonal to the Z axis at distance f (focal length) from the nodal point, then an image point $\mathbf{r} = (x, y, f)$ and its corresponding scene point $\mathbf{R}$ are related by $\mathbf{r} = \frac{f}{\mathbf{R} \cdot \mathbf{z}_0}\mathbf{R}$, where $\mathbf{z}_0$ is a unit vector in the direction of the Z axis. The motion field becomes

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



(a) (b)

Figure 1: Image formation on the plane (a) and on the sphere (b). The system moves with a rigid motion with translational velocity $\mathbf{t}$ and rotational velocity $\boldsymbol{\omega}$. Scene points $\mathbf{R}$ project onto image points $\mathbf{r}$ and the 3D velocity $\dot{\mathbf{R}}$ of a scene point is observed in the image as image velocity $\dot{\mathbf{r}}$.

with $Z = \mathbf{R} \cdot \mathbf{z}_0$ representing the depth. If the image is formed on a sphere of radius f (Figure 1b) having the center of projection as its origin, the image $\mathbf{r}$ of any point $\mathbf{R}$ is $\mathbf{r} = \frac{\mathbf{R}f}{|\mathbf{R}|}$, with R being the norm of $\mathbf{R}$ (the range), and the image motion is

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

The motion field is the sum of two components, one, $\mathbf{u}_{\mathrm{tr}}$, due to translation and the other, $\mathbf{u}_{\mathrm{rot}}$, due to rotation. The depth Z or range R of a scene point is inversely proportional to the translational flow, while the rotational flow is independent of the scene in view. As can be seen from (1) and (2), the effects of translation and scene depth cannot be separated, so only the direction of translation, $\mathbf{t}/|\mathbf{t}|$, can be computed. We can thus choose the length of $\mathbf{t}$; throughout the following analysis f is set to 1, and the length of $\mathbf{t}$ is assumed to be 1 on the sphere and the Z-component of $\mathbf{t}$ to be 1 on the plane. The problem of 3D motion estimation then amounts to finding the scaled vector $\mathbf{t}$ and the vector $\boldsymbol{\omega}$ from a representation of the motion field. In the following, to make the analysis easier, for the camera-type eye we will employ a non-vector notation. Then, the first two coordinates of $\mathbf{r}$ denote the image point in the Cartesian system oxy, with $ox||OX$, $oy||OY$ and o the intersection of OZ with the image plane. Denote by $(x_0, y_0) = (U/W, V/W)$ the image point representing the direction of the translation vector $\mathbf{t}$ (referred to as the Focus of Expansion (FOE) or Focus of Contraction (FOC) depending on whether W is positive or negative). Equations (1) become then the well known equations expressing the flow measurement $\mathbf{v}(x,y) = (u(x,y), v(x,y))$:

$$\begin{aligned} u(x,y) &= \frac{u_{\mathrm{tr}}(x,y)}{Z} + u_{\mathrm{rot}}(x,y) = (x - x_0)\frac{W}{Z} + \alpha xy - \beta\left(x^2+1\right) + \gamma y \\ v(x,y) &= \frac{v_{\mathrm{tr}}(x,y)}{Z} + v_{\mathrm{rot}}(x,y) = (y - y_0)\frac{W}{Z} + \alpha\left(y+1\right) - \beta xy - \gamma x \end{aligned} \tag{3}$$

where $\frac{\mathbf{v}_{\mathrm{tr}}(x,y)}{Z} = (\frac{u_{\mathrm{tr}}(x,y)}{Z}, \frac{v_{\mathrm{tr}}(x,y)}{Z})$ and $\mathbf{v}_{\mathrm{rot}}(x,y) = (u_{\mathrm{rot}}(x,y), v_{\mathrm{rot}}(x,y))$ are the translational and rotational flow components, respectively.

Regarding the value of the normal flow, if $\mathbf{n}$ is a unit vector at an image point denoting the orientation of the gradient at that point, the normal flow v_n satisfies

$$v_n = \dot{\mathbf{r}} \cdot \mathbf{n} \tag{4}$$

Finally, the following convention is employed throughout the paper. We use letters with hat signs to represent estimated quantities, unmarked letters to represent the actual quantities and the subscript "ϵ" to denote errors, where the error quantity is defined as the actual quantity minus the estimated one. For example, $\mathbf{u}_{\mathrm{rot}}(\boldsymbol{\omega})$ represents actual rotational flow, $\mathbf{u}_{\mathrm{rot}}(\hat{\boldsymbol{\omega}})$ estimated rotational flow, $\mathbf{t}_\epsilon$ the translational error vector, $x_{0_\epsilon} = x_0 - \hat{x}_0$, $\alpha_\epsilon = \alpha - \hat{\alpha}$, etc.

3.2 The model

The classic approach to 3D motion estimation is to minimize the deviation from the epipolar constraint. This constraint is obtained by eliminating depth (or range) from equation (1) (or (2)). For both planar and spherical eyes it is

$$(\mathbf{t} \times \mathbf{r}) \cdot (\dot{\mathbf{r}} + \boldsymbol{\omega} \times \mathbf{r}) = 0. \tag{5}$$

Equating image motion with optic flow, this constraint allows for the derivation of 3D rigid motion on the basis of optic flow measurements. One is interested in the estimates of translation $\hat{\mathbf{t}}$ and rotation $\hat{\boldsymbol{\omega}}$ which best satisfy the epipolar constraint at every point $\mathbf{r}$ according to some criterion of deviation. The Euclidean norm is usually used, leading to the minimization [25, 34] of the function[1]

$$M_{ep} = \iint\limits_{\text{image}} \left[(\hat{\mathbf{t}} \times \mathbf{r}) \cdot (\dot{\mathbf{r}} + \hat{\boldsymbol{\omega}} \times \mathbf{r}) \right]^2 d\mathbf{r}. \tag{6}$$

On the other hand, if normal flow is given, the vector equations (1) and (2) cannot be used directly. The only constraint is scalar equation (4), along with the inequality $Z > 0$ which states that since the surface in view is in front of the eye its depth must be positive. Substituting (1) or (2) into (4) and solving for the estimated depth $\hat{Z}$ or range $\hat{R}$, we obtain for a given estimate $\hat{\mathbf{t}}, \hat{\boldsymbol{\omega}}$ at each point $\mathbf{r}$:

$$\hat{Z}(\text{or } \hat{R}) = \frac{\mathbf{u}_{\mathrm{tr}}(\hat{\mathbf{t}}) \cdot \mathbf{n}}{(\dot{\mathbf{r}} - \mathbf{u}_{\mathrm{rot}}(\hat{\boldsymbol{\omega}})) \cdot \mathbf{n}}. \tag{7}$$

If the numerator and denominator of (7) have opposite signs, negative depth is computed. Thus, to utilize the positivity constraint one must search for the motion $\hat{\mathbf{t}}, \hat{\boldsymbol{\omega}}$ that produces a minimum number of negative depth estimates. Formally, if $\mathbf{r}$ is an image point, define the indicator function

$$I_{nd}(\mathbf{r}) = \begin{cases} 1 \text{ for } \left(\mathbf{u}_{\mathrm{tr}}(\hat{\mathbf{t}}) \cdot \mathbf{n}\right) \left((\dot{\mathbf{r}} - \mathbf{u}_{\mathrm{rot}}(\hat{\boldsymbol{\omega}})) \cdot \mathbf{n}\right) < 0 \\ 0 \text{ for } \left(\mathbf{u}_{\mathrm{tr}}(\hat{\mathbf{t}}) \cdot \mathbf{n}\right) \left((\dot{\mathbf{r}} - \mathbf{u}_{\mathrm{rot}}(\hat{\boldsymbol{\omega}})) \cdot \mathbf{n}\right) > 0 \end{cases}.$$

Then estimation of 3D motion from normal flow amounts to minimizing [19, 20, 27] the function

$$M_{nd} = \iint\limits_{\text{image}} I_{nd}(\mathbf{r}) d\mathbf{r}. \tag{8}$$

Expressing $\dot{\mathbf{r}}$ in terms of the real motion from (1) and (2), functions (6) and (8) can be expressed in terms of the actual and estimated motion parameters $\mathbf{t}$, $\boldsymbol{\omega}$, $\hat{\mathbf{t}}$ and $\hat{\boldsymbol{\omega}}$ (or, equivalently, the actual motion parameters $\mathbf{t}, \boldsymbol{\omega}$ and the errors $\mathbf{t}_\epsilon = \mathbf{t} - \hat{\mathbf{t}}$, $\boldsymbol{\omega}_\epsilon = \boldsymbol{\omega} - \hat{\boldsymbol{\omega}}$) and the depth Z (or range R) of the viewed scene. To conduct any analysis, a model for the scene is needed. We are interested in the statistically expected values of the motion estimates resulting from all possible scenes. Thus, as our probabilistic model we assume that the depth values of the scene are uniformly distributed between two arbitrary values $Z_{\min}$(or $R_{\min}$) and $Z_{\max}$(or $R_{\max}$) ($0 < Z_{\min} < Z_{\max}$). For the minimization of negative depth values, we further assume that the directions in which flow measurements

[1] Because $\mathbf{t} \times \mathbf{r}$ introduces the sine of the angle between $\mathbf{t}$ and $\mathbf{r}$, the minimization prefers vectors $\mathbf{t}$ close to the center of gravity of the points $\mathbf{r}$. This bias has been recognized [40] and alternatives have been proposed that reduce this bias, but without eliminating the confusion between rotation and translation.

are made are uniformly distributed in every direction for every depth. Parameterizing $\mathbf{n}$ by ψ, the angle between $\mathbf{n}$ and the x axis, we thus obtain the following two functions:

$$E_{ep} = \int\limits_{Z=Z_{\min}}^{Z_{\max}} M_{ep} dZ, \qquad (9) \qquad\qquad E_{nd} = \int\limits_{\psi=0}^{\pi} \int\limits_{Z=Z_{\min}}^{Z_{\max}} M_{nd} dZ\, d\psi, \qquad (10)$$

measuring deviation from the epipolar constraint and the amount of negative depth, respectively. Functions (9) and (10) are five-dimensional surfaces in $\mathbf{t}_\epsilon, \boldsymbol{\omega}_\epsilon$, the errors in the motion parameters. Finally, since for the scene in view we employ a probabilistic model, the results are of a statistical nature, that is, the geometric constraints between $\mathbf{t}_\epsilon, \boldsymbol{\omega}_\epsilon$ at the minima of (9) and (10) that we shall uncover should be interpreted as being likely to occur.

3.3 Negative depth and depth distortion

This section contains a few technical prerequisites needed for the study of negative depth minimization and a geometric observation that show the relationship of epipolar minimization to minimization of negative depth.

Equation (7) shows the estimated depth $\hat{Z}$ (or range $\hat{R}$) given normal flow $\dot{\mathbf{r}} \cdot \mathbf{n}$ and estimates $\hat{\mathbf{t}}, \hat{\boldsymbol{\omega}}$ of the motion. It can be further written as:

$$\hat{Z}(\text{or } \hat{R}) = Z(\text{or } R)D \qquad (11)$$

with D in the case of noiseless motion measurements (that is, $\dot{\mathbf{r}} = \frac{\mathbf{u}_{\text{tr}}(\mathbf{t})}{Z} + \mathbf{u}_{\text{rot}}(\boldsymbol{\omega})$) of the form

$$D = \frac{\mathbf{u}_{\text{tr}}(\hat{\mathbf{t}}) \cdot \mathbf{n}}{(\mathbf{u}_{\text{tr}}(\mathbf{t}) + Z\mathbf{u}_{\text{rot}}(\boldsymbol{\omega}_\epsilon)) \cdot \mathbf{n}} \qquad (12)$$

Equation (11) shows how wrong depth estimates are produced due to inaccurate 3D motion values. The distortion factor, D, multiplies the real depth value to produce the estimate. Equation (12), for a fixed value of D and $\mathbf{n}$ describes a surface in $(\mathbf{r}, Z)$ (or $(\mathbf{r}, R)$) space which is called an iso-distortion surface. Any such surface is to be understood as the locus of points in space which are distorted in depth by the same multiplicative factor if the image measurements are in direction $\mathbf{n}$. If we fix $\mathbf{n}$ and vary D, the iso-distortion surfaces of the resulting family change continuously as D varies. Thus all scene points giving rise to negative depth estimates lie between the 0 and $-\infty$ distortion surfaces. The integral over all points (for all directions) giving rise to negative depth estimates we call the negative depth volume. In Section 5 we will make use of the iso-distortion surfaces and the negative depth volume to study in a geometric way function (10) resulting from minimization of the negative depth values.

Let us now examine the two different minimizations from a geometric perspective. When deriving the deviation from the epipolar constraint we consider integration over all points and depth values of the expression

$$[(\hat{\mathbf{t}} \times \mathbf{r}) \cdot (\dot{\mathbf{r}} + \hat{\boldsymbol{\omega}} \times \mathbf{r})]^2$$

or, equivalently,

$$[(\dot{\mathbf{r}} - \mathbf{u}_{\text{rot}}(\hat{\boldsymbol{\omega}})) \cdot \mathbf{u}_{\text{tr}}^{\perp}(\hat{\mathbf{t}})]^2$$

where $\mathbf{u}_{\text{tr}}^{\perp}(\hat{\mathbf{t}})$ denotes the vector perpendicular to $\mathbf{u}_{\text{tr}}(\hat{\mathbf{t}})$ in the plane of the image coordinates. When deriving the number of negative depth values we consider integration over all points and depth values of the angle α between the two vectors

$$\hat{\mathbf{u}}_{\text{tr}}(\hat{\mathbf{t}}) \qquad \text{and} \qquad \dot{\mathbf{r}} - \mathbf{u}_{\text{rot}}(\hat{\boldsymbol{\omega}})$$

An illustration is given in Figure 2. The two measures considered in the two different minimizations are clearly related to each other. In the case of negative depth we consider the angle α, whereas in the case of the epipolar constraint we consider the squared distance a^2, which amounts to $(\sin\alpha \|\mathbf{u}_{\text{tr}}(\hat{\mathbf{t}})\| \cdot \|\dot{\mathbf{r}} - \mathbf{u}_{\text{rot}}(\hat{\boldsymbol{\omega}})\|)^2$.

The major difference in using the two measures arises for angles $\alpha > 90°$. The measure for negative depth is monotonic in α, but the measure for the deviation from the epipolar constraint is not, because it does not differentiate between depth estimates of positive and negative value.

The fact that the computed depth has to be positive has not been considered in past approaches employing minimization of the epipolar constraint. As this fact provides an additional constraint which could be utilized in

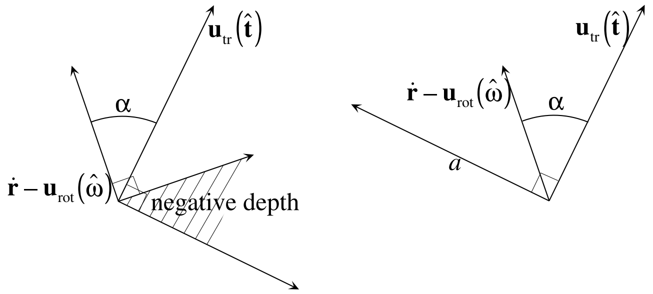


Figure 2: Different measures used in minimization constraints. (a) $\angle\alpha$ in the minimization of negative depth values from projections of motion fields. (b) a^2 in the epipolar constraint.

the future, maybe in conjunction with the epipolar constraint, we are interested in the influence of this constraint on 3D motion estimation as well. Thus, in addition to minimization of the functions (9) and (10) discussed above, we study the minimization of a third function. This function amounts to the number of values giving rise to negative depth when full flow measurements (optical flow) are assumed and the analysis is presented in Appendix B.

3.4 A model for the noise in the flow

In the analysis on the plane we will also consider noise in the image measurements, that is, we will consider the flow values of the form $(u + N_x, v + N_y) = (\frac{u_{\rm tr}}{Z} + u_{\rm rot} + N_x, \frac{v_{\rm tr}}{Z} + v_{\rm rot} + N_y)$. The choice of the noise model is motivated by the following considerations: First, the noise should be such that no specific directions are favored. Second, assuming that noise is both additive and multiplicative, there should be a dependence between noise and depth, because the translational flow component is proportional to the inverse depth. Therefore we define noise using two stochastic variables $\boldsymbol{\epsilon} = (\epsilon_x, \epsilon_y)$ and $\boldsymbol{\delta} = (\delta_x, \delta_y)$:

$$(N_x, N_y) = (\epsilon_x, \epsilon_y)\frac{1}{Z} + (\delta_x, \delta_y) \tag{13}$$

with the first and second moments being

$$\begin{array}{rrrr} E(\epsilon_x) = & E(\epsilon_y) = & E(\delta_x) = & E(\delta_y) = 0 \\ E(\delta_x^2) = & E(\delta_y^2) = & \sigma_\delta^2 & E(\delta_x\delta_y) = 0 \\ E(\epsilon_x^2) = & E(\epsilon_y^2) = & \sigma_\epsilon^2 & E(\epsilon_x\epsilon_y) = 0 \end{array}$$

and all stochastic variables being independent of each other, independent of image position, and independent of the depth.

3.5 Overview of the paper, summary of results and related work

Our approach expresses functions (9) and (10) in terms of $\mathbf{t}$, $\boldsymbol{\omega}$, $\mathbf{t}_\epsilon$ and $\boldsymbol{\omega}_\epsilon$ and finds the conditions that $\mathbf{t}_\epsilon$ and $\boldsymbol{\omega}_\epsilon$ satisfy at local minima which represent solutions of the different estimation algorithms. Procedures for estimating 3D motion can be classified into those estimating either the translation or rotation as a first step and the remaining component (that is, the rotation or translation) as a second step, and those estimating all components simultaneously. Procedures of the former kind result when systems utilize inertial sensors which provide them with estimates of one of the components of the motion, or when two-step motion estimation algorithms are used.

Thus, three cases need to be studied: the case were no prior information about 3D motion is available and the cases where an estimate of translation or rotation is available with some error. Imagine that somehow the rotation has been estimated, with an error $\boldsymbol{\omega}_\epsilon$. Then our functions become two-dimensional in the variables $\mathbf{t}_\epsilon$ and represent the space of translational error parameters corresponding to a fixed rotational error. Similarly, given a translational error $\mathbf{t}_\epsilon$, the functions become three-dimensional in the variables $\boldsymbol{\omega}_\epsilon$ and represent the space of rotational errors corresponding to a fixed translational error. To study the general case, one needs to consider the lowest valleys of the functions in 2D subspaces which pass through 0. In the image processing literature, such local minima are often referred to as ravine lines or courses. Each of the three cases is studied for four optimizations: epipolar minimization for the sphere and the plane (full field of view and restricted field of view vision) and minimization of negative depth for the sphere and the plane. Thus, there are twelve (four times three) cases, but since the effects of rotation on the image are independent of depth, it makes no sense to perform minimization of negative depth assuming an estimate of translation is available. Thus, we are left with ten different cases which are studied below. These ten cases represent all the possible, meaningful motion estimation procedures.

The analysis shows that:

1. In the case of a camera-type eye (restricted field of view) for algorithms in both classes which estimate all motion parameters simultaneously (i.e., there is no prior information), the obtained solution will have an error $(x_{0_\epsilon}, y_{0_\epsilon})$ in the translation and $(\alpha_\epsilon, \beta_\epsilon, \gamma_\epsilon)$ in the rotation such that $\frac{x_{0_\epsilon}}{y_{0_\epsilon}} = \frac{-\beta_\epsilon}{\alpha_\epsilon}$ and $\gamma_\epsilon = 0$. This means that the projections of the translational and rotational errors on the image are perpendicular to each other and that the rotation around the Z axis has the least ambiguity. We refer to this constraint as the "orthogonality constraint." In addition, the estimated translation $(\hat{x}_0, \hat{y}_0)$ will lie on a line passing from the origin of the image and the real translation (x_0, y_0), i.e., $\frac{x_0}{y_0} = \frac{\hat{x_0}}{\hat{y_0}} = \frac{x_{0_\epsilon}}{y_{0_\epsilon}}$. We refer to this second constraint as the "line constraint." Similar results are achieved for the case where translation is estimated first and on that basis rotation is subsequently found, while the case where rotation is first estimated and subsequently translation provides different results. The work is performed both in the absence of error in the image measurements—in which case it becomes a geometric analysis of the inherent confusion between rotation and translation—and in the case where the image measurements are corrupted by noise that satisfies the model from Section 3.4. In this case, we derive the expected values of the local and global minima. As will be shown, the noise does not alter the local minima, and the global minima fall within the valleys of the function without noise. Thus, the noise does not alter the functions' overall structure.

2. In the case of panoramic vision, for algorithms in both classes which estimate all motion parameters simultaneously, the obtained solution will have no error in the translation ($\mathbf{t}_\epsilon = 0$) and the rotational error will be perpendicular to the translation. In addition, for the case of epipolar minimization from optic flow, given a translational error $\mathbf{t}_\epsilon$, the obtained solution will have no error in the rotation ($\boldsymbol{\omega}_\epsilon = 0$), while for the case of negative depth minimization from normal flow, given a rotational error the obtained solution will have no error in the translation. In other cases ambiguities remain. In the spherical eye case the analysis is simply performed for noiseless flow.

   A large number of error analyses have been carried out [2, 11, 12, 15, 29, 39, 41, 42] in the past for a camera-type eye, while there is no published research of this kind for the full field of view case. None of the existing studies, however, has attempted a topographic characterization of the function to be minimized for the purpose of analyzing different motion techniques. All the studies consider optical flow or correspondence as image measurements and investigate minimizations based on the epipolar constraint. Often, restrictive assumptions about the structure of the scene or the estimator have been made, but the main results obtained are in accordance with our findings. In particular, the following two results already occur in the literature.

   (a) Translation along the x axis can be easily confounded with rotation around the y axis, and translation along the y axis can be easily confounded with rotation around the x axis, for small fields of view and insufficient depth variation. This fact has long been known from experimental observation, and has been proved for planar scene structures and unbiased estimators [10]. The orthogonality constraint found here confirms these findings, and imposes even more restrictive constraints. It shows, in addition, that the x-translation and y-rotation, and y-translation and x-rotation, are not decoupled. Furthermore, we have found that rotation around the Z axis can be most easily distinguished from the other motion components.

   (b) Maybank [32, 33] and Jepson and Heeger [29] established before the line constraint, but under much more restrictive assumptions. In particular, they showed that for a small field of view, a translation $\mathbf{t}$ far away from the image center and an irregular surface, the function in (9) has its minima along a line in the space of translation directions which passes through the true translation and the viewing direction. Under fixation the viewing direction becomes the image center.

4 Epipolar Minimization: Camera-type Eye

Let $\mathbf{t} = (x_0, y_0, 1)$ and $\boldsymbol{\omega} = (\alpha, \beta, \gamma)$. Since the field of view is small, the quadratic terms in the image coordinates are very small relative to the linear and constant terms, and are therefore ignored. All the computations are carried out with the symbolic algebraic computation software Maple, and, for abbreviation, intermediate results are not given. First the case of noise-free flow is studied.

Considering a circular aperture of radius e, setting the focal length $f = 1$, $W = 1$ and $\hat{W} = 1$, the function in (9) becomes

$$E_1 = E_{ep} = \int_{Z=Z_{\min}}^{Z_{\max}} \int_{r=0}^{e} \int_{\phi=0}^{2\pi} \left\{ \left(\left(\frac{x - x_0}{Z} - \beta_\epsilon + \gamma_\epsilon y + x \right) (y - \hat{y}_0) \right.\right.$$
$$\left.\left. - \left(\frac{y - y_0}{Z} + \alpha_\epsilon - \gamma_\epsilon x + y \right) (x - \hat{x}_0) \right)^2 r \right\} dr\, d\phi\, dZ$$

where (r, ϕ) are polar coordinates ($x = r\cos\phi$, $y = r\sin\phi$). Performing the integration, one obtains

$$\begin{aligned} E_1 \;=\; & \pi e^2 \Bigg((Z_{\max} - Z_{\min}) \left(\frac{1}{3}\gamma_\epsilon^2 e^4 + \frac{1}{4}\left(\gamma_\epsilon^2 \left(\hat{x}_0^2 + \hat{y}_0^2\right) + 6\gamma_\epsilon \left(\hat{x}_0\alpha_\epsilon + \hat{y}_0\beta_\epsilon\right) + \alpha_\epsilon^2 + \beta_\epsilon^2 \right) e^2 + \right. \\ & \left. (\hat{x}_0\alpha_\epsilon + \hat{y}_0\beta_\epsilon)^2 \right) + \left(\ln(Z_{\max}) - \ln(Z_{\min})\right) \left(\frac{1}{2}(3\gamma_\epsilon(x_{0_\epsilon} y_0 - y_{0_\epsilon} x_0) + x_{0_\epsilon}\beta_\epsilon - y_{0_\epsilon}\alpha_\epsilon)e^2 + \right. \\ & \left. 2\left(x_{0_\epsilon} y_0 - y_{0_\epsilon} x_0\right)\left(\hat{x}_0\alpha_\epsilon + \hat{y}_0\beta_\epsilon\right) \right) + \\ & \left(\frac{1}{Z_{\min}} - \frac{1}{Z_{\max}} \right) \left(\frac{1}{4}\left(y_{0_\epsilon}^2 + x_{0_\epsilon}^2\right) e^2 + \left(x_{0_\epsilon} y_0 - y_{0_\epsilon} x_0\right)^2 \right) \Bigg) \end{aligned} \tag{14}$$

(a) Assume that the translation has been estimated with a certain error $\mathbf{t}_\epsilon = (x_{0_\epsilon}, y_{0_\epsilon}, 0)$. Then the relationship among the errors in 3D motion at the minima of (14) is obtained from the first-order conditions $\frac{\partial E_1}{\partial \alpha_\epsilon} = \frac{\partial E_1}{\partial \beta_\epsilon} = \frac{\partial E_1}{\partial \gamma_\epsilon} = 0$, which yield

$$\alpha_\epsilon = \frac{y_{0_\epsilon}\left(\ln(Z_{\max}) - \ln(Z_{\min})\right)}{Z_{\max} - Z_{\min}} \qquad \beta_\epsilon = \frac{-x_{0_\epsilon}\left(\ln(Z_{\max}) - \ln(Z_{\min})\right)}{Z_{\max} - Z_{\min}} \qquad \gamma_\epsilon = 0 \tag{15}$$

It follows that $\alpha_\epsilon/\beta_\epsilon = -x_{0_\epsilon}/y_{0_\epsilon}$, $\gamma_\epsilon = 0$.

(b) Assuming that rotation has been estimated with an error $(\alpha_\epsilon, \beta_\epsilon, \gamma_\epsilon)$, the relationship among the errors is obtained from $\frac{\partial E_1}{\partial x_{0_\epsilon}} = \frac{\partial E_1}{\partial y_{0_\epsilon}} = 0$. In this case, the relationship is very elaborate and the translational error depends on all the other parameters—that is, the rotational error, the actual translation, the image size and the depth interval. See Appendix A.

(c) In the general case, we need to study the subspaces in which E_1 changes least at its absolute minimum; that is, we are interested in the direction of the smallest second derivative at 0, the point where the motion errors are zero. To find this direction, we compute the Hessian at 0, that is the matrix of the second derivatives of E_1 with respect to the five motion error parameters, and compute the eigenvector corresponding to the smallest eigenvalue. The scaled components of this vector amount to

$$\begin{aligned} & x_{0_\epsilon} = x_0 \qquad y_{0_\epsilon} = y_0 \qquad \beta_\epsilon = -\alpha_\epsilon \tfrac{x_0}{y_0} \qquad \gamma_\epsilon = 0 \\ & \alpha_\epsilon = \; 2y_0 Z_{\min} Z_{\max} \left(\ln(Z_{\max}) - \ln(Z_{\min})\right) / \\ & \qquad \Big((Z_{\max} - Z_{\min})(Z_{\max} Z_{\min} - 1) \\ & \qquad + \left((Z_{\max} - Z_{\min})^2 (Z_{\max} Z_{\min} - 1)^2 + 4Z_{\max}^2 Z_{\min}^2 \left(\ln(Z_{\max}) - \ln(Z_{\min})\right)^2 \right)^{1/2} \Big) \end{aligned}$$

As can be seen, for points defined by this direction, the translational and rotational errors are characterized by the "orthogonality constraint" $\alpha_\epsilon/\beta_\epsilon = -x_{0_\epsilon}/y_{0_\epsilon}$ and by the "line constraint" $x_0/y_0 = \hat{x}_0/\hat{y}_0$.

Next we consider noise in the flow measurements. We model the noise (N_x, N_y) as defined in Section 3.4 and derive $E(E_{ep})$, the expected value of E_{ep}, which amounts to

$$E(E_{ep}) = E_1 + \int\limits_{Z=Z_{\min}}^{Z_{\max}} \int\limits_{\phi=0}^{2\pi} \int\limits_{r=0}^{e} \left\{ \left(\sigma_\epsilon^2 \frac{1}{Z^2} \left[(\hat{y}_0 - y)^2 + (\hat{x}_0 - x)^2 \right] \right. \right.$$
$$\left. \left. + \sigma_\delta^2 \left[(\hat{y}_0 - y)^2 + (\hat{x}_0 - x)^2 \right] \right) r \right\} dr\, d\phi\, dZ \qquad (16)$$

and thus

$$E(E_{ep}) = E_1 + e^2 \pi \left(\sigma_\epsilon^2 \left(\frac{1}{Z_{\min}} - \frac{1}{Z_{\max}} \right) + \sigma_\delta^2 (Z_{\max} - Z_{\min}) \right) \left(\hat{y}_0^2 + \hat{x}_0^2 + \frac{e^2}{2} \right)$$

(a) If we fix x_{0_ϵ} and y_{0_ϵ} and solve for

$$\frac{\partial E(E_{ep})}{\partial \alpha_\epsilon} = 0 \qquad \frac{\partial E(E_{ep})}{\partial \beta_\epsilon} = 0 \qquad \frac{\partial E(E_{ep})}{\partial \gamma_\epsilon} = 0$$

we obtain the same relationship as for noiseless flow described in (15).

This shows that the noise does not alter the expected behavior of techniques which in a first step minimize the translation.

(b) If we fix α_ϵ, β_ϵ, and γ_ϵ, and solve for

$$\frac{\partial E(E_{ep})}{\partial x_{0_\epsilon}} = 0 \qquad \frac{\partial E(E_{ep})}{\partial y_{0_\epsilon}} = 0$$

we obtain as before a complicated relationship between the translational error, the actual translation, and the rotational error.

(c) To analyze the behavior of techniques which minimize for all 3D motion parameters, we study the global minimum of $E(E)$. From minimization with regard to the rotational parameters, we obtained (15) (the orthogonality constraint and $\gamma_\epsilon = 0$). Substituting (15) into (16) and solving for

$$\frac{\partial E(E_{ep})}{\partial x_{0_\epsilon}} = 0 \qquad \frac{\partial E(E_{ep})}{\partial y_{0_\epsilon}} = 0$$

we get, in addition,

$$\begin{aligned} x_{0_\epsilon} &= 4x_0 (Z_{\max} - Z_{\min})^2 \left(\sigma_\epsilon^2 + \sigma_\delta^2 Z_{\max} Z_{\min} \right) / \\ &\quad \left((Z_{\max} - Z_{\min})^2 \left(4\sigma_\epsilon^2 + 4\sigma_\delta^2 Z_{\max} Z_{\min} + e^2 \right) \right. \\ &\quad \left. - Z_{\max} Z_{\min} e^2 \left(\ln(Z_{\max}) - \ln(Z_{\min}) \right)^2 \right) \\ y_{0_\epsilon} &= x_{0_\epsilon} \frac{y_0}{x_0} \end{aligned}$$

Thus the absolute minimum of $E(E_{ep})$ is to be found in the direction of smallest increase in E_1, and is described by the constraint $\gamma_\epsilon = 0$, by the orthogonality constraint, and by the line constraint.

5 Minimization of Negative Depth Volume: Camera-type Eye

In the following analysis we study the function describing the negative depth values geometrically, by means of the negative depth volumes, that is, the points corresponding to negative depth distortion as defined in Section 3.3. This allows us to incrementally derive properties of the function without considering it with respect to all its parameters at once. For simplicity, we assume that the FOE and the estimated FOE are inside the image, and we do not consider the exact effects resulting from volumes of negative depth in different directions being outside the field of view. We first concentrate on the noiseless case. If, as before, we ignore terms quadratic in the image coordinates, the 0 distortion surface (from equation (11)) becomes

$$(x - \hat{x}_0)\, n_x + (y - \hat{y}_0)\, n_y = 0 \tag{17}$$

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

$$(x - x_0)n_x + (y - y_0)n_y + Z\left(\left(-\beta_\epsilon + \gamma_\epsilon y\right) n_x + \left(\alpha_\epsilon - \gamma_\epsilon x\right) n_y\right) = 0 \tag{18}$$

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 $[x', y']^T = R[x, y]^T$, $[x_0', y_0']^T = R[x_0, y_0]^T$, $[\hat{x}_0', \hat{y}_0']^T = R[\hat{x}_0, \hat{y}_0]^T$, $[\alpha_\epsilon', \beta_\epsilon']^T = R[\alpha_\epsilon, \beta_\epsilon]^T$, where $R = \begin{bmatrix} \cos\psi & \sin\psi \\ -\sin\psi & \cos\psi \end{bmatrix}$.

Equations (17) and (18) thus become $(x' - \hat{x}_0') = 0$ and $(x' - x_0') + Z\left(-\beta_\epsilon' + \gamma_\epsilon y'\right) = 0$.

(a) To investigate techniques which, as a first step, estimate the rotation, we first study the case of $\gamma_\epsilon = 0$ and then extend the analysis to the general case of $\gamma_\epsilon \neq 0$.

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

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

The equation $(x' - \hat{x}_0') = 0$ describes a plane parallel to the $y'Z$ plane at distance $\hat{x}_0'$ from the origin, and $(x' - x_0') - \beta_\epsilon' Z = 0$ describes a plane parallel to the y' axis of slope $\frac{1}{\beta_\epsilon'}$, which intersects the $x'y'$ plane in the x' coordinate x_0'. Thus we obtain a wedge-shaped volume parallel to the y' axis. Figure 3 illustrates the volume through a slice parallel to the $x'Z$ plane.

The scene in view extends between the depth values $Z_{\min}$ and $Z_{\max}$. We denote by A_ψ the area of the cross section parallel to the $x'Z$ plane through the negative depth volume in direction ψ.

As can be seen from Figure 3, to obtain the minimum, $\hat{x}_0'$ has to lie between $x_0' + \beta_\epsilon' Z_{\min}$ and $x_0' + \beta_\epsilon' Z_{\max}$; then $A_\psi = \int\limits_{Z=Z_{\min}}^{Z_{\max}} \left| x_0' - \hat{x}_0' + \beta_\epsilon' Z \right| dZ$ which amounts to

$$A_\psi = \left| x_{0_\epsilon}' \left(Z_{\max} + Z_{\min}\right) + \frac{\beta_\epsilon'}{2}\left({Z_{\max}}^2 + {Z_{\min}}^2\right) + \frac{{x_{0_\epsilon}'}^2}{\beta_\epsilon} \right| \tag{19}$$

If we fix β_ϵ' and solve for x_{0_ϵ}' $\frac{\partial A_\psi}{\partial x_{0_\epsilon}'} = 0$, we obtain

$$x_{0_\epsilon}' = -\frac{\beta_\epsilon'}{2}\left(Z_{\max} + Z_{\min}\right) \tag{20}$$

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

$\frac{x_{0_\epsilon}'}{\beta_\epsilon'}$ depends only on the depth interval, and thus is independent of the direction, ψ. Therefore, the negative depth volume is minimized if (20) holds for every direction. Since $\beta_\epsilon' = \cos\psi\beta_\epsilon - \sin\psi\alpha_\epsilon$ and $x_{0_\epsilon}' = \cos\psi x_{0_\epsilon} + \sin\psi y_{0_\epsilon}$, we obtain $\frac{x_{0_\epsilon}}{y_{0_\epsilon}} = \frac{-\beta_\epsilon}{\alpha_\epsilon}$.

If $\gamma_\epsilon \neq 0$, the $-\infty$ distortion surface becomes $(x' - x_0{}') + Z\left(-\beta_\epsilon{}' + \gamma_\epsilon y'\right) = 0$.

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 at $x' = x_0{}' + \beta_\epsilon{}' Z$ (see Figure 4a). For any given Z the slopes of the lines in different directions are the same. An illustration of the volume of negative depth is given in Figure 4b.

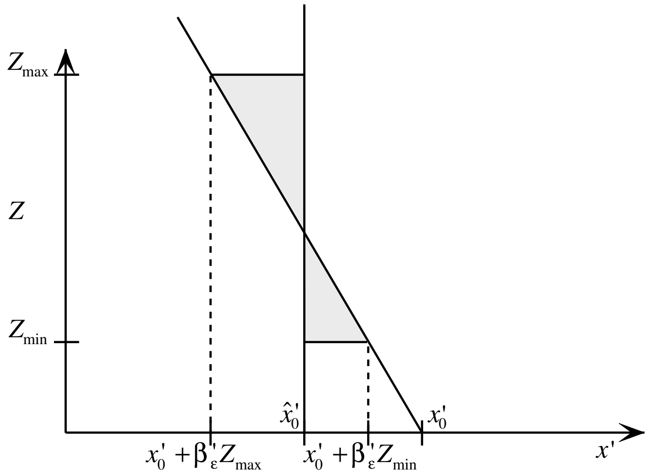
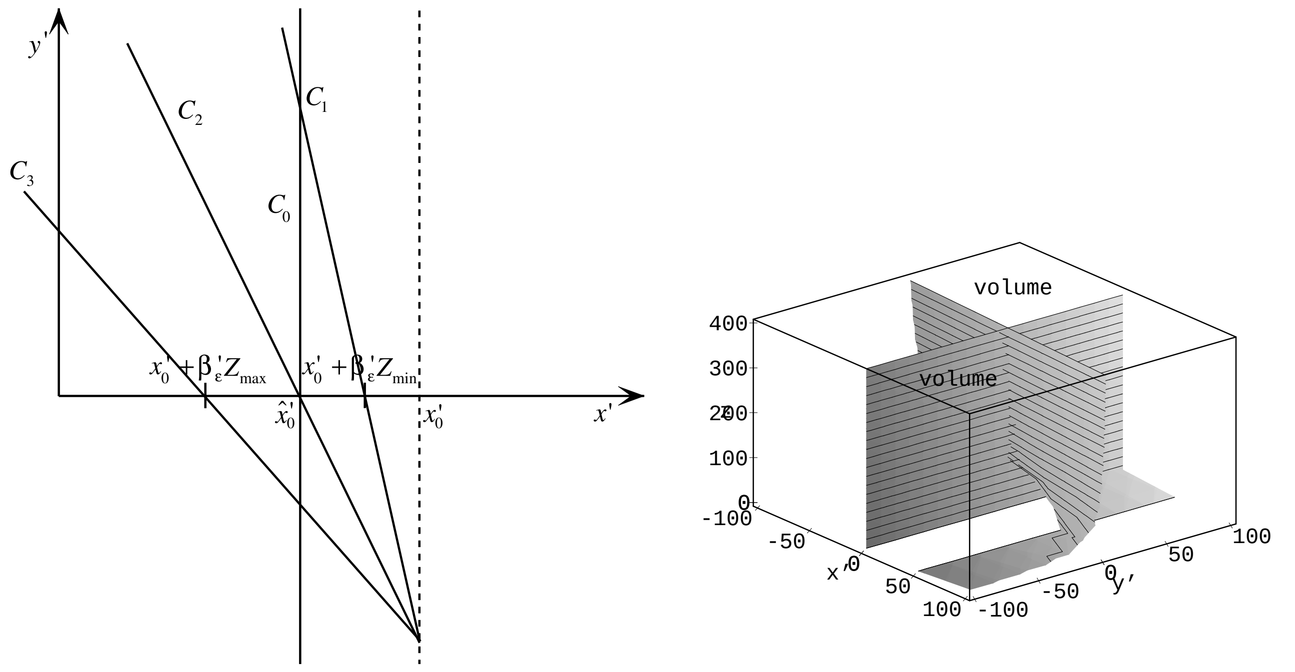


Figure 3: Slice parallel to the $x'Z$ plane through the volume of negative estimated depth for a single direction.



(a) (b)

Figure 4: (a) 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), $Z = -\frac{x'_{0_\epsilon}}{\beta_\epsilon{}'}$ (C_2), and $Z = Z_{\max}$ (C_3). (b) $\gamma_\epsilon \neq 0$: volume of negative depth values between the 0 and $-\infty$ distortion surfaces.

Let us express the value found for $\hat{x}_0'$ in the case of $\gamma_\epsilon = 0$ as $\hat{x}_0' = x_0{}' + \beta_\epsilon{}' Z_I$. In order to derive the position of $\hat{x}'_0$ that minimizes the negative depth volume for the general case of $\gamma_\epsilon \neq 0$, we study the change of volume as $\hat{x}'_0$ changes from $x_0{}' + \beta_\epsilon{}' Z_I$.

Referring to Figure 5, 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 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 $x' = x_0{}' + \beta_\epsilon{}' Z_I$ and $x' = x_0{}' + \beta_\epsilon{}' Z_I + d$. These coordinates are

$$\begin{aligned} & y'_1 &= \tfrac{\beta_\epsilon{}'(Z-Z_I)}{\gamma_\epsilon Z} \\ \text{and} \quad & y'_2 &= \tfrac{\beta_\epsilon{}'(Z-Z_I)-d}{\gamma_\epsilon Z} \end{aligned}$$

Therefore

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

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

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

which amounts to

$$V_c = -\mathrm{sgn}\left(\gamma_\epsilon\right) \frac{d}{\gamma_\epsilon} \left(2\beta_\epsilon{}' \left(Z_{\max} - Z_{\min}\right) - \left(2\beta_\epsilon{}' Z_I + d\right) \left(\ln\left(Z_{\max}\right) - \ln\left(Z_{\min}\right)\right)\right) \tag{21}$$

Substituting for $Z_I = \frac{Z_{\min} + Z_{\max}}{2}$, it can be verified that in order for V_c to be negative, we must have $\mathrm{sgn}(\beta_\epsilon{}') = -\mathrm{sgn}(d)$.

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

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

we obtain

$$d = \beta_\epsilon{}' \left(\frac{Z_{\max} - Z_{\min}}{\ln\left(Z_{\max}\right) - \ln\left(Z_{\min}\right)} - Z_I \right)$$

Thus

$$x'_{0_\epsilon} = -\beta'_\epsilon \frac{\left(Z_{\max} - Z_{\min}\right)}{\ln\left(Z_{\max}\right) - \ln\left(Z_{\min}\right)} \tag{22}$$

Since $\frac{x'_{0_\epsilon}}{\beta'_\epsilon}$ depends only on the depth interval, the total negative depth volume is obtained if the volume in every direction is minimized. Therefore, for any rotational error $(\alpha_\epsilon, \beta_\epsilon, \gamma_\epsilon)$ and independent of γ_ϵ, we have the orthogonality constraint:

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

A 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 inside 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 directions 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 inside the image can (22) be used to describe the value of x'_{0_ϵ} as a function of β'_ϵ and the interval of depth values of the scene in view.

(b) Next we investigate techniques which minimize both the translation and rotation at once. First consider a certain translational error and change the value of γ_ϵ. An increase in γ_ϵ decreases the slope of the $-\infty$ distortion surfaces, and thus, as can be inferred from Figure 4a, the area of negative depth values for every direction ψ and every depth Z increases. Thus $\gamma_\epsilon = 0$. In addition, from above, we know that $\frac{x_{0_\epsilon}}{y_{0_\epsilon}} = \frac{-\beta_\epsilon}{\alpha_\epsilon}$. The exact relationship of β'_ϵ to x'_{0_ϵ} is characterized by the locations of local minima of the function A_ψ, which in the image processing literature are often referred to as "courses." (These are not the courses in the topographical sense [30].) To be more precise, we are interested in the local minima of A_ψ in the direction corresponding to the largest second derivative. We compute the largest eigenvalue, λ_1, of the Hessian, H, of A_ψ, that is the matrix of the second derivatives of A_ψ with respect to x'_{0_ϵ} and β'_ϵ, which amounts to $H = \begin{bmatrix} \frac{2}{\beta'_\epsilon} & \frac{-2{x_{0_\epsilon}}'}{\beta'^2_\epsilon} \\ \frac{-2{x_{0_\epsilon}}'}{\beta'^2_\epsilon} & \frac{2{x_{0_\epsilon}}'^2}{\beta'^3_\epsilon} \end{bmatrix}$ and obtain $\lambda_1 = \frac{2}{\beta'_\epsilon} + \frac{2x'^2_{0_\epsilon}}{\beta'^3_\epsilon}$. The corresponding eigenvector, $\mathbf{e}_1$, is $(\beta'_\epsilon, -x'_{0_\epsilon})$. Solving for

$$\mathbf{e}_1 \cdot \nabla A_\psi = 0$$

we obtain

$$\begin{aligned} x'_{0_\epsilon} &= \beta'_\epsilon 6^{-2/3}\left(\left(-18\left(Z_{\max}+Z_{\min}\right)+6^{1/2}\left(54\left(Z_{\max}+Z_{\min}\right)^2-\left(Z^2_{\max}+Z^2_{\min}-4\right)^3\right)^{1/2}\right)^{2/3}\right. \\ &\quad \left. -\,6^{1/3}\left(Z^2_{\max}+Z^2_{\min}-4\right)\right) / \\ &\quad \left(-18\left(Z_{\max}+Z_{\min}\right)+6^{1/2}\left(54\left(Z_{\max}+Z_{\min}\right)^2-\left(Z^2_{\max}+Z^2_{\min}-4\right)\right)^{1/2}\right)^{1/3} \end{aligned} \tag{23}$$

Last in an analysis of the noiseless case, we consider the effects due to the finiteness of the aperture. As before, we consider a circular aperture. We assume a certain amount of translational error $\sqrt{x^2_{0_\epsilon}+y^2_{0_\epsilon}}$, and we seek the direction of translational error that results in the smallest negative depth volume.

Independent of the direction of translation, (23) describes the relationship of x_{0_ϵ} and β_ϵ for the smallest negative depth volume. Substituting (23) into (19), we obtain the cross-sections through the negative depth volume 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 line constraint holds, that is, $\frac{x_0}{y_0} = \frac{\hat{x_0}}{\hat{y_0}}$ (see Figure 6).

It remains to be shown that noise in the flow measurements does not alter the qualitative characteristics of the negative depth volume and thus the results obtained.

First we analyze the orthogonality constraint. The analysis is carried out without considering the size of the aperture; as will be shown, this analysis leads to the orthogonality constraint.

First, let $\gamma_\epsilon = 0$. Ignoring the image size, we are interested in $E(V)$, the expected value of the integral of the cross sections A_ψ, which amounts to

$$E(V) = E\left(\int_{Z=Z_{\min}}^{Z_{\max}}\int_{\psi=-\pi/2}^{\pi/2}\left|(x_{0_\epsilon}+\beta_\epsilon Z-\delta_x Z-\epsilon_x)\cos\psi+(y_{0_\epsilon}-\alpha_\epsilon Z-\delta_y Z-\epsilon_y)\sin\psi\right| d\psi\, dZ\right)$$

or

$$E(V) = E\left(2\int_{Z=Z_{\min}}^{Z_{\max}}\left((x_{0_\epsilon}+Z\beta_\epsilon-Z\delta_x-\epsilon_x)^2+(y_{0_\epsilon}-Z\alpha_\epsilon-Z\delta_y-\epsilon_y)^2\right)^{1/2} dZ\right)$$

We approximate the expectation of the integrand by performing a Taylor expansion at 0 up to second order, which gives

$$E\left((x_{0_\epsilon}+Z\beta_\epsilon-Z\delta_x-\epsilon_x)^2+(y_{0_\epsilon}-Z\alpha_\epsilon-Z\delta_y-\epsilon_y)^2\right)^{1/2} \sim$$

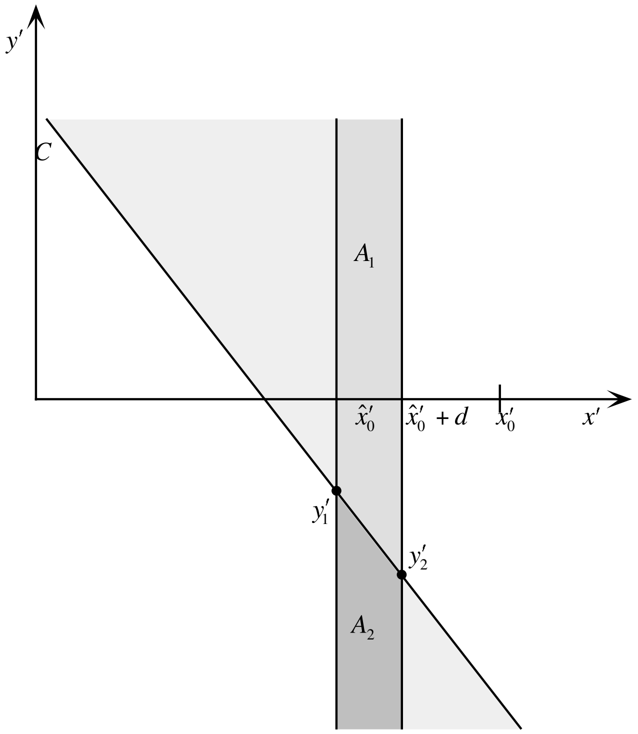


Figure 5: A change of $\hat{x}'_0$ to $\hat{x}'_0 + d$ causes the area of negative depth values A_c to increase by A_1 and to decrease by A_2. This change amounts to $A_c = -(y'_1 + y'_2) d \,\mathrm{sgn}(\gamma_\epsilon)$.

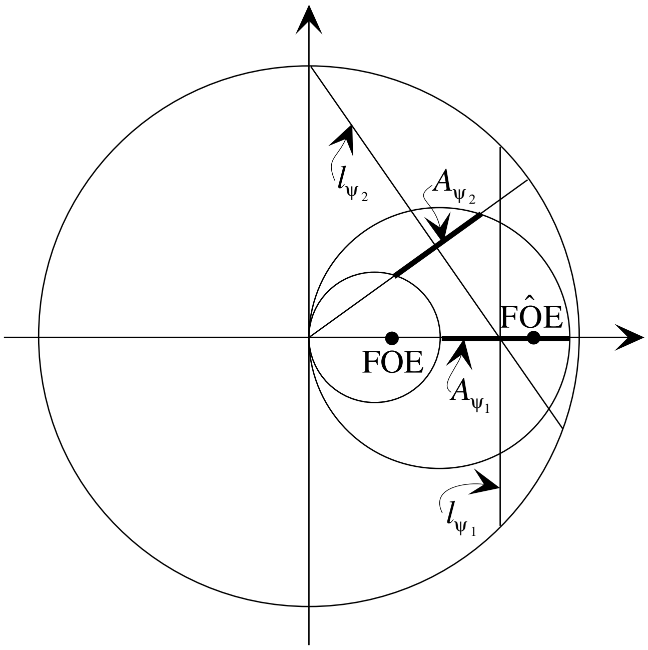


Figure 6: Cross-sectional view of the wedge-shaped negative depth volumes in a circular aperture. The minimization of the negative depth volume for a given amount of translational error occurs when $\frac{x_0}{y_0} = \frac{\hat{x_0}}{\hat{y_0}}$. A_{ψ_i} and l_{ψ_i} denote the areas of the cross sections and average extents respectively, for two angles ψ_1 and ψ_2. The two circles bounding A_{ψ_i} are given by the equations $x' = x'_{0_\epsilon} + \beta'_\epsilon Z_{\min}$ and $x' = x'_{0_\epsilon} + \beta'_\epsilon Z_{\max}$.

$$\sqrt{(x_{0_\epsilon} + Z\beta_\epsilon)^2 + (y_{0_\epsilon} - Z\alpha_\epsilon)^2} + \left(\frac{Z^2\sigma_\delta^2 + \sigma_\epsilon^2}{2}\right)$$
$$\left(\frac{2}{\sqrt{(x_{0_\epsilon} + Z\beta_\epsilon)^2 + (y_{0_\epsilon} - Z\alpha_\epsilon)^2}} - \sqrt{(x_{0_\epsilon} + Z\beta_\epsilon)^2 + (y_{0_\epsilon} - Z\alpha_\epsilon)^2}\right)$$

We are interested in the angle μ between the translational and rotational error which minimizes the negative depth volume. If we align the translational error with the x axis, that is, $y_{0_\epsilon} = 0$, and if we express the rotational error as $(-\beta_\epsilon, \alpha_\epsilon) = (K\cos\mu, K\sin\mu)$, we obtain

$$E(V) = 2\int_{Z=Z_{\min}}^{Z_{\max}} \sqrt{x_{0_\epsilon}^2 + Z^2K^2 - 2Zx_{0_\epsilon}K\cos\mu} + \frac{(Z^2\sigma_\delta^2 + \sigma_\epsilon^2)}{2}$$
$$\left(\frac{2}{\sqrt{x_{0_\epsilon}^2 + Z^2K^2 - 2Zx_{0_\epsilon}K\cos\mu}} - \sqrt{x_{0_\epsilon}^2 + Z^2K^2 - 2Zx_{0_\epsilon}K\cos\mu}\right)$$

Solving for $\frac{\partial E(V)}{\partial \mu} = 0$ we obtain $\mu = 0$, which shows$(x_{0_\epsilon}, y_{0_\epsilon})$ to be perpendicular to $(\alpha_\epsilon, \beta_\epsilon)$, if $\gamma_\epsilon = 0$, for all minimizations considered.

Next, we allow γ_ϵ to be different from zero. For the case of a fixed translational error, again, the volume increases as γ_ϵ increases, and thus the smallest negative depth volume occurs for $\gamma_\epsilon = 0$. For the case of a fixed rotational error we have to extend the previous analysis (studying the change of volume when changing the estimated translation) to noisy motion fields: The $-\infty$ distortion surface becomes $x' - x'_0 - Z\beta'_\epsilon + Z\gamma_\epsilon y' + Z\delta' + \epsilon'$ and the 0 distortion surfaces remain the same. Therefore, V_C becomes

$$-\operatorname{sgn}(\gamma_\epsilon)\frac{d}{\gamma_\epsilon}\left(2\left(\beta'_\epsilon - \delta'\right)\left(Z_{\max} - Z_{\min}\right) - \left(2\beta'_\epsilon Z_I + 2\epsilon' + d\right)\left(\ln\left(Z_{\max}\right) - \ln\left(Z_{\min}\right)\right)\right)$$

and $E(V_C)$ takes the same form as V_C in (21). We thus obtain the orthogonality constraint for $\gamma_\epsilon \neq 0$.

Finally, we take into account the limited extent of a circular aperture for the case of global minimization. As noise does not change the structure of the iso-distortion surfaces, as shown in Figure 6, in the presence of noise, too, the smallest negative depth volume is obtained if the FOE and the estimated FOE lie on a line passing through the image center. This proves the orthogonality constraint for the full model as well as the line constraint. The global minimum of the negative depth volume is thus described by the constraint $\gamma_\epsilon = 0$, the orthogonality constraint and the line constraint.

6 Epipolar Minimization: Spherical Eye

The function representing deviation from the epipolar constraint on the sphere takes the simple form

$$E_{ep} = \int\limits_{R_{\min}}^{R_{\max}} \iint\limits_{\text{sphere}} \left\{\left(\frac{\mathbf{r}\times(\mathbf{r}\times\mathbf{t})}{R} - (\boldsymbol{\omega}_\epsilon \times \mathbf{r})\right)\cdot(\hat{\mathbf{t}}\times\mathbf{r})\right\}^2 dA\, dR$$

where A refers to a surface element. Due to the sphere's symmetry, for each point $\mathbf{r}$ on the sphere, there exists a point with coordinates $-\mathbf{r}$. Since $\mathbf{u}_{\rm tr}(\mathbf{r}) = \mathbf{u}_{\rm tr}(-\mathbf{r})$ and $\mathbf{u}_{\rm rot}(\mathbf{r}) = -\mathbf{u}_{\rm rot}(-\mathbf{r})$, when the integrand is expanded the product terms integrated over the sphere vanish. Thus

$$E_{ep} = \int\limits_{R_{\min}}^{R_{\max}} \iint\limits_{\text{sphere}} \left\{\frac{\left((\mathbf{t}\times\hat{\mathbf{t}})\cdot\mathbf{r}\right)^2}{R^2} + \left((\boldsymbol{\omega}_\epsilon\times\mathbf{r})\cdot(\hat{\mathbf{t}}\times\mathbf{r})\right)^2\right\} dA\, dR$$

(a) Assuming that translation $\hat{\mathbf{t}}$ has been estimated, the $\boldsymbol{\omega}_\epsilon$ that minimizes E_{ep} is $\boldsymbol{\omega}_\epsilon = 0$, since the resulting function is non-negative quadratic in $\boldsymbol{\omega}_\epsilon$ (minimum at zero). The difference between sphere and plane is already clear. In the spherical case, as shown here, if an error in the translation is made we do not need to compensate for it by making an error in the rotation ($\boldsymbol{\omega}_\epsilon = 0$), while in the planar case we need to compensate to ensure that the orthogonality constraint is satisfied!

(b) Assuming that rotation has been estimated with an error $\boldsymbol{\omega}_\epsilon$, what is the translation $\hat{\mathbf{t}}$ that minimizes E_{ep}? Since R is uniformly distributed, integrating over R does not alter the form of the error in the optimization. Thus, E_{ep} consists of the sum of two terms:

$$K = K_1 \iint\limits_{\text{sphere}} \left((\mathbf{t}\times\hat{\mathbf{t}})\cdot\mathbf{r}\right)^2 dA \quad \text{and} \quad L = L_1 \iint\limits_{\text{sphere}} \left((\boldsymbol{\omega}_\epsilon\times\mathbf{r})\cdot(\hat{\mathbf{t}}\times\mathbf{r})\right)^2 dA,$$

where K_1, L_1 are multiplicative factors depending only on $R_{\min}$ and $R_{\max}$. For angles between $\mathbf{t}, \hat{\mathbf{t}}$ and $\hat{\mathbf{t}}, \boldsymbol{\omega}_\epsilon$ in the range of 0 to $\pi/2$, K and L are monotonic functions. K attains its minimum when $\mathbf{t} = \hat{\mathbf{t}}$ and L when $\hat{\mathbf{t}} \perp \boldsymbol{\omega}_\epsilon$. Fix the distance between $\mathbf{t}$ and $\hat{\mathbf{t}}$ leading to a certain value K, and change the position of $\hat{\mathbf{t}}$. L takes its minimum when $(\mathbf{t}\times\hat{\mathbf{t}})\cdot\boldsymbol{\omega}_\epsilon = 0$, as follows from the cosine theorem. Thus E_{ep} achieves its minimum when $\hat{\mathbf{t}}$ lies on the great circle passing through $\mathbf{t}$ and $\boldsymbol{\omega}_\epsilon$, with the exact position depending on $|\boldsymbol{\omega}_\epsilon|$ and the scene in view.

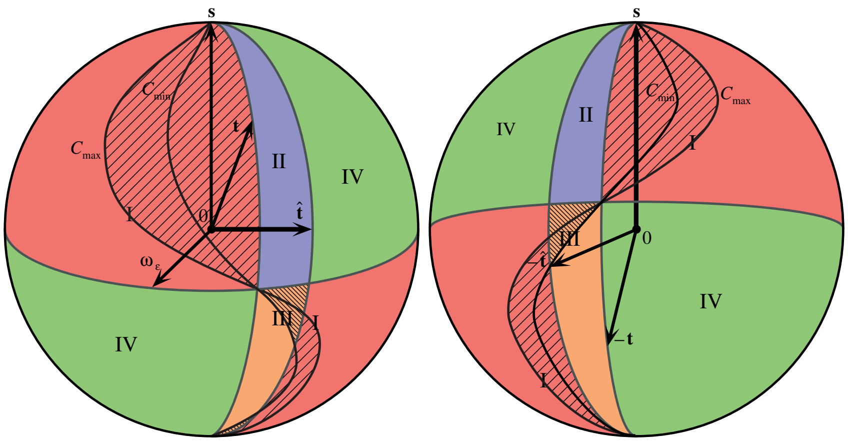
(c) For the general case where no information about rotation or translation is available, we study the subspaces where E_{ep} changes the least at its absolute minimum, i.e., we are again interested in the direction of the smallest second derivative at 0. For points defined by this direction we calculate, using Maple, $\mathbf{t} = \hat{\mathbf{t}}$ and $\boldsymbol{\omega}_\epsilon \perp \mathbf{t}$.

7 Minimizing Negative Depth Volume on the Sphere

(a) Assuming that the rotation has been estimated with an error $\boldsymbol{\omega}_\epsilon$, what is the optimal translation $\hat{\mathbf{t}}$ that minimizes the negative depth volume?

Since the motion field along different orientations $\mathbf{n}$ is considered, a parameterization is needed to express all possible orientations on the sphere. This is achieved by selecting an arbitrary vector $\mathbf{s}$; then, at each point $\mathbf{r}$ of the sphere, $\frac{\mathbf{s}\times\mathbf{r}}{\|\mathbf{s}\times\mathbf{r}\|}$ defines a direction in the tangent plane. As $\mathbf{s}$ moves along half a circle, $\frac{\mathbf{s}\times\mathbf{r}}{\|\mathbf{s}\times\mathbf{r}\|}$ takes on every possible orientation (with the exception of the points $\mathbf{r}$ lying on the great circle of $\mathbf{s}$). Let us pick $\boldsymbol{\omega}_\epsilon$ perpendicular to $\mathbf{s}$ ($\mathbf{s}\cdot\boldsymbol{\omega}_\epsilon = 0$).

We are interested in the points in space with estimated negative range values $\hat{R}$. Since $\mathbf{n} = \frac{\mathbf{s}\times\mathbf{r}}{\|\mathbf{s}\times\mathbf{r}\|}$ and $\mathbf{s}\cdot\boldsymbol{\omega}_\epsilon = 0$, the estimated range $\hat{R}$ amounts to $\hat{R} = R\frac{(\hat{\mathbf{t}}\times\mathbf{s})\cdot\mathbf{r}}{(\mathbf{t}\times\mathbf{s})\cdot\mathbf{r}-R(\boldsymbol{\omega}_\epsilon\cdot\mathbf{r})(\mathbf{s}\cdot\mathbf{r})}$. $\hat{R} < 0$ if $\mathrm{sgn}[(\hat{\mathbf{t}}\times\mathbf{s})\cdot\mathbf{r}] = -\mathrm{sgn}[(\mathbf{t}\times\mathbf{s})\cdot\mathbf{r}-R(\boldsymbol{\omega}_\epsilon\cdot\mathbf{r})(\mathbf{s}\cdot\mathbf{r})]$, where $\mathrm{sgn}(x)$ provides the sign of x. This constraint divides the surface of the sphere into four areas, I to IV, whose locations are defined by the signs of 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 shown in Figure 7.



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

Figure 7: Classification of image points according to constraints on R. The four areas are marked by different colors. The textured parts (parallel lines) in areas I and III denote the image points for which negative depth values exist if the scene is bounded. 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.

For any direction $\mathbf{n}$ a volume of negative range values is obtained consisting of the volumes above areas I, II and III. Areas II and III cover the same amount of area between the 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, $R \in [0, +\infty]$, there is for every $\mathbf{r}$ a range of values above areas I and III which result in negative depth estimates; in area I the volume at each point $\mathbf{r}$ is bounded from below by $R = \frac{(\mathbf{t}\times\mathbf{s})\cdot\mathbf{r}}{(\boldsymbol{\omega}_\epsilon\cdot\mathbf{r})(\mathbf{s}\cdot\mathbf{r})}$, and in area III it is bounded from above by $R = \frac{(\mathbf{t}\times\mathbf{s})\cdot\mathbf{r}}{(\boldsymbol{\omega}_\epsilon\cdot\mathbf{r})(\mathbf{s}\cdot\mathbf{r})}$. If there exist lower and upper bounds $R_{\min}$ and $R_{\max}$ in the scene, we obtain two additional curves $C_{\min}$ and $C_{\max}$ with $C_{\min} = (\mathbf{t}\times\mathbf{s})\cdot\mathbf{r} - R_{\min}(\boldsymbol{\omega}_\epsilon\cdot\mathbf{r})(\mathbf{s}\cdot\mathbf{r}) = 0$ and $C_{\max} = (\mathbf{t}\times\mathbf{s})\cdot\mathbf{r} - R_{\max}(\boldsymbol{\omega}_\epsilon\cdot\mathbf{r})(\mathbf{s}\cdot\mathbf{r}) = 0$, and we obtain negative depth values in area I only between $C_{\max}$ and $(\mathbf{t}\times\mathbf{s})\cdot\mathbf{r} = 0$ and in area III only between $C_{\min}$ and $(\boldsymbol{\omega}_\epsilon\times\mathbf{r})(\mathbf{s}\times\mathbf{r}) = 0$. We are given $\boldsymbol{\omega}_\epsilon$ and $\mathbf{t}$, and we are interested in the $\hat{\mathbf{t}}$ 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 through $\mathbf{t}$ and $\mathbf{s}$, that is, $(\mathbf{t}\times\mathbf{s})\cdot\hat{\mathbf{t}} = 0$, as will be shown next.

Let us consider a $\hat{\mathbf{t}}$ such that $(\mathbf{t}\times\mathbf{s})\cdot\hat{\mathbf{t}} \neq 0$ and let us change $\hat{\mathbf{t}}$ so that $(\mathbf{t}\times\mathbf{s})\cdot\hat{\mathbf{t}} = 0$. As $\hat{\mathbf{t}}$ changes, the area of type II becomes an area of type IV and the area of type III becomes an area of type I. The negative depth volume is changed as follows: It is decreased by the spaces above area II and area III, and it is increased by the space above area I (which changed from type III to type I). Clearly, the decrease is larger than the increase, which implies that the smallest volume is obtained for $\mathbf{s}, \mathbf{t}, \hat{\mathbf{t}}$ lying on a great circle. Since this is true for any $\mathbf{s}$, the minimum negative depth volume is attained for $\mathbf{t} = \hat{\mathbf{t}}$.[2]

(b) Next, assume that no prior knowledge about the 3D motion is available. We want to know for which configurations of $\hat{\mathbf{t}}$ and $\boldsymbol{\omega}_\epsilon$ the negative depth values change the least in the neighborhood of the absolute minimum, that is, at $\mathbf{t}_\epsilon = \boldsymbol{\omega}_\epsilon = 0$. From the analysis above, it is known that for any $\boldsymbol{\omega}_\epsilon \neq 0$, $\mathbf{t} = \hat{\mathbf{t}}$. Next, we show that $\boldsymbol{\omega}_\epsilon$ is indeed different from zero: Take $\mathbf{t} \neq \hat{\mathbf{t}}$ on the great circle of $\mathbf{s}$ and let $\boldsymbol{\omega}_\epsilon$, as before, be perpendicular to $\mathbf{s}$.

Since $(\mathbf{t}\times\mathbf{s})\times\boldsymbol{\omega}_\epsilon = 0$, the curves $C_{\max}$ and $C_{\min}$ can be expressed as $C_{\max(\min)} = (\boldsymbol{\omega}_\epsilon\cdot\mathbf{r})(\frac{\sin\angle(\mathbf{t},\mathbf{s})}{|\boldsymbol{\omega}_\epsilon| R_{\max(\min)}} - (\mathbf{s}\cdot\mathbf{r})) = 0$, where $\sin\angle(\mathbf{t},\mathbf{s})$ denotes the angle between vectors $\mathbf{t}$ and $\mathbf{s}$. These curves consist of the great circle $\boldsymbol{\omega}_\epsilon\cdot\mathbf{r} = 0$ and the circle $\frac{\sin\angle(\mathbf{t},\mathbf{s})}{|\boldsymbol{\omega}_\epsilon| R_{\max(\min)}} - (\mathbf{s}\cdot\mathbf{r}) = 0$ parallel to the great circle $(\mathbf{s}\cdot\mathbf{r}) = 0$ (see Figure 8). If $\frac{\sin\angle(\mathbf{t},\mathbf{s})}{|\boldsymbol{\omega}_\epsilon| R_{\max(\min)}} > 1$, this circle disappears.

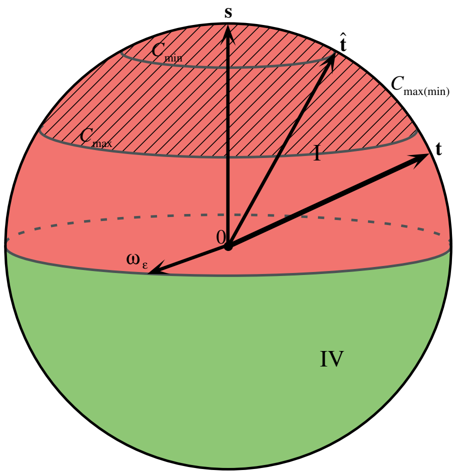


Figure 8: Configuration for $\mathbf{t}$ and $\hat{\mathbf{t}}$ on the great circle of $\mathbf{s}$ and $\boldsymbol{\omega}_\epsilon$ perpendicular to $\mathbf{s}$. The textured part of area I denotes image points for which negative depth values exist if the scene is bounded.

Consider next two flow directions defined by vectors $\mathbf{s}_1$ and $\mathbf{s}_2$ with $(\mathbf{s}_1\times\mathbf{t}) = -(\mathbf{s}_2\times\mathbf{t})$ and $\mathbf{s}_1$ between $\mathbf{t}$ and $\hat{\mathbf{t}}$.

For every point $\mathbf{r}_1$ in area III defined by $\mathbf{s}_1$ there exists a point $\mathbf{r}_2$ in area I defined by $\mathbf{s}_2$ such that the negative estimated ranges above $\mathbf{r}_1$ and $\mathbf{r}_2$ add up to $R_{\max} - R_{\min}$. Thus the volume of negative range obtained from $\mathbf{s}_1$ and $\mathbf{s}_2$ amounts to the area of the sphere times $(R_{\max} - R_{\min})$ (area II of $\mathbf{s}_1$ contributes a hemisphere; area III of $\mathbf{s}_1$ and area I of $\mathbf{s}_2$ together contribute a hemisphere). The total negative range volume can be decomposed into

[2] A word of caution about the parameterization used for directions $\mathbf{n} = \frac{\mathbf{s}\times\mathbf{r}}{\|\mathbf{s}\times\mathbf{r}\|}$ is needed. It 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). Thus to obtain a uniform distribution, normalization is necessary. The normalization factors, however, do not affect the previous proof, due to symmetry.

three components: a component V_1 originating from the set of $\mathbf{s}$ between $\mathbf{t}$ and $\hat{\mathbf{t}}$, a component V_2 originating from the set of $\mathbf{s}$ symmetric in $\mathbf{t}$ to the set in V_1, and a component V_3 corresponding to the remaining $\mathbf{s}$, which consists of range values above areas of type I only. If for all $\mathbf{s}$ in V_3, $\frac{\sin \angle(\mathbf{t},\mathbf{s})}{R_{\max}|\boldsymbol{\omega}_\epsilon|} \geq 1$, V_3 becomes zero. Thus for all $|\boldsymbol{\omega}_\epsilon|$ with $|\boldsymbol{\omega}_\epsilon| \leq \frac{\sin \angle(\mathbf{t},\hat{\mathbf{t}})}{R_{\max}}$, the negative range volume is equally large and amounts to the area on the sphere times $(R_{\max} - R_{\min})$ times $\angle(\mathbf{t}, \hat{\mathbf{t}})$. Unless $R_{\max} = \infty$, $|\boldsymbol{\omega}_\epsilon|$ takes on values different from zero.

This shows that for any $\mathbf{t}_\epsilon \neq 0$, there exist vectors $\boldsymbol{\omega}_\epsilon \neq 0$ which give rise to the same negative depth volume as $\boldsymbol{\omega}_\epsilon = 0$. However, for any such $\boldsymbol{\omega}_\epsilon \neq 0$ this volume is larger than the volume obtained by setting $\mathbf{t}_\epsilon = 0$. It follows that $\mathbf{t} = \hat{\mathbf{t}}$. From Figure 7, it can furthermore be deduced that for a given $\boldsymbol{\omega}_\epsilon$ the negative depth volume, which for $\mathbf{t} = \hat{\mathbf{t}}$ only lies above areas of type I, decreases as $\mathbf{t}$ moves along a great circle away from $\boldsymbol{\omega}_\epsilon$, as the areas between $C_{\min}$ and $C_{\max}$ and between $C_{\min}$ and $(\mathbf{t} \times \mathbf{s}) \cdot \mathbf{r} = 0$ decrease. This proves that in addition to $\mathbf{t} = \hat{\mathbf{t}}$, $\mathbf{t} \perp \boldsymbol{\omega}_\epsilon$.

8 Conclusions

The preceding results constitute a geometric statistical investigation of the observability of 3D motion from images. On their basis, a number of striking conclusions can be achieved. First, they clearly demonstrate the advantages of panoramic vision in the process of 3D motion estimation. Table 1 lists the eight out of ten cases which lead to clearly defined error configurations. It shows that 3D motion can be estimated more accurately with spherical eyes. Depending on the estimation procedure used—and systems might use different procedures for different tasks—either the translation or the rotation can be estimated very accurately. For planar eyes, this is not the case, as for all possible procedures there exists confusion between the translation and rotation. The error configurations also allow systems with inertial sensors to use more efficient estimation procedures. If a system utilizes a gyrosensor which provides an approximate estimate of its rotation, it can employ a simple algorithm based on the positive depth constraint for only translational motion fields to derive its translation and obtain a very accurate estimate. Such algorithms are much easier to implement than algorithms designed for completely unknown rigid motions, as they amount to searches in 2D as opposed to 5D spaces [19]. Similarly, there exist computational advantages for systems with translational inertial sensors in estimating the remaining unknown rotation.

Since the positive depth constraint turns out to be very powerful and since epipolar minimization does not consider depth positivity, an interesting research question that arises for the future is how to couple the epipolar constraint with the positive depth constraint. We attempted a first investigation into this problem through a study of negative depth on the basis of optic flow for the plane in Appendix B, which gave very interesting results. Specifically, we found that estimating all motion parameters simultaneously by minimizing negative depth from optic flow provides a solution with no error in the translation. However, the rotation cannot be decoupled from the translation, which makes it clear that for cameras with restricted fields of view the problem of rotation/translation confusion cannot be escaped.

Camera-type eyes are found in nature in systems that walk and perform sophisticated manipulation because such systems have a need for very accurate segmentation and shape estimation and thus high resolution in a limited field of view. Panoramic vision, either through compound eyes or a pair of camera-type eyes positioned on opposite sides of the head is usually found in flying systems which have the obvious need for a larger field of view but also rely on accurate 3D motion estimation as they always move in an unconstrained way. When we face the task of equipping robots with visual sensors, we do not have to necessarily copy nature, and we also do not have to necessarily use what is commercially available. Instead, we could construct new, powerful eyes by taking advantage of both the panoramic vision of flying systems and the high-resolution vision of primates. An eye like the one in Figure 9, assembled from a few video cameras arranged on the surface of a sphere, can easily estimate 3D motion since, while it is moving, it is sampling a spherical motion field!

Such an eye not only has panoramic properties, allowing very accurate determination of the transformations relating multiple views, but it has the unexpected benefit of making it easy to estimate image motion with high accuracy. Any two cameras with overlapping fields of view also provide high-resolution stereo vision, and this collection of stereo systems makes it possible to locate a large number of depth discontinuities. Given scene discontinuities, image motion can be estimated very accurately. As a consequence, having accurate 3D motion

Table 1: Summary of results

	I Spherical Eye	II Camera-type Eye
Epipolar minimization, given optic flow	(a) Given a translational error $\mathbf{t}_\epsilon$, the rotational error $\boldsymbol{\omega}_\epsilon = 0$ (b) Without any prior information, $\mathbf{t}_\epsilon = 0$ and $\boldsymbol{\omega}_\epsilon \perp \mathbf{t}$	(a) For a fixed translational error $(x_{0_\epsilon}, y_{0_\epsilon})$, the rotational error $(\alpha_\epsilon, \beta_\epsilon, \gamma_\epsilon)$ is of the form $\gamma_\epsilon = 0$, $\alpha_\epsilon/\beta_\epsilon = -x_{0_\epsilon}/y_{0_\epsilon}$ (b) Without any a priori information about the motion, the errors satisfy $\gamma_\epsilon = 0$, $\alpha_\epsilon/\beta_\epsilon = -x_{0_\epsilon}/y_{0_\epsilon}$, $x_0/y_0 = x_{0_\epsilon}/y_{0_\epsilon}$
Minimization of negative depth volume, given normal flow	(a) Given a rotational error $\boldsymbol{\omega}_\epsilon$, the translational error $\mathbf{t}_\epsilon = 0$ (b) Without any prior information, $\mathbf{t}_\epsilon = 0$ and $\boldsymbol{\omega}_\epsilon \perp \mathbf{t}$	(a) Given a rotational error, the translational error is of the form $-x_{0_\epsilon}/y_{0_\epsilon} = \alpha_\epsilon/\beta_\epsilon$ (b) Without any error information, the errors satisfy $\gamma_\epsilon = 0$, $\alpha_\epsilon/\beta_\epsilon = -x_{0_\epsilon}/y_{0_\epsilon}$, $x_0/y_0 = x_{0_\epsilon}/y_{0_\epsilon}$

and image motion, the eye in Figure 9 is very well suited to developing accurate models of the world necessary for many robotic/servoing applications.

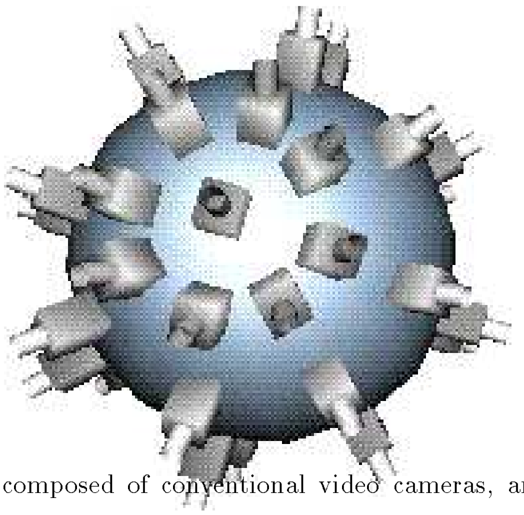


Figure 9: A compound-like eye composed of conventional video cameras, arranged on a sphere and looking outward.

References

[1] G. Adiv. Determining 3D motion and structure from optical flow generated by several moving objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 7:384–401, 1985.

[2] G. Adiv. 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, 1989.

[3] J. Aloimonos, I. Weiss, and A. Bandopadhay. Active vision. *International Journal of Computer Vision*, 2:333–356, 1988.

[4] J. Y. Aloimonos. Purposive and qualitative active vision. In *Proc. DARPA Image Understanding Workshop*, pages 816–828, 1990.

[5] R. Bajcsy. Active perception. *Proceedings of the IEEE*, 76:996–1005, 1988.

[6] D. H. Ballard and C. M. Brown. Principles of animate vision. *CVGIP: Image Understanding: Special Issue on Purposive, Qualitative, Active Vision, Y. Aloimonos (Ed.)*, 56:3–21, 1992.

[7] D. H. Ballard and O. A. Kimball. Rigid body motion from depth and optical flow. *Computer Vision, Graphics, and Image Processing*, 22:95–115, 1983.

[8] A. Bandopadhay. *A Computational Approach to Visual Motion Perception.* PhD thesis, Department of Computer Science, University of Rochester, 1986.

[9] W. Burger and B. Bhanu. Estimating 3-D egomotion from perspective image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:1040–1058, 1990.

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

[11] K. Daniilidis and H.-H. Nagel. Analytical results on error sensitivity of motion estimation from two views. *Image and Vision Computing*, 8:297–303, 1990.

[12] K. Daniilidis and M. E. Spetsakis. Understanding noise sensitivity in structure from motion. In *Visual Navigation: From Biological Systems to Unmanned Ground Vehicles*, chapter 4. Lawrence Erlbaum Associates, Mahwah, NJ, 1997.

[13] T. Dean and M. Wellman. *Planning and Control.* Morgan Kaufmann, 1991.

[14] F. Dornaika and R. Horaud. Simultaneous robot-world and hand-eye calibration. *IEEE Transactions on Robotics and Automation*, 14:617–622, 1998.

[15] R. Dutta and M. Snyder. Robustness of correspondence-based structure from motion. In *Proc. International Conference on Computer Vision*, pages 106–110, 1990.

[16] B. Espiau, F. Chaumette, and P. Rives. A new approach to visual servoing in robotics. *IEEE Transactions on Robotics and Automation*, 8:313–326, 1992.

[17] O. D. Faugeras, F. Lustman, and G. Toscani. Motion and structure from motion from point and line matches. In *Proc. International Conference on Computer Vision*, pages 25–34, 1987.

[18] J. T. Feddema, C. S. G. Lee, and O. R. Mitchell. Feature-based visual servoing of robotic systems. In K. Hashimoto, editor, *Visual Servoing*, pages 105–138. World Scientific, New York, 1993.

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

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

[21] G. D. Hager, G. Grunwald, and G. Hirzinger. Feature-based visual servoing and its application to telerobotics. In *Proc. IEEE/RSJ/GI International Conference on Intelligent Robots Systems*, volume 1, pages 164–171, 1994.

[22] D. J. Heeger and A. D. Jepson. Subspace methods for recovering rigid motion I: Algorithm and implementation. *International Journal of Computer Vision*, 7:95–117, 1992.

[23] R. Horaud and F. Dornaika. Hand-eye calibration. *International Journal of Robotics Research*, 14:195–210, 1995.

[24] R. Horaud, F. Dornaika, and B. Espiau. Visual guided object grasping. *IEEE Transactions on Robotics and Automation*, 14:525–532, 1998.

[25] B. K. P. Horn. *Robot Vision*. McGraw Hill, New York, 1986.

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

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

[28] S. Hutchinson, G. D. Hager, and P. I. Corke. A tutorial on visual servo control. *IEEE Transactions on Robotics and Automation*, 12:651–670, 1996.

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

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

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

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

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

[34] S. J. Maybank. *Theory of Reconstruction from Image Motion*. Springer, Berlin, 1993.

[35] J. Philip. Estimation of three-dimensional motion of rigid objects from noisy observations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13:61–66, 1991.

[36] K. Prazdny. Egomotion and relative depth map from optical flow. *Biological Cybernetics*, 36:87–102, 1980.

[37] K. Prazdny. Determining instantaneous direction of motion from optical flow generated by a curvilinear moving observer. *Computer Vision, Graphics, and Image Processing*, 17:238–248, 1981.

[38] J. H. Rieger and D. T. Lawton. Processing differential image motion. *Journal of the Optical Society of America A*, 2:354–359, 1985.

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

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

[41] 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*, pages 691–695, 1993.

[42] G. S. Young and R. Chellappa. Statistical analysis of inherent ambiguities in recovering 3-D motion from a noisy flow field. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14:995–1013, 1992.

[43] X. Zhuang, T. S. Huang, and R. M. Haralick. Two-view motion analysis: A unified algorithm. *Journal of the Optical Society of America A*, 3:1492–1500, 1986.

A Translational Error Given Estimate of Rotation: Epipolar Minimization on the Plane

Minimization of E_{NF}, as given in (14), with respect to the translational error parameters is obtained by solving

$$\frac{\partial E_{\mathrm{NF}}}{\partial x_{0_\epsilon}} = 0 \qquad \frac{\partial E_{\mathrm{NF}}}{\partial y_{0_\epsilon}} = 0$$

which yields

$$x_{0_\epsilon} = \frac{K_3}{K_4}\left(\begin{array}{ll} & e^4\left((3\alpha+\gamma x_0)\gamma\left(\gamma^2K_3+1\right)K_1^2-(3\gamma y_0+\beta)\left(\gamma^2K_3+1\right)K_1K_2\right) \\ + & 4e^2\left(\begin{array}{ll} & 3x_0\gamma\left(x_0\alpha+y_0\beta\right)\left(K_1^2-K_2^2K_3\right) \\ - & \beta\left(\alpha^2+\beta^2\right)K_1K_2K_3+2x_0\gamma^2\left(x_0\beta-y_0\alpha\right)K_1K_2K_3 \\ + & \gamma^2x_0\left(\alpha^2+\beta^2\right)K_1^2K_3+\left(y_0\alpha\beta-x_0\alpha^2-2x_0\beta^2\right)K_2^2K_3 \\ + & \alpha\left(y_0\beta+x_0\alpha\right)K_1^2-\beta\left(x_0^2+y_0^2\right)K_1K_2+x_0\gamma^2\left(x_0^2+y_0^2\right)K_1^2 \end{array}\right) \\ + & 16x_0\left(x_0\alpha+y_0\beta\right)^2\left(K_1^2-K_2^2K_3\right) \end{array}\right)$$

and

$$y_{0_\epsilon} = \frac{K_3}{K_4}\left(\begin{array}{ll} & e^4\left((3\beta+\gamma y_0)\gamma\left(\gamma^2K_3+1\right)K_1^2+(3\gamma x_0+\alpha)\left(\gamma^2K_3+1\right)K_1K_2\right) \\ + & 4e^2\left(\begin{array}{ll} & 3y_0\gamma\left(x_0\alpha+y_0\beta\right)\left(K_1^2-K_2^2K_3\right) \\ + & \alpha\left(\alpha^2+\beta^2\right)K_1K_2K_3+2y_0\gamma^2\left(x_0\beta-y_0\alpha\right)K_1K_2K_3 \\ + & \gamma^2y_0\left(\alpha^2+\beta^2\right)K_1^2K_3+\left(x_0\alpha\beta-y_0\beta^2-2y_0\alpha^2\right)K_2^2K_3 \\ + & \beta\left(y_0\beta+x_0\alpha\right)K_1^2+\alpha\left(x_0^2+y_0^2\right)K_1K_2+y_0\gamma^2\left(x_0^2+y_0^2\right)K_1^2 \end{array}\right) \\ + & 16y_0\left(x_0\alpha+y_0\beta\right)^2\left(K_1^2-K_2^2K_3\right) \end{array}\right)$$

with

$$\begin{aligned} K_1 &= Z_{\max}-Z_{\min} \\ K_2 &= \ln(Z_{\max})-\ln(Z_{\min}) \\ K_3 &= Z_{\max}Z_{\min} \\ K_4 &= \begin{array}{ll} & e^4\left(K_3\gamma^2+1\right)K_1^2 \\ + & 4e^2\left(\begin{array}{ll} & \left(\alpha^2+\beta^2\right)\left(\gamma^2K_3+1\right)K_1^2K_3+2\left(x_0\beta-y_0\alpha\right)\left(\gamma^2K_3+1\right)K_1K_2K_3 \\ + & \left(x_0^2+y_0^2\right)\left(\gamma^2K_3+1\right) \end{array}\right) \\ + & 16\left(x_0\alpha+y_0\beta\right)\left(K_1^2-K_2^2K_3\right)K_3 \end{array} \end{aligned}$$

B Minimization of Negative Depth from Optical Flow

In the following analysis we study the function arising from all flow values on the image plane yielding negative depth, that is, we do not consider a finite aperture. As it turns out, the functions due to noiseless and noisy flow have the same local and global minima that interest us; we thus consider only flow with noise. Negative depth is derived if the component obtained by subtracting the estimated rotational flow from the flow and then projecting it onto the estimated translational component, is negative, that is, when (using equations (3) in the presence of noise):

$$\left(u_{\mathrm{tr}}+Zu_{\mathrm{r}}+ZN_x\right)\hat{u}_{\mathrm{tr}}+\left(v_{\mathrm{tr}}+Z\mathbf{v}_{\mathrm{rot}_\epsilon}+ZN_y\right)\hat{v}_{\mathrm{tr}}<0$$

or

$$\left(x-x_0+\epsilon_x+Z\left(-\beta_\epsilon+\gamma_\epsilon y+\delta_x\right)\right)\left(x-\hat{x}_0\right)+\left(y-y_0+\epsilon_y+Z\left(\alpha_\epsilon-\gamma_\epsilon x+\delta_y\right)\right)\left(y-\hat{y}_0\right)<0$$

which can be written alternately as

$$
\begin{aligned}
&(x-(x_0+\hat{x}_0-\epsilon_x+Z(\beta_\epsilon-\gamma_\epsilon\hat{y}_0-\delta_x))/2)^2\\
&+\quad(y-(y_0+\hat{y}_0-\epsilon_y+Z(-\alpha_\epsilon+\gamma_\epsilon\hat{x}_0-\delta_y))/2)^2\\
&-\quad(x_0-\hat{x}_0-\epsilon_x+Z(\beta_\epsilon-\gamma_\epsilon\hat{y}_0-\delta_x))^2/4\\
&-\quad(y_0-\hat{y}_0-\epsilon_y+Z(-\alpha_\epsilon+\gamma_\epsilon\hat{x}_0-\delta_y))^2/4<0
\end{aligned}
\tag{24}
$$

From (24) it can be seen that for any Z the points with coordinates (x,y) yielding negative depth lie within a circle, and the number of values yielding negative depth, which form a volume V in space, is thus

$$V=\frac{\pi}{4}\int_{Z=Z_{\min}}^{Z_{\max}}(x_{0_\epsilon}-\epsilon_x+Z(\beta_\epsilon-\gamma_\epsilon\hat{y}_0-\delta_x))^2+(y_{0_\epsilon}-\epsilon_y+Z(-\alpha_\epsilon+\gamma_\epsilon\hat{x}_0-\delta_y))^2$$

Its expected value, $E(V)$, amounts to

$$
\begin{aligned}
E(V) &= \frac{\pi}{12}\Big(\Big((\beta_\epsilon-\gamma_\epsilon\hat{y}_0)^2+(-\alpha_\epsilon+\gamma_\epsilon\hat{x}_0)^2+2\sigma_\delta^2\Big)\left(Z_{\max}^3-Z_{\min}^3\right)\\
&+\ 3\left(x_{0_\epsilon}(\beta_\epsilon-\gamma_\epsilon\hat{y}_0)+y_{0_\epsilon}(-\alpha_\epsilon+\gamma_\epsilon\hat{x}_0)\right)\left(Z_{\max}^2-Z_{\min}^2\right)\\
&+\ 3\left(x_{0_\epsilon}^2+y_{0_\epsilon}^2+2\sigma_\epsilon^2\right)(Z_{\max}-Z_{\min})\Big)
\end{aligned}
$$

(a) Given a fixed rotational error $(\alpha_\epsilon,\beta_\epsilon,\gamma_\epsilon)$, the local minima found by solving for

$$\frac{\partial V}{\partial x_{0_\epsilon}}=0\quad\text{and}\quad\frac{\partial V}{\partial y_{0_\epsilon}}=0$$

are described by

$$x_{0_\epsilon}=\frac{\gamma_\epsilon(\gamma_\epsilon x_0-\alpha_\epsilon)\left(Z_{\max}^2+Z_{\min}^2+Z_{\max}Z_{\min}\right)+\frac{3}{2}(\gamma_\epsilon y_0-\beta_\epsilon)(Z_{\max}+Z_{\min})}{\gamma_\epsilon^2\left(Z_{\max}^2+Z_{\min}^2+Z_{\max}Z_{\min}\right)+3}$$

and

$$y_{0_\epsilon}=\frac{\gamma_\epsilon(\gamma_\epsilon y_0-\beta_\epsilon)\left(Z_{\max}^2+Z_{\min}^2+Z_{\max}Z_{\min}\right)-\frac{3}{2}(\gamma_\epsilon x_0-\alpha_\epsilon)(Z_{\max}+Z_{\min})}{\gamma_\epsilon^2\left(Z_{\max}^2+Z_{\min}^2+Z_{\max}Z_{\min}\right)+3}$$

As can be seen, in general, the translational error depends on all the rotational error parameters and on the actual translation. For the special case $\gamma_\epsilon=0$, we obtain the orthogonality constraint.

(b) Given a fixed translational error $(x_{0_\epsilon},y_{0_\epsilon})$, the local minima found by solving for

$$\frac{\partial E(V)}{\partial\alpha_\epsilon}=0\quad\frac{\partial E(V)}{\partial\beta_\epsilon}=0\quad\frac{\partial E(V)}{\partial\gamma_\epsilon}=0$$

are described by the constraints

$$\alpha_\epsilon=\frac{\gamma_\epsilon\hat{x}_0\left(Z_{\max}^2+Z_{\min}^2+Z_{\max}Z_{\min}\right)+\frac{3}{2}y_{0_\epsilon}(Z_{\max}+Z_{\min})}{\left(Z_{\max}^2+Z_{\min}^2+Z_{\max}Z_{\min}\right)}$$

and

$$\beta_\epsilon=\frac{\gamma_\epsilon\hat{y}_0\left(Z_{\max}^2+Z_{\min}^2+Z_{\max}Z_{\min}\right)-\frac{3}{2}x_{0_\epsilon}(Z_{\max}+Z_{\min})}{\left(Z_{\max}^2+Z_{\min}^2+Z_{\max}Z_{\min}\right)}$$

This means that the rotational error is constrained to a line, which contains a point with $\gamma_\epsilon=0$. The orthogonality constraint holds for this point.

(c) Independent of the amount of noise, the function also has global minima different from zero. Solving for

$$\frac{\partial E(V)}{\partial x_{0_\epsilon}} = \frac{\partial E(V)}{\partial y_{0_\epsilon}} = \frac{\partial E(V)}{\partial \alpha_\epsilon} = \frac{\partial E(V)}{\partial \beta_\epsilon} = \frac{\partial E(V)}{\partial \gamma_\epsilon} = 0$$

we obtain

$$x_{0_\epsilon} = 0 \quad y_{0_\epsilon} = 0 \quad \alpha_\epsilon = \gamma_\epsilon x_0 \quad \text{and} \quad \beta_\epsilon = \gamma_\epsilon y_0$$

which shows that, using negative depth from optical flow as a criterion in 3D motion estimation, the translation can be derived exactly, but the rotation cannot be decoupled from the translation.