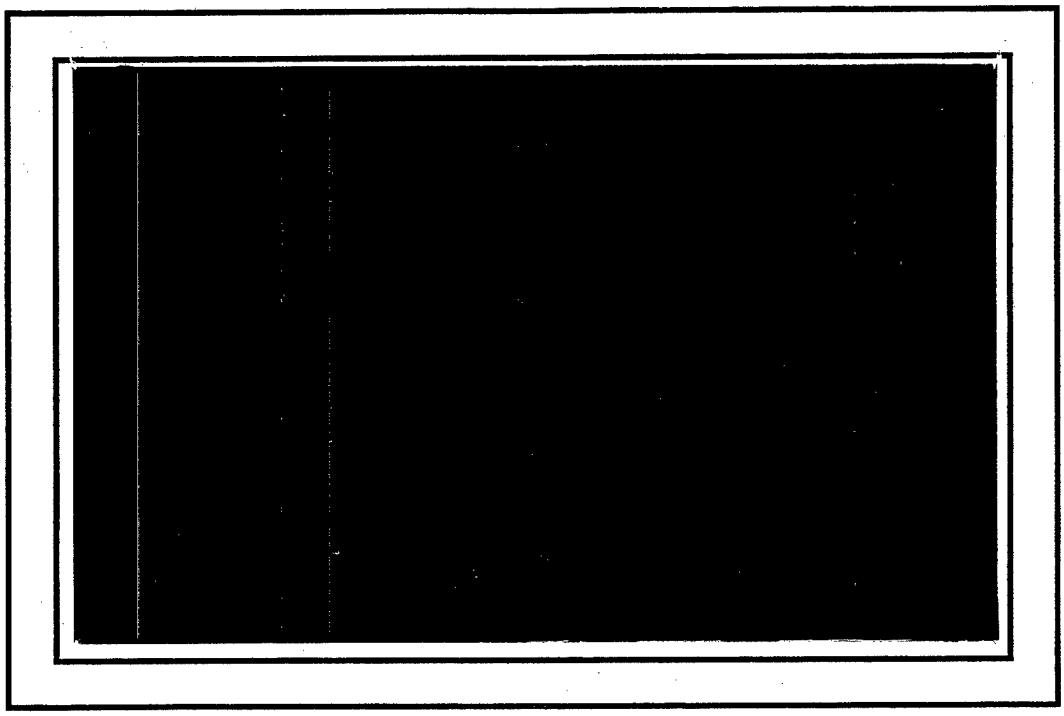
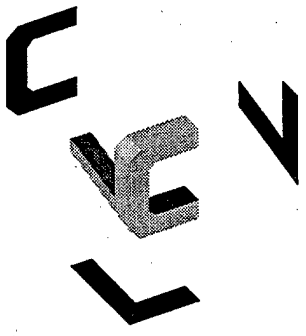


COMPUTER VISION LABORATORY



CENTER FOR AUTOMATION RESEARCH

UNIVERSITY OF MARYLAND

COLLEGE PARK, MARYLAND
20742-3275

CAR-TR-901
CS-TR-3963

N00014-96-1-0587
December 1998

Geometry of Eye Design: Biology and Technology

Cornelia Fermüller and Yiannis Aloimonos
Computer Vision Laboratory
Center for Automation Research
University of Maryland, College Park, MD 20742-3275

Abstract

Natural or artificial vision systems process the images that they collect with their eyes or cameras in order to derive information for performing tasks related to navigation and recognition. Since the way images are acquired determines how difficult it is to perform a visual task, and since systems have to cope with limited resources, the eyes used by a specific system should be designed to optimize subsequent image processing as it relates to particular tasks. Different ways of sampling light, i.e., different eyes, may be less or more powerful with respect to particular competences. This seems intuitively evident in view of the variety of eye designs in the biological world. It is shown here that a spherical eye (an eye or system of eyes providing panoramic vision) is superior to a camera-type eye (an eye with restricted field of view) as regards the competence of three-dimensional motion estimation. This result is derived from a statistical analysis of all the possible computational models that can be used for estimating 3D motion from an image sequence. The findings explain biological design in a mathematical manner, by showing that systems that fly and thus need good estimates of 3D motion gain advantages from panoramic vision. Also, insights obtained from this study point to new ways of constructing powerful imaging devices that suit particular tasks in robotics, visualization and virtual reality better than conventional cameras, thus leading to a new camera technology.

Special thanks to Sara Larson for her editorial and graphics assistance. The support of the Office of Naval Research under Grant N00014-96-1-0587 is gratefully acknowledged.

When classifying eye designs in biological systems, one can differentiate between the different ways of gathering light at the retina, whether single or multiple lenses are used, the spatial distribution of the photoreceptors, the shapes of the imaging surfaces, and what geometrical and physical properties of light are measured (frequency, polarization). A landscape of eye evolution is provided by Michael Land in [3]. Considering evolution as a mountain, with the lower hills representing earlier steps in the evolutionary ladder, and the highest peaks representing later stages of evolution, the situation is pictured in Figure 1. At the higher levels of evolution one finds the compound eyes of insects and crustaceans and the camera-type eyes such as the corneal eyes of land vertebrates and fish. These two categories constitute two fundamentally different designs. Fundamental differences also arise from the positions in the head where camera-type eyes are placed, for example, close to each other as in humans and primates, or on opposite sides of the head as in birds and fish, providing panoramic vision. It appears that the eyes of an organism evolve in a way that best serves that organism in carrying out its tasks. Thus, the success of an eye design should not be judged in an anthropicanic manner, i.e., by how accurately it forms an image of the outside world; rather, it should be judged in a purposive sense. A successful eye design is one that makes the performance of the visual tasks a system is confronted with as easy as possible (fast and robust) [18]. The discovery of principles relating eye design to system behavior will shed light on the problem of evolution in general, and on the structure and function of the brain in particular. At the same time, it will contribute to the development of alternative camera technologies; cameras replace eyes in artificial systems and different camera designs will be more or less appropriate for different tasks. Cameras used in alarm systems, inspection processes, virtual reality systems and human augmentation tasks need not be the same; they should be designed to facilitate the tasks at hand. This paper represents a first effort to introduce structure into the landscape of eyes as it relates to tasks that systems perform.

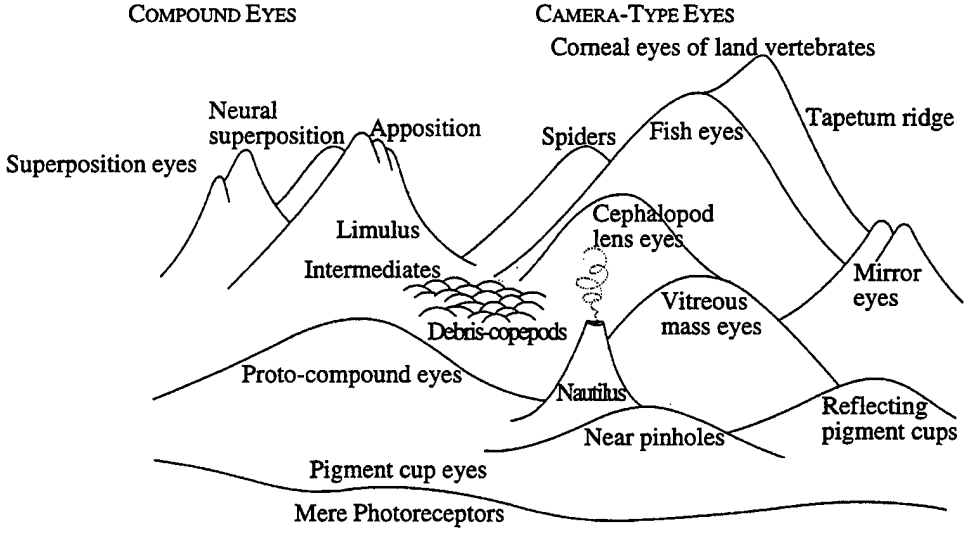


Figure 1: Michael Land's landscape of eye evolution (from [3]).

Although the space of tasks or behaviors performed by vision systems is difficult to formalize, there exist a few tasks that are performed by the whole spectrum of vision systems. All systems with vision move in their environments. As they move, they need to continuously make sense of the moving images they receive on their retinae and they need to solve problems related to

navigation; in particular, they need to know how they themselves are moving [1, 4, 20]. Inertial sensors can help in this task, but it is vision that can provide accurate answers. Regardless of the way in which a system moves (walks, crawls, flies, etc.), its eyes move rigidly. This rigid motion can be described by a translation and a rotation; knowing how a system moves amounts to knowing the parameters describing its instantaneous velocity. This is not to say, of course, that a vision system has an explicit representation of the parameters of the rigid motion that its eyes undergo. This knowledge could be implicit in the circuits that perform specific tasks, such as stabilization, landing, pursuit, etc. [9, 14, 26, 28], but successful completion of navigation-related tasks presupposes some knowledge of the egomotion parameters or subsets of them. Thus, a comparison of eyes with regard to egomotion estimation should lead to a better understanding of one of the most basic visual competences.

Two fundamentally different eye designs are compared here, a spherical eye and a planar, camera-type eye (Figure 2). Spherical eyes model the compound eyes of insects, while planar eyes model the corneal eyes of land vertebrates as well as fish. In addition, the panoramic vision of some organisms, achieved by placing camera-type eyes on opposite sides of the head, is approximated well by a spherical eye. The essential difference between a spherical and a planar eye lies in the field of view, 360 degrees in the spherical case and a restricted field in the planar case. The comparison performed here demonstrates that spherical eyes are superior to planar eyes for 3D motion estimation. "Superior" here means that the ambiguities inherent in deriving 3D motion from planar image sequences are not present in the spherical case. Specifically, a geometrical/statistical analysis is conducted to investigate the functions that can be used to estimate 3D motion, relating 2D image measurements to the 3D scene. These functions are expressed in terms of errors in the 3D motion parameters and they can be understood as multi-dimensional surfaces in those parameters. 3D motion estimation amounts to a minimization problem; thus, our approach is to study the relationships among the parameters of the errors in the estimated 3D motion at the minima of the surfaces, because these locations provide insight into the behaviors of the estimation procedures. It is shown that, at the locations of the minima, the errors in the estimates of both the translation and rotation are non-zero in the planar case, while in the spherical case either the translational or rotational error becomes zero. Intuitively, with a camera-type eye there is an unavoidable confusion between translation and rotation, as well as between translational errors and the actual translation. This confusion does not occur with a spherical eye. The implication is that visual navigation tasks involving 3D motion parameter estimation are easier to solve with spherical eyes than with planar eyes.

The basic geometry of image motion is well understood. As a system moves in its environment, every point of the environment has a velocity vector relative to the system. The projections of these 3D velocity vectors on the retina of the system's eye constitutes the motion field. For an eye moving with translation $\mathbf{t}$ and rotation $\boldsymbol{\omega}$ in a stationary environment, each scene point $\mathbf{R} = (X, Y, Z)$ measured with respect to a coordinate system $OXYZ$ fixed to the nodal point of the eye has velocity $\dot{\mathbf{R}} = -\mathbf{t} - \boldsymbol{\omega} \times \mathbf{R}$. 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 2a) 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

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

with $Z = \mathbf{R}\cdot\mathbf{z}_0$ representing the depth. If the image is formed on a sphere of radius f (Figure 2b) 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

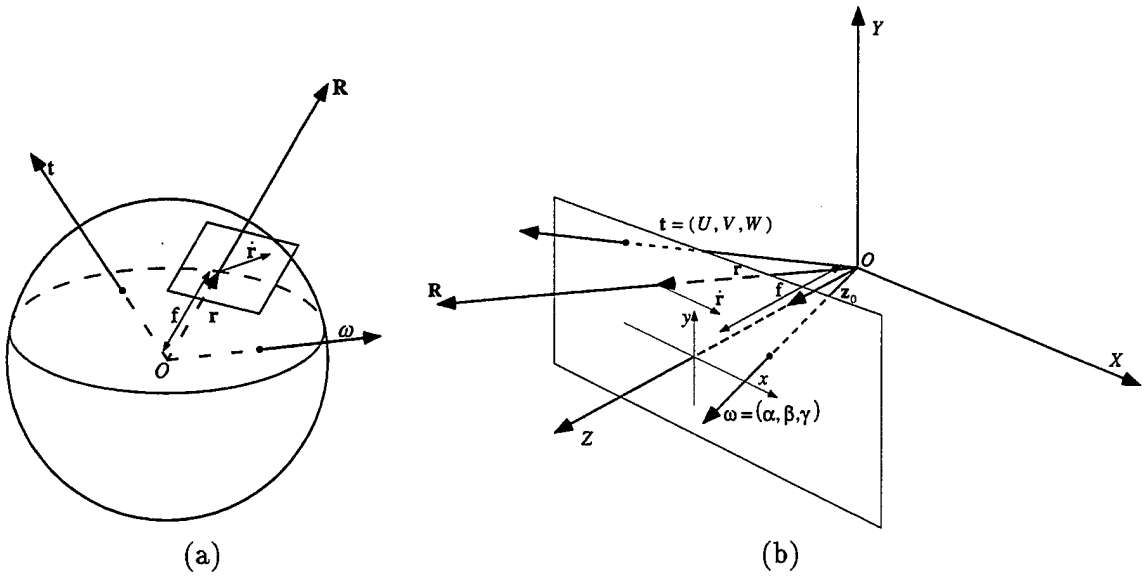


Figure 2: Image formation on the sphere (a) and on the plane (b). The system moves with a rigid motion with translational velocity $\mathbf{t}$ and rotational velocity $\boldsymbol{\omega}$. Scene points $\mathbf{R}$ project onto image points $\mathbf{r}$ and the 3D velocity $\dot{\mathbf{R}}$ of a scene point is observed in the image as image velocity $\dot{\mathbf{r}}$.

being the norm of $\mathbf{R}$ (the range), and the image motion is

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

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 egomotion then amounts to finding the scaled vector $\mathbf{t}$ and the vector $\boldsymbol{\omega}$ from a representation of the motion field.

To set up mathematical formulations for 3D motion estimation, the following questions should be answered. The first question to be addressed is, what description containing information about 3D motion does a system use to represent the image sequence? One might envision a sophisticated system that could attempt to estimate the motion field, termed the optic flow field [15]. On the other hand, it is also easy to envision a system that does not have the capacity to estimate the motion field, but only to obtain a partial description of it. An example of a description containing minimal information about image motion is the normal motion field. This amounts to the projection of the motion field onto the direction of the image gradient at each point, and represents the movement of each local edge element in the direction perpendicular to itself. Normal flow can be estimated from local spatiotemporal information in the image [22–24, 27]. 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

$$v_n = \dot{\mathbf{r}}\cdot\mathbf{n}. \tag{3}$$

Unlike normal flow, the estimation of optic flow is a difficult problem because information from different image neighborhoods must be compared and used in a smoothing scheme to account for discontinuities [10, 12]. Although it is not yet known exactly what kinds of image representations different visual systems recover, it is clear that such descriptions should lie somewhere between normal flow fields and optic flow fields. Thus, when comparing eye designs with regard to 3D motion estimation, one must consider both kinds of flow fields.

The second question to be addressed is, through what geometric laws or constraints is 3D motion coded into image motion? The constraints are easily observed from (1–3). Equations (1) and (2) show how the motions of image points are related to 3D rigid motion and to scene depth. By eliminating depth from these equations, one obtains the well known epipolar constraint [19]; for both planar and spherical eyes it is

$$(\mathbf{t} \times \mathbf{r}) \cdot (\dot{\mathbf{r}} + \boldsymbol{\omega} \times \mathbf{r}) = 0. \tag{4}$$

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 [11, 21] of the function[1]

$$M_{ep} = \iint_{\text{image}} \left[\left(\hat{\mathbf{t}} \times \mathbf{r} \right) \cdot (\dot{\mathbf{r}} + \hat{\boldsymbol{\omega}} \times \mathbf{r}) \right]^2 d\mathbf{r}. \tag{5}$$

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 (3), 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 (3) 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 $\mathbf{r}$:

$$\hat{Z}(\text{or } \hat{R}) = \frac{\mathbf{u}_{\text{tr}}(\hat{\mathbf{t}}) \cdot \mathbf{n}}{(\dot{\mathbf{r}} - \mathbf{u}_{\text{rot}}(\hat{\boldsymbol{\omega}})) \cdot \mathbf{n}}. \tag{6}$$

If the numerator and denominator of (6) 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

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

Then estimation of 3D motion from normal flow amounts to minimizing [4, 5, 13] the function

$$M_{nd} = \iint_{\text{image}} I_{nd}(\mathbf{r}) d\mathbf{r}. \tag{7}$$

Expressing $\dot{\mathbf{r}}$ in terms of the real motion from (1) and (2), functions (5) and (7) 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

[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 [25] and alternatives have been proposed that reduce this bias, but without eliminating the confusion between rotation and translation.

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 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:

$$E_{ep} = \int_{Z=Z_{\min}}^{Z_{\max}} M_{ep} dZ, \qquad (8) \qquad E_{nd} = \int_{\psi=0}^{\pi} \int_{Z=Z_{\min}}^{Z_{\max}} M_{nd} dZ \, d\psi, \qquad (9)$$

measuring deviation from the epipolar constraint and the amount of negative depth, respectively. Functions (8) and (9) are five-dimensional surfaces in $\mathbf{t}_\epsilon, \omega_\epsilon$, the errors in the motion parameters.

We are interested in the topographic structure of these surfaces, in particular, in the relationships among the errors and the relationships of the errors to the actual motion parameters at the minima of the functions. 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.

Independent of the particular algorithm, 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, 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 ω_ϵ. 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 ω_ϵ 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.[2] Each of the three cases is studied for four optimizations: epipolar minimization for the sphere and the plane 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 on the plane and sphere.

[2] One may wish to study the problem in the presence of noise in the flow measurements and derive instead the expected values of the local and global minima. It has been shown, however, that noise which is of no particular bias does not alter the local minima, and the global minima fall within the valleys of the function without noise. In particular, we considered in [7] noise N of the form $N = \epsilon\frac{1}{Z} + \delta$, with ϵ, δ 2D, independent, stochastic error vectors. As such noise does not alter the function's overall structure, it won't be considered here; the interested reader is referred to [7].

Epipolar Minimization on the Plane Denote estimated quantities by letters with hat signs, actual quantities by unmarked letters, and the differences between actual and estimated quantities (the errors) by the subscript "ϵ." Furthermore, let $\mathbf{t} = (x_0, y_0, 1)$ and $\omega = (\alpha, \beta, \gamma)$. 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.

Considering a circular aperture of radius e, setting the focal length $f = 1$, $W = 1$ and $\hat{W} = 1$, the function in (8) becomes

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

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

$$\begin{aligned} E_{ep} \;=\; & \pi e^2 \left((Z_{\max} - Z_{\min}) \left(\frac{1}{3}\gamma_\epsilon^2 e^4 + \frac{1}{4}\left(\gamma_\epsilon^2 \left(\hat{x}_0^2 + \hat{y}_0^2 \right) + 6\gamma_\epsilon (\hat{x}_0\alpha_\epsilon + \hat{y}_0\beta_\epsilon) + \alpha_\epsilon^2 + \beta_\epsilon^2 \right) e^2 + \right. \right. \\ & \left. \left. (\hat{x}_0\alpha_\epsilon + \hat{y}_0\beta_\epsilon)^2 \right) + (\ln(Z_{\max}) - \ln(Z_{\min})) \left(\frac{1}{2}(3\gamma_\epsilon(x_{0_\epsilon} y_0 - y_{0_\epsilon} x_0) + x_{0_\epsilon}\beta_\epsilon - y_{0_\epsilon}\alpha_\epsilon) e^2 + \right. \\ & \left. 2(x_{0_\epsilon} y_0 - y_{0_\epsilon} x_0)(\hat{x}_0\alpha_\epsilon + \hat{y}_0\beta_\epsilon) \right) + \\ & \left(\frac{1}{Z_{\min}} - \frac{1}{Z_{\max}} \right) \left(\frac{1}{4}\left(y_{0_\epsilon}^2 + x_{0_\epsilon}^2 \right) e^2 + (x_{0_\epsilon} y_0 - y_{0_\epsilon} x_0)^2 \right) \Big) \end{aligned} \tag{10}$$

(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 (10) is obtained from the first-order conditions $\frac{\partial E_{ep}}{\partial \alpha_\epsilon} = \frac{\partial E_{ep}}{\partial \beta_\epsilon} = \frac{\partial E_{ep}}{\partial \gamma_\epsilon} = 0$, which yield

$$\alpha_\epsilon = \frac{y_{0_\epsilon}(\ln(Z_{\max}) - \ln(Z_{\min}))}{Z_{\max} - Z_{\min}} \quad \beta_\epsilon = \frac{-x_{0_\epsilon}(\ln(Z_{\max}) - \ln(Z_{\min}))}{Z_{\max} - Z_{\min}} \quad \gamma_\epsilon = 0 \tag{11}$$

It follows that $\alpha_\epsilon/\beta_\epsilon = -x_{0_\epsilon}/y_{0_\epsilon}, \gamma_\epsilon = 0$, which means that there is no error in γ and the projection of the translational error on the image is perpendicular to the projection of the rotational error. This constraint is called the "orthogonality constraint."

(b) Assuming that rotation has been estimated with an error $(\alpha_\epsilon, \beta_\epsilon, \gamma_\epsilon)$, the relationship among the errors is obtained from $\frac{\partial E_{ep}}{\partial x_{0_\epsilon}} = \frac{\partial E_{ep}}{\partial y_{0_\epsilon}} = 0$. 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.

(c) In the general case, we need to study the subspaces in which E_{ep} 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_{ep} 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

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

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 constraint $x_0/y_0 = \hat{x}_0/\hat{y}_0$; that is, the projection of the actual translation and the projection of the estimated translation lie on a line passing through the image center. We refer to this second constraint as the "line constraint." These results are in accordance with previous studies [2, 21], which found that the translational components along the x and y axes are confused with rotation around the y and x axes, respectively, and the "line constraint" under a set of restrictive assumptions.

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

$$
E_{ep} = \int_{R_{\min}}^{R_{\max}} \iint_{\text{sphere}} \left\{ \left(\frac{\mathbf{r} \times (\mathbf{r} \times \mathbf{t})}{R} - (\omega_\epsilon \times \mathbf{r}) \right) \cdot \left(\hat{\mathbf{t}} \times \mathbf{r} \right) \right\}^2 dA \, dR
$$

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 $\mathbf{u}_{\text{tr}}(\mathbf{r}) = \mathbf{u}_{\text{tr}}(-\mathbf{r})$ and $\mathbf{u}_{\text{rot}}(\mathbf{r}) = -\mathbf{u}_{\text{rot}}(-\mathbf{r})$, when the integrand is expanded the product terms integrated over the sphere vanish. Thus

$$
E_{ep} = \int_{R_{\min}}^{R_{\max}} \iint_{\text{sphere}} \left\{ \frac{\left(\left(\mathbf{t} \times \hat{\mathbf{t}} \right) \cdot \mathbf{r} \right)^2}{R^2} + \left((\omega_\epsilon \times \mathbf{r}) \cdot \left(\hat{\mathbf{t}} \times \mathbf{r} \right) \right)^2 \right\} dA \, dR
$$

(a) Assuming that translation $\hat{\mathbf{t}}$ has been estimated, the ω_ϵ that minimizes E_{ep} is $\omega_\epsilon = 0$, since the resulting function is non-negative quadratic in ω_ϵ (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 ($\omega_\epsilon = 0$), 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 ω_ϵ, 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:

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

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 $\hat{\mathbf{t}}, \omega_\epsilon$ 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 $\hat{\mathbf{t}} \perp \omega_\epsilon$. Consider a certain 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 $(\mathbf{t} \times \hat{\mathbf{t}}) \cdot \omega_\epsilon = 0$, 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 ω_ϵ, with the exact position depending on $|\omega_\epsilon|$ 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 $\mathbf{t} = \hat{\mathbf{t}}$ and $\omega_\epsilon \perp \mathbf{t}$.

To study the negative depth values described by function (9) a more geometric interpretation is needed. Substituting into (6) the value of $\dot{\mathbf{r}}$ from (1) or (2) gives

$$\hat{Z}(\text{or } \hat{R}) = \frac{\mathbf{u}_{\text{tr}}(\hat{\mathbf{t}}) \cdot \mathbf{n}}{\left(\frac{\mathbf{u}_{\text{tr}}(\mathbf{t})}{Z(\text{or } R)} - \mathbf{u}_{\text{rot}}(\omega_\epsilon)\right) \cdot \mathbf{n}}$$

This equation shows that for every $\mathbf{n}$ and $\mathbf{r}$ a range of values for Z (or R) is obtained which result in negative estimates of $\hat{Z}$ (or $\hat{R}$). Thus for each direction $\mathbf{n}$, considering all image points $\mathbf{r}$, we obtain a volume in space corresponding to negative depth estimates. The sum of all these volumes for all directions is termed the "negative depth" volume, and calculating 3D motion in this case amounts to minimizing this volume. Minimization of this volume provides conditions for the errors in the motion parameters.

Minimizing Negative Depth Volume on the Plane This analysis is given in [6]. The findings are summarized here:

(a) Assume that rotation has been estimated with an error $(\alpha_\epsilon, \beta_\epsilon, \gamma_\epsilon)$. Then the error $(x_{0_\epsilon}, y_{0_\epsilon})$ that minimizes the negative depth volume satisfies the orthogonality constraint $x_{0_\epsilon}/y_{0_\epsilon} = -\beta_\epsilon/\alpha_\epsilon$.

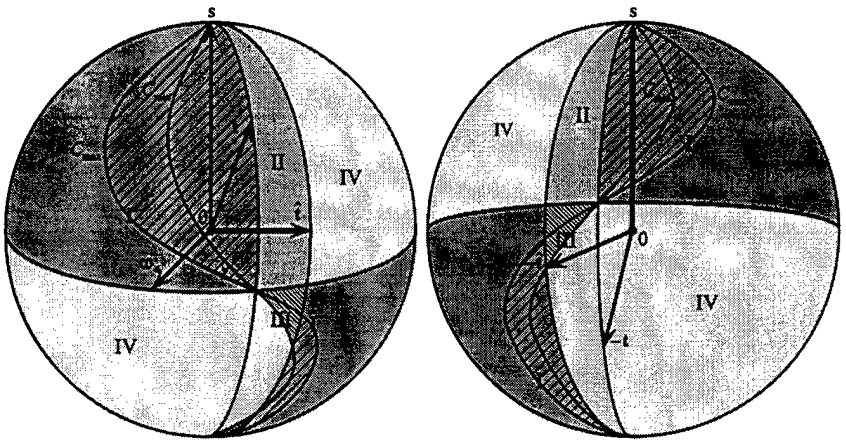
(b) In the absence of any prior information about the 3D motion, the solution obtained by minimizing the negative depth volume has errors that satisfy the orthogonality constraint $x_{0_\epsilon}/y_{0_\epsilon} = -\beta_\epsilon/\alpha_\epsilon$, the line constraint $x_0/y_0 = \hat{x}_0/\hat{y}_0$ and $\gamma_\epsilon = 0$

Minimizing Negative Depth Volume on the Sphere

(a) Assuming that the rotation has been estimated with an error ω_ϵ, 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 ω_ϵ perpendicular to $\mathbf{s}$ $(\mathbf{s} \cdot \omega_\epsilon = 0)$.

We are interested in the points in space with estimated negative range values $\hat{R}$. Since $\mathbf{n} = \frac{\mathbf{s}\times\mathbf{r}}{\|\mathbf{s}\times\mathbf{r}\|}$, $\mathbf{s} \cdot \omega_\epsilon = 0$, the estimated range $\hat{R}$ amounts to $\hat{R} = R\frac{(\hat{\mathbf{t}}\times\mathbf{s})\cdot\mathbf{r}}{(\mathbf{t}\times\mathbf{s})\cdot\mathbf{r} - R(\omega_\epsilon\cdot\mathbf{r})(\mathbf{s}\cdot\mathbf{r})}$. $\hat{R} < 0$ if $\text{sgn}[(\hat{\mathbf{t}} \times \mathbf{s}) \cdot \mathbf{r}] = -\text{sgn}[(\mathbf{t} \times \mathbf{s}) \cdot \mathbf{r} - R(\omega_\epsilon \cdot \mathbf{r})(\mathbf{s} \cdot \mathbf{r})]$, where $\text{sgn}(x)$ provides the sign of x. This



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

Figure 3: 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.

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 3.

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 $\mathbf{t} = \hat{\mathbf{t}}$.[3]

(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 ω_ϵ the negative depth values change the least in the neighborhood of the absolute minimum, that is, at $\mathbf{t}_\epsilon = \omega_\epsilon = 0$. From the analysis above, it is known that for any $\omega_\epsilon \neq 0$, $\mathbf{t} = \hat{\mathbf{t}}$. Next, we show that ω_ϵ is indeed different from zero: Take $\mathbf{t} \neq \hat{\mathbf{t}}$ on the great circle of $\mathbf{s}$ and let ω_ϵ, as before, be perpendicular to $\mathbf{s}$.

Since $(\mathbf{t} \times \mathbf{s}) \times \omega_\epsilon = 0$, the curves $C_{\max}$ and $C_{\min}$ can be expressed as $C_{\max(\min)} = (\omega_\epsilon \cdot \mathbf{r})(\frac{\sin \angle(\mathbf{t},\mathbf{s})}{|\omega_\epsilon| R_{\max(\min)}} - (\mathbf{s} \cdot \mathbf{r})) = 0$, where $\sin \angle(\mathbf{t}, \mathbf{s})$ denotes the angle between vectors $\mathbf{t}$ and $\mathbf{s}$. These curves consist of the great circle $\omega_\epsilon \cdot \mathbf{r} = 0$ and the circle $\frac{\sin \angle(\mathbf{t},\mathbf{s})}{|\omega_\epsilon| R_{\max(\min)}} - (\mathbf{s} \cdot \mathbf{r}) = 0$ parallel to the great circle $(\mathbf{s} \cdot \mathbf{r}) = 0$ (see Figure 4). If $\frac{\sin \angle(\mathbf{t},\mathbf{s})}{|\omega_\epsilon| R_{\max(\min)}} > 1$, this circle disappears.

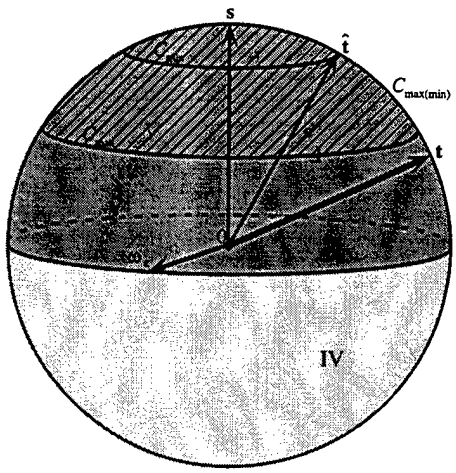


Figure 4: Configuration for $\mathbf{t}$ and $\hat{\mathbf{t}}$ on the great circle of $\mathbf{s}$ and ω_ϵ perpendicular to $\mathbf{s}$. The textured part of area I denotes image points for which negative depth values exist if the scene is bounded.

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 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, $\frac{\sin \angle(\mathbf{t},\mathbf{s})}{R_{\max}|\omega_\epsilon|} \geq 1$, V_3

[3] 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.

becomes zero. Thus for all $|\omega_\epsilon|$ with $|\omega_\epsilon| \leq \frac{\sin \angle(\mathbf{t},\hat{\mathbf{t}})}{R_{\max}}$, 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 $R_{\max} = \infty$, $|\omega_\epsilon|$ takes on values different from zero.

This shows that for any $\mathbf{t}_\epsilon \neq 0$, there exist vectors $\omega_\epsilon \neq 0$ which give rise to the same negative depth volume as $\omega_\epsilon = 0$. However, for any such $\omega_\epsilon \neq 0$ this volume is larger than the volume obtained by setting $\mathbf{t}_\epsilon = 0$. It follows that $\mathbf{t} = \hat{\mathbf{t}}$. From Figure 3, it can furthermore be deduced that for a given ω_ϵ 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 ω_ϵ, 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 $\mathbf{t} = \hat{\mathbf{t}}$, $\mathbf{t} \perp \omega_\epsilon$.

The preceding results demonstrate the advantages of spherical eyes for 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 negative 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 [4]. Similarly, there exist computational advantages for systems with translational inertial sensors in estimating the remaining unknown rotation.

In nature, systems that walk and perform sophisticated manipulation have camera-type eyes, and systems that fly usually have panoramic vision, either through compound eyes or a combination of camera-type eyes. The obvious explanation for this difference is the need for a larger field of view in flying species, and the need for very accurate segmentation and shape estimation, and thus high resolution in a limited field of view, for land-walking species. As shown in this paper, the geometry of the sphere also provides a computational advantage; it allows for more efficient and accurate egomotion estimation (even at the expense of trading off resolution in some systems, for example, in insects), and this is much more necessary for systems flying and thus moving with all six degrees of freedom than for systems moving with usually limited rigid motion on surfaces.

The above results also point to ways of constructing 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 5, assembled from a few video cameras arranged on the surface of a sphere,[4] can easily estimate 3D motion since, while it is moving, it is sampling a spherical motion field! Even more important for today's applications is the reconstruction of the shape of an object or scene in a very accurate manner. Accurate shape models are needed in many applications dealing with visualization, as in video editing/manipulation or in virtual reality settings [16, 17]. To obtain accurate shape reconstruction, both the 3D transformation relating two views and the 2D transformation relating two images are needed with good precision. Given accurate 3D motion $(\mathbf{t},\omega)$ and image motion $(\dot{\mathbf{r}})$, equations (1–3) can be used in a straightforward manner to estimate depth (Z) or range (R) and thus object shape. An eye like the one in Figure 5 not only has panoramic properties, eliminating the rotation/translation

[4] Like a compound eye with video cameras replacing ommatidia

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 $\omega_\epsilon = 0$ (b) Without any prior information, $\mathbf{t}_\epsilon = 0$ and $\omega_\epsilon \perp \mathbf{t}$	(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 ω_ϵ, the translational error $\mathbf{t}_\epsilon = 0$ (b) Without any prior information, $\mathbf{t}_\epsilon = 0$ and $\omega_\epsilon \perp \mathbf{t}$	(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 $\gamma_\epsilon = 0$, $\alpha_\epsilon/\beta_\epsilon = -x_{0_\epsilon}/y_{0_\epsilon}$, $x_0/y_0 = x_{0_\epsilon}/y_{0_\epsilon}$

confusion, 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 [8]. As a consequence, the eye in Figure 5 is very well suited to developing accurate models of the world, and many experiments have confirmed this finding. However, such an eye, although appropriate for a moving robotic system, may be impractical to use in a laboratory. Fortuitously from a mathematical viewpoint, it makes no difference whether the cameras are looking inward or outward! Consider, then, a "negative" spherical eye like the one in Figure 6, where video cameras are arranged on the surface of a sphere pointing toward its center. Imaging a moving rigid object at the center of the sphere creates image motion fields at the center of each camera which are the same as the ones that would be created if the whole spherical dome were moving with the opposite rigid motion! Thus, utilizing information from all the cameras, the 3D motion of the object inside the sphere can be accurately estimated, and at the same time accurate shape models can be obtained from the motion field of each

camera. The negative spherical eye also allows for accurate recovery of models of action, such as human movement, because putting together motion and shape, sequences of 3D motion fields representing the motion inside the dome can be estimated. Such action models will find many applications in telereality, graphics and recognition. The above described configurations are examples of alternative sensors, and they also demonstrate that multiple-view vision has great potential. Different arrangements best suited for other problems can be imagined. This was perhaps foreseen in ancient Greek mythology, which has Argus, the hundred-eyed guardian of Hera, the goddess of Olympus, defeating a whole army of Cyclopes, one-eyed giants!

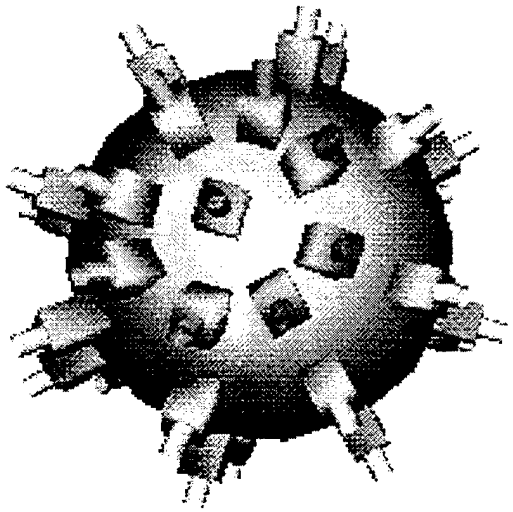


Figure 5: A compound-like eye composed of conventional video cameras, arranged on a sphere and looking outward.

References

[1] Y. Aloimonos and A. Rosenfeld. Computer vision. *Science*, 253:1249–1254, 1991.

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

[3] R. Dawkins. *Climbing Mount Improbable.* Norton, New York, 1996.

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

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

[6] C. Fermüller and Y. Aloimonos. Ambiguity in structure from motion: Sphere versus plane. *International Journal of Computer Vision*, 28:137–154, 1998.

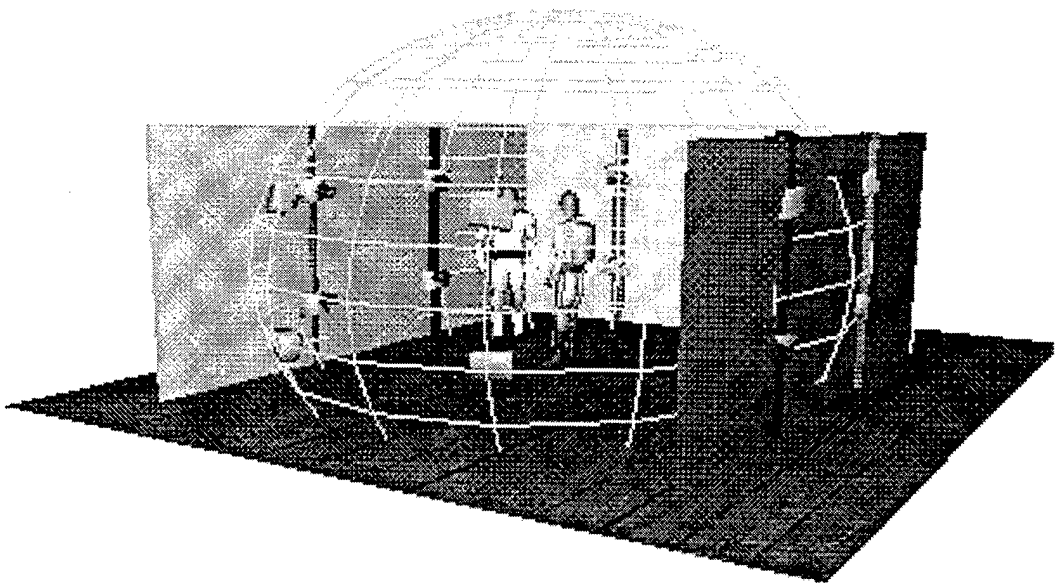


Figure 6: A "negative" spherical eye, consisting of conventional video cameras arranged on a sphere and pointing inward.

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

[8] C. Fermüller, F. Defoort, and Y. Aloimonos. Motion segmentation for a binocular observer. Technical Report CS-TR-3893, Center for Automation Research, University of Maryland, April 1998.

[9] T. Hamada. Vision, action, and navigation in animals. In Y. Aloimonos, editor, *Visual Navigation: From Biological Systems to Unmanned Ground Vehicles*, chapter 2. Lawrence Erlbaum Associates, Mahwah, NJ, 1997.

[10] E. Hildreth. Computations underlying the measurement of visual motion. *Artificial Intelligence*, 23:309–354, 1984.

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

[12] B. K. P. Horn and B. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

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

[14] G. A. Horridge. What can engineers learn from insect vision? *Proc. Royal Society, London B*, August 1992.

[15] J. J. Koenderink. Optic flow. *Vision Research*, 26:161–180, 1986.

[16] K. N. Kutulakos and S. M. Seitz. What do N photographs tell us about 3D shape? Computer Science Technical Report 692, University of Rochester, 1998.

[17] K. N. Kutulakos and J. R. Vallino. Calibration-free augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, 4:1–20, 1998.

[18] M. Land and R. D. Fernald. The evolution of eyes. *Annual Review of Neuroscience*, 15:1–29, 1992.

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

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

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

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

[23] W. Reichardt. Autocorrelation, a principle for evaluation of sensory information by the central nervous system. In W. A. Rosenblith, editor, *Sensory Communication: Contributions to the Symposium on Principles of Sensory Communication*, pages 303–317. John Wiley and Sons, New York, 1961.

[24] W. Reichardt. Evaluation of optical motion information by movement detectors. *Journal of Comparative Physiology*, 161:533–547, 1987.

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

[26] M. V. Srinivasan, S. W. Zhang, and K. C. Shekara. Evidence for two distinct movement detecting mechanisms in insect vision. *Naturwissenschaften*, 80:38–41, 1993.

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

[28] S. M. Zeki. *A Vision of the Brain.* Blackwell Scientific Publications, 1993.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	December 1998	Technical Report

4. TITLE AND SUBTITLE

Geometry of Eye Design: Biology and Technology

5. FUNDING NUMBERS

N00014-96-1-0587

6. AUTHOR(S)

Cornelia Fermuller and Yiannis Aloimonos

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

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

8. PERFORMING ORGANIZATION REPORT NUMBER

CAR-TR-901
CS-TR-3963

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

Office of Naval Research, 800 North Quincy Street, Arlington, VA 22217-5660

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Navy position, policy, or decision, unless so designated by other documentation.

12a. DISTRIBUTION/AVAILABILITY STATEMENT

Approved for public release.
Distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

Natural or artificial vision systems process the images that they collect with their eyes or cameras in order to derive information for performing tasks related to navigation and recognition. Since the way images are acquired determines how difficult it is to perform a visual task, and since systems have to cope with limited resources, the eyes used by a specific system should be designed to optimize subsequent image processing as it relates to particular tasks. Different ways of sampling light, i.e., different eyes, may be less or more powerful with respect to particular competences. This seems intuitively evident in view of the variety of eye designs in the biological world. It is shown here that a spherical eye (an eye or system of eyes providing panoramic vision) is superior to a camera-type eye (an eye with restricted field of view) as regards the competence of three-dimensional motion estimation. This result is derived from a statistical analysis of all the possible computational models that can be used for estimating 3D motion from an image sequence. The findings explain biological design in a mathematical manner, by showing that systems that fly and thus need good estimates of 3D motion gain advantages from panoramic vision. Also, insights obtained from this study point to new ways of constructing powerful imaging devices that suit particular tasks in robotics, visualization and virtual reality better than conventional cameras, thus leading to a new camera technology.

14. SUBJECT TERMS

Computer vision, Eye, Navigation, Planar projection, Recognition, Spherical projection, Vision

15. NUMBER OF PAGES

17

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL