

Department of Computer & Information Science

Technical Reports (CIS)

University of Pennsylvania

Year 1995

Fixation and a 180 Degree View Simplify
Ego Motion Estimation

Inigo Thomas
University of Pennsylvania,

Eero Simoncelli
University of Pennsylvania,

Ruzena Bajcsy
University of Pennsylvania,

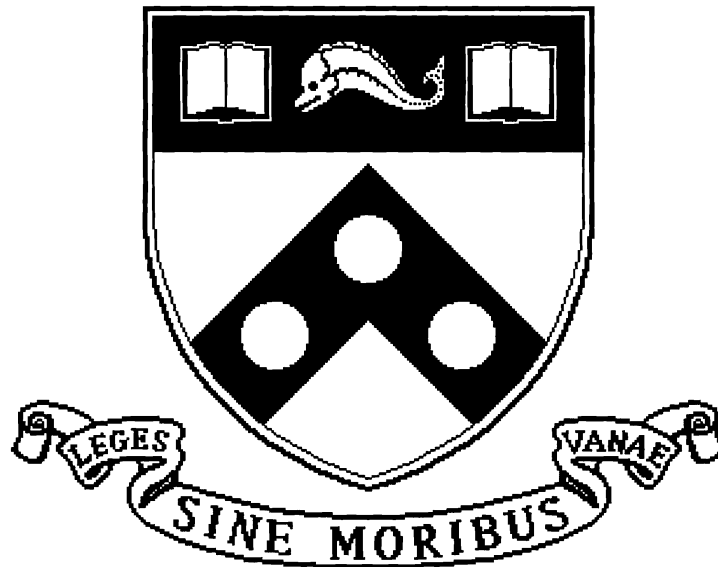
University of Pennsylvania Department of Computer and Information Science Technical Report No. MS-CIS-95-15.

This paper is posted at ScholarlyCommons@Penn.
http://repository.upenn.edu/cis_reports/212

Fixation and a 180 Degree View Simplify Ego Motion Estimation

MS-CIS-95-15
GRASP 390

Inigo Thomas
Eero Simoncelli
Ruzena Bajcsy



University of Pennsylvania
School of Engineering and Applied Science
Computer and Information Science Department
Philadelphia, PA 19104-6389

March 1995

Fixation and a 180 Degree View Simplify Ego Motion Estimation

Inigo Thomas Eero Simoncelli Ruzena Bajcsy

GRASP Laboratory
University of Pennsylvania
3401 Walnut Street, Philadelphia, PA 19104

email: {inigo,eero,bajcsy}@grip.cis.upenn.edu

Abstract

Although moving human observers actively fixate their eyes on points in the world, computer vision algorithms designed for the estimation of structure-from-motion or egomotion typically do not make use of this constraint. The main contribution of this work is to precisely specify the form of the optical flow field for a fixating observer. In particular, we show theoretically that the use of a hemispherical (retinal) imaging surface generates an optical flow field of a particularly simple form. The predictions of this theory are tested using the first actual hemispherical lens-camera system in computer vision, involving a 180 degree field of view lens. A further contribution is the finding that the *sign* of flow at the retinal periphery can be used to predict collisions.

1 Introduction

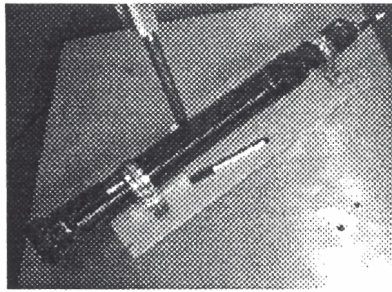
Introspection reveals that when a human observer moves, the eye continually fixates on targets in the world. Although both fixation and very large field-of-view images are involved in human vision, the computational advantage of either has not yet been established. The main contributions of this paper include (i) formalizing retinal flow for a fixating observer, (ii) using a hemispherical retina/imager to simplify patterns of flow, and (iii) thereby isolating very simple patterns of flow at the retinal periphery that are useful in determining an observer's direction of heading. A major contribution of this work is the introduction of the first 180 degree field of view imaging system. This system is used to test some of our theoretical predictions. We expect such imagers, when installed in actively fixating head/eye systems [7], to play an important role in future motion related research.

Active Vision – a paradigm introduced by Bajcsy [2, 3] – provides constraints that allow certain ill-posed problems in Computer Vision to be converted into well-posed

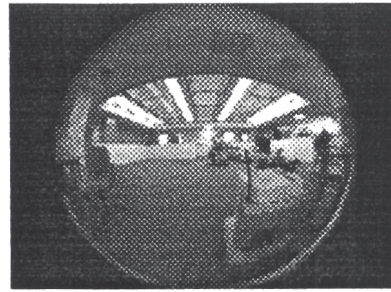
ones. These constraints arise due to simplifying assumptions about the world and its visual appearance that are satisfied when the camera moves actively. In the tradition of such active vision research the present article discusses the constraints on optical flow fields that are imposed by the process of active fixation. The crucial difference between the present work and traditional optical flow analyses (cf. [11] for a review) is *fixation*. In the computer vision literature, a previous analysis of optical flow field under fixation has been provided by Raviv and Herman [8]. As in this paper, they determine the loci of zero flow. They also describe the temporal evolution of these loci of zero flow. In contrast, our development is done in a different coordinate system using a different representation of optical flow, we analyze the pattern of flow across the *entire* hemispherical retina, and we use the *sign* of a particular projection of flow vectors in order to determine direction of heading.

The notion of *sign of flow* has been previously exploited in work by Fermueller [4], who analyzes components of the flow field (so-called *normal* flow) in terms of sign. Due to the assumption of normal flow she arrives at very different patterns from ours. The simple four-way distinction which we isolate at the periphery cannot be observed using normal flow, which gives rise to areas bounded by conic sections. Furthermore, the search over the solution space that is required by her algorithm may make it unrealistic for implementation on a mobile robot. In section 8, we demonstrate an implementation of our closed-form solution being used to guide a robot to a target.

The spherical geometry we employ is a camera model used rarely in motion related research. An exception is that of Nelson and Aloimonos [6]. Spherical projection has the advantage of treating all viewing directions homogeneously, and it allows us to write the optical flow in a particularly simple form. A planar perspective image can be directly converted into an image obtained under spherical projection and vice versa, assuming a calibrated camera (i.e., known focal length and optical axis). In this paper we report results from a 180 degree field of view (FOV) camera, which corresponds to having a truly hemispherical imager. Such a camera is novel to motion research in particular, as well as computer vision in general, offering tremendous potential for realistic obstacle avoidance and autonomous navigation. An image of the 180 degree FOV camera/lens arrangement is shown in Figure 1, along with an image obtained by the camera.



180° FOV Camera



180° Image

Figure 1: 180 Degree Field of View Camera. On the left is shown the camera and lens arrangement that obtains a 180 degree FOV image. On the right is depicted an image of the GRASP Lab obtained by this 180 degree FOV camera. See Section 7 for more details.

In this work we make the following assumptions: (i) the observer moves with respect to the world and fixates on a target; (ii) the world is rigid, with no independently moving elements; and (iii) the possible rotation axes of the eye lie on a plane orthogonal to the direction of gaze. Assumption (i) is a behavioral one that is observed in humans and animals. Assumptions (ii) and (iii) will allow us to write the retinal flow field in a particularly simple form; (iii) also corresponds closely to Listing’s Plane for the human eye.

We first define retinal flow for a 2D universe and then extend it to the full 3D case; the flow in 2D provides one component of the flow in 3D. The retinal flow in 3D is decomposed into longitudinal and latitudinal flow, and we show that the longitudinal component depends only on the translational direction of heading and not the rotation of the eye. We demonstrate that longitudinal flow, especially at the periphery, can be analyzed to determine the direction of heading. Finally we test the performance of our theory in experiments involving a simulated environment, the 180 degree FOV lens, as well as a mobile robot platform.

2 Retinal Flow in a Rigid 2D Universe

For ease of exposition, we first consider a reduced case of a 2D universe in which we define the flow on the retina for any given point in the universe; as the observer moves the flow determines how each point projected on the retina moves.

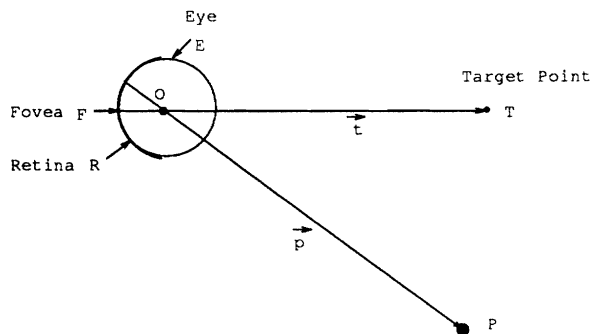


Figure 2: Model of the eye in a 2D universe.

2.1 Calculating Retinal Flow

In a 2D planar universe, the eye of the observer corresponds to a circle (E in Fig. 2), and the retina corresponds to a semicircle (R).¹ As the observer moves, the center of the eye (O in Fig. 2) translates on the 2D plane. In addition, the eye may also rotate about its center (O). A combination of these two types of motion is sufficient to capture all possible rigid movements of the eye in this 2D universe.

When the observer fixates on a target point (such as a corner of an object) this point – by definition – remains projected at the center of the retina, i.e., on the fovea (F in Fig. 2). In order to maintain fixation while moving, the observer has to rotate the eye about its center (O). Although the target point on which the observer fixates (T) is stationary at the fovea, the remainder of the retinal image (e.g. P) can be expected to change; this change will be precisely defined below.

The *instantaneous* movement of the retinal image will be referred to as *retinal flow*, and it will be defined in the present formalization in terms of angular coordinates. We parameterize the retinal flow of a point P as an angular velocity, i.e., the temporal derivative of an angle between two rays: (i) the direction of gaze (ray OT in Fig. 2)

¹The eye and the retina considered in this paper will only correspond to the human eye in terms of optics and not in terms of the actual physical structure.

and (ii) the ray from the point in the world to the center of the eye (OP in Fig. 2).

The retinal flow may be decomposed into two components: one due to observer *translation* and the other due to the fixating *rotation*. The first component is written as:

$$\omega_1 = \frac{|\vec{v} \times \vec{p}|}{|\vec{p}|^2} \quad (1)$$

where \vec{v} is the translational velocity of the center of the eye, \vec{p} is the vector from the center of the eye to an arbitrary point in the world P, and \times indicates a standard vector cross-product.

The second component corresponds to the fixation rotation about the center of the eye:

$$\omega_2 = -\frac{|\vec{v} \times \vec{t}|}{|\vec{t}|^2} \quad (2)$$

where \vec{t} is the vector from the center of the eye to the target point (T in Fig. 2); the direction of \vec{t} is also called the direction of gaze, or optical axis. The negative sign indicates that when the eye rotates in one direction, the points on the retina move in the opposite direction.

Finally, the resultant retinal angular velocity of a point is the sum of the two angular velocities², i.e.

$$\begin{aligned} \omega &= \omega_1 + \omega_2 \\ &= \frac{|\vec{v} \times \vec{p}|}{|\vec{p}|^2} - \frac{|\vec{v} \times \vec{t}|}{|\vec{t}|^2} \end{aligned} \quad (3)$$

Given the observer translation velocity \vec{v} , the vector representing gaze \vec{t} , and the vector from the eye to any point in the world \vec{p} , Equation (3) defines the retinal flow of that point in a 2D universe.

2.2 Level Sets of Retinal Flow

In this section we will consider the points in the 2D universe that give rise to the same value of retinal flow. Let us first isolate those points in the 2D universe that

²In general, the sum of two rotation vectors does *not* produce a rotation vector corresponding to the composition of rotations. The angular velocities may be added here because they are *instantaneous* measurements.

correspond to zero retinal flow. The projections of such points on the retina come to rest (for an instant) while the observer moves and fixates. To find such points, we set Equation (3) to zero and solve for \vec{p} :

$$\frac{|\vec{v} \times \vec{p}|}{|\vec{p}|^2} - \frac{|\vec{v} \times \vec{t}|}{|\vec{t}|^2} = 0 \quad (4)$$

Factoring the magnitude of \vec{v} from both terms results in:

$$\frac{|\hat{v} \times \vec{p}|}{|\vec{p}|^2} - \frac{|\hat{v} \times \vec{t}|}{|\vec{t}|^2} = 0 \quad (5)$$

where \hat{v} is the unit vector in the direction of \vec{v} .

Although the points in the 2D universe satisfying the above equation lie on a simple curve, reducing the solution to a recognizable form requires further vector algebraic manipulations. To this effect, let us introduce a unit vector \hat{u} perpendicular to \hat{v} ; then $|\hat{v} \times \vec{t}| = \hat{u} \cdot \vec{t}$. We can now rewrite equation 5 in terms of \hat{u} , removing \hat{v} .

$$\frac{\hat{u} \cdot \vec{p}}{|\vec{p}|^2} - \frac{\hat{u} \cdot \vec{t}}{|\vec{t}|^2} = 0 \quad (6)$$

This can be rewritten as:

$$\vec{p} \cdot \vec{p} - \hat{u} \cdot \vec{p} \left(\frac{\vec{t} \cdot \vec{t}}{\hat{u} \cdot \vec{t}} \right) = 0 \quad (7)$$

and further (by completing the square) as:

$$\left\| \vec{p} - \frac{1}{2} \left(\frac{\vec{t} \cdot \vec{t}}{\hat{u} \cdot \vec{t}} \right) \hat{u} \right\|^2 = \left[\frac{1}{2} \left(\frac{\vec{t} \cdot \vec{t}}{\hat{u} \cdot \vec{t}} \right) \right]^2 \quad (8)$$

For $r = \frac{1}{2} \left(\frac{\vec{t} \cdot \vec{t}}{\hat{u} \cdot \vec{t}} \right)$ and $\vec{c} = r \hat{u}$, we obtain the familiar equation of a circle:

$$(\vec{p} - \vec{c})^2 = r^2 \quad (9)$$

The vector \vec{c} is the center of this circle (with respect to the center of the eye). Recall that \hat{u} is a vector perpendicular to the velocity of the observer \vec{v} ; this means

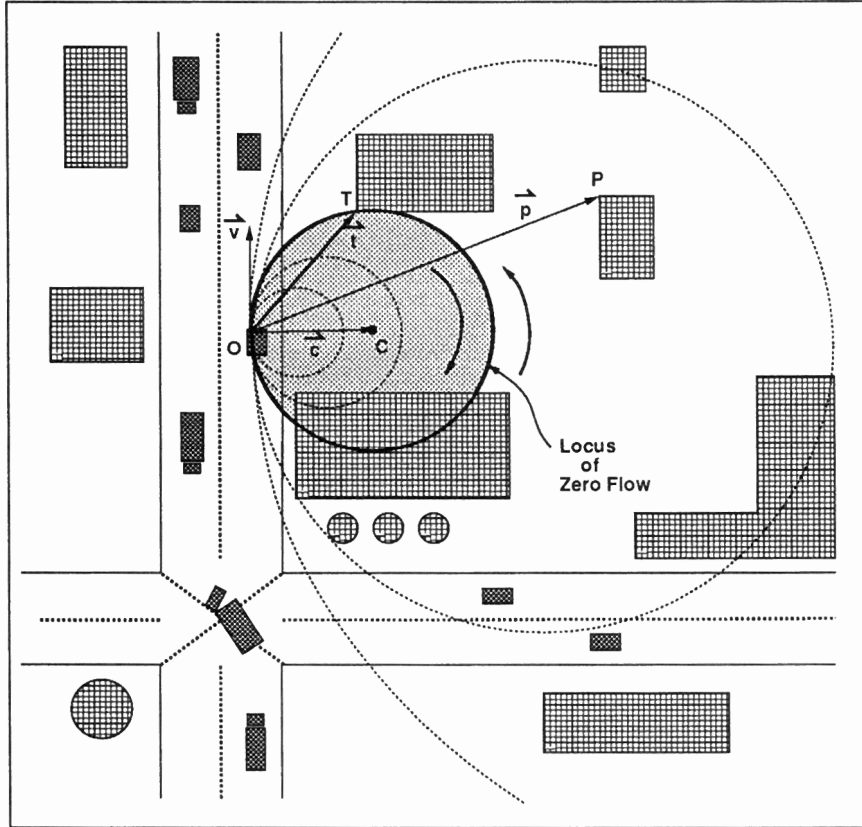


Figure 3: The level sets of retinal flow in a 2D universe. This scene depicts a traffic intersection, where the observer is moving along the road with translational velocity \vec{v} , and fixates on a corner of a building T . O is the center of the eye of the observer, and P is an example point in the universe, the retinal velocity of which is being calculated. The points with zero retinal flow lie on the solid circle with center C , which passes through the target. The other circles correspond to level sets of flow of non-zero values. Note that the centers of these circles lie on the horizontal line through O . Points in the shaded area have flow in the clockwise direction while points in the unshaded area have flow in the counter-clockwise direction.

that the center of the circle lies in a direction perpendicular to the direction of movement. This circle, corresponding to zero flow in the retina, is depicted by the solid line in Fig. 3. The circle passes through the target point of fixation and through the center of the eye. All points on the circle, including these two points, behave in the same way: momentarily, they are stationary. Furthermore, the points *within* this circle all move in the same direction on the retina, whereas the points *outside* of this circle move in the opposite direction.³

A similar analysis can be performed for any other value of retinal flow besides the zero flow. For each such value, the result corresponds to a circle of points in the 2D universe; the radius of the circle varies depending on the particular value chosen. Sample circles which correspond to points with the equal retinal flow are shown as dotted curves in Fig. 3. Note that the centers of all such circles lie on a straight line.

An interesting boundary case involves fixating straight ahead. In this case the direction of fixation coincides with the direction of observer movement. The resulting circle of zero flow has an infinite radius (i.e., it is a *line*), as illustrated in Fig. 4. In this situation, points lying to one side of this line move in one direction, while points on the remaining half plane move in the opposite direction. This result fits intuition: when looking and moving straight ahead, points on the left half of the visual field move left, and points on the right half move right.

3 Retinal Flow in a Rigid 3D Universe

The case of moving and fixating in a 3D universe is clearly more complicated than the 2D case. However, the 3D case can be elegantly decomposed into two modules: one involving the retinal flow just as in the 2D case, and the other involving a new component.

3.1 Calculating Retinal Flow

In a 3D universe, the eye corresponds to a sphere (rather than a circle) with a center O (in Fig. 5) and the retina involves a hemisphere (rather than a semicircle). As in

³The retinal flow associated with each point in the world forms a vector field or dynamical system [1] with a *separatrix* corresponding to the circle of zero flow which separates the two regions of opposite flow. This suggests an interesting connection between the present analysis and the dynamics of well known systems.

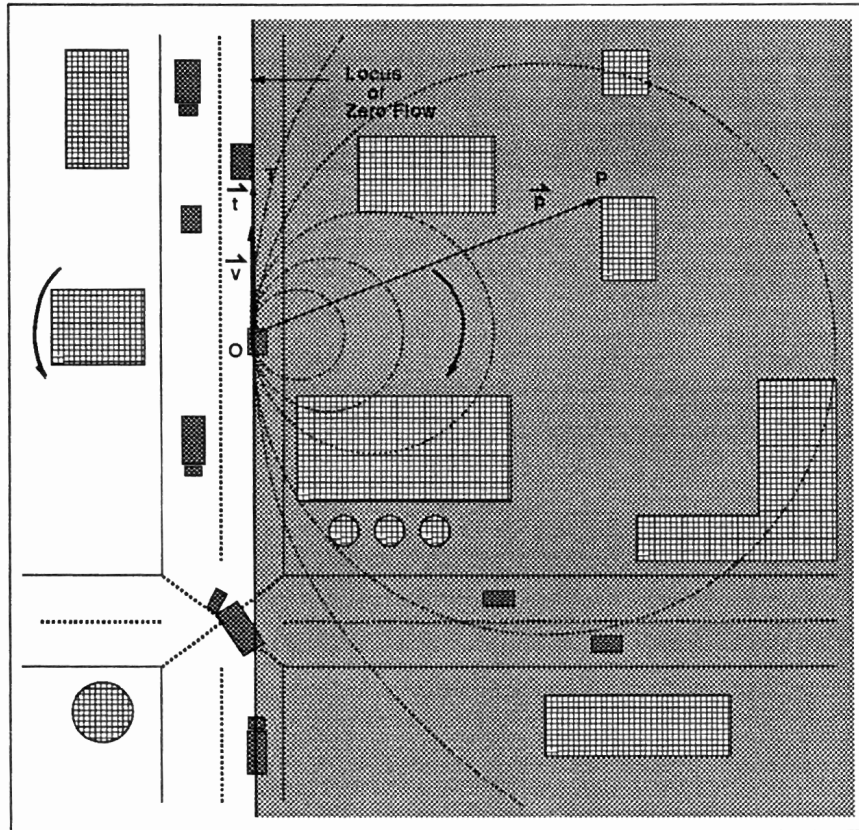


Figure 4: The special case of moving and fixating in the same direction; cf. previous figure for explanation of symbols.

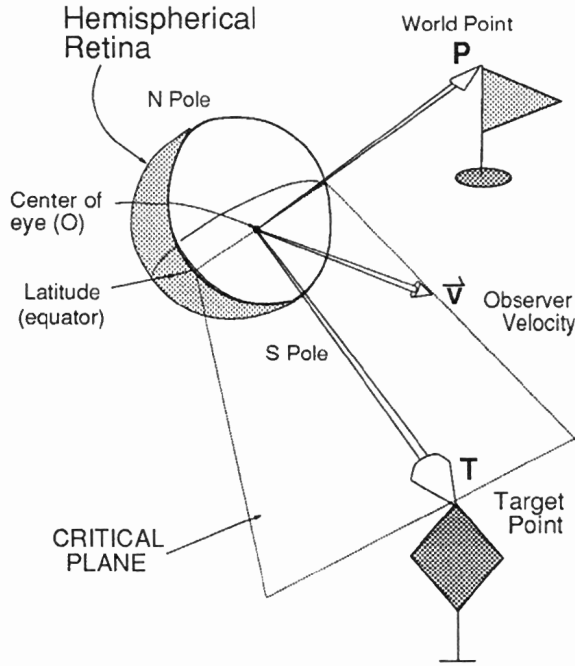


Figure 5: Model of the eye in a 3D universe. The critical plane contains the center of the eye, the target point and the velocity direction. The eye rotates about an axis passing through the North and South Poles in order to fixate.

the 2D case, the eye rotates in order to fixate on a target. However, the 3D rotation that accomplishes fixation is no longer unique.⁴

In order to make the problem manageable, we constrain the way in which the eye can rotate in order to fixate. The constraint we impose is that the axis about which the eye rotates is always perpendicular to the direction of gaze; i.e., the possible rotation axes lie on a plane. Although this is an arbitrary constraint, the physiology of the eye suggests that a similar constraint operates in humans (involving the so-called Listing's Plane [12]). Furthermore, this particular formulation of the constraint allows us to decompose the retinal flow into two components.

In order to represent the two components of retinal flow in a 3D universe we impose a grid of longitudes and latitudes on the hemispherical retina. These longitudes and latitudes are comparable to the standard grid used to specify coordinates on the earth. In the present analysis, we wish to fix this grid on the retina in a such a way that the center of the eye, the target point (that is being fixated on), and the direction of

⁴For any fixating rotation, an additional instantaneous rotation about the optical axis can be added without loss of fixation. The family of such fixation rotation axes are obtained by varying the amount of this additional rotation.

movement all fall on a plane containing the equator. This plane will be referred to as the *critical plane* (cf. Fig. 5). The other latitudes are semicircles on the retina lying on planes parallel to the critical plane. The longitudes are semicircles⁵ on the retina starting at the North Pole and ending at the South Pole. As is standard, the North Pole, center of the globe (eye) and the South Pole lie on an axis perpendicular to the equatorial (critical) plane. Note that this line is the axis about which the eye rotates in order to fixate the chosen target.

The retinal flow corresponding to points lying on the critical plane is identical to the retinal flow of points in the 2D universe, and was given by Equation (3). As before we will use angular velocity to represent retinal flow. However, in the 3D case angular velocity is no longer unique: any component of angular velocity in the direction of \vec{p} is not observable on the retina. In order to achieve uniqueness of representation, we constrain angular velocity of a point \vec{p} to be perpendicular to \vec{p} :

$$\vec{\omega} = \frac{\vec{v} \times \vec{p}}{|\vec{p}|^2} - \left(\left(\frac{\vec{v} \times \vec{t}}{|\vec{t}|^2} \right) \times \hat{p} \right) \times \hat{p} \quad (10)$$

$\vec{\omega}$ is the retinal flow corresponding to a 3D world point P , represented as an *angular velocity*. This equation is a generalization of the 2D result given in Equation (3). The double cross-products with \hat{p} in Equation (10) ensure that $\vec{\omega}$ (for any given world point) is always perpendicular to \hat{p} , the unit vector in the direction of that world point.

3.2 Points with Zero Flow in the 3D Universe

The points with zero flow are interesting boundary cases that enhance an intuitive understanding of how flow is related to positions in the world. The points that are momentarily stationary on the retina are those that have neither latitudinal nor longitudinal flow.

In order to determine the points in the world which have no retinal flow we rewrite Equation (10) as:

$$\vec{\omega} = \left[\frac{\vec{v}}{|\vec{p}|} - \frac{(\vec{v} \times \vec{t}) \times \hat{p}}{|\vec{t}|^2} \right] \times \hat{p} \quad (11)$$

⁵These semicircles are half of the so-called great circles.

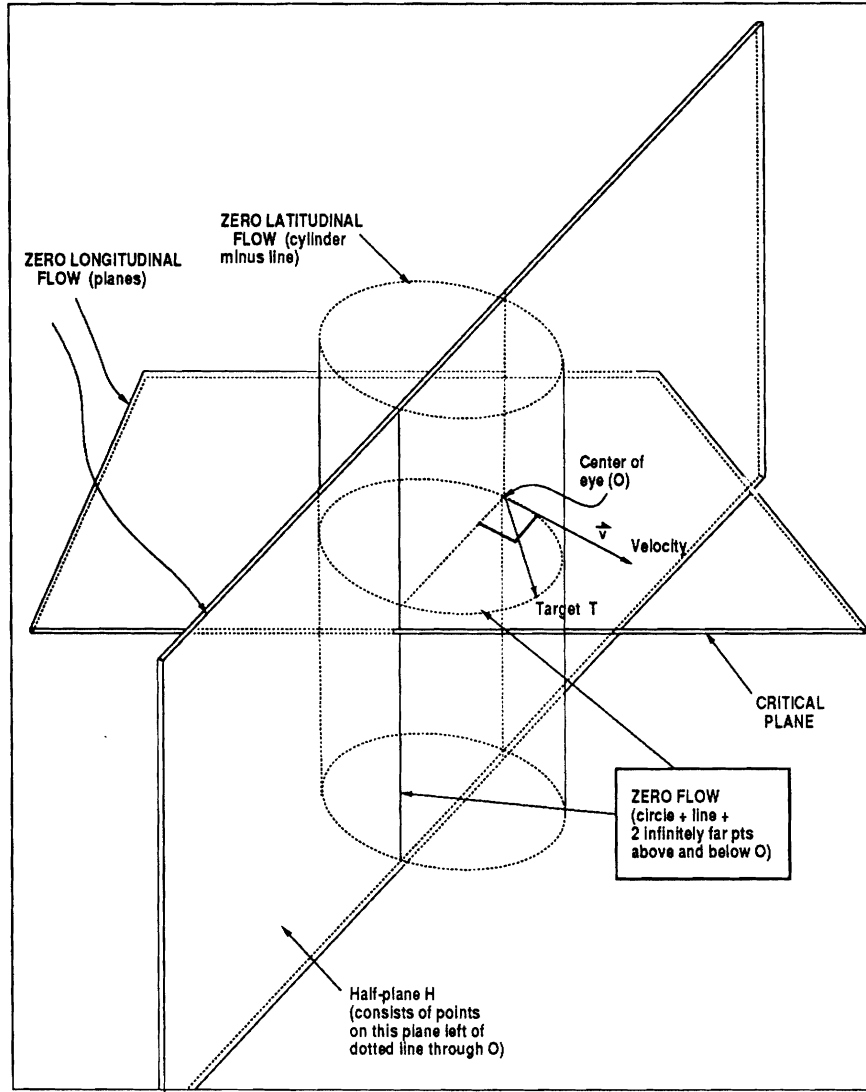


Figure 6: Points with zero flow

When \vec{v} coincides with \vec{t} , $\vec{v} \times \vec{t} = 0$ and there is no rotation due to fixation. In such a special case, all infinitely far away points have zero flow, as well as points along the line of sight. In what follows we concentrate on the more typical case with rotation due to fixation (i.e. $\vec{v} \times \vec{t} \neq 0$).

Assuming that $\vec{v} \times \vec{t} \neq 0$, the points with zero flow are those where $\vec{\omega}$ is identical to zero, i.e. when $\left[\frac{\vec{v}}{|\vec{p}|} - \frac{(\vec{v} \times \vec{t}) \times \hat{p}}{|\vec{t}|^2} \right]$ is either (a) equal to zero or (b) parallel to \hat{p} ; we shall denote this term by $\vec{m}(\vec{p})$.

Case (a): $\vec{m}(\vec{p}) = 0$

From this condition it follows:

$$\frac{\vec{v} \times \vec{t}}{|\vec{t}|^2} \times \hat{p} = \frac{\vec{v}}{|\vec{p}|} \quad (12)$$

The above vector equation holds only when the vectors on both sides of the equation have the same direction and magnitude. One solution for \vec{p} is either of a pair of points infinitely far away in the direction of the north and south pole (i.e. in the direction of $\vec{v} \times \vec{t}$).

An additional set of solutions arises may be derived as follows. The left hand side gives rise to vectors orthogonal to \vec{p} . If these vectors are to be in the direction of \vec{v} , then \vec{p} must lie on the plane perpendicular to \vec{v} (labeled H in Figure 6). Furthermore, for any given direction of \vec{p} in this plane, there is one particular magnitude that will satisfy the magnitude component of the above equation. This magnitude is such that

$$\frac{|(\vec{v} \times \vec{t}) \times \vec{p}|}{|\vec{t}|^2} = |\vec{v}| \quad (13)$$

By simple vector algebra we get

$$\hat{v}_p \cdot \vec{p} = \frac{|\vec{t} \cdot \vec{t}|}{|\hat{v} \times \vec{t}|} \quad (14)$$

where \hat{v}_p is a unit vector perpendicular to \vec{v} and $\vec{v} \times \vec{t}$. The magnitudes of \vec{p} that satisfy 14 correspond to a line parallel to $\vec{v} \times \vec{t}$ at a distance $\frac{|\vec{t} \cdot \vec{t}|}{|\hat{v} \times \vec{t}|}$ (on the plane H), as shown in Fig. 6.

Case (b): $\vec{m}(\vec{p})$ parallel to \hat{p}

The second term of $\vec{m}(\vec{p})$ always corresponds to vectors on the plane perpendicular to $\vec{v} \times \vec{t}$; i.e., these vectors lie on the critical plane spanned by \vec{v} and \vec{t} (Fig. 5). By definition, the first term of $\vec{m}(\vec{p})$ also lies on the critical plane, and hence the difference of the two terms of $\vec{m}(\vec{p})$ lies on the critical plane. If this difference vector $\vec{m}(\vec{p})$ is to be parallel to \hat{p} , it is necessary that \hat{p} lie on the critical plane. Thus, the

situation reduces to the 2D case already discussed in Section 2.2. As in the 2D case, the points with zero flow lie on a circle (defined by Equation (8)).

In conclusion, apart from the special case without rotation due to fixation, the points with zero retinal flow are:

1. the two infinitely far away points in the direction of the north and south poles,
2. the points on the line parallel to the rotation axis (Equation (14)), and
3. the circle on the critical plane.⁶

4 Latitudinal and Longitudinal Flow

4.1 Calculating Latitudinal and Longitudinal Flow

In general, the retinal flow may be decomposed into two components: latitudinal flow and longitudinal flow. It is crucial for our purposes later in the chapter to isolate the longitudinal component of the flow in Equation (10). The second term of this equation is already in the correct form: since $\frac{\vec{v} \times \vec{t}}{|\vec{t}|^2}$ is a vector perpendicular to the critical plane, the resulting flow is entirely latitudinal.

In order to perform the desired decomposition of the second term of Equation (10), we decompose the observer's velocity \vec{v} into two components: (i) \vec{v}' , which is the projection of \vec{v} onto a plane containing P, the North Pole and the center of the eye (this plane will be referred to as the *longitudinal plane*) and (ii) \vec{v}'' , which is perpendicular to the longitudinal plane, such that:

$$\vec{v} = \vec{v}' + \vec{v}'' \tag{15}$$

Since \vec{v}' lies in the longitudinal plane, it generates flow only within that plane (i.e., longitudinal flow). On the other hand, \vec{v}'' is orthogonal to the longitudinal plane, and thus produces flow out of that plane (i.e., latitudinal flow).

⁶Under a different formulation, Raviv and Herman [8] identify the line and the circle as involving zero flow. However, our decomposition in the remainder of this paper is unique and distinct from Raviv and Herman. The level sets of our latitudinal flow do not correspond to any of their level sets.

Thus, $\vec{\omega}$ can be decomposed into $\vec{\omega}_x$, the latitudinal flow, and $\vec{\omega}_y$, the longitudinal flow, as follows:

$$\vec{\omega}_x = \frac{\vec{v}'' \times \vec{p}}{|\vec{p}|^2} - \frac{\vec{v} \times \vec{t}}{|\vec{t}|^2} \times \hat{p} \times \hat{p} \quad (16)$$

$$\vec{\omega}_y = \frac{\vec{v}' \times \vec{p}}{|\vec{p}|^2} \quad (17)$$

As Equation (17) shows, the longitudinal flow of a point P depends only on the movement of the observer. The direct relationship between the observer motion and longitudinal flow is exploited in this paper; this relationship is a result of the particular choice of the latitudes and longitudes in this formalization.

4.2 Points with Zero Longitudinal Flow in the 3D Universe

Although we have already obtained expressions for the world points that produce zero flow, we can gain additional intuition by considering those points that generate either zero latitudinal flow or zero longitudinal flow.

In the case of longitudinal flow, all points with zero flow lie on either of two planes (cf. Figure 6). One such plane is the critical plane. All points on the critical plane project onto a single latitude, the equator. Any translation within or rotation perpendicular to the critical plane will not induce the points to change latitudes; i.e. the points remain on the equator regardless of observer motion or fixation. Since the points remain on the same latitude, they have zero longitudinal flow.⁷ The second plane with zero longitudinal flow is perpendicular to the direction of the observer's velocity \vec{v} and passes through the center of the eye. For points on this plane, the modified velocity (\vec{v}' in Equation (17)) is zero, resulting in no longitudinal flow.

4.3 Points with Zero Latitudinal Flow in the 3D Universe

In the case of latitudinal flow, all points with zero flow lie on a cylinder but not all points on the cylinder have zero flow. The points with non-zero flow on this cylinder

⁷For points on the critical plane, note that \vec{v}' and \vec{p} are in the same direction, leading to a zero cross-product term in equation 17.

form an open interval line passing through the north and south poles (see Fig. 6). This can be shown by first writing an explicit expression for \vec{v}'' :

$$\vec{v}'' = \frac{[\vec{v} \cdot ((\vec{v} \times \vec{t}) \times \hat{p})](\vec{v} \times \vec{t}) \times \hat{p}}{|(\vec{v} \times \vec{t}) \times \hat{p}|^2} \quad (18)$$

After substituting this expression for \vec{v}'' into Equation (16) and performing some simple algebraic manipulations, we arrive at the following constraint on the world points \vec{p} with zero latitudinal:

$$\frac{\vec{v} \cdot [(\vec{v} \times \vec{t}) \times \vec{p}]}{|(\vec{v} \times \vec{t}) \times \vec{p}|^2} - \frac{1}{|\vec{t}|^2} = 0 \quad (19)$$

Note that in the above equation only that component of \vec{p} which is orthogonal to $\vec{v} \times \vec{t}$ matters. Hence the circle of zero flow on the critical plane extends to a cylinder of points above and below the critical plane.

5 A Systematic Pattern at the Periphery

We will show that when a target is fixated on, the retinal *periphery* has a unique invariant property: it is the only longitude that is a constant across all possible rotations of the eye. This makes the retinal periphery an interesting location to look at for certain visual tasks. We define the periphery to be that longitude (or great circle) lying in a plane perpendicular to the direction of gaze.

Let us consider a situation where the moving observer has to decide whether he/she is heading towards the fixated target or not. In the former case, the observer will *hit* the target if he/she *continues* in the current direction of motion, whereas in the latter case the observer will *miss* the fixated target.⁸ Such an ability to predict *hit* and *miss* situations (assuming that the current direction of movement is maintained, and assuming that the target does not move) should turn out to be useful in navigation.

An analysis of the retinal flow at the periphery of our model eye indicates that the characteristics of the longitudinal flow distinguish *hit* from *miss* situations (in the sense described above). The magnitude of the longitudinal flow on the periphery

⁸This holds for an idealized case where the observer is a point. For a practical situation in which the observer has finite size, determining a *hit* situation is more involved.

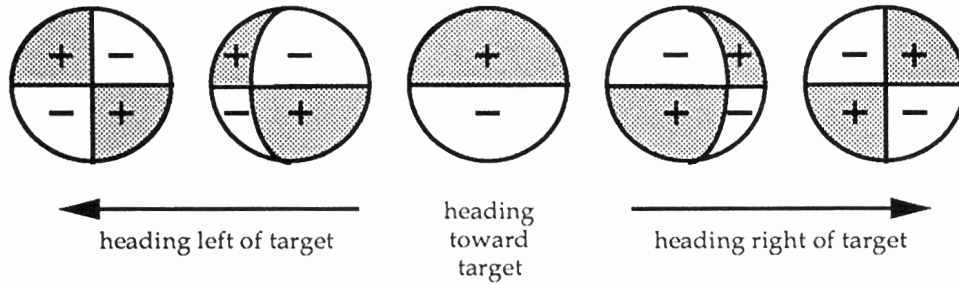


Figure 7: Sign of longitudinal flow on the hemispherical retina, projected (for illustration purposes) onto the plane perpendicular to the viewing direction. The pattern at the center depicts the *hit* situation and the four other patterns depict *miss* situations. From left to right the directions of heading are: -90° -45° , 0° (direct ahead), 45° and 90° .

(from a point P in the world) depends on how far the point is from the eye as well as how fast the observer moves. On the other hand, the direction – or sign – of the longitudinal flow *within a quadrant of the retina* only depends on whether the observer is heading towards the target. The quadrants are defined by the location of the North and South Poles on the periphery.

In the *miss* situation, the sign of the longitudinal flow (sign of $\vec{\omega}_y$ in Equation (17)) switches exactly four times as one traces along the periphery (i.e., once for each quadrant). Figure 7 indicates the sign of the longitudinal flow along the periphery as well as across the entire retina for the two most extreme *miss* situations as well as two other intermediate ones.

Although Equation (17) could be analyzed to derive the number of sign changes, we will turn to Figure 6 for a more intuitive proof. The two planes of zero longitudinal flow divide the entire universe into four quadrants; within each quadrant the longitudinal flow is of the same sign, whereas sign switches at the planes. For example, in the quadrant spanned between the observer’s direction of heading and the North Pole the longitudinal flow is towards the North Pole, whereas in the quadrant on the other side of the critical plane (and on the same side as the observer’s direction) the longitudinal flow is towards the South Pole. When the observer is not moving directly ahead, these two planes will cut the hemispherical retina into four quadrants, thereby producing the four-way distinction in the sign of the longitudinal flow.

As we continuously move from a *miss* situation towards a *hit* situation, the translation direction gets closer to the direction of gaze. When the translation direction

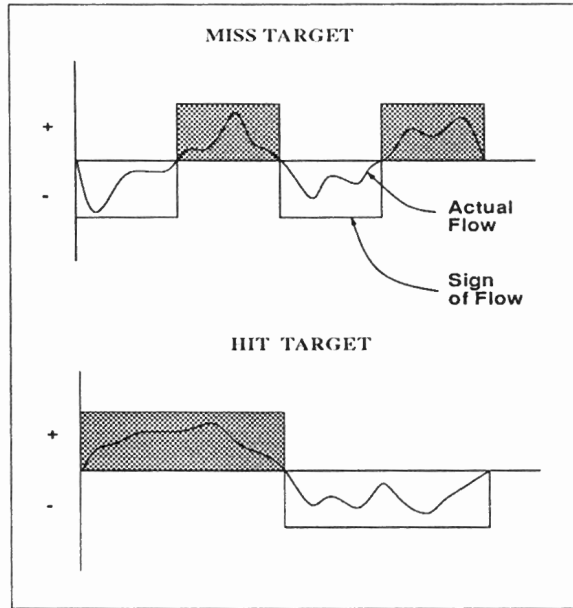


Figure 8: The sign of the retinal flow at the periphery. (See text for description of “at the periphery”.)

and the direction of gaze coincide, one of the planes with zero longitudinal flow (in Fig. 6) contains the entire periphery. Thus the longitudinal flow of all the points on the periphery is zero in the *hit* situation. Immediately adjacent to this infinitely thin periphery, a different picture emerges. As one traces the longitudinal flow just inside of the periphery, the sign of the longitudinal flow changes *exactly twice*; see central pattern in Figure 7.

Thus, in theory, the infinitely thin line of the periphery will contain the four-way change for the *miss* situation and zero flow for the *hit* situation. However, in practice, the observable difference at a periphery with finite thickness involves a four-way change in the *miss* situation and a two-way change in the *hit* situation. This situation is illustrated in Fig. 8.

6 Experiment I: Simulated Image Sequence

In a simulated experiment we attempted to obtain the sign of the longitudinal flow on the retina as in Fig. 7. The number of times the sign changes along the periphery (two vs. four times) allows us to distinguish between the *hit* and the *miss* situations.

In the simulation we created a rigid world consisting of 1024 points. The points were uniformly and randomly distributed in all directions around the initial position

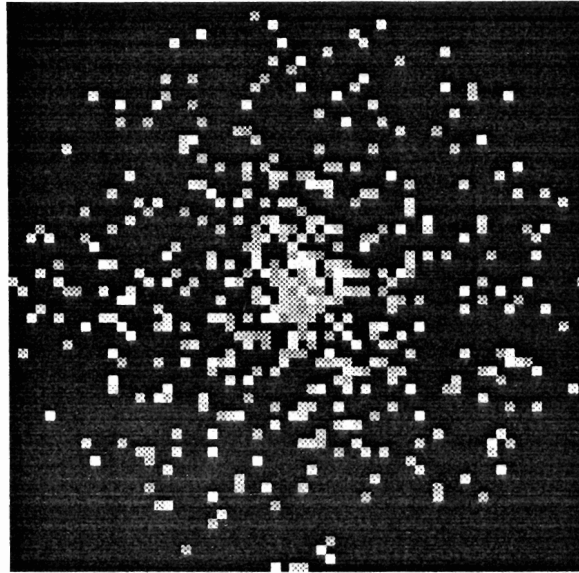


Figure 9: The first image of the image sequence generated in the simulation. The world consists of 3D points randomly distributed in position (cf. text for details). During the simulated motion, the camera was made to fixate on the same dot that projects to the center of the image.

of the eye. The points were also randomly (but nonuniformly) distributed in distance, with more points closer to the simulated eye.⁹ Due to the hemispheric nature of the retina only those points in front of the eye were projected on the retina; the visual field was 180° along any diameter of the retina.

Two pairs of images were generated containing the projections of points in the world onto the retina. The target on which the eye fixated was located at a distance equal to thrice the radius of the eye, directly ahead of the initial position of the eye. The two pairs contained the same first image: a view from an initial eye position. The second image of the first pair was generated to simulate a sideways movement of one fourth of the radius of the eye (*miss* situation). The second image of the second pair simulated a movement of one fourth of the radius of the eye towards the target (*hit* situation) from the initial position. The two image sequences are shown in Fig. 9.

In order to test the sign of the longitudinal flow on the retina, the optical flow in both image sequences was estimated, using a multi-scale gradient-based algorithm described in [9] [10]. The longitudinal flow was then extracted from the total optical

⁹The distance (d) of the points was chosen randomly such that $\frac{1}{d}$ was uniformly distributed in the range $[0, 0.5]$.

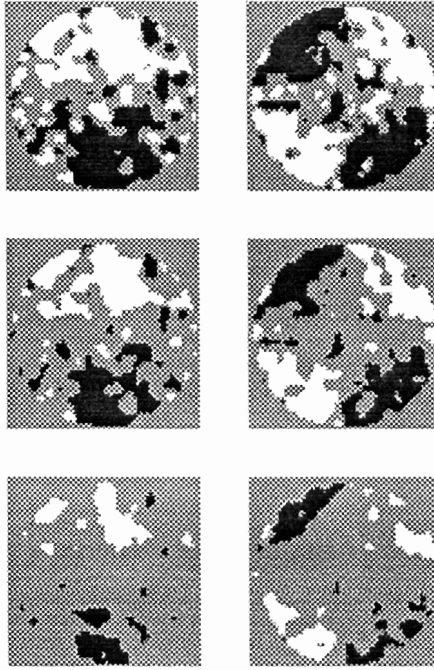


Figure 10: Simulated Experiment: Sign of the Longitudinal Flow for *hit* (left) and *miss* (right) situations. In the first row the points with less that 0.125 pixels of longitudinal flow were ignored. The comparable figure for the second row was 0.25 pixels and for the third row was 0.5 pixels.

flow.¹⁰

Fig. 10 shows the sign of the longitudinal flow across the entire retina for the two sequences. The longitudinal flow in one direction is shown in white, and in the opposite direction in black. For the *hit* situation these regions approximately divide the retina into two halves – shown on the left in Fig. 10 – while in the *miss* situation the retina is divided into quarters, shown on the right. Each row corresponds to a different range of values of longitudinal flow that was ignored (shown grey in figure).

Although the overall pattern of the sign of the longitudinal flow on the simulated retina is in agreement with the theoretical prediction (cf. Fig. 7), there are certain regions on the retina that have an unexpected sign. We attribute this error due to imperfect optical flow; a set of simulations with perfect optical flow obtained perfect patterns as shown in Figure 11. The errors present in the optical flow estimates are due to blank regions of the images and undersampling in time.

¹⁰The optical flow contained both the latitudinal and longitudinal flow. For the purposes of this demonstration only the longitudinal flow was needed. Simoncelli’s optical flow algorithm was easily modified to obtain the longitudinal component of flow.

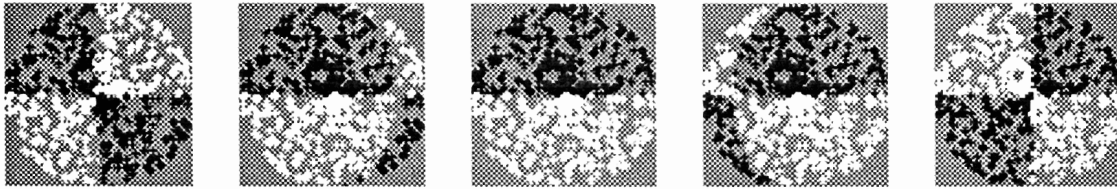


Figure 11: Simulated Experiment with perfect optical flow. The sign of the longitudinal flow is depicted for different directions of heading. From left to right, the directions of heading are -90° , -45° , 0° (direct ahead), 45° and 90° .

7 Experiment II: Real 180 Degree FOV Image Sequences

This experiment involves the first truly 180 degree field of view camera used in computer vision. Recall that the camera–lens system and a typical image obtained by the camera are depicted in Figure 1.

The lens system is 50 cm long and consists of a fish–eye lens at the front, followed by re–imaging optics. At the rear of the lens system is mounted a standard Sony CCD camera. The image obtained on the camera is modeled by the so–called Equidistant Projection [5]. In order to define this projection we will employ spherical coordinates for points in the 3D world, and polar coordinates for points on the image. Let us consider a 3D point P in the world at a distance d from the center of the fish–eye lens; let the ray to the point from the origin make an angle θ with the optical axis, and let the projection of the ray onto the plane normal to the optical axis make an angle ϕ w.r.t to some arbitrary direction on the plane (say, the x–axis). If the polar image coordinates of the point P in the 2D image are (r, α) , then these coordinates according to Equidistant Projection are:

$$\begin{aligned} r &= c\theta \\ \alpha &= \phi \end{aligned} \tag{20}$$

where c is a constant.

Since the camera–lens system is large in dimensions, mounting it onto the turntable (in order to perform fixation) on the mobile robot is not a trivial problem as is being worked on. The following experiment instead involves manual motion of the camera on a planar workspace using guides that ensured linear camera motion. The camera

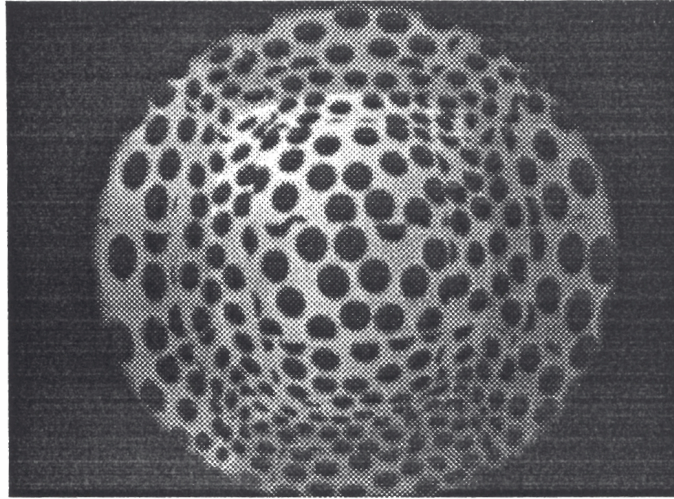


Figure 12: A typical image obtained from the 180 degree FOV camera. The camera moved inside a box whose walls were pasted with white paper containing black dots. The texture was provided to ensure good performance of the optical flow algorithm so that the errors in the pattern due to noisy flow would be minimal.

motion – apart from ubiquitously present small rotations – is comparable to fixating at a target at infinity. Further, since optical flow is typically noisy we presented a sufficiently well textured environment of black dots on a white background stuck on the walls of a box; one typical image is depicted in Figure 12.

The camera was moved in five directions as described in Figure 11: sideways (left and right), diagonally (left and right) and directly forward. A sequence of three images was collected for each direction of motion, and the same optical flow algorithm used in Experiment I (due to Simoncelli) was applied. Figure 13 depicts the sign of the optical flow along the longitudes. The results are a strong corroboration of the theoretically predicted patterns (cf. Figure 7), and look like the patterns obtained using perfect optical flow in the simulated experiments (cf. Figure 11).¹¹

Note that the patterns obtained here using real images seem better than those obtained from the simulations. The main reason is that the simulated world consisted of a sparse sets of points, and optical flow algorithms are noisy on images with such sparse texture. In contrast the 180 degree images were obtained from a real world with good texture and lighting conditions, thereby giving rise to better flow.

¹¹Note that in Figure 11 the optical flow is sparse, since the simulated world consisted of a finite number of randomly located 3D points.



Figure 13: Real 180 Degree FOV Image Sequences. The sign of the longitudinal flow is depicted for different directions of heading; compare with the figure in the simulated experiment (Fig. 11). From left to right, the directions of heading are -90° , -45° , 0° (direct ahead), 45° and 90° .

8 Experiment III: Servoing to a target

This experiment demonstrates that the sign of even noisy retinal flow is sufficient to determine the direction of heading of a moving robot that fixates on a target. The task is to guide the robot toward the target from any starting position and velocity, using only visual information. As the robot moves, its direction of heading is continuously computed based on the sign of longitudinal flow. Based on the current direction of heading, the robot’s direction of motion is continuously changed until it heads directly towards the target.¹² It is assumed that there are no obstacles, since obstacle avoidance is beyond the scope of this work.

The experiment reported in this section was done prior to the acquisition of the 180 degree FOV lens system. We are currently in the process of installing the 180 degree lens system onto the mobile platform. However, in order to simulate a wide field of view, we employed two cameras in the present work : (i) a camera pointed at the target, which we denote as the *fovea*, and (ii) a large field-of-view camera pointed upwards in a direction orthogonal to the foveal camera, which we denote as the *peripheral* camera. Note that although the second camera obtains a view of just a portion of the periphery, this portion is sufficient to calculate the direction of heading (as indicated by e.g. the top portion of the different situations in Figure 11). The two cameras are held rigid with respect to each other.

The purpose of the foveal camera is to track the target such that the target is always in the center of the image. While the robot is in motion, and the target moves from the center of the image, the fovea is rotated to compensate for the target motion,

¹²Note that fixation corresponds merely to having the robot *look* in the direction of the target. We must determine which way the robot is heading relative to the target. Although such information could potentially be obtained through positional encoders (corresponding to vestibular or muscular feedback in humans) the model developed in this paper solves for direction of heading using visual information alone.

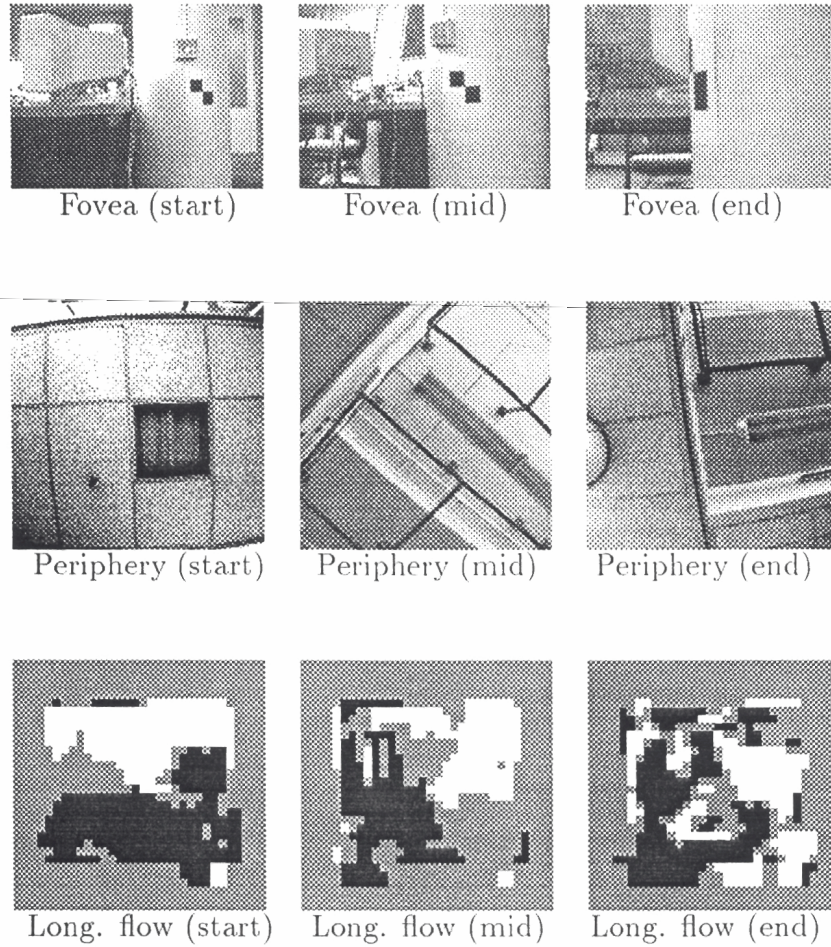


Figure 14: Each column of images corresponds to one position in the robot's path towards the target (a black-and-white checkerboard pattern hung on a pillar). The first column is at the robot's starting position, the last column is at the robot's finish position and the middle column is about midway between start and finish. At each position, the first row is the foveal image, the second row is the peripheral image. The third row is the sign of the longitudinal flow in the periphery image: these images are used to move the robot towards the target. The robot starts off heading in a direction perpendicular to the direction of the target.

thereby achieving fixation. The fovea is mounted on a turntable whose stepper motors are constantly servoed using the amount of slip of the target from the image center. The first column in Figure 14 shows the initial position of the robot, and the target being fixated. The target in this case is chosen to be a simple pattern that can be easily tracked in order to maintain fixation. However, since the fixation is performed by a course-to-fine correlation scheme, the target could be any arbitrary textured pattern.

The sign of the longitudinal flow provided by the peripheral camera is used to compute the direction of heading. Recall that for this particular camera arrangement, the peripheral camera provides sign of flow only in a circular patch around the North Pole. The problem of finding the direction of heading is reduced to the problem of locating the line that divides the positive flow from the negative flow at the North Pole; as was seen in Figure 11 this line rotates about the North Pole. In practice, the orientation of this line is computed as the phase of the first harmonic of the flow pattern. As is standard, the phase is computed by first projecting the pattern onto two functions, a sine and a cosine, and then calculating the arctangent of the ratio of the two projections.

The third row in Figure 14 shows the sign of the longitudinal flow at three different positions of the robot along its trajectory towards the target (white is one direction and black the opposite direction of sign). The first row of Figure 14 shows the images obtained from the fovea and the second row from the periphery at the three robot locations. The computed value of the direction of heading (w.r.t to the direction of gaze/target) at any given position of the robot is used to servo the robot to move towards its direction of gaze. Note that the last row of Figure 14 is noisy: the pattern does not consist of uniform white and black areas. This is expected since flow was computed at only one scale (as opposed to multi-scale), and since the scene texture was unfavorable (e.g. the blank ceiling). In the previous experiment with the 180 degree FOV images, the flow is extremely good owing to the controlled lighting, good texture and multi-scale flow algorithm.

The current implementation of the system is in C running on a Sun Workstation. Each run of the experiment takes under 2 minutes, and the initial distance from the robot to the target is 3.1 meters. The experiment was repeated 10 successive

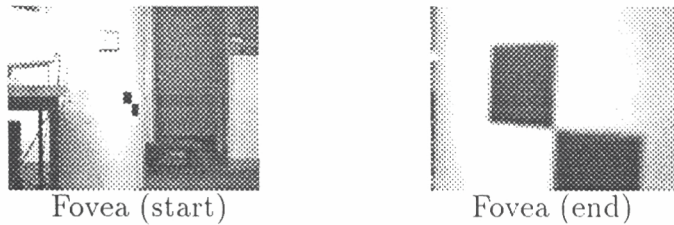


Figure 15: Foveal images for a typical uninterrupted run of the robot, where the robot reaches the target.

times. In every trial the robot started by heading in a direction perpendicular to the target direction. In two of the ten trials the robot arrived within 40 cm of the pillar containing the target; in the remaining eight trials the base of the robot ended up in contact with the pillar, directly underneath the target. Figure 14 is an example of a case where the robot's base does not touch the pillar. Due to the fact that the robot was stopped in the middle of its run in order to take the mid pictures in Figure 14, the servoing algorithm in effect had to start again, resulting in overshooting. Pictures from the fovea of a typical uninterrupted sequence – where the robot base touched the pillar – are provided in Figure 15.

9 Conclusion

In this paper we have formulated the precise relationship between retinal flow, the movement of a fixating observer and the geometry of the physical world in a simple and systematic way. The simplicity of the formulation derives from the use of a spherical imaging surface, coupled with the constraints imposed by fixation.

Our theoretical analysis and subsequent experiments reveal that the information along the periphery of the retina appears to be sufficient for determining whether the observer will eventually hit a target if it continues moving in its current direction (assuming the target will not move). This is possible precisely because the observer actively fixates on the target while moving. We believe this formulation will prove useful in many problems involving optical flow computation. Finally we have demonstrated strongly corroborating results from the first 180 degree FOV camera employed in motion research as well as in computer vision.

Acknowledgments:

The first author is indebted to Prof. V. Aurich, Computer Science/Univ. of Düsseldorf, for triggering this research by bringing to the author's attention interesting patterns on the retinal periphery. Thanks to U. Cahn von Seelen for the tracking algorithm, to J. Kosecka for the robot control program, and to Dr. A. Vainikka who helped with the style and logic of presentation, This work was supported by ARPA Grant N00014-92-J-1647, ARO Grant DAAL03-89-C-0031PRI and NSF Grant CISE/CDA-88-22719 and NSF Grant STC SBR8920230.

References

- [1] R. H. Abraham and C. D. Shaw. *Dynamics: The geometry of behavior*. Addison-Wesley, Redwood City, CA, 1993.
- [2] R. Bajcsy. Active perception vs. passive perception. In *Proc. IEEE III Workshop on Computer Vision: Representation and Control*, pages 55–59, Bellaire, MI, October 1985.
- [3] R. Bajcsy. Active perception. *Proceedings of the IEEE, Special Issue on Computer Vision*, 76(8):996–1005, August 1988. Invited Paper.
- [4] C. Fermueller. Navigational preliminaries. In Y. Aloimonos, editor, *Active Perception*, chapter 3, pages 103–150. Lawrence Erlbaum, Hillsdale, NJ, 1993.
- [5] M. M. Fleck. Shape and the wide-angle image. 1994. Technial Report 94-04, Department of Computer Science, University of Iowa. Iowa City.
- [6] R. Nelson and J. Aloimonos. Finding motion parameters from spherical motion fields (or the advantages of having eyes in the back of your head). *Biological Cybernetics*, 58:261–273, 1988.
- [7] K. Pahlavan, T. Uhlin, and J. O. Eklundh. Dynamic fixation. In *Prof. Fourth International Conference on Computer Vision*, pages 412–419, 1993.
- [8] D. Raviv and M. Herman. Towards an understanding of camera fixation. In *Proc. 1990 IEEE International Conference on Robotics and Automation*, pages 28–33, Cincinnati, OH, 1990.

- [9] E. Simoncelli. *Distributed Representation and Analysis of Visual Motion*. PhD thesis, Department of Electrical Engineering, Massachusetts Institute of Technology, Cambridge, MA, 1993. Vision and Modeling Group Technical Report 209, MIT Media Laboratory.
- [10] E. P. Simoncelli, E. H. Adelson, and D. J. Heeger. Probability distributions of optical flow. In *IEEE Conference on Computer Vision and Pattern Recognition*, Maui, Hawaii, June 1991.
- [11] I. Thomas. *Reducing Noise in 3D Models Recovered from a Sequence of 2D Images*. PhD thesis, Department of Computer Science, University of Massachusetts, Amherst, MA, 1993.
- [12] A. Yarbus. *Eye Movements and Vision*. Plenum Press, New York, 1967.