NISTIR 4865

Binocular Spherical Disparity: A Study in Representation for a Forward Translating Camera

Don J. Orser
Sensory Intelligent Group

*U.S. DEPARTMENT OF COMMERCE
Technology Administration
National Institute of Standards
and Technology
Robot Systems Division
Bldg. 220 Rm. B124
Gaithersburg, MD 20899

NIST

NISTIR 4865

Binocular Spherical Disparity: A Study in Representation for a Forward Translating Camera

Don J. Orser
Sensory Intelligent Group

*U.S. DEPARTMENT OF COMMERCE
Technology Administration
National Institute of Standards
and Technology
Robot Systems Division
Bldg. 220 Rm. B124
Gaithersburg, MD 20899

June 1992

U.S. DEPARTMENT OF COMMERCE
Barbara Hackman Franklin, Secretary

TECHNOLOGY ADMINISTRATION
Robert M. White, Under Secretary for Technology

NATIONAL INSTITUTE OF STANDARDS AND TECHNOLOGY
John W. Lyons, Director

BINOCULAR SPHERICAL DISPARITY: A STUDY IN REPRESENTATION FOR A FORWARD TRANSLATING CAMERA

Don J. Orser

Robot Systems Division
Manufacturing Engineering Laboratory
National Institute of Standards and Technology
Gaithersburg, Md 20899

ABSTRACT

The problem of interpreting optical flow and binocular disparities for a forward translating camera is addressed. A solution is offered in the form of image remappings which convert the images to the analogous well understood case for a laterally translating camera.

After reviewing this latter case, a binocular camera-retina imaging model utilizing spherical projection and foveal peripheral resolution is described for analyzing both binocular disparity and optical flow. The result provides the basis for analyzing both types of disparities within a single framework for the purpose of understanding how these "orthogonal" sources of information can be exploited in a computational model.

The process of image remapping, called "normalization," is then defined for four 1-D parameterizations of 3-D space: range, depth, looming and clearance. These latter parameterizations are based on work by Raviv. It is shown that normalization transforms optical flow into a form analogous to that for the laterally translating camera. In addition, it is shown how to obtain these same normalizations from standard planar projection images.

A binocular wire frame scene simulator is used to experimentally verify the ideas. In addition a program for normalizing real iconic planar projection imagery is applied to several example images and the results demonstrated.

We conclude that while the spherical projection model has advantages for purposes of analysis, computational equivalency can be had using planar projection images.

BINOCULAR SPHERICAL DISPARITY: A STUDY IN REPRESENTATION FOR A FORWARD TRANSLATING CAMERA

Don J. Orser

Robot Systems Division
Manufacturing Engineering Laboratory
National Institute of Standards and Technology
Gaithersburg, Md 20899

TABLE OF CONTENTS

"If sight were to deceive us as to the position and distance of external objects, we should at at once become aware of the delusion on attempting to grasp or approach them. This daily verification by our other senses of the impressions we receive by sight produces so firm a conviction of its absolute and complete truth that the exceptions taken by philosophy or physiology, however well grounded they may seem, have no power to shake it."

Herman von Helmholtz, 1869

1. INTRODUCTION

This report describes research carried out at the Robot Systems Division of the National Institute of Standards and Technology (NIST). This research has had as its objective the development of a camera-retina imaging model motivated by considerations going beyond those found in the standard monocular planar projection model. This research is part of a larger effort of the Robot Systems Division whose goal is the creation of a general purpose low-level artificial vision system incorporating both pre-attentive and attentive components.

1.1. STRUCTURE AND SUMMARY OF THIS REPORT

The work reported on in this report addresses the problem of extracting and interpreting binocular and optical flow disparity for a forward translating camera. The problem is looked at as one of iconic image representation, and in particular, the question of whether spherical projection provides any additional utility over conventional representations is addressed.

In section 2, optical flow is reviewed, and the problem of two dimensional extraction and interpretation of optical flow disparity for a laterally translating camera viewing a dynamic scene described in terms of the aperture problem. An algorithm is described which provides for segmenting the image into regions of differential rigid body motion.

Section 3 defines a binocular spherical projection model for which both optical flow and binocular disparity can be characterized. Foveal-peripheral resolution is introduced and shown to provide as a by product a representation, the logarithmic isometric plane, which transforms the radial optical flow for a forward translating camera into a form analogous to that for a laterally translating camera. Hence the techniques and algorithm for that case, as described in section 2, become applicable for converting optical flow into range. (In this report, we will refer to the Euclidean distance to a point as range, and will use the term depth to refer to the forward component of range.)

This concept is extended to three other 1-dimensional parameterizations of 3-dimensional space: depth, looming and clearance and is based on work by Raviv [RAVIV3].

In the last half of section 3, the relationship between spherical projection and planar projection is used to derive the mathematical relationships for computing the logarithmic isometric representations from planar projections. Hence the question of whether spherical projection has some intrinsic merit over planar projection is answered in the negative.

Section 4 contains a description of a binocular spherical projection wire frame scene simulator modeling the binocular spherical projection model developed in the first half of section 3. Several examples of its use are given. It generates both optical flow and binocular disparity in several representations.

In section 5, several experiments, utilizing the simulator, and demonstrating the theoretical correctness of the ideas developed in section 3 are described. In addition, sample computations mapping planar projection images to the logarithmic isometric plane are given.

A summary of applications and conclusions is given in section 6, which is readable by itself.

1.2. BACKGROUND

The mental reconstruction of the physical world via the sense of vision has been the source of wonder and speculation among philosophers, scientists and young children from time immemorial. The advent of the computer has turned this passive speculation into creative action directed toward the creation of electro-optical devices capable of performing this act of "vision". This activity has gone on for some thirty-five years, and while some success has been achieved in specialized tasks, markedly little progress has been obtained, either in understanding the principles by which our own vision system operates or in our attempts at creating an artificial vision system.

Progress has been made, however, in understanding the naivete of some of the original attempts in computer vision. With no regard for incorporating perspective, image to image correspondence or other basic components of an imaging model, optimistic researchers imagined themselves the beneficiaries of a tireless omniscient processor of visual information, but were in fact forced to retreat to minimal claims for overly complicated systems performing simple tasks under highly controlled conditions.

While we make no claim to special knowledge, we do have the advantage of twenty-twenty hindsight directed at past attempts at a general purpose artificial vision system. In doing this, one of the conclusions one comes to is that past progress has in large part been linked to advances in the model used to describe the image formation process and its implicit geometry, e. g., the incorporation of a projection model, the realization that an image sequence should be the source of input and not isolated frames, etc.

We suggest that additional elaboration of the geometry of imaging and initial low-level iconic image representation will result in even more progress. While elaboration of vision tasks and image processing architectures is important, it is the detailed understanding of the precise nature of the information potentially available to us in an image sequence which sets the bounds within which these subsequent activities must operate.

This report will address these issues within the context of low level vision processing, and in particular, will address the issues in the context of the extraction and geometric interpretation of disparity in binocular image sequences.

Initially, those working toward the goal of an artificial vision system took their ideas from the emerging field of computer science and were for the most part preoccupied with applying techniques which were successful in related applications of that technology. In particular, digital image processing [PRATT], the successful use of the computer for performing image restoration, enhancement, bandwidth compression, etc., was the precursor of many techniques applied to the generally unappreciatedly more difficult problem of computer vision. Success, for the most part, was achieved when the ultimate consumer of the image was a human perception system, but the automation of this last step has yet to be realized.

Historically, the viewer of a "picture" was a human being using a visual processing system adapted to survival in a physically hostile world. Because it was the drawer's intent that these pictures be processed by a human visual system, they took on the specific and peculiar geometry dictated by the human imaging system. This system was adapted to the perception of rigidity of 3-D shape from continuously changing 2-D retinal projections. As a result, cave drawings, and the subsequent mathematical description of planar projection during the Renaissance by Alberti and its application to the perfection of this geometry, has created a momentum which has expressed itself in the form of the modern camera. As a result, the engineering design decision of how imagery is to be formed and represented has, for the most part, been made by default through the use of cameras emulating the geometry arrived at for a completely different set of circumstances. We believe that this decision needs to be reexamined and consciously made against a backdrop of other possibilities.

For those interested in exploring this decision, it is of interest to note that attention was paid to this by intellectual heirs of the first cave painters, the Cubists (and others), during the first part of this century. This attention, in part, resulted from the needs of the creative artist to come to terms with the burgeoning popularity of photography, and in fact was a demonstration of the limitations of the planar projection camera. Alternative choices have been popularized by Picasso and others, albeit usually to the mystification of the general public.

In fact, the issue addressed by the Cubists is one of central importance to the artificial vision research community: What is the proper relationship between invariant solid shape and a 2-D representation of it, so as to maximize the predictive capacity of the 2-D representation under a set of rules devised by the artist [GOMBRICH]?

More recently, the experimental psychologist J. J. Gibson [GIBSON] has addressed this question from the biological information processing point of view: How is it that organisms react to the visually invariant objective properties of the physical world rather than to their constantly changing retinal projections? His introduction of the concept of visual flow enabled him to provide a qualitative model in which the cues for the perception of e. g., solid shape, are contained in the invariance properties of the relationship between the 3-D visual world and their ever changing retinal images.

The implication is that the organism creates an internal model of the 3-D portion of the world in its visual field, and via the predictive capacity of that model, is able to provide a temporally invariant interpretation of those fleeting retinal images consistent with the world. Although Gibson's work is lacking in quantitative detail, his concept of visual, or optical flow, and the cues available from it has provided the inspiration for much recent work in visual motion understanding [ALBUS, RAVIV1], and provides, in part, the motivation for what is contained in this report.

Initially, research in artificial vision was relatively far removed from parallel research in understanding the vision of natural organisms, including primates and man. However, with the apparent lack of progress made toward the achievement of artificial vision, more attention is now being paid to what is understood (and not understood) about biological vision [NAKAYAMA, KAUFMAN, SCHWARTZ]. In this regard, the German physiologist and physicist Herman L. F. Von Helmholtz (1821-1894), founder of the science of perceptual physiology, can hardly go unmentioned, as a very large amount of what is known about the functional capacities of human vision are due to his conceptual and experimental genius. His *Treatise on Physiological Optics* [HELMHOLTZ] is still relevant today.

Clearly, the massively parallel organization of the brain contrasts sharply with the serial organization of the Von Neumann computer model. The significance of this has not been lost on those pursuing multiple instruction/multiple data architectures, neural net architectures, optical computing architectures and other architectures exhibiting a high degree of connectivity in which the operations are brought to the data. This "in-place" massively parallel architecture seems particularly relevant for computer vision, and it is with these potential architectures in mind that we present the ideas in this report.

Other aspects of past mainstream computer vision research also contrast sharply with biological vision. In particular, biological vision is binocular and the

exploitation of the resultant binocular disparity, in conjunction with the temporal disparity of optical flow, provides a potentially richer environment within which computational algorithms may be sought. Recognition of this is now developing within the computer vision research community. (Note that the practice of stereo photogrammetry in remote sensing is a related, but different problem.)

A natural adjunct to a binocular camera-retina is the idea of "active vision". Here, the need for visual fixation of a moving target, gaze direction control etc., is accomplished by the introduction of a predict/compare control servo loop for controlling binocular divergence and convergence (vergence), focus and iris control, etc. In this way, the images are not passively acquired, but are acquired actively so as to maintain a "visual equilibrium" with respect to some visual subtask. While these ideas are obviously modeled on biological vision, they also seem to have appropriate counterparts in a design for an artificial vision system.

The modern camera has inherited two features which contrast sharply with biological vision. One of these is the uniform resolution of the imaging process, leaving it to the human viewer to apply variable resolution via the radially decreasing density of the rods and cones of his saccading retina when viewing the picture. The discovery of the logarithmic spiral nature of this decreasing density in conjunction with the "digitization" of the field of view was first made by Schultze and published in 1866 [SCHULTZE]. This highly regular tessellation of the retina and its subsequent mapping to the visual cortex has been given considerable attention and a large literature now exists on the subject from the point of view of descriptive anatomy [DOW, MAUNSELL, NAKAYAMA, SCHWARTZ], electrical engineering [ESSEN], prescriptions for artificial vision [FISHER, ROJER, WEIMAN], and other points of view.

One may imagine a human visual field with a constant resolution, and hence the reason that it isn't may be only one of economy. However, the body of this report will describe an imaging model in which varying resolution is a natural product of other design decisions.

The second sharp difference between biological imaging and the modern camera concerns the geometry of the surface, or manifold, on which the projected image is formed. In the modern camera, this manifold is flat or planar, while for the retina it is, to a first approximation, spherical. The significance of this is still unclear, but as will be elaborated on later in this report, for purposes of analysis, the spherical projection model provides new insights. With respect to computation, we will show how planar projection images can be transformed to spherical projections, as mapped to the plane, and hence the spherical projection in principle provides no information not found in a planar projection.

While the modern camera is capable of precisely recording color images, the exploitation of color information is not within the objectives of the work reported on

here. Rather, the images we will be concerned with may be thought of as either black and white, or colored, and it will be left to future research to understand how exploitation of color may be accomplished.

This report will elaborate on the above ideas in conjunction with their application to a camera-retina imaging model and attempts to integrate them into a consistent whole. However, this should in no way be construed to mean that we propose to model any portion of biological vision.

1.3. THE OBJECTIVES OF THIS RESEARCH

Research of necessity takes place within a larger context than just that one aspect worked and reported on. The objective of the larger context and our research methodology are as follows:

Objective: The realization of general purpose real-time low-level (pre and attentive) artificial vision. Such a system would provide a robust geometric interpretation of the information inherent in iconic imagery, i. e., pictures, and would serve as the precursor for higher level symbolic processing associated with specialized tasks.

Methodology: The creation of a detailed (mathematical) model of binocular camera-retina imaging structures together with low level iconic image, i. e., picture, processing, whose purpose is to go beyond the standard static monocular planar projection imaging model. In particular, we are interested in a non-standard iconic image representation motivated by the need to simplify optical flow and stereo disparity extraction and interpretation, as we believe these provide fundamental cues to biological vision systems.

1.4. THE METHODOLOGICAL COMPONENTS OF THE VISION RESEARCH REPORTED ON HERE

The research reported on here may be thought of as being part of several complimentary components. These components are described in the following (1) through (3) elements.

(1) **Analytic/Geometric Model:** The analytic/geometric model is concerned with the description of how iconic images and iconic image sequences are formed, transformed, represented and interact, together with mathematically idealized lenses, manifolds, coordinate frames etc., for describing the model.

A dynamic control portion is concerned with the description of what is required to control binocular convergence/divergence (vergence), iris control, gaze direction, compensating camera-retina motion control for fixation etc.

In this report we will be primarily concerned with the geometric portion of the spherical projection model, and will refer the reader to readily available descriptions

of planar projection [HARALICK1].

The projection model also contains additional non-standard components characterized by the following:

- The ''ego-sphere'' and its modeling in the form of spherical projection.
- Foveal-peripheral resolution in which the ''resolution'' of the image falls off radially.
- The ''normalization'' of optical flow so as to make its extraction and interpretation computationally simpler.
- The integration of two camera-retinas in support of a binocular visual field in which the integration of stereo disparity and optical flow disparity can be ''fused'' into a single monocular interpretation.

(2) **Computational Model:** A massively parallel computational model, assumed to be operating within an implementation of the geometric and dynamic control model, for making explicit such information as the spatial and temporal gradients, disparities etc. It provides the raw information for maintaining the internal representation, e. g., world model, at the lowest level.

We have indicated only the most rudimentary kinds of information in the computational model, but understand that information extracted from imagery must also be combined with task information, integrated past imagery etc. For the purpose of this report, we will be primarily concerned with the camera-retina imaging representations and associated low-level processing for optical flow and stereo disparity extraction and interpretation.

(3) **Model Simulator:** A computer simulation of (1), the geometric model, for the purpose of

- Demonstrating concepts
- Performing experiments
- Acting as a design tool for both the geometric model (1) and the computational model (2).
- Suggesting new relations as a result of its being specific and detailed.

In section 4 a computer simulator for the analytic/geometric model is described. It will be used to demonstrate the ideas and properties of the geometry with respect to the extraction and geometric interpretation of optical flow and stereo disparity.

Historically, the goal of vision research has been to develop techniques for inferring a 3-D dynamic description of the world from two dimensional projected sequences [THORPE]. While this may still be an objective for certain applications, it seems to us that interaction with task behavior must be emphasized: an iterated perception followed by action cycle. e. g., an object is grasped by iteratively decreasing the difference between the non-grasped state of affairs and its desired

grasped state. In this way the world acts as its own model, accessed by active camera-retinas, so that within limits, the need for a detailed internal model of the world is minimized.

This is in contrast to calculating an a priori trajectory from a detailed internally generated 3-D map followed by its subsequent execution.

We invite the reader to infer the basic nature of the vision task from the following.

At the lowest level, the internal behavioral task is to control vergence, fixation, focus, gaze direction etc., in support of searching and accessing the external world. At the next level, the external behavioral task is e. g., obstacle detection, and hence avoidance, based on "looming" and/or "time to collision", detection, vergence and fixation of moving "point" targets, and the determination of rigid body motions at discrete ranges, all in support of spatially directed mobility.

The design of an artificial vision system to achieve these tasks seems feasible given that we can extract and interpret optical flow and stereo disparity, and their linear combinations, from camera-retinas which can be controlled, and would seem to be basic for a self-guiding vehicle operating in an unstructured (no road) environment, for example.

2. REVIEW OF OPTICAL FLOW

The temporal change of incident light intensities making up the projection of the 3-D world on a camera-retina is informally called the optical flow.

More precisely, one should distinguish between *motion flow* constituting the actual motion of the surfaces of the objects in the scene from the *optical flow* stemming from the change in location of incident light features. For example, the motion flow of a rotating reflective sphere will result in a zero optical flow, while if the sphere remains stationary while a (point) source of illumination moves, the optical flow will be nonzero. A motion picture or television screen operates by exploiting the latter. Unfortunately, the only visual cue available for detecting actual motion, i. e., motion flow, is optical flow. Fortunately, for most uncontrived benevolent scenes, optical flow and motion flow are highly correlated, and for purposes of this report will be considered identical. Note that optical flow and motion flow are both two dimensional vector fields defined on the imaging plane.

The theoretical study of optical flow has two aspects: (1), the characterization of induced optical flow given camera-retina relative motion and 3-D scene geometry, i. e., the optical flow field equations, and (2), the converse, the numerical extraction and geometric interpretation of optical flow given camera-retina projective imagery. The first is part of the geometric model, while the second is part of the computational model.

In practical engineering applications, e. g., for a camera on a self-guiding vehicle, camera motion and gaze direction will be available in the computation model whose goal might be the detection and avoidance of obstacles. Other combinations of known and desired unknown quantities come readily to mind in such applications [HERMAN].

2.1. THE PLANAR PROJECTION OPTICAL FLOW FIELD EQUATIONS

We briefly describe the equations characterizing the induced optical flow resulting from motion of a camera using planar projection in order that it may be contrasted in a latter section with the spherical projection analog. A more detailed derivation is provided in Appendix A1, or is readily available in the literature [BRUSS, HORN2].

For concreteness, we may imagine a panning, tilting planar projection camera mounted on a vehicle for which the primary motion is one of translation in the direction of the optical axis of the camera, and further, that the environment through which the vehicle moves is static. A camera centered $X-Y-Z$ coordinate system, in which the optical axis is aligned with the X-axis and the $y-z$ image plane is located at $X = f$ results in the planar projection equations

$$y = f\frac{Y}{X}, \text{ and } z = f\frac{Z}{X}, \qquad 2.1.1$$

where f is the focal length. See figure C1.

The resulting instantaneous velocity $\dot{y} \equiv \frac{dy}{dt}$ and $\dot{z} \equiv \frac{dz}{dt}$ of the intensity patterns, i. e., the optical flow, in the image is then the sum of the translational and rotational components in the y and z directions of the image: $\dot{y} = \dot{y}_t + \dot{y}_r$ and $\dot{z} = \dot{z}_t + \dot{z}_r$, where

$$\frac{\dot{y}_t}{f} = \left[\frac{-V + y\,\frac{U}{f}}{X}\right], \quad \text{and} \quad \frac{\dot{y}_r}{f} = B\,y\,z - C\,(y^2 + f) + \frac{A}{f}\,y \qquad 2.1.2a$$

$$\frac{\dot{z}_t}{f} = \left[\frac{-W + z\,\frac{U}{f}}{X}\right], \quad \text{and} \quad \frac{\dot{z}_r}{f} = B\,(z^2 + f) - C\,y\,z - \frac{A}{f}\,y. \qquad 2.1.2b$$

In these equations, U, V and W are the instantaneous translational velocities along, and A, B and C are the instantaneous angular velocities about, the X, Y and Z axes, respectively.

Globally, we may think of the point P as lying on a surface defined by a "depth" function" $X(Y, Z)$, which is positive for all values of Y and Z. Given such a surface, we can associate with a camera motion an optical flow defined in the image plane by 2.1.2, and think of this optical flow as being generated by this surface and the camera motion.

Several intuitively experienced observations concerning 2.1.2 can be made:

(1) Under general motion, the induced flow field has the property that it is dependent on image location, and not just on the motion parameters and depth function $X(Y, Z)$. This has the implication that the interpretation of optical flow as depth is a global operation requiring knowledge of where in the image the optical flow is located, and hence complicates an in place parallel algorithm.

(2) The rotational component $[\dot{y}_r, \dot{z}_r]$ is independent of the depth X. For example, if the camera pans or tilts with zero translation, then the resulting optical flow is independent of the distance to any object in front of the camera.

(3) As the depth becomes very large, optical flow due to the translation component $[\dot{y}_t, \dot{z}_t]$ goes to zero.

(4) Setting $[\dot{y}_t, \dot{z}_t] = 0$ and solving, one obtains the coordinates $(y = V/U, z = W/U)$ of the point where the translational component of optical flow is zero. This point is called the *focus of expansion*. Similarly, one can solve $[\dot{y}_r, \dot{z}_r] = 0$ and $[\dot{y}, \dot{z}] = 0$ as is done in [RAVIV1] to obtain *zero-flow circles*.

PLANAR PROJECTION CAMERA IMAGING MODEL

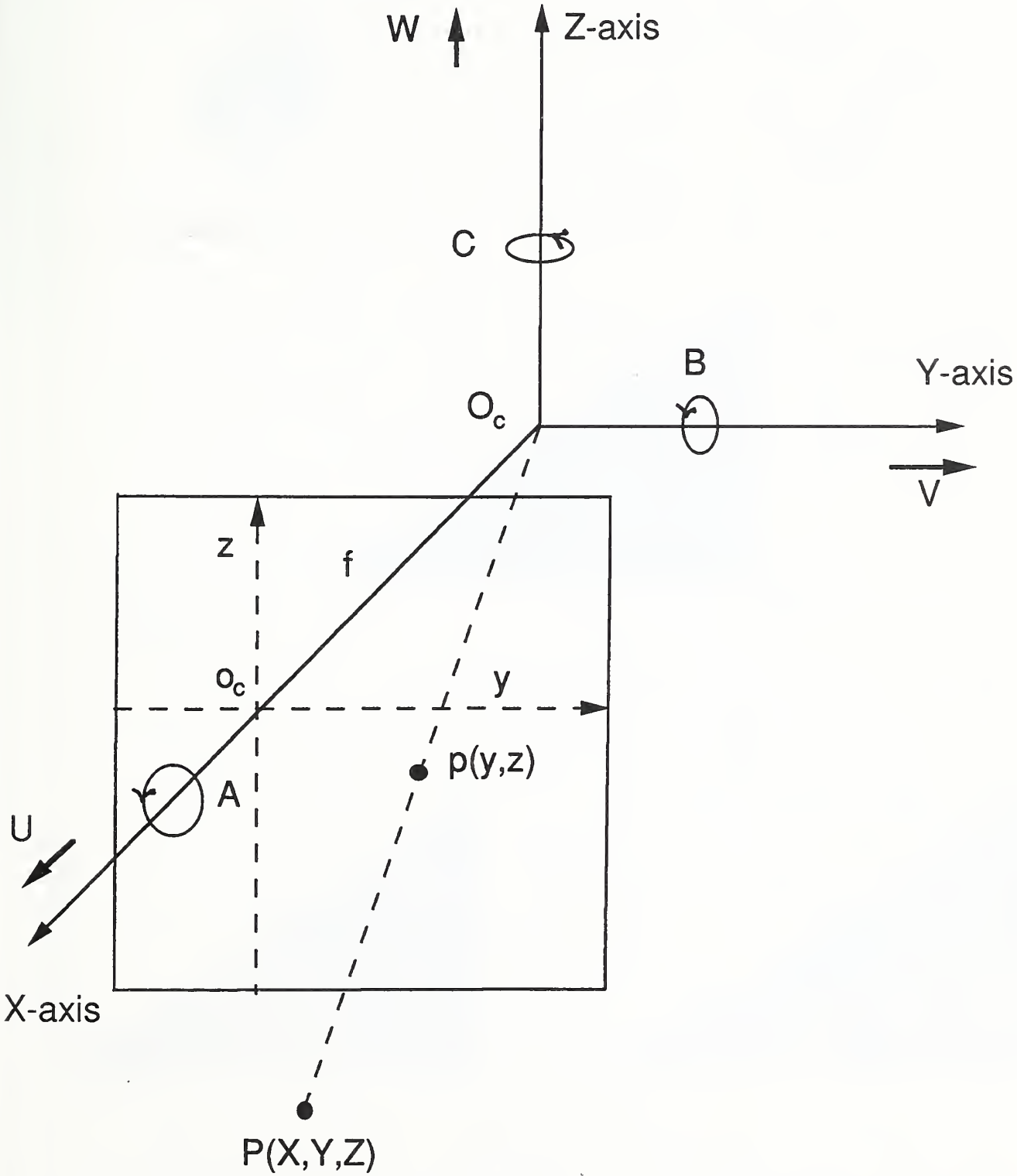


Figure C1: The camera centered coordinate system used for planar projection optical flow. U, V, W are the translation velocities, and A, B, C the angular velocities. The world coordinate point $P(X, Y, Z)$ projects to the image point $p(y, z)$.

APERTURE PROBLEM

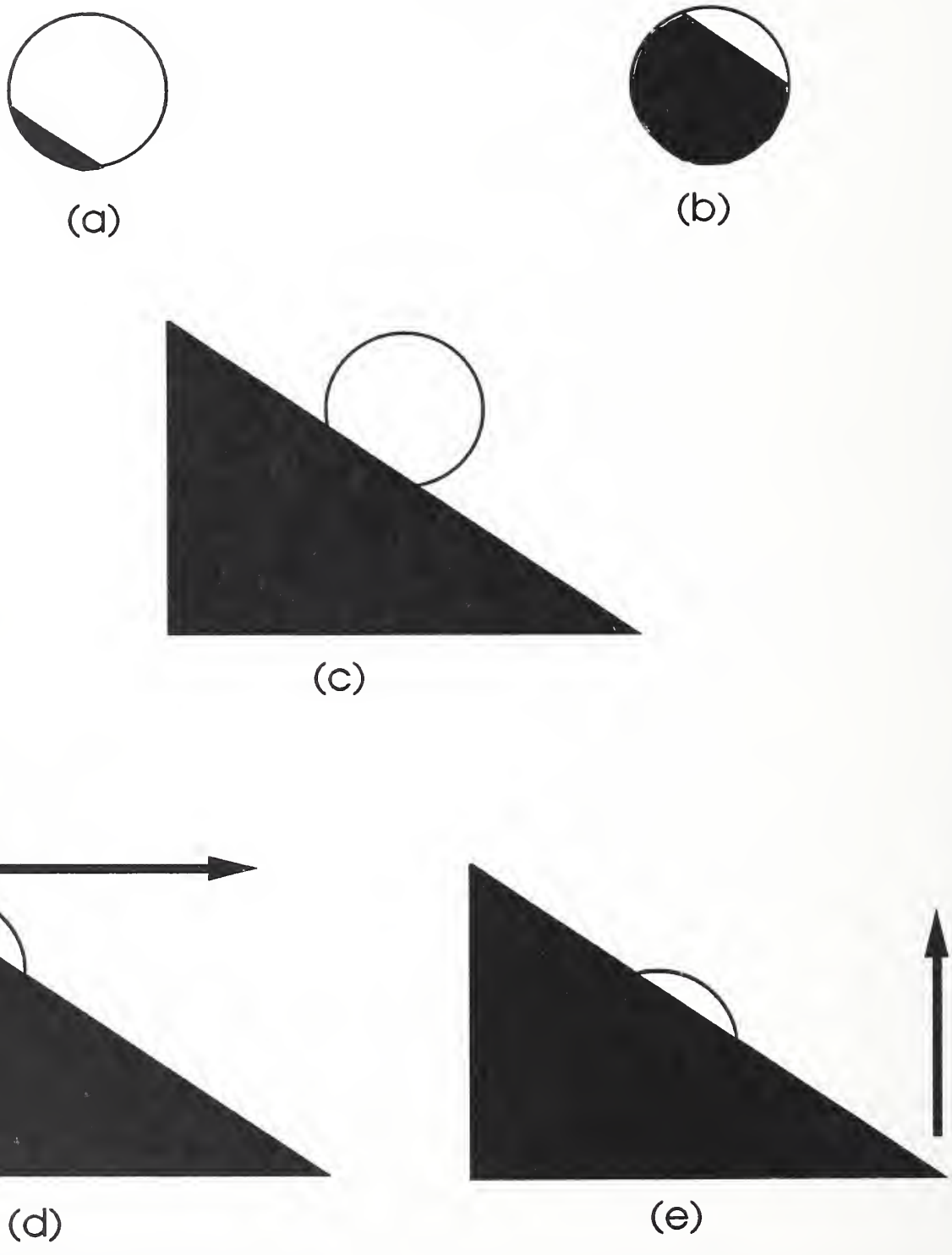


Figure C2: (a) and (b) are the before and after situations in looking through an aperture. If (c) is the actual situation seen in (a), then two motions, either (d) or (e) or a combination of the two, can account for what is seen in (b).

(5) Camera motion and the depth enter into generating optical flow only as a ratio between the component of motion along the depth, and the depth, i. e., U/X. If two translational motions generate the same optical flow, then at best it can only be said that one scene is a scaled version of the other. This means at least one of U or X must be known absolutely to determine the other.

For the case of pure rotation, it can be shown that two distinct rotations will give rise to two distinct optical flow fields. This can be seen by assuming the contrary and equating the respective X and Y components, from which one immediately concludes that the rotations are the same.

The case in which the camera translates in the plane perpendicular to its optical axis, e. g., $U = A = B = C = 0$, $V \neq 0$, $W \neq 0$, as would be the case for a camera looking straight down from an airplane, results in a particularly simple flow field, one which is independent of y and z, and hence one in which when these assumptions hold, has a simple geometric interpretation in the computation model in terms of the depth function $X(Y, Z)$. This case will be elaborated on later in this section.

2.2. THE NUMERICAL EXTRACTION OF OPTICAL FLOW

Optical flow for a video camera is induced by the image frame to frame difference of some invariant, i. e., "constant", feature as it moves relative to the camera.

The fundamental problem in numerically extracting optical flow from an image sequence is the problem of *correspondence:* given two or more images, determine the image coordinates of each which correspond to the same real-world feature. The resultant frame to frame coordinate differences, or *temporal disparity*, in principle allow one to compute the distance to the feature point by a simple triangulation. This is the so called "correspondence based method", which for a small number of feature points is a feasible solution. However, explicitly calculating correspondences for n such features results in an $O(n^2)$ computation, and hence will be computationally expensive for computing a dense depth map.

One method of finessing the correspondence problem is to make the time difference between image frames so small that "adjacent" features in three dimensional *spatio-temporal space* may be assumed to be the same real-world feature. This of course places a limit on the velocity that features may have for a given frame frequency, typically not more than that needed to induce a disparity of one picture element per frame. These methods include, among others, gradient based [HORN1], spatio-temporal filters [HEEGER] and correlation [RAVIV2] of various flavors.

These "non-correspondence" methods lend themselves to massively parallel algorithms acting locally. However, as a result, they are much more dependent upon local manifold geometry for both extracting and interpreting optical flow. It is this dependence we wish to review here.

2.2.1. THE OPTICAL FLOW CONSTRAINT EQUATION

The non-correspondence based methods have as their basis the *visual flow constraint equation* [SCHUNK] defined on spatio-temporal space:

$$\frac{\partial I}{\partial y}\frac{dy}{dt} + \frac{\partial I}{\partial z}\frac{dz}{dt} = -\frac{\partial I}{\partial t}. \qquad 2.2.1.1$$

In words, given some motion invariant distinguished image feature I at image coordinates y, z, I's instantaneous trajectory in spatio-temporal space is constrained to lie along a continuous path obeying the above equation. The terms $\frac{dy}{dt}$ and $\frac{dz}{dt}$, i. e., $\dot{y}$ and $\dot{z}$, are in fact identified with the coordinate components of the resultant optical flow vector field as given by equations 2.1.2.

An important point concerning this constraint equation is that it does not determine the two velocity components, but constrains them to a single degree of freedom. This is because only the component of motion in the direction of the light intensity change can be extracted. This problem, known as the aperture problem, and its solution will be elaborated on in the next section.

In order that the inverse problem of optical flow extraction not be confused with the description of the optical flow field equations of the previous section, we use the notation

$$\Delta I_y u + \Delta I_z v = -\Delta I_t, \qquad 2.2.1.2$$

where ΔI_y and ΔI_z are the numerically estimated components of the spatial gradient, ΔI_t is the estimated temporal gradient and u and v are the possible y and z components of optical flow in the observed image.

The direct application of the optical flow constraint equation results in the gradient based method for extracting optical flow, i. e., the component of optical flow $|u, v|$ in the direction of the intensity change is given by

$$|u, v| = \frac{\Delta I_t}{\sqrt{\Delta I_y^2 + \Delta I_z^2}}. \qquad 2.2.1.3$$

(For a derivation of the individual components u and v as well as this equation see Appendix A3.)

The application of this equation to real imagery is complicated by several issues: spatial, temporal and intensity discretization, and by the fact that typically, real world features do not always remain intensity invariant under changing illumination or changing point of view. e. g., specular reflections. As a result, a number of algorithms have been explored by a number of researchers in which the "I" in equation 2.2.3 has taken on different interpretations. These have typically been weighted sequences of images: weighted sequences of intensity as well as temporal and/or spatial convolutions with one, two and three dimensional Gaussian distributions and its derivatives.

The extraction algorithms will typically have limitations: within the context of using gradient based algorithms, the numerically extracted optical flow magnitude $|u, v|$ must lie, a priori, within the range $0 < |u, v| \leq 1$, where the unit is pixels per frame.

In this report we will not be concerned so much with the particular method used to extract optical flow as we will the geometric nature of the underlying flow field and its impact on the resulting extraction and interpretation geometry.

2.3. THE GEOMETRIC INTERPRETATION OF OPTICAL FLOW

The extraction of optical flow is followed by its geometric interpretation. While the actual computations may be relatively unrelated, they are implicitly related through the geometry of the manifold on which the iconic imagery is represented.

We may imagine a camera viewing a scene, possibly containing other translating objects, while translating laterally to the scene, as would a camera viewing the ground from a plane, or pointed out a side window of a forward translating car. The resulting extracted optical flow, as recorded by a planar projection camera, has a particularly simple geometric interpretation owing to its linear nature.

In this section we elaborate this situation, and in particular, show that the general case is intrinsically a two dimensional problem. This is due to the fact that optical flow extracted from edges, as opposed to computing optical flow from point to point correspondences, does not uniquely determine relative motion.

This is elaborated here at length as it represents an ideal situation for their interpretation, and is fundamental to understanding both optical flow and binocular disparity.

2.3.1. THE APERTURE PROBLEM

When a moving point feature is viewed through a camera, it is in principle possible to calculate the relative displacement of the point in two successive images, and hence calculate the optical flow for that point. For straight edges, the situation is more complex. If the relative motion of the camera with respect to the edge is not known, then a continuum of possible motions could account for what is observed. This is known as the *aperture problem,* and is depicted in figure C2. While its implications must eventually be understood for all combinations of camera motion, we will present it here primarily in terms of the simple case in which a camera-retina is translating, without rotation, parallel to its focal plane.

The solution to this problem is to observe that only the *normal component* of the edge motion can be unambiguously calculated from two successive images. This then resolves the problem into two cases: (1), that for a camera viewing a static scene in which typically the camera direction of translation is known, and (2), the case in which the camera is viewing independently translating objects with no a priori knowledge of their relative magnitude and direction of motion.

For the first case, the known direction of camera motion may be used to calculate the point to point optical flow of the edge by compensating for the fact that the direction of camera motion is at a known angle with the edge.

More precisely, denote the known camera velocity direction (with respect to the image plane of the camera) by ϕ and the magnitude of the extracted normal component by $|u, v| \equiv \sqrt{u^2 + v^2}$. This latter magnitude will have image coordinate components u and v and direction within the image of $\theta = \tan^{-1} \frac{v}{u}$, where the image coordinate axis for u is aligned with the direction of camera motion. Then the true optical flow magnitude, i. e., as would be calculated for a point, $|U, V|$ is related to the observed magnitude $|u, v|$ by

$$|u, v| = |U, V| \cos(\phi - \theta). \qquad 2.3.1.1$$

Hence numerically extracted optical flow is always less than or equal in magnitude to the actual magnitude as would be determined by a point to point correspondence method.

For example, assuming that ϕ is measured from the y axis, with the camera translating parallel to the y axis, ϕ will be zero, and an edge will be foreshortened by cos θ, where θ is the angle the edge normal makes with the y axis. Hence, if the edge is perpendicular to the motion, the edge normal makes an angle $\theta = 0$ and the extracted and actual are the same. However, an edge parallel to the motion has

$\theta = 90°$ and hence has no apparent observed motion in the image.

Applying this to the computation of equation 2.2.1.3 implementing the gradient paradigm of optical flow extraction, results in computing the "true" optical flow, $|U, V|$ or what will be called the *compensated normal component:*

$$|U, V| = \frac{|u, v|}{\cos(\phi - \theta)} = \frac{u^2 + v^2}{u \cos\phi + v \sin\phi} \qquad 2.3.1.2$$

Equation 2.3.1.2 tells us how to calculate this true optical flow when the direction of relative motion is known.

For the case of a camera translating parallel to the Y axis viewing a static scene, i. e., case (1) above, ϕ is constant, and without loss of generality ϕ can be aligned with, i. e., measured from, the Y axis, and hence equation 2.3.1.2 becomes

$$|U, V| = \frac{-\Delta I_t}{\Delta I_y}. \qquad 2.3.1.3$$

This computation can be performed using one dimensional correlation, numerical gradients etc.

Hence, the problem of extracting optical flow for the case of a translating camera viewing a static scene is not affected by the aperture problem, and does not require the more complex computation required for the second case. This latter problem is two dimensional in an essential way.

The second case, involving a camera viewing a scene containing multiple translating objects, results in several unknown relative motions, and hence a nonunique ϕ. However, because the optical flow fields linearly superimpose, it is still possible to determine respective relative motions from two or more optical flow vectors. This will be elaborated on in the next section.

2.3.2. THE CASE OF A LATERALLY TRANSLATING CAMERA VIEWING A DYNAMIC SCENE

The optical flow generated by a laterally translating camera is referred to as being "linear". This is due to its simple inverse relationship to depth: two equal optical flow vectors have the same interpretation if and only if they are projections of two points having the same depth. This property makes its interpretation particularly simple, even for a dynamic scene.

The geometric properties of such linear optical flow will be elaborated in the next several subsections together with a robust algorithm for exploiting this property. Later, when the case for a forward translating camera is considered, it will be seen that this linear property can be achieved for nonlinear images by "warping" them appropriately.

Assume again, as in section 2.1, a $X-Y-Z$ world model coordinate system centered at the camera with the camera's optical axis aligned with the X axis. Then the planar perspective image plane coordinates y and z are related to the camera coordinates by

$$y = f\,\frac{Y(t)}{X(t)} \quad \text{and} \quad z = f\,\frac{Z(t)}{X(t).} \qquad 2.3.2.1$$

Referring back to equations 2.1.2, and assuming no rotation, i. e., $A = B = C = 0$, and further, $\dot{X}(t) = U = 0$, and denoting $\dot{Y}$ and $\dot{Z}$ by V and W respectively, the induced optical flow vector field has components $\dot{y}$ and $\dot{z}$ at any point in the image given by

$$\dot{y} = f\,\frac{\dot{Y}}{X} \quad \text{and} \quad \dot{z} = f\,\frac{\dot{Z}}{X}. \qquad 2.3.2.2$$

For such a translating camera, an edge in the scene, either background or a translating object, will be at some unknown arbitrary angle θ, $0 \le \theta \le 180°$, with respect to the direction $\phi = \tan^{-1}\frac{\dot{Z}}{\dot{Y}}$ of camera translation.

For the purpose of computing relative motion from optical flow it is the magnitude and direction of these vectors which is important, not their location in the image. To this end we plot magnitude versus orientation: the extracted optical flow vector magnitude $m \equiv |\mu, \nu|$ is plotted on the graph's horizontal axis M, while its orientation θ is plotted on the vertical axis Ω, $-180° \le \theta \le 180°$. We arbitrarily locate the camera direction $\phi = 0$ at $\Omega = 0$, so that scene background edges perpendicular to the camera motion, i. e., having $\theta = 0$, will map to $\Omega = 0$.

In this *magnitude-orientation plot*, as it will be referred to, each optical flow vector will result in a point which corresponds to its magnitude and direction. (In fact, the value stored may be thought of as being proportional to the total number of vectors in the scene of that magnitude and orientation, so that what results is a two dimensional histogram.)

A camera translating relative to and viewing a circular disk will generate an image in which every angle an edge could make with respect to the translating direction is represented. It thus provides a contour whose optical flow normal

component magnitude varies continuously from some maximal amount at its two points whose normals are parallel to the relative direction of motion, down to a magnitude of zero at the points 90° from these points. Figure D1-(a) depicts this situation for a disk translating to the right with respect to the camera.

Figure D1-(b) depicts the magnitude-orientation plot of the optical flow for the disk. The characteristic shape that results, i. e., the cosine curve cos ($\phi - \theta$), $\phi = 0$, is fundamental to the interpretation of numerically extracted optical flow for a laterally translating camera: all normal components extracted from background objects and at the same depth will fall somewhere along this *loci of normal components* with respect to $\phi = 0$.

For an arbitrarily shaped stationary object, its optical flow normal components will fall with some shape characteristic varying density along it. In fact, any optical flow vector computed from parallel edges and having the same relative motion and depth, irrespective of image location, will map to the same point on the locus of normal components. Stationary objects at a differing depth will map to a second distinct locus having a differing maximum magnitude, but all centered on $\Omega = 0$.

A static point feature will move in the image in a manner prescribed by $\dot{x}$ and $\dot{y}$, while the optical flow components of an object translating in the scene will be $\dot{y} + \dot{y}_{obj}$ and $\dot{z} + \dot{z}_{obj}$, the sum of the known optical flow components induced by the camera and the unknown object translation.

Because of the linear nature of the flow field, the numerically extracted optical flow components will still reflect the relative motion, but with the location of the locus of normal components shifted along the Ω axis by the angle ω between the camera direction and the translating object's direction.

This situation, using actual data from a laterally translating camera viewing three disks at different depths, one of which is moving at right angles for a relative direction of 45°, is shown in figure D1-(c)

The task of determining these objects against a background of other features is that of locating the "peaks" of the cosine curves, i. e., the points of largest magnitude. In general, there may not even be data points at this location if there is no contour perpendicular to the relative motion.

In the section following the next an algorithm for locating these peaks is described.

Binocular disparity complements optical flow disparity in the following way: two cameras translating in parallel and offset from each other along a line perpendicular to their motion, will generate four sets of disparities offset by 45° in

orientation in the magnitude-orientation plot. These disparities result from computing the differences between image 1 at time 1 and image 1 at time 2 (optical flow disparities), image 1 and image 2, both at time 1 (binocular disparities), and the "mixed disparities": image 1 at time 1 and image 2 at time 2, and vice versa. Good edges will reinforce one another, while noisy ones will not.

2.3.3. THE GEOMETRY OF LINEAR OPTICAL FLOW FIELDS

In this subsection the informal discussion of the previous section concerning the interpretation of optical flow for a translating camera viewing a dynamic scene is summarized more formally. The reader may skip to the next section without loss.

Optical Flow Normal Component Theorem: *Let* $|\dot{x}, \dot{y}|$ *be the optical flow field induced by a laterally translating camera in direction* ϕ. *Then all optical flow normal components* $|u, v|$ *resulting from features at a fixed distance from the camera must lie along the loci of points parameterized by edge normal directions* θ, $\phi - 90° < \theta < \phi + 90°$,

$$|u, v| = |\dot{x}, \dot{y}| \cos(\phi - \theta), \qquad 2.3.3.1$$

or in terms of its X–axis and Y–axis components,

$$u = |\dot{x}, \dot{y}| \cos\theta \cos(\phi - \theta) \qquad 2.3.3.2$$

$$v = |\dot{x}, \dot{y}| \sin\theta \cos(\phi - \theta).$$

COROLLARY: *Relative camera velocity/depth* $|U, V|$ *and direction* $\Phi = \tan^{-1}\frac{V}{U}$ *is uniquely determined by two distinct optical flow vectors.*

Proof: Solve the following simultaneous equations for $|U, V|$ and Φ:

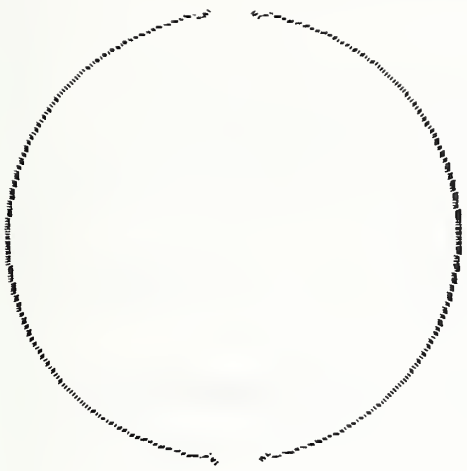
$$|U, V| = \frac{|u_1, v_1|}{\cos(\Phi - \theta_1)} \qquad |U, V| = \frac{|u_2, v_2|}{\cos(\Phi - \theta_2)}$$

Let the two optical normal component data points be denoted by $m_1 \equiv |u_1, v_1|$, $\theta_1 = \tan^{-1}\frac{v_1}{u_1}$ and $m_2 \equiv |u_2, v_2|$, $\theta_2 = \tan^{-1}\frac{v_2}{u_2}$.

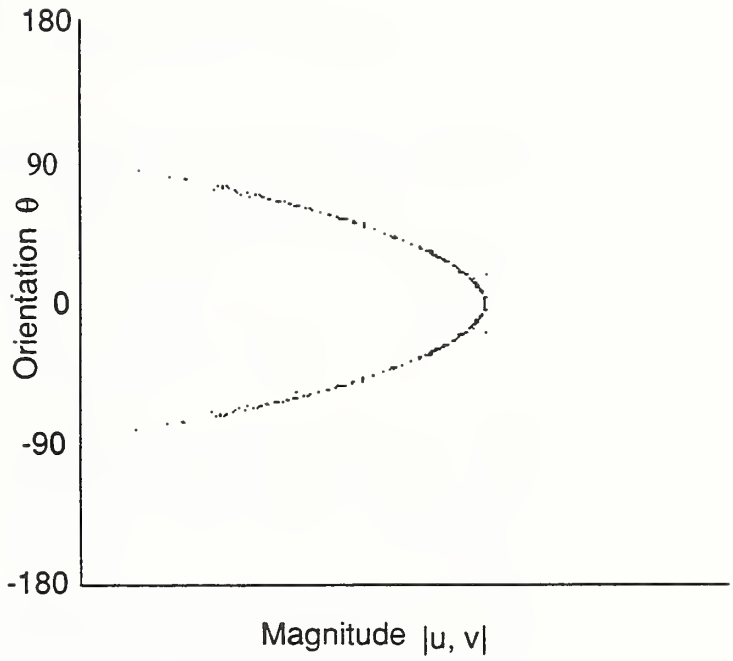
Solving for $|U, V|$ and Φ yields

$$|U, V| = \frac{m_i}{\cos(\phi - \theta_i)}, \qquad i = 1 \text{ or } 2 \qquad 2.3.3.3(a)$$

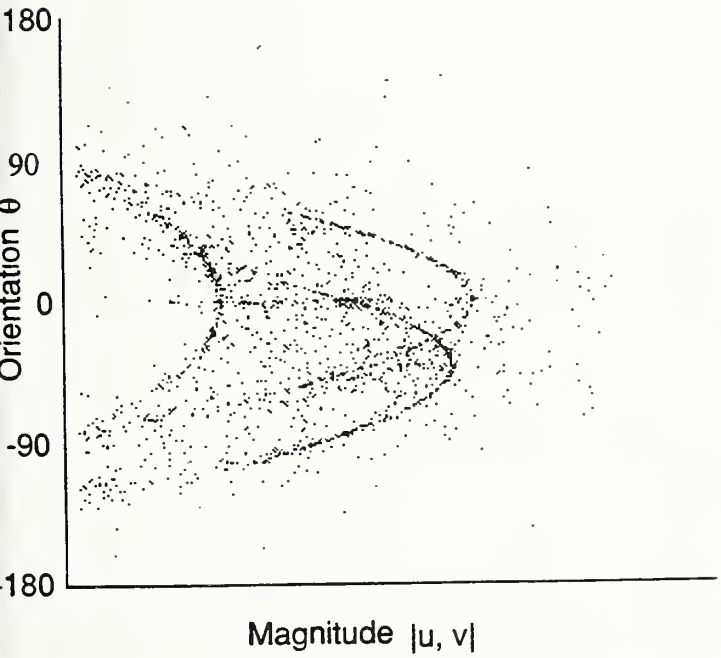
$$\Phi = \tan^{-1}\left\{\frac{m_1 \cos\theta_2 - m_2 \cos\theta_1}{m_2 \sin\theta_1 - m_1 \sin\theta_2}\right\}, \qquad 2.3.3.3(b)$$



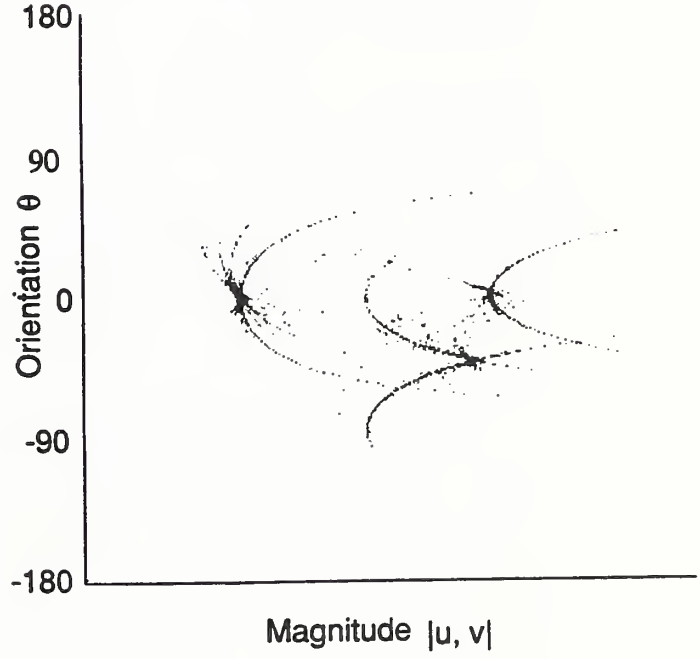
(a)



(b)



(c)



(d)

Figure D1: In (a) the optical flow is shown for a circular disk which is translating to the right, while in (b), the optical flow is plotted in magnitude-orientation space. In (c), the magnitude-orientation plot for three translating disks, two in the same direction but with differing magnitudes and the third at 45°, is plotted. (d) is the Hough transform space of (c), in which clusters indicate the magnitude and orientation of the relative motions.

or in terms of the x and y components,

$$|U, V| = \frac{\sqrt{(u_1^2 + v_1^2)(u_2^2 + v_2^2)[(u_1 - v_1)^2 + (u_2 - v_2)^2 - 2u_1v_1 - 2u_2v_2]}}{u_2v_1 - u_1v_2} \quad 2.3.3.4(a)$$

$$\Phi = \tan^{-1}\left[\frac{u_2(u_1^2 + v_1^2) - u_1(u_2^2 + v_2^2)}{v_1(u_2^2 + v_2^2) - v_2(u_1^2 + v_1^2)}\right] \quad 2.3.3.4(b)$$

The solution to these two equations, $|U, V|$ and ϕ, are the magnitude and orientation of the optical flow field which would give rise to these two particular normal components.

COROLLARY: *Moving curved edges generate more distinct pairs of optical flow vectors and hence yield more information than straight edges.*

COROLLARY: *A loci of normal components is characterized by two parameters, camera velocity magnitude* $M = |U, V|$ *and direction* Φ.

COROLLARY: *Two distinct loci intersect in one point if and only if* $0° < \Phi_1 - \Phi_2 < 90°$.

2.3.4. AN ALGORITHM FOR SEGMENTING LINEAR OPTICAL FLOW FIELDS

The interpretation of optical flow following its numerical extraction using gradient or correlation methods requires the use of clustering or similar techniques in order that the highly overconstrained system of resulting vectors leads to robust solutions. In addition, real scenes often result in many noisy vectors whose lack of mutual support must be used as a basis for their being discarded.

We briefly describe here an algorithm, based on a form of Hough [DUDA] or Radon [DOUGHERTY] transform, to segment an image using optical flow extracted from a laterally translating camera viewing a dynamic scene containing multiple translating objects at the same or differing depths. Components of motion along the camera optical axis are ignored, as would be the case for a downward directed camera in an aircraft viewing vehicles translating on the ground with no such component.

In such a scenario, the objects are relatively small compared with the depths involved, so that the task of image segmentation into regions of relative rigid motion at distinct depths is appropriate.

The vision task is to (1), determine the number n of moving objects, as determined by similar motion and depth, (2), for each object determine its translation

vector (as projected in the camera image plane),and (3), segment the image by classifying each pixel, for which optical flow can be computed, as to which of the $n+1$ objects and background it belongs to in the scene.

This is the situation shown in figure D1-(c), where the optical flow for three objects, moving relative to the camera against a background of noise, is depicted in a magnitude-orientation plot.

The optical flow normal components extracted induce a natural two parameter Hough transform. For each "point" of normal component magnitude $|u, v|$ and orientation $\theta = \tan^{-1}\frac{v}{u}$, compute the "Hough transform" curve, from equation 2.3.1.2, given by

$$m_i = \frac{u^2 + v^2}{u \cos \phi_i + v \sin \phi_i}. \qquad \theta - 90° < \phi_i < \theta + 90° \qquad 2.3.4.1$$

This curve, plotted as a two dimensional histogram in the same magnitude-orientation (m-ϕ) space, represents all possible motions and relative normal orientations which could give rise to the given normal component. Two normal components coming from the same rigid body motion, irrespective of normal direction, will intersect at a common point, the magnitude and orientation of the relative motion common to both. A cluster of such solutions generates a peak in the histogram and represents many such common solutions all coming from the same rigid body motion. The number of such distinct clusters is the number of translating bodies. Noise is plotted randomly and is not reinforced. Figure D1-(d) is the Hough transform corresponding to Figure D1-(c). Note the three clusters at locations corresponding to the cosine peaks of figure D1-(c).

Segmentation of the original image is performed by a second step: each optical flow vector is classified by computing its compensated value for all solutions, and determining that one which is closest to accounting for this data point. Those vectors which exceed a given threshold for the best solution may be assumed to be noise.

The resulting pixel-wise classification may then used to segment the original image by relative motion classification, or to collect like moving features with the intent of interpreting them as a single partially occluded object, etc. A more detailed description of the algorithm is given in appendix 3.

We have included this algorithm here because it is capable of generating robust solutions to the interpretation of linear optical flow disparity fields. By linear is meant that two disparity vectors can be of equal magnitude and orientation if and

only if they have the same interpretation as relative motion.

This result will be used in a later section to segment images created by a forward translating camera, but "warped" by a nonlinear sampling procedure in such a way as to make the optical flow linear in the above sense. This process will be called *normalization*.

3. THE BINOCULAR SPHERICAL PROJECTION CAMERA-RETINA IMAGING MODEL

In the natural world both predators and prey have binocular vision. While it can be argued that this binocularity was originally the product of bilateral symmetry, it is clear that higher life forms have exploited the resulting mix of geometry, optics, and control above and beyond that needed for control of two distinct monocular eyes.

The problem of exploiting the information inherent in the differences between two images formed by two eyes parallels the problem of exploiting the information inherent in the differences over successive instants of time for a monocularly formed image. In fact, we have every reason to believe that natural life forms exploit the mixing of both spatial (binocular) differences and temporal (optical flow) differences. We will argue here that an artificial vision system must aim to do the same.

In human vision we ''see'' monocularly in the region of convergence (normally within the fovea) though we have two eyes. For nearby feature points at a slightly different range, the resulting binocular disparity is perceived as differential range (in a proportional manner up to a point), after which this monocular vision degrades into ''double vision''.

Historically, the development of techniques for understanding how to extract and interpret binocular images, i. e., stereo, was developed as part of quantitative photogrammetry. There, the primary objective was to support the economic demands of cartography, and hence the techniques developed were subservient to the manual processes of map making. In particular, the problem of correspondence, i. e., the problem of identifying the same real world feature in two images, is performed manually, point by point. As a result, photogrammetry has not been a precursor of artificial vision and is viewed by most vision researchers as relatively irrelevant.

However, photogrammetry has been very successful in solving the technical problems that have arisen in understanding the geometric interpretation of stereo disparity under many geometrically differing conditions. While these techniques are not applicable to vision research for the most part, this success has been in large part due to the highly developed analytical models of the stereo imaging process [GHOSH].

Artificial vision researchers have developed their own ''stereo'' imaging models, based primarily on planar projection. Work has also been done in the development of methods for simultaneously exploiting both stereo and optical flow [WAXMAN, GROSSO].

In this section we will describe an analytical imaging model which is motivated by the need to characterize both optical flow and binocular disparity within a single framework. Further, it is based, not on planar projection as is usual, but on spherical projection. It is this basic framework which will be described here, and we view it, or something like it, as a precursor to any characterization of these vector fields, this latter being a much larger research topic [HOFFMAN, KOENDERINK, NELSON, RAVIV1].

We first describe a binocular spherical projection imaging model in which binocular (stereo) disparity plays a key role. The geometry of the resulting iconic images is that of a spherical manifold. As is well known, a sphere is not easily mapped to conventional computer memory, whose underlying differential geometry is that of the plane. Since there is no way to perform this mapping without distorting the Euclidean metric (keeping the distance between two points independent of their location in the image), our solution is to abandon this metric entirely, and provide for a mapping which has other desirable properties, e. g., we introduce a foveal-peripheral resolution.

In the next subsection we describe within a single context three types of camera-retina centered coordinate frames.

(1) The standard spherical projection coordinates, symmetrically placed about the optical axis, will be used to define and compute optical flow for each camera-retina.

(2) A second spherical projection, in elevation and azimuth coordinates, will be used to define control variables for vergence and gaze control for each camera-retina, as well as define and compute binocular disparity.

(3) The third will be a binocular gaze direction and vergence centered coordinate system, one which provides a one-to-one mapping between points in space and the combined degrees of freedom of pan and tilt for the two camera-retinas.

3.1. BINOCULAR CAMERA-RETINA GEOMETRY

Let two camera-retinas, labeled L and R respectively, be centered at $\pm d$ along the Y axis of a right handed Cartesian coordinate system $X-Y-Z$, as shown in figure C3. By camera-retinas we mean hemispheres of radius r oriented so that their concave sides, i. e., retinas, are directed toward the positive X axis. A lens is used to focus light from the environment onto the retinas through their respective idealized nodal points $\pm d$, thus creating spherical projection images on the retinas. We will also refer to these nodal point locations as L and R. (The lenses will not be developed here.) We will refer to these as *(binocular) camera-retinas*. In this report

the value of r will be made 1, and all other distances will be given implicitly as a multiplicative factor of it.

Given a point $P(X, Y, Z)$, the plane containing it and the Y axis will be called the *plane of elevation,* and will intersect both retinas in identical great circles of elevation. This plane will make an angle, referred to as the *angle of elevation,* with the $X-Y$ *horizontal plane.* The $X-Z$ plane will be called the *median plane.*

For each of the two retinas we define its optical axis to be collinear with an auxiliary Y axis located at L and R. That is, we define Y_r and Y_l, to be parallel to Y, but offset from it by $\pm d$. Variables measured in these Y coordinates will have a subscript of r or l, or r/l when used as a variable subscript. They are related to the Y axis variable Y by $Y_r = Y + d$, and $Y_l = Y - d$.

The points L and R will serve as the centers of rotation for the camera-retinas. However, for the moment, one should think of them as fixed, "staring straight ahead".

We next define two "spherical projections" via two spherical coordinate systems.

3.1.1. BI-SPHERICAL COORDINATES

Given a point $P(X, Y, Z)$, its spherical coordinates azimuth $\phi_{r/l}$, eccentricity $\theta_{r/l}$ and range $R_{r/l}$ will be defined by (See figure C4):

$$\tan \phi_{r/l} = \frac{Z}{Y_{r/l}}, \quad \tan \theta_{r/l} = \frac{\sqrt{Y_{r/l}^2 + Z^2}}{X} \quad \text{and} \quad R_{r/l} = \sqrt{X^2 + Y_{r/l}^2 + Z^2} \qquad 3.1.1.1$$

Note that the "optical axis", the direction in which eccentricity θ is zero, is aligned with the X axis.

In figure C3 we have traced out on the camera-retina R lines of constant spherical azimuth and eccentricity. The azimuth ϕ, $0 \le \phi \le 360°$, is measured about the optical axis clockwise (viewed down the positive X axis), and the eccentricity θ, $0 \le \theta \le 90°$, is the off optical axis angle.

The inverse relations are given by

$$\begin{aligned} X &= R_{r/l} \cos \theta_{r/l}, \\ Y_{r/l} &= R_{r/l} \sin \theta_{r/l} \cos \phi_{r/l}, \\ Z &= R_{r/l} \sin \theta_{r/l} \sin \phi_{r/l}. \end{aligned} \qquad 3.1.1.2$$

BINOCULAR SPHERICAL PROJECTION

CAMERA-RETINAS

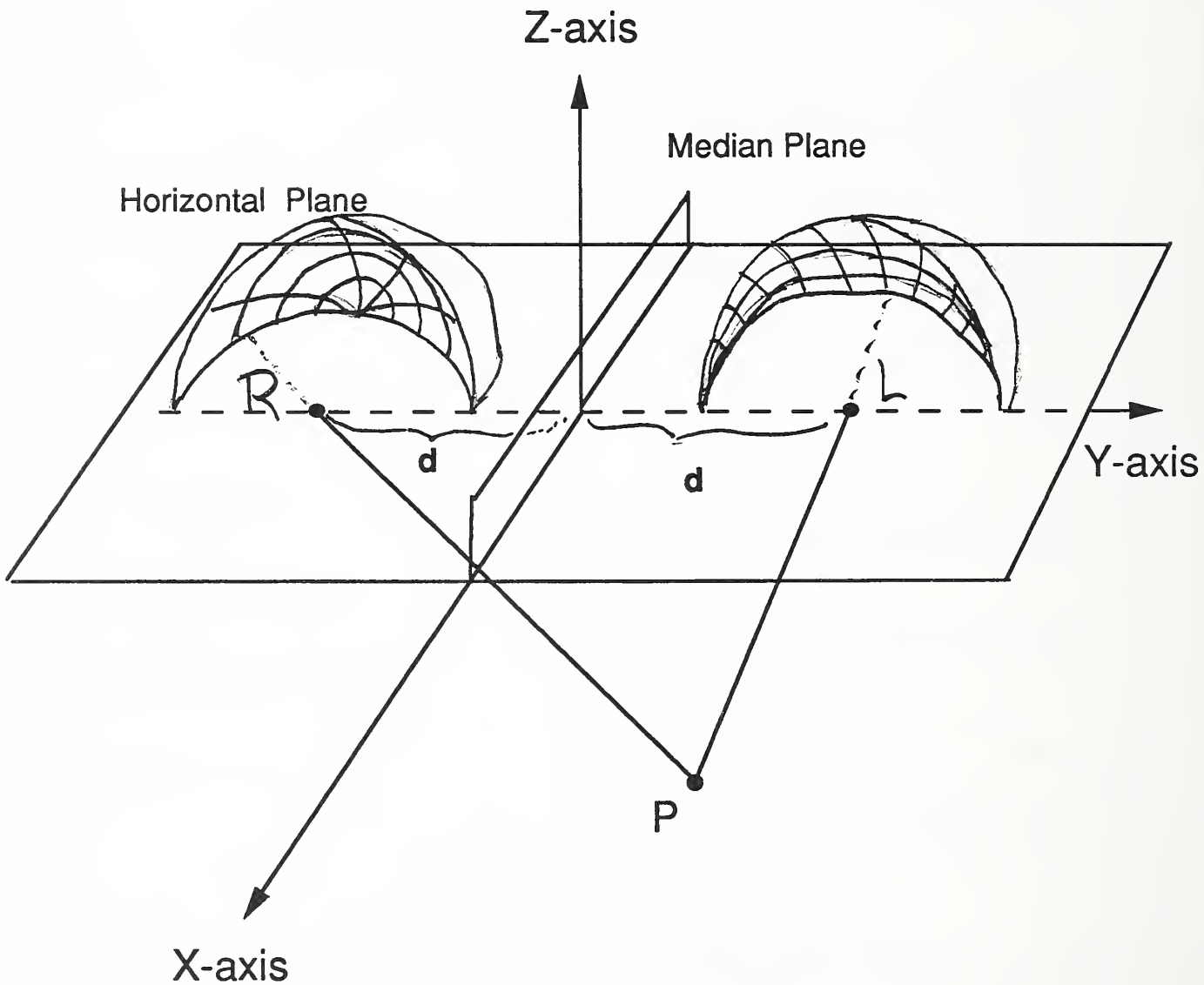


Figure C3: Two hemispherical camera-retinas, labeled R and L, are positioned with their centers located $\pm d$ to either side of the origin. Curves of constant spherical azimuth and eccentricity are shown on R, while curves of constant bi-retinal azimuth and elevation are shown on L.

SPHERICAL COORDINATES

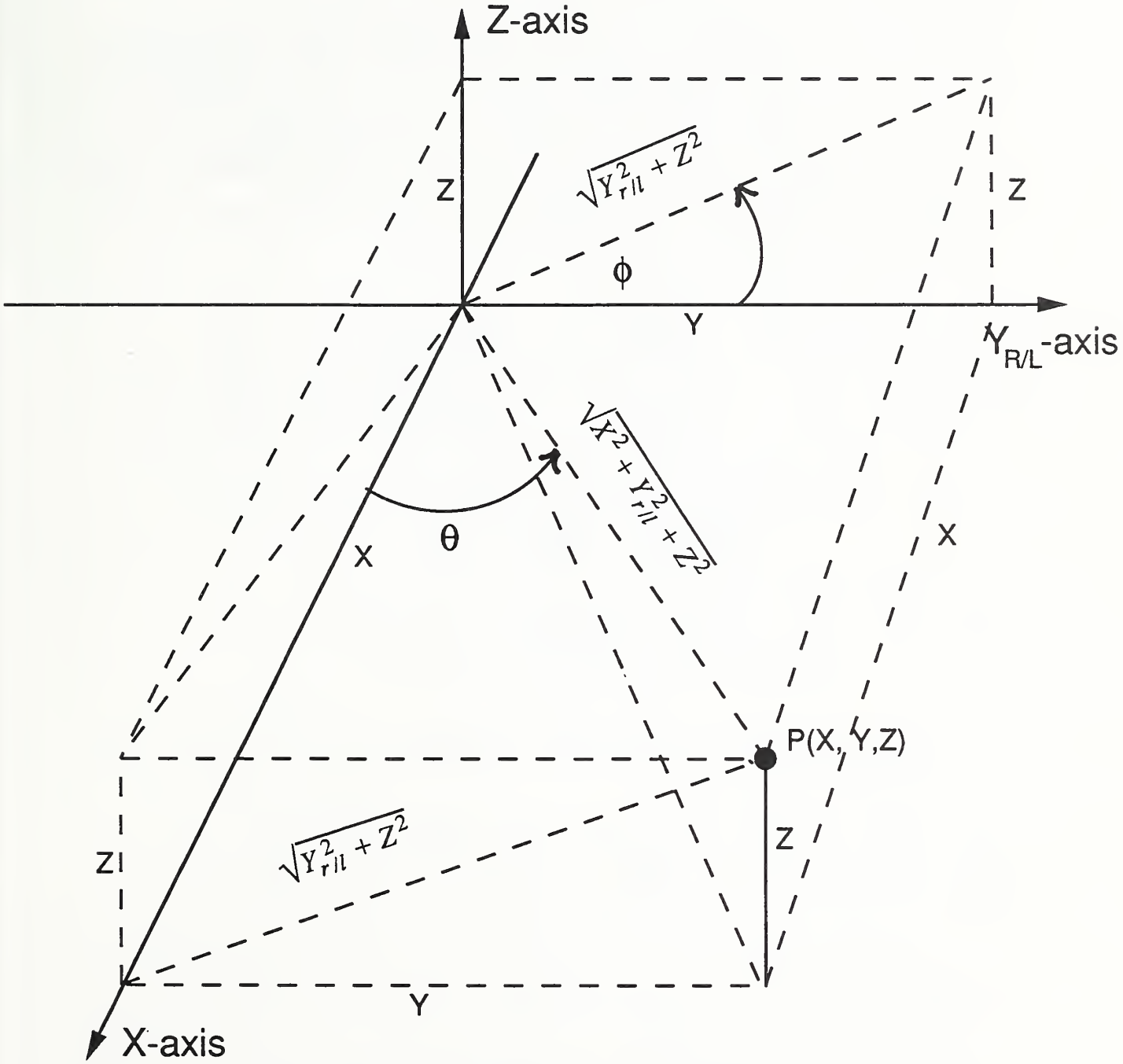


Figure C4: The relationship between a world coordinate point $P(X, Y, Z)$ and its spherical coordinates is shown. Note that the tangents for eccentricity θ and spherical azimuth ϕ can be easily read from the figure.

These equations define, respectively, the right and left bi-spherical coordinates for the camera-retinas. In addition, they define the right and left camera-retina bi-spherical projections, $p[\phi_r, \theta_r]_r$ and $p[\phi_l, \theta_l]_l$ of the point $P(X, Y, Z)$.

Bi-spherical coordinates will be used to characterize optical flow, which will now be done in analogy with section 2.1. As there, let U, V and W be the instantaneous translational velocities along, and A, B and C be the instantaneous rotational velocities about the X, Y and Z axis respectively. Further, denote by $\dot{\theta} = d\theta/dt$ and $\dot{\phi} = d\phi/dt$ the resultant optical flow, in spherical coordinates, on the camera retina. (Note that the right/left subscript will be dropped for the moment, since the resultant characterization is identical for both.)

Then the optical flow is given by

$$\begin{bmatrix}\dot{\theta}\\ \dot{\phi}\end{bmatrix} = \begin{bmatrix} \dfrac{-\sqrt{Y^2+Z^2}}{R^2} & \dfrac{XY}{R^2\sqrt{Y^2+Z^2}} & \dfrac{XZ}{R^2\sqrt{Y^2+Z^2}} \\ 0 & \dfrac{-Z}{Y^2+Z^2} & \dfrac{Y}{Y^2+Z^2} \end{bmatrix} \begin{bmatrix} U - BZ + CY \\ V - CX + AZ \\ W - AY + BX \end{bmatrix} \qquad 3.1.1.3$$

The X, Y, Z terms of the center matrix are easily found using the chain rule, i. e.,

$$\frac{d\theta}{dt} = \frac{\partial\theta}{\partial x}\frac{dx}{dt} + \frac{\partial\theta}{\partial y}\frac{dy}{dt} + \frac{\partial\theta}{\partial z}\frac{dz}{dt},$$

so that for example, the upper left hand entry is obtained by

$$\frac{d\theta}{dx} = \frac{d\ \tan^{-1}}{dx}\left[\frac{\sqrt{Y^2+Z^2}}{X}\right] = \frac{-\sqrt{Y^2+Z^2}}{R^2}.$$

Now substituting the right hand sides of definition 3.1.1.2 for X, Y and Z into 3.1.1.3 and simplifying, we have,

$$\begin{bmatrix}\dot{\theta}\\ \dot{\phi}\end{bmatrix} = \frac{1}{R}\begin{bmatrix} -\sin\theta & \cos\theta\cos\phi & \cos\theta\sin\phi \\ 0 & \dfrac{-\sin\phi}{\sin\theta} & \dfrac{\cos\phi}{\sin\theta} \end{bmatrix} \begin{bmatrix} U - BZ + CY \\ V - CX + AZ \\ W - AY + BX \end{bmatrix} \qquad 3.1.1.4$$

The most noteworthy fact of these expressions is that when optical flow is expressed as in equation 3.1.1.4 it is readily seen that the angular velocities of a moving point in space are identical to its angular velocity on the camera-retina. This is due to the property of spherical projection whereby a points angular coordinates in 2-D image space are the same as its angular coordinates in 3-D space.

3.1.2. BI-RETINAL COORDINATES

If figure C3, we have drawn another set of lines of constant elevation and azimuth on the left camera-retina L. These are the second coordinate system we will define, the bi-retinal coordinates of azimuth ψ and elevation δ. This also defines a spherical projection, which to avoid confusion, we will refer to as the *azimuthal (spherical) projection.* (Note that both coordinate frames of reference for both types of spherical projection are being defined on both camera-retinas.)

More specifically, we have the following, referring to figure C5.

Given a point $P(X, Y, Z)$, the plane of elevation containing it defines P's *angle of elevation* δ, $-90° \le \delta \le 90°$. This angle will serve as the projection elevation coordinate for both retinas for the azimuthal projection of the point P.

Unlike the elevation, which is identical for both retinas, the azimuths ψ_r and ψ_l, $-90 \le \psi_r, \psi_l \le 90°$, are distinct. They are defined to be the angle, measured in the plane of elevation, between the median plane and the vertical plane containing the point P and the respective retina center. Note that ψ_r and ψ_l are measured in opposite directions, so that if the point P is at infinity, $\psi_r + \psi_l = 0$.

Referring again to figure C5, we have the following relationships:

$$X = \frac{2\,d\,\cos\delta}{\tan\psi_r + \tan\psi_l}, \qquad \tan\psi_r = \frac{d + Y}{\sqrt{X^2 + Z^2}}$$

$$Y = d\,\frac{\tan\psi_r - \tan\psi_l}{\tan\psi_r + \tan\psi_l}, \qquad \tan\psi_l = \frac{d - Y}{\sqrt{X^2 + Z^2}} \qquad 3.1.2.1$$

$$Z = \frac{2\,d\,\sin\delta}{\tan\psi_r + \tan\psi_l}, \qquad \tan\delta = \frac{Z}{X}$$

(The tangents can be read off figure C5, while the Cartesian coordinates must be algebraically derived.)

These equations define, respectively, the right and left bi-retinal coordinates of the point $P(X, Y, Z)$. In addition, they also define the right and left camera-retina azimuthal spherical projection coordinates $p[\psi_r, \delta]_r$ and $p[\psi_l, \delta]_l$ for the point $P(X, Y, Z)$.

The relationship between bi-spherical and bi-retinal projection coordinates is given by

$$\tan\phi_{r/l} = \frac{\sin\delta}{\tan\psi_{r/l}} \quad \text{and} \quad \tan\theta_{r/l} = \frac{\sqrt{\sin^2\delta + \tan^2\psi_{r/l}}}{\cos\delta}, \qquad 3.1.2.2$$

and the inverse by

BI-RETINAL COORDINATES

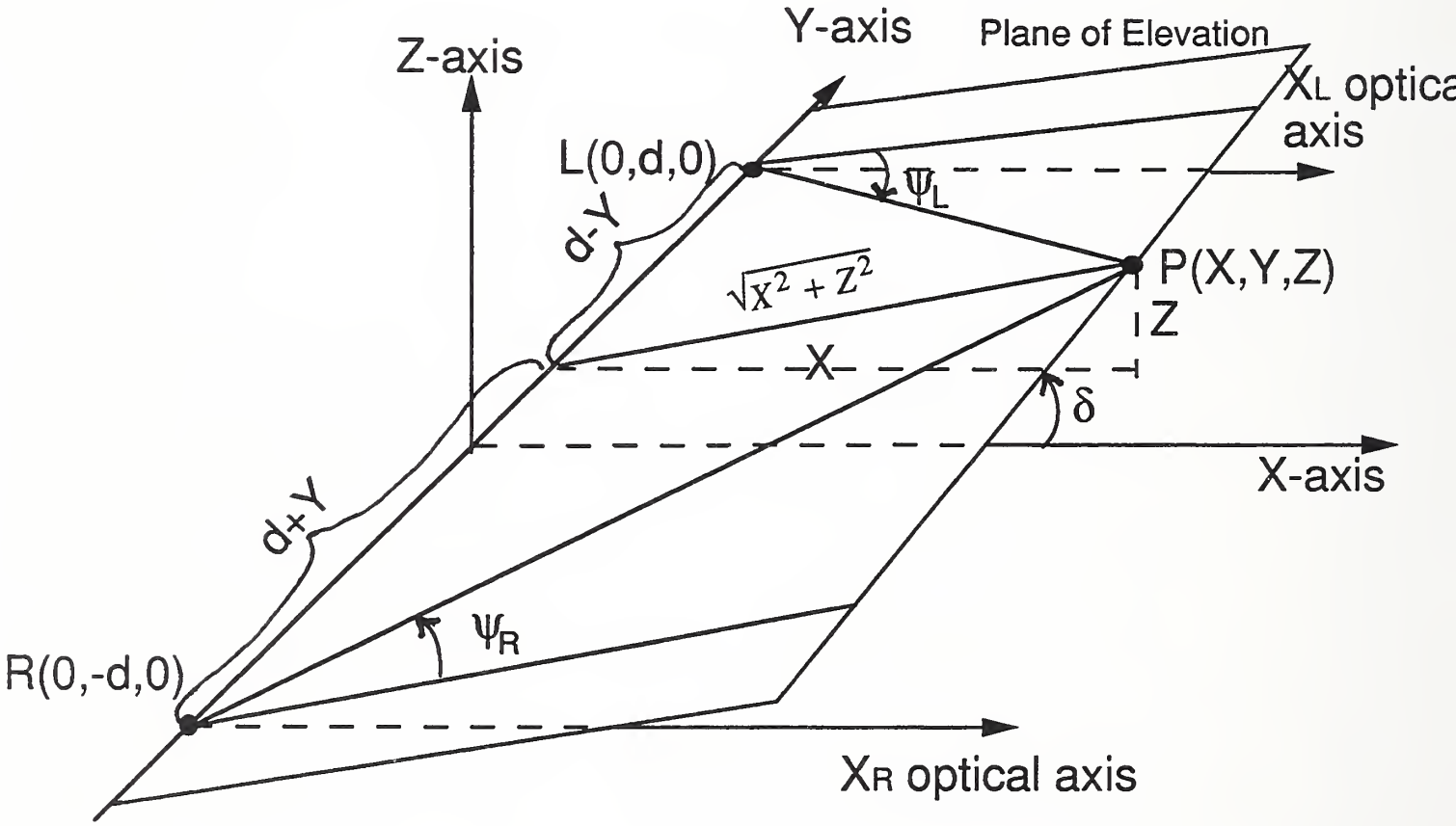


Figure C5: Bi-retinal azimuths $\psi_{r/l}$ are measured in the plane of elevation, making angle δ with the X-Y plane. The plane of elevation is determined by the world point $P(X, Y, Z)$ and the Y-axis. Note that the elevation angle δ is the same for both camera-retinas.

BINOCULAR COORDINATES

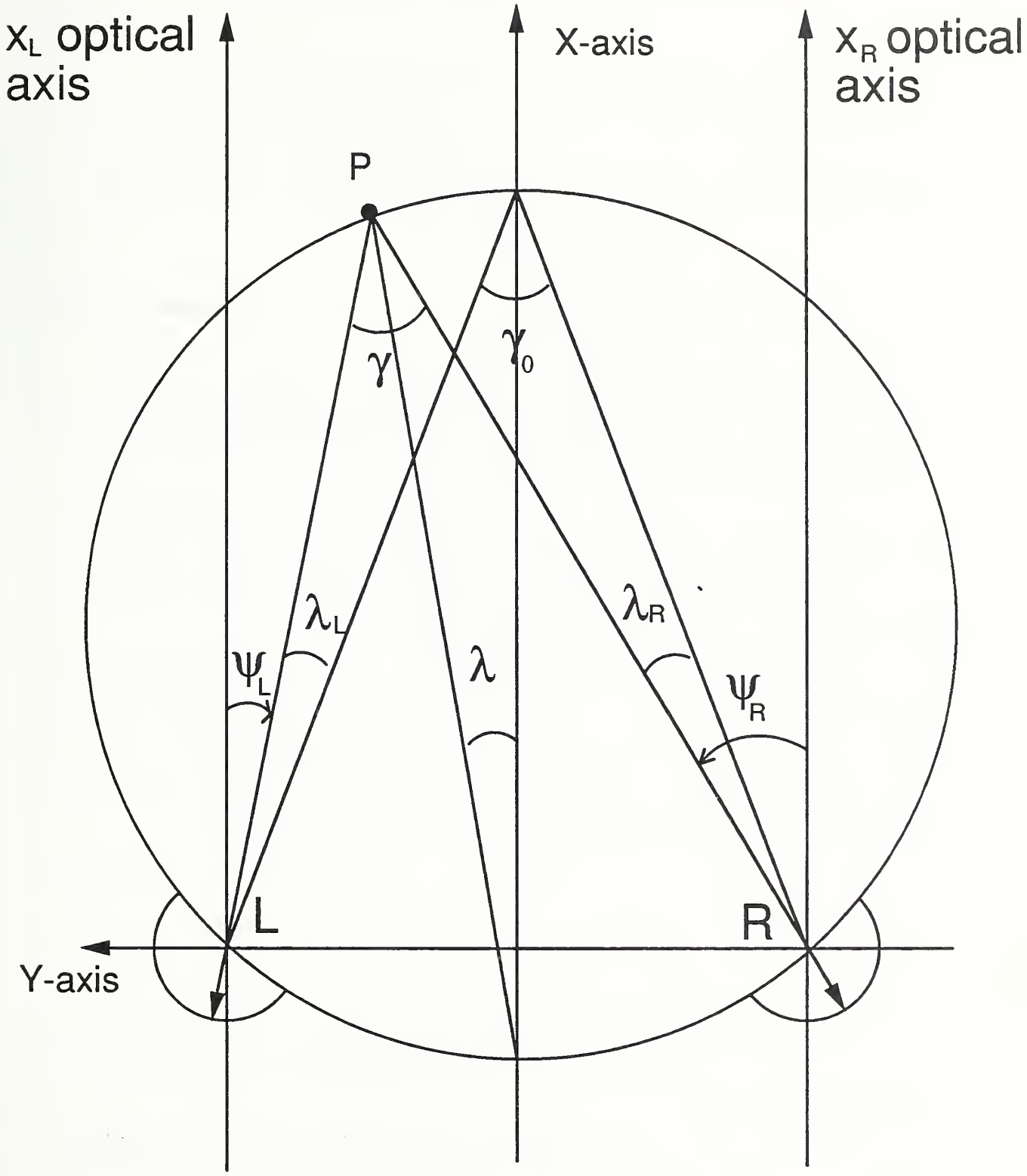


Figure C6: The binocular disparity γ is constant for all points $X > 0$ of the circle, while the binocular azimuth λ is the same for both camera-retinas, e. g., in the figure, $\gamma = \gamma_0$, and $\lambda_l = \lambda_r = \lambda$.

$$\tan\delta = \sin\phi_{r/l}\,\tan\theta_{r/l} \quad \text{and} \quad \tan\psi_{r/l} = \frac{\tan\theta_{r/l}\cos\phi_{r/l}}{\sqrt{\sin^2\phi_{r/l}\,\tan^2\theta_{r/l}+1}}. \qquad 3.1.2.3$$

For example, to see that the bi-retinal coordinates are a spherical projection, use equations 3.1.2.2 to find cos θ, cos ϕ etc., and substitute into 3.1.1.2 to obtain

$$\begin{aligned} X &= R_{r/l}\cos\delta\cos\psi_{r/l}, \\ Y &= R_{r/l}\sin\psi_{r/l}. \qquad 3.1.2.4 \\ Z &= R_{r/l}\sin\delta\cos\psi_{r/l}, \end{aligned}$$

a ''standard'' spherical projection with its ''optical axis'' rotated 90°.

Let the camera-retinas converge at infinity, i. e., assume both optical axes are parallel to the X-axis. The problem of translating binocular disparity into 3-D information may then be stated as the following:

Assume by some means that we have identified a point p which appears in the projections of both camera-retinas with coordinates $p[\psi_r, \delta]$ and $p[\psi_l, \delta]$. Then we may apply the left hand side of equation set 3.1.2.1 to compute X, Y, Z, thus locating the point P in 3-D space from its two projections.

The significance of the bi-retinal coordinates, as compared to bi-spherical coordinates, is that in order that two projected points, p_r and p_l be candidates for being the projection of the same point P, they must have the same elevation angle δ. Hence, given a projection point $p[\psi, \delta]$ in one retina, the search to find its corresponding *conjugate projection* $p'[\psi', \delta]$ in the other retina may be restricted to the line of constant elevation δ, thus restricting searches to one coordinate.

For the half-plane $X > 0$, equation 3.1.2.1 defines a constructive one-to-one mapping between a point $P(X, Y, Z)$ and its two projections $p[\psi_r, \psi_l, \delta]$. We may refer to a points coordinates either in terms of X–Y–Z, or in terms of its ψ_r–ψ_l–δ coordinates.

By rotating both camera-retinas through the angle δ, R through the angle ψ_r, and L through the angle ψ_l, the point P will be projected to $p[0,0]_r$ and $p[0,0]_l$. and we will say that the two retinas are *verged* (contraction of converged and diverged) on the point P.

For vergence, the camera retinas rotate around their respective vertical axes independently, but are required to rotate in unison about the Y axis. Hence their optical axes will will always lie in the plane of elevation for the point verged to.

Remark: Note that the rotation by δ about the Y axis results in new $\hat{X}$ and $\hat{Z}$ coordinates given by

$$\hat{X} = X \cos\delta + Z \sin\delta \quad \text{and} \quad \hat{Z} = -X \cos\delta + Z \sin\delta \equiv 0. \qquad 3.1.2.5$$

Hence one might think that the azimuths $\psi_{r/l}$ may have changed. That this is not the case can be verified by computing

$$\tan\psi_{r/l} = \frac{Y_{r/l}}{\sqrt{\hat{X}^2 + \hat{Z}^2}} = \frac{Y_{r/l}}{\sqrt{X^2 + Z^2}}. \qquad 3.1.2.6$$

The idea of vergence may be used to define the pan and tilt state vectors of the camera-retinas: When the right and left camera-retinas have been tilted by δ and panned by an amount ψ_r and ψ_l, respectively, then their optical axes will intersect at a unique point P, whose projection will be at $p[0, 0]_r$ and $p[0, 0]_l$, i. e., the respective origins of the camera-retinas. This establishes a one-to-one mapping between all camera-retina states $[pan_r, pan_l, tilt]$, and points $P(X, Y, Z), X > 0$.

In table T are listed a number of additional identities relating Cartesian, bi-spherical and bi-retinal coordinates.

In summary, bi-retinal coordinates provide, as compared to "standard" spherical coordinates, a two fold advantage in dealing with binocular disparity:

(1) Within the analytic model, they provide a geometrically simpler relationship between a projected point's retinal coordinates and its 3-D coordinates, and hence,

(2) Within the computational model, provide a simpler computation for interpreting binocular disparity.

In addition, projected feature points on the retina may be easily brought to the fovea and verged in terms of the bi-retinal coordinates. This is due to the orthogonality of pan and tilt.

3.1.3. BINOCULAR COORDINATES

We have defined $\psi_{r/l}$ for a pair conjugate projections p_r and p_l in such a way that the quantity $\psi_r + \psi_l$ is the horizontal component of the retinal disparity. When the optical axes of the two camera-retinas are parallel, i. e., verged at infinity, this quantity is called the *absolute binocular disparity* for point P.

Assume that two points, P and P', are in the visual field of the camera-retinas, and assume further that the camera-retinas are verged on point P. Then the disparity for the projections of P', $\psi'_r + \psi'_l$ is called P''s *relative disparity* with respect to P.

Relative disparity is sensed as differential range in humans. World 3-D feature points P are actively sought out and verged on, thereby making adjacent

positive/negative disparities interpretable as being either nearer or farther than the point P. Vergence takes place with respect to feature points within the fovea, so that as an adjunct of this process, the feature point of interest must be brought into this region. In this way disparities are calculated relative to the most highly resolved feature points within the image.

When relative disparity is positive, P' is closer to the origin than P and the disparity is said to be *uncrossed;* otherwise it is said to be *crossed.* This is reflective of whether the optical axes "cross" in front of or behind the point P' in verging on P.

The panning and tilting of the camera-retinas to achieve vergence is best done in coordinates reflective of the role of binocular disparity. In addition, we are interested in a coordinate system which is independent of either camera-retina, and becomes "monocular". To this end we define *binocular coordinates.* These coordinates are motivated by Luneberg's mathematical description of binocular geometry as described in [LUNEBURG, HOFFMAN].

Figure C6 depicts the geometry of these coordinates: the camera-retinas are converged at infinity, the point P has azimuths ψ_r and ψ_l and a circle is drawn through the three points R, L and P in the plane of elevation δ.

Geometrically, the binocular disparity $\psi_r + \psi_l$ is the amount by which the lines from the camera retinas differ from being parallel and hence it will be called the binocular disparity γ:

$$\gamma = \psi_r + \psi_l. \qquad 3.1.3.1$$

By a well known theorem of elementary geometry, this disparity angle is the same for all points lying on the circle, and more particularly, for the point of intersection of the circle and the X axis, i. e., in figure C6, $\gamma = \gamma_0$.

We define the *binocular azimuth* λ to be the angle made by the X axis and the line connecting the point P to the intersection of the circle and the negative X axis. This line is a bisector of the angle $\psi_r - \psi_l$ and hence is given by

$$\lambda = \frac{\psi_r - \psi_l}{2}. \qquad 3.1.3.2$$

Again by application of the same theorem, this angle is the same for both cameras as measured from their respective optical axes to the line connecting their centers to the intersection of the circle and the X axis. In figure C6, $\lambda_L = \lambda_R = \lambda$, and hence we are justified in making this definition.

$$\sin \phi_r \tan \theta_r = \tan \delta = \sin \phi_l \tan \theta_l \qquad \text{I1}$$

$$\begin{cases} X = \dfrac{\sin \phi}{\sqrt{\sin^2\phi + \tan^2\delta}} \\ Y = \dfrac{\tan \delta}{\sqrt{\sin^2\phi + \tan^2\delta}} \cos \phi \\ Z = \dfrac{\tan \delta}{\sqrt{\sin^2\phi + \tan^2\delta}} \sin \phi \end{cases} \qquad \text{I2}$$

$$\tan \phi_r \tan \psi_r = \sin \delta = \tan \phi_l \tan \psi_l \qquad \text{I3}$$

$$\frac{\tan \psi_r}{\tan \psi_l} = \frac{d + Y}{d - Y} = \frac{\tan \phi_l}{\tan \phi_r} \qquad \text{I4}$$

$$\sin \phi_r \tan \theta_r = \tan \delta = \sin \phi_l \tan \theta_l \qquad \text{I5}$$

$$\begin{cases} X = 2d \cot \delta \left[\dfrac{\tan \phi_r \tan \phi_l}{\tan \phi_r + \tan \phi_l} \right] \\ Y = d \left[\dfrac{\tan \phi_r \tan \phi_l}{\tan \phi_r + \tan \phi_l} \right] \\ Z = 2d \left[\dfrac{\tan \phi_r \tan \phi_l}{\tan \phi_r + \tan \phi_l} \right] \end{cases} \qquad \text{I6}$$

$$\tan \psi_r + \tan \psi_l = \frac{2d}{\sqrt{X^2 + Z^2}} \qquad \text{I7}$$

$$\tan \psi_r - \tan \psi_l = \frac{2Y}{\sqrt{X^2 + Z^2}} \qquad \text{I8}$$

$$\frac{\tan \psi_r - \tan \psi_l}{\tan \psi_r + \tan \psi_l} = \frac{Y}{d} \qquad \text{I9}$$

$$\tan \phi_r + \tan \phi_l = \frac{2dZ}{d^2 - Y^2} \qquad \text{I10}$$

$$\tan \phi_r - \tan \phi_l = \frac{2YZ}{d^2 - Y^2} \qquad \text{I11}$$

$$\frac{\tan \phi_r - \tan \phi_l}{\tan \phi_r + \tan \phi_l} = \frac{Y}{d} \qquad \text{I12}$$

TABLE T: Identities relating Cartesian coordinates, bi-spherical coordinates and bi-retinal coordinates.

$$\psi_r = \frac{\gamma}{2} + \lambda \quad \text{and} \quad \psi_l = \frac{\gamma}{2} - \lambda \qquad 3.1.3.3$$

The third coordinate for the binocular coordinates will remain the bi-retinal angle of elevation δ.

By substituting 3.1.3.3 into 3.1.2.1, and simplifying we obtain

$$X = d\,\frac{\cos 2\lambda + \cos\gamma}{\sin\gamma}\cos\delta, \qquad \tan\gamma = \frac{2\,d\,\sqrt{X^2+Z^2}}{X^2+Y^2+Z^2-d^2}$$

$$Y = d\,\frac{\sin 2\lambda}{\sin\gamma}, \qquad \tan 2\lambda = \frac{2\,Y\,\sqrt{X^2+Z^2}}{X^2-Y^2+Z^2+d^2} \qquad 3.1.3.4$$

$$Z = d\,\frac{\cos 2\lambda + \cos\gamma}{\sin\gamma}\sin\delta, \qquad \tan\delta = \frac{Z}{X}$$

In the horizontal plane, $\delta = 0$, we have

$$X = d\,\frac{\cos 2\lambda + \cos\gamma}{\sin\gamma}, \qquad \tan\gamma = \frac{2\,d\,X}{X^2+Y^2-d^2} \qquad 3.1.3.6$$

$$Y = d\,\frac{\sin 2\lambda}{\sin\gamma}, \qquad \tan 2\lambda = \frac{2\,X\,Y}{X^2-Y^2+d^2} \qquad 3.1.3.5$$

The locus of points having constant disparity is obtained by setting $\gamma = constant$, and yield the *Veith-Muller circles* [OGLE] through the camera retinas:

$$(X - d\ cot\ \gamma)^2 + Y^2 = \frac{d^2}{\sin^2\gamma}, \qquad 3.1.3.6$$

a circle located at $[X = d\ cot\ \gamma,\ Y = 0]$, with radius $d/\sin\gamma$. This circle is the one depicted in figure C6. The locus of points which are perceived by a particular person as having the same range, (and hence nominally zero disparity on the retina), while fixating at some point on the X-axis results in the *horopter* [OGLE] of Helmholtz, and in fact deviates from the theoretical Veith-Muller circles. This has been attributed to everything from slight image scale differences between the left and right eyes, to an as yet unknown need for perceptual space to be non-Euclidean, and has been the subject of much research in perceptual physiology. This discrepancy should not occur with artificial camera retinas.

If δ is allowed to vary, a particular Veith-Muller circle becomes the surface of a torus.

Similarly, setting $\lambda = constant$ yields the locus of points at a constant off axis angle, and is given by

$$Y^2 - X^2 + 2\,X\,Y\ cot\ 2\lambda = d^2, \qquad 3.1.3.7$$

an hyperbola whose asymptotes are given by

$$Y = \pm d\ X\ \tan \lambda. \qquad 3.1.3.8$$

For large distances, these asymptotic lines are a reasonable approximation to the points of convergence for the two camera retinas as a function of binocular azimuth.

A differential line element i. e., a short edge, $(\Delta x, \Delta y, \Delta z)$ will project to a line element $(\Delta\psi_r, \Delta\delta)$ on camera retina R, and to $(-\Delta\psi_l, \Delta\delta)$ on camera retina L.

The vertical component of disparity, $\Delta\delta$ is not extractable for reasons which are exactly analogous to the aperture problem for optical flow. The projection of the differential line element will make an angle $\theta = \tan^{-1}\Delta Y/\Delta Z$ with respect to the horizontal, and hence its horizontal disparity will be shortened by the factor $\cos\theta$. In order to determine the vertical disparity $\Delta\delta$, a third camera-retina would be needed, positioned in such a way that a vertical component could be extracted, i. e., laterally to the first two. The temporal disparity produced by optical flow may provide this.

The next section elaborates on the geometry of optical flow disparity.

3.2. SPHERICAL PROJECTION, OPTICAL FLOW AND ICONIC IMAGE REPRESENTATION

The objective of this section is to present a representation of iconic imagery which will facilitate the algorithmic extraction and interpretation of optical flow (temporal) disparities for a forward translating camera. More specifically, we are interested in optical flow as mapped by spherical projection and subsequently mapped to the plane, say for example, as would be the case in which such a lens was used with a digitizing "video chip" camera-retina.

In the case of a laterally translating camera, addressed in section 2, it was seen that the interpretation of optical flow was facilitated by the fact that the flow field was uniform or linear. That is, the geometric interpretation of a temporal disparity vector was independent of its location in the image.

For a forward moving camera, this is not the case. Instead, for a particular range, the resultant optical flow is dependent on where in the image the object falls. This true for both planar and spherical projection.

The log polar transform has been used for linearizing the planar projection of a forward translating camera [WEIMAN, MESSNER, FISHER]. Here, the analogous mapping for spherical projection is developed, along with the natural foveal-peripheral resolution which comes with it.

Here, we are interested in a 180° field of view hemispherical lens in which the resolution decreases in an analogous manner to that of the log polar mapping, and hence makes disparities both easy to extract and interpret, i. e., in a manner that is linear in camera velocity and range (or "invariant", so that disparity is not a function of image location, but only range), across the entire image.

In the top half of figure C7 we show the geometry of spherical projection: Each point $P(R, \Theta, \Phi)$ of space takes on spherical projection coordinates eccentricity θ and azimuth ϕ determined by the point on the sphere intersected by the straight line connecting the point P and the center of the sphere. The camera-retinas are hemispheres, so that the spherical projection defined here is restricted to $X > 0$.

Spherical projection lenses are readily available commercially, though mostly sized for a standard 35 mm camera. They are known generically as "fish-eye lenses". This name came about as a result of photographs published in 1906 by Robert W. Wood, [WOOD] , of Johns Hopkins University. His photographs were made using a box, emulating a "pin-hole" camera, filled with water, and due to the differences in indices of refraction for air and water, resulted in 180° field of view images, and are what he imagined a fish viewing the world of air might see.

More technically, these fisheye lenses are known as equidistant spherical projections and are designed to be used with a planar focal plane, e. g., a flat piece of 35 mm film. (See bottom half of figure C7.) Hence, there are really two projections being performed and combined into one.

In a standard 180° field of view fisheye lens a 3D point $P(R, \Theta, \Phi)$ located at range R, eccentricity Θ and azimuth Φ, $0 \le \Theta \le 90°$, $0 \le \Phi \le 360°$, is first "mapped" under spherical projection p_s to a point $(\theta, \phi) = p_s(\Theta, \Phi)$ on the sphere, which in turn is mapped by p_f to the point $(\hat{\theta}, \hat{\phi})$ in the image space circle given by

$$p_f: \quad \hat{\theta} = f\,\theta \qquad \text{and} \qquad \hat{\phi} = \phi. \tag{3.2.1}$$

This results in an image whose "magnification" or "power", or reciprocally, its "resolution", is constant along a radial line, as determined by the constant f, and hence the name "equidistant". However, the image is geometrically "distorted" when compared to the original spherical image or to a planar projection image.

As a result, these are not "true" spherical projections, these latter being available only on a portion of a sphere. In this report we will refer to these equidistant projections as *polar spherical* projections since they treat θ and ϕ as polar coordinates.

Another way of viewing this projection to the plane is to note that $d\hat{\theta}/d\theta = f$, a constant. In the next subsection this constant relationship is changed so as to

provide for a radially decreasing resolution.

3.2.1. SPHERICAL PROJECTION AND FOVEAL-PERIPHERAL RESOLUTION

In a previous section we defined the bi-spherical projection coördinates for each camera-retina. Since these are identical, in this section we will treat both in the same terms, and will refer to a single camera-retina etc.

The radial mapping of θ to $\hat{\theta}$ which we will define will provide for a central region of high constant resolution, i. e., a "fovea", with the resolution outside this region falling off monotonically as a function of eccentricity θ. For the moment we will arbitrarily introduce this mapping and justify it in the next section by showing how it solves the problem of "linearizing" optical flow for a forward translating camera-retina.

The function will be denoted by $\hat{\theta}(\theta)$, and its value by $\hat{\theta}$: (ln denotes the natural logarithm)

$$\hat{\theta}(\theta) = \begin{cases} \dfrac{\theta}{\sin F} & \theta \le F \\ \ln \tan\dfrac{\theta}{2} - \ln \tan\dfrac{F}{2} + \dfrac{F}{\sin F} & \theta \ge F \end{cases} \qquad 3.2.1.1$$

Figure R1 is the graph of this retinal mapping. Points within a circle of radius F about the optical axis constitute the "fovea", a region having both a constant and the highest resolution. Note that an angle θ becomes a distance $\hat{\theta}$ in the image plane under the mapping. To algebraically simplify the expression $\ln \tan\dfrac{\theta}{2}$ F has arbitrarily been assigned the value

$$F \equiv \tan^{-1} e^{-e} \sim 3.77° \qquad 3.2.1.2$$

in computing this graph.

In the figure there are two curves. The solid curve is the graph of $\hat{\theta}(\theta)$ and is the radial displacement of $\hat{\theta}$ as a function of eccentricity. It has been arbitrarily normalized by dividing it by its maximum, $\hat{\theta}(90°)$, whose value is

$$-\ln\frac{F}{2} + \frac{F}{\sin F} = 57.2962. \qquad 3.2.1.3$$

The dotted curve is the graph of the derivative of $\hat{\theta}$ with respect to θ. This function is

$$\frac{d\hat{\theta}}{d\theta} = \begin{cases} \dfrac{1}{\sin F} & \theta \le F \\ \dfrac{1}{\sin \theta/2} & \theta \ge F \end{cases} \qquad 3.2.1.4$$

EQUIDISTANT/POLAR SPHERICAL PROJECTION

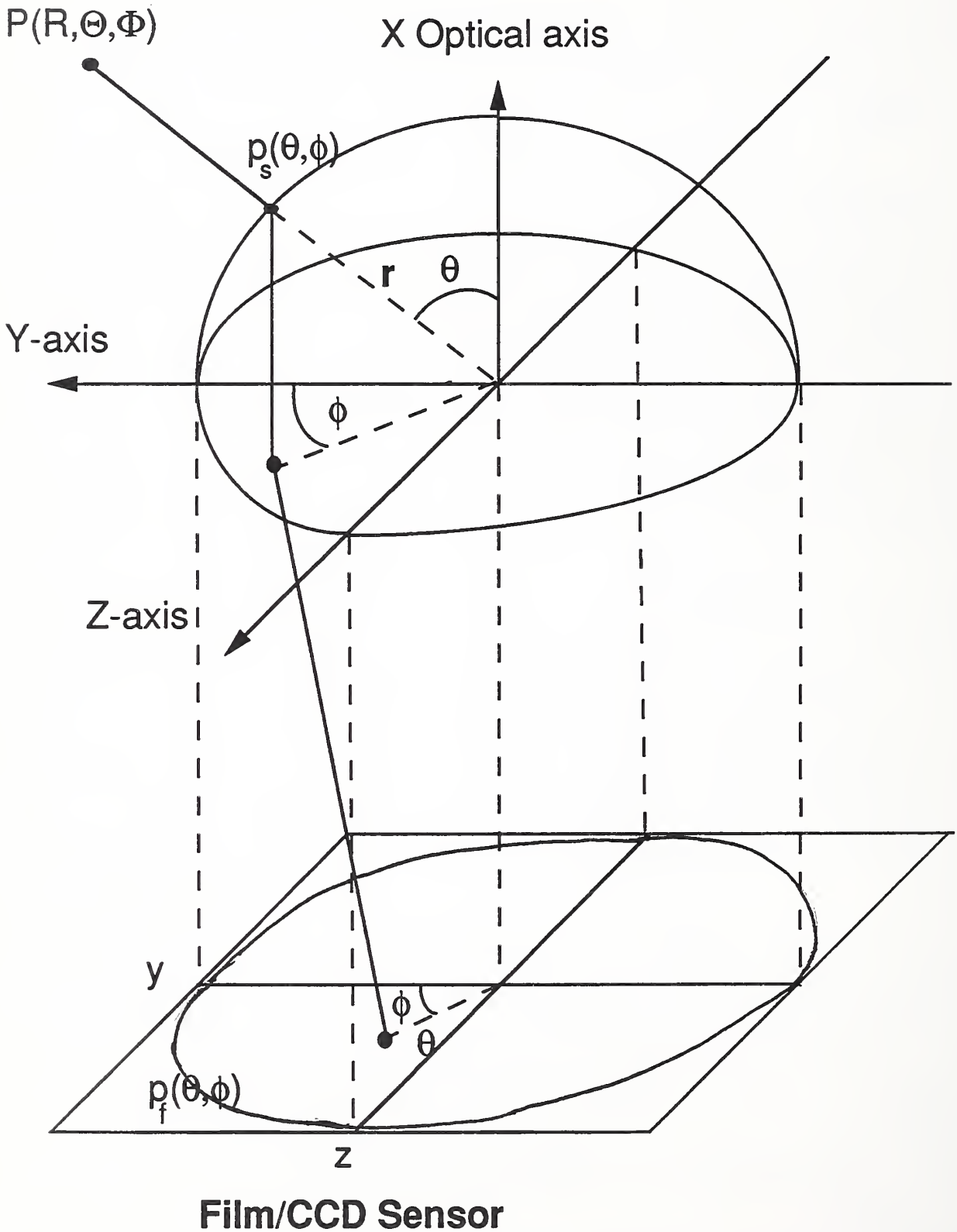


Figure C7: The projection geometry of a "fisheye" lens is shown. At top the spherical projection is shown which in turn is then mapped to the flat plane of the film. It is not a true spherical projection.

F = ATAN E**−E; S = 1.0/SIN F;
θ`=IF θ<F THEN S*θ ELSE LOG[TAN θ/2 / TAN F/2]+S*F

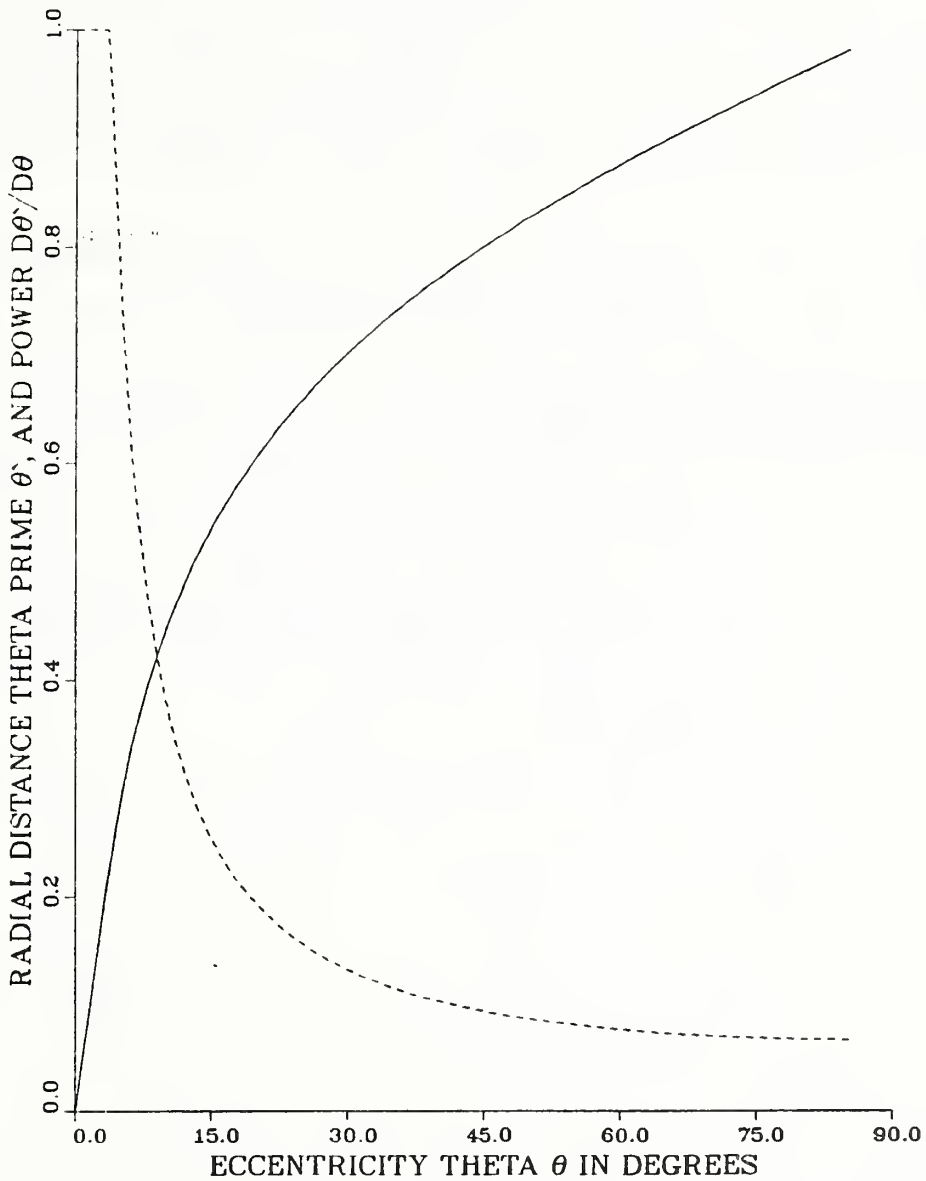


Figure R1: Graph of the foveal-peripheral retinal mapping. The solid line is the radial displacement of $\hat{\theta}$ as a function of eccentricity θ. The broken line is the derivative (normalized from 0 to 1), which has the interpretation as being the number of resolution elements per unit of viewing angle eccentricity. (Note that $\hat{\theta}$ is denoted θ ` above.)

Note that both curves are continuous at the boundary of the fovea, $\theta = F$.

The value of this derivative has the interpretation of being the number of resolution elements per degree of eccentricity, i. e., the "power" of the lens. For most lenses, this is a constant and is the reciprocal of the focal length f. This assumes that the resolution elements are uniformly located on the film, or in the case of a digitizing video camera chip, that the pixel elements are of a constant size and uniformly located.

Again, the value of the plot is normalized to a resolution of 1.0 within the fovea, and as shown in the graph, the magnification power drops off radially.

3.2.2. THE REPRESENTATION OF OPTICAL FLOW FOR A FORWARD TRANSLATING CAMERA-RETINA

Under spherical projection, translation along the optical axis with velocity v results in an optical flow field on the hemispherical camera-retina which can be identified with the change in eccentricity θ and azimuth ϕ with time, i. e., $\dot{\theta} \equiv d\theta / dt$ and $\dot{\phi} \equiv d\phi/dt$, and is related to the camera velocity v, range R and eccentricity θ by (See equation 3.1.1.4, in which $v \equiv U$ and $V \equiv W \equiv A \equiv B \equiv C \equiv 0$)

$$\dot{\theta} = \frac{v \sin \theta}{R} \quad \text{and} \quad \dot{\phi} = 0. \qquad 3.2.2.1$$

Hence, range R is related to numerically extracted optical flow by

$$R = \frac{v \sin \theta}{\dot{\theta}}, \qquad 3.2.2.2$$

and conversion of extracted optical flow to range is dependent on location θ in the image. It is the goal of what follows to demonstrate that it is possible to "warp" the image in a manner which removes this image location dependence.

Figure F2-(a) shows a polar spherical projection of the optical flow in which the camera-retina is translating toward a field of 324 points uniformly positioned on a hemisphere and hence all at the same range. The arrows connect locations in the projection at time 1 to locations at time 2, i. e., the optical flow. The flow is radially outward, along lines of constant azimuth ϕ, and their magnitude is seen to increase radially, as would be indicated by the sin θ factor in 3.2.2.1.

It is this sin θ factor which we propose to get rid of by choosing an appropriate $\hat{\theta}$ foveal peripheral mapping. That is, we want to find $\hat{\theta} = \hat{\theta}(\theta)$ so that

$$R = g(v, \dot{\hat{\theta}}) \qquad 3.2.2.3$$

for as some yet undetermined relationship g.

This can be accomplished by rewriting 3.2.2.2 as

$$\frac{d\theta/dt}{\sin\theta} = \frac{v}{R}, \qquad 3.2.2.4$$

where each side is constant for a fixed R, and hence by integrating each side,

$$\int_0^{\theta} \frac{d\theta}{\sin\theta} = \int_0^{t} \frac{v}{R} dt$$

$$\ln \tan\frac{\theta}{2} - \ln \tan\frac{\theta_0}{2} = \frac{v(t - t_0)}{R}, \qquad 3.2.2.5$$

we have the displacement on the retina which will keep v/R constant.

This mapping cancels out the $\sin\theta$ term and the resultant relation between range and extracted optical flow is then given by

$$R = \frac{v}{\dot{\hat{\theta}}} \qquad 3.2.2.6$$

This is easily verified. Noting that

$$\dot{\theta} = \dot{\hat{\theta}}/\frac{d\hat{\theta}}{d\theta} \quad \text{and} \quad \frac{d\hat{\theta}}{d\theta} = \frac{1}{\sin\theta}, \qquad 3.2.2.7$$

we have by substituting this into 3.2.2.2

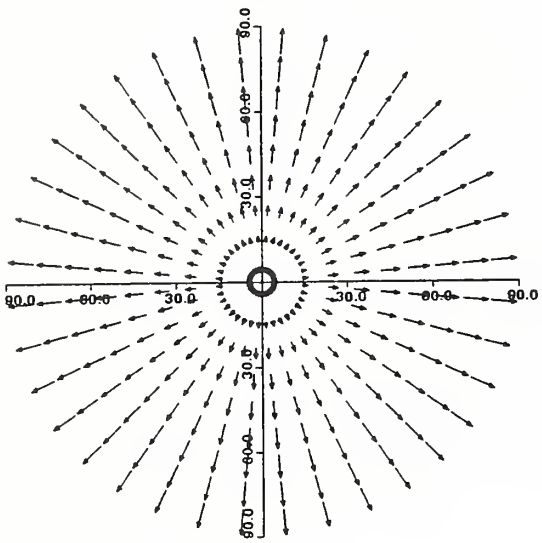
$$R = \frac{v \sin\theta}{\dot{\theta}}$$

$$= \frac{v \sin\theta}{\dot{\hat{\theta}}/\frac{d\hat{\theta}}{d\theta}}$$

$$= \frac{v}{\dot{\hat{\theta}}}. \qquad 3.2.2.8$$

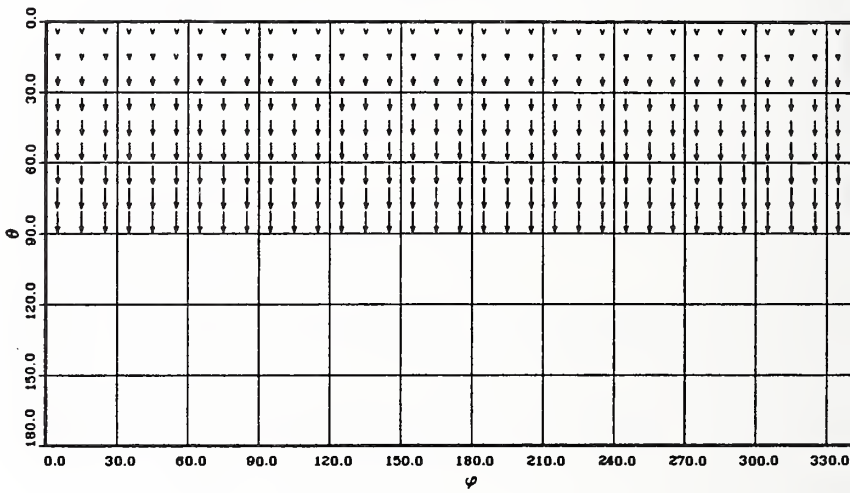
Within the region of the fovea, the mapping can be linear, as is the case for the primate and human fovea. The constant of proportionality is determined by requiring that the resolution, i. e., the derivative of $\hat{\theta}$ with respect to θ, be equal on both sides of the boundary of the fovea. This, by straightforward calculation, is $1/\sin F$. The remaining value, the amount by which the mapping must be translated, is determined by requiring that the displacement also be equal across the fovea boundary. Again, by straightforward calculation, this turns out to be $-ln\ F/2 + F/\sin F$. Putting this all together, the result is

POLAR SPHERICAL IMAGE OF OPTICAL FLOW
FOR POINTS ON FORWARD TRANSLATING SPHERE



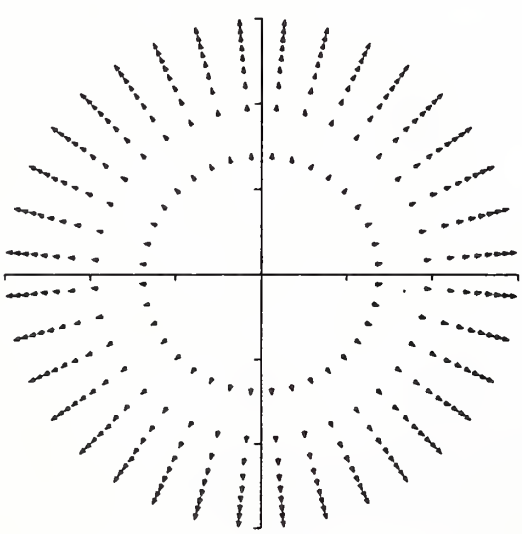
(a)

ISOMETRIC IMAGE OF OPTICAL FLOW
FOR POINTS ON FORWARD TRANSLATING SPHERE



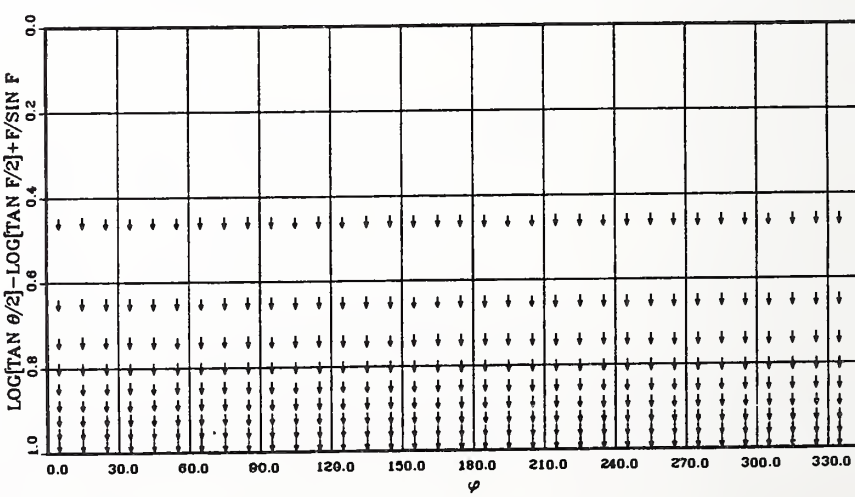
(b)

LOGARITHMIC SPHERICAL IMAGE OF OPTICAL FLOW
FOR POINTS ON FORWARD TRANSLATING SPHERE



(c)

RANGE NORMALIZATION OF OPTICAL FLOW
FOR POINTS ON FORWARD TRANSLATING SPHERE



(d)

Figure F2: The radial optical flow for a forward translating spherical field of points is shown in (a), and below it in (c), the corresponding range normalized mapping. Opposite them, (b) and (d), are the corresponding isometric representations. Note that in the normalized mappings, all optical flow vectors have the same magnitude.

$$\hat{\theta}(\theta) = \begin{cases} \dfrac{\theta}{\sin F} & \theta \le F \\ \ln \tan\dfrac{\theta}{2} - \ln \tan \dfrac{F}{2} + \dfrac{F}{\sin F} & \theta \ge F \end{cases} \qquad 3.2.2.9$$

the same as was given in section 3.2.1.

Note that for small θ, ln tan $\theta/2 \sim \theta$, so that in the region of the retina, radial flow is also linearized.

Figure F2-(c), labeled as a logarithmic spherical projection, shows a remapping of figure F2-(a) in which θ has been remapped according to equation 3.2.1.9.

We call this remapping *range normalization,* since the magnitude of optical flow disparities will be equal if and only if they are generated by points having the same range. This can also be said about optical flow extracted from contours which have been compensated by dividing by the cosine of the angle between its normal and the angle θ, in a manner analogous to that for the laterally translating camera.

The question naturally arises as to whether their might be other remappings which normalize optical flow for some other significant subset of 3-D points. The answer is yes. For example, as will be developed in the next subsection, there exists retina remappings which renormalize depth, i. e., the component of range along the optical axis of the camera retina. However, before doing this, we address the matter of image representation.

The logarithmic spherical projection is an abstract mathematical mapping and is not a representation. By representation is meant a discrete sampling, or tessellation, whose geometry (shape of the pixels) is implicitly representable in some manner, e. g., as regularly spaced samples in θ and ϕ, and hence easily mapped to conventional computer memory.

In figure F2-(c) we have depicted the image using the same polar plot of as for figure F2-(a). This representation has a non-rectangular non-regular tessellation in $\hat{\theta}$ and ϕ coordinates and hence is not easily mapped to the rectangular regular tessellation geometry of the computer. This latter pixel geometry is well suited to numerical computing of gradients, convolution operators etc., while the former is not.

Another aspect of the spherical manifold, if left unmapped to orthogonal planar coordinates, is that the spherical gradient must be used in calculating the gradient. In particular, the visual flow constraint equation for orthogonal coordinates, given in equation 2.2.1.1, must be recast in the form

$$\frac{\partial I}{\partial \theta}\frac{d\theta}{dt} + \frac{1}{\sin\theta}\frac{\partial I}{\partial \phi}\frac{d\phi}{dt} = -\frac{\partial I}{\partial t}. \qquad 3.2.2.10$$

This results in a two dimensional gradient calculation, i. e., the analog of equation 2.2.1.3, of the form

$$|\mu_\theta, \nu_\phi| = \frac{\Delta I_t}{\sqrt{(\Delta I_\theta)^2 + (\frac{\Delta I_\phi}{\sin \theta})^2}} \qquad 3.2.2.11$$

The remapping of the eccentricity θ explicitly indicates the radially decreasing resolution when mapped to a linearly addressed manifold, e. g., the graphical output shown in figure F2-(d). This is not the case for the spherical azimuth ϕ since the arc length decreases toward the fovea for a constant differential angle $\Delta\phi$.

For these reasons, the *isometric plane* [KRAKIWSKY] representation of the sphere is appropriate. In this representation the spherical azimuth ϕ is plotted at right angles to the eccentricity θ. Figure F2-(b) is the isometric plane representation of figure F2-(a). Note that for the unnormalized case, eccentricity θ varies from 0° at the top (fovea) to 180° at the bottom, which is to the rear of the camera-retina. In this report, eccentricity angles greater than $\theta = 90°$ will not be used.

Figure F2-(d) is the normalized optical flow corresponding to figure F2-(c), but in the isometric plane representation, or as we shall refer to it, the *logarithmic isometric plane,* when it is a normalized image. For the normalized case, the eccentricity domain $0° \le \theta \le 90°$ is arbitrarily mapped to the normalized range $0 \le \hat{\theta} \le 1$ for purposes of plotting.

In the logarithmic isometric plane representation, optical flow (both magnitude and orientation) for two points is equal if and only if it is the projection of 3-D points having the same range. In section 5.1.1 an experiment using the wire frame scene simulator is described demonstrating range normalization.

The isometric and logarithmic isometric representations expand the arc lengths for differential $\Delta\phi$ so that when plotted on a linearly addressed manifold, differential area is proportional to differential ϕ times differential θ. This is what is needed in order to treat all pixels as geometric equivalents.

This solution appears (very roughly) to be (one of many) used in the map from the primate retina to the visual cortex and for the same reason: resolution elements, represented by neurons all of the same size, require area proportional to their number, and hence if forced to an approximate rectangular region, will take on the form of the logarithmic isometric plane in order to minimize total inter cell linkage length [SCHWARTZ]. It is logarithmic because the cones/rods are denser in the fovea. A retina whose density of rods/cones were constant would map to the isometric plane. It is not yet known whether this is of any consequence for

biological vision.

Another important property of the logarithmic isometric plane is its "rotation and scale invariance": rotation about the optical axis results in a circular translation along the ϕ axis of the isometric representation, while a scale change, in which the camera-retina acts as the origin, results in a translation along the $\hat{\theta}$ axis.

Similar ideas have been documented in a number of places, e. g., [WEIMAN, FISHER], for the so called "log-polar" transform. The log-polar transform is the planar projection image "normalized for depth", as opposed to range as discussed above for spherical projection. The log-polar transform is the stereographic projection of the range normalized spherical projection of the same scene. In the next section, where several other normalizations will be discussed, the exact analog for the log-polar transform, "depth normalization", will be given for a forward translating spherical camera-retina.

3.2.3. OPTICAL FLOW NORMALIZATION

The numerical extraction and geometric interpretation of optical flow is facilitated by having an iconic image representation in which the underlying manifold of the representation does not enter into the computation.

This was exemplified in the previous section, where the range normalization mapping and subsequent representation as the logarithmic isometric plane, created a manifold in which two optical flow vectors are equal if and only if they come from points having the same range.

By replacing range with some other set of 3-D points we can ask the same question as was asked at the beginning of the last section: Is there some mapping of the spherical projection which will cause these points to be normalized so as to have a simple geometric interpretation, i. e., inverse mapping back to these points, based on optical flow?

In this section three additional such subsets will be described and the analogous mappings defined. These sets and their properties are based on work done by [LEE, RAVIV] where the relevant geometric properties are called "invariants".

These 1-D parameterizations of 3-D space are (the constant range points of the previous section are included for completeness):

(1) *Constant Range:* These points are characterized by being at a constant range from the camera-retina, i. e., points lying on a sphere centered at the camera-retina.

(2) *Constant Depth:* These points are characterized by lying in the plane perpendicular to the optical axis of the camera-retina, and are also referred to as constant *time to contact* points in the literature [LEE, RAVIV].

(3) *Constant Looming:* Looming is defined [RAVIV] as $-\dot{R}/R$ and is a measure of an obstacle's collision threat. Points of constant looming lie on a sphere passing through the center of the camera-retina and whose diameter is coincident with the translational vector of the camera-retina.

(4) *Constant Clearance:* These are points lying on a cylinder whose axis is coincident with the translational vector and represent points which have a constant lateral depth or clearance [RAVIV, ALBUS].

These subsets of points are easily understood in terms of the equation relating optical flow, range and camera-retina translational velocity, i. e., equation 3.2.2.4:

$$\frac{v}{R} = \frac{\dot{\theta}}{\sin\theta} \qquad 3.2.3.1$$

The instantaneous velocity v and range R are constant, while θ, or some function of θ, becomes a 1-D parameter for the subset of 3-D space. Constraining this through the use of 3.2.3.1 results in a differential relationship which upon integration yields the normalizing map in which two optical flow vectors are equal if and only if the are generated by points in this 3-D subset.

Each of these will be briefly treated, starting with (2), since (1) was covered in the previous section.

3.2.3.1. DEPTH NORMALIZATION

Instead of keeping R constant, as in the proceeding section, it may be desirable to keep depth constant. Points at a common depth X from the camera-retina are parameterized by θ and in terms of range R are given by :

$$X = R\cos\theta \qquad 0 \le \theta \le 90°. \qquad 3.2.3.1.1$$

Hence by 3.2.3.1 we have, dividing equation 3.2.3.1 by cos θ,

$$\frac{v}{R\cos\theta} = \frac{1}{\cos\theta}\frac{\dot{\theta}}{\sin\theta}$$

$$\frac{v}{R\cos\theta}dt = \frac{2}{\sin 2\theta}d\theta$$

$$\int_0^t \frac{v}{R\cos\theta}dt = \int_0^\theta \frac{2}{\sin 2\theta}d\theta$$

$$\frac{v}{R\cos\theta}(t - t_0) = \ln\tan\theta - \ln\tan\theta_0 \qquad 3.2.3.1.2$$

This last equation is the radial displacement which will keep $v/R \cos\theta$ constant. Introducing a fovea of radius F, and requiring that both displacement and the resolution across the boundary of the fovea be equal results in the *depth normalization* retinal mapping:

$$\hat{\theta}(\theta) = \begin{cases} \dfrac{2\theta}{\sin 2F} & \theta \le F \\ \ln\tan\theta - \ln\tan F + \dfrac{2F}{\sin 2F} & F \le \theta \le 90° \end{cases} \qquad 3.2.3.1.3$$

Again, it is easily verified that the set of points at a constant depth, $R \cos\theta$, is given by

$$R\cos\theta = \frac{v \sin\theta\cos\theta}{\dot{\theta}} = \frac{v}{\dot{\hat{\theta}}}. \qquad 3.2.3.1.4$$

Figure R4-(b) shows the graph of the depth normalization mapping along with its derivative. Both have been "normalized" to lie within the range 0 to 1.

In section 5.1.2 an experiment using the wire frame scene simulator is described demonstrating depth normalization.

This depth normalization is the spherical projection analog of the log-polar transform [WEIMAN, FISHER]. In section 3.3.1 this relationship is elaborated on.

3.2.3.2. LOOMING NORMALIZATION

Spheres of constant looming have been discussed in [RAVIV] and refer to points at range R lying on a circle whose diameter is given by $R/\cos\theta$. *Looming normalization* refers to the mapping of the spherical projection in such a way that two optical flow vectors are equal if and only if they lie on the same sphere of constant looming.

These points of the sphere can be parameterized by the reciprocal of the radius of the sphere in terms of θ and range R as:

$$\frac{1}{radius} = \frac{\cos\theta}{R}. \qquad 0 \le \theta \le 90°. \qquad 3.2.3.2.1$$

Hence by 3.2.3.1 we have, multiplying equation 3.2.3.1 by $\cos\theta$,

$$\frac{v\cos\theta}{R} = \cos\theta\frac{\dot{\theta}}{\sin\theta}$$

$$\frac{v\cos\theta}{R}dt = \frac{d\theta}{\tan\theta}$$

$$\int_0^t \frac{v\cos\theta}{R}dt = \int_0^\theta \frac{d\theta}{\tan\theta}$$

$$\frac{v \cos\theta}{R}(t - t_0) = \ln \sin\theta - \ln \sin\theta_0 \qquad 3.2.3.2.2$$

This last equation is the radial displacement which will keep radius curvature $\cos\theta/R$ constant. Introducing a fovea of radius F, and again requiring that both displacement and the resolution across the boundary of the fovea be equal results in the *looming normalization* retinal mapping:

$$\hat{\theta}(\theta) = \begin{cases} \dfrac{\theta}{\tan F} & \theta \le F \\ \ln \sin\theta - \ln \sin F + \dfrac{F}{\tan F} & F \le \theta \le 90° \end{cases} \qquad 3.2.3.2.3$$

Again, it is easily verified that the set of points on one of these circles whose radius is $R/\cos\theta$, is given by

$$\frac{R}{\cos\theta} = \frac{v \tan\theta}{\dot{\theta}} = \frac{v}{\dot{\hat{\theta}}}. \qquad 3.2.3.2.4$$

Figure R4-(c) shows the graph of the depth normalization mapping along with its derivative. Both have been "normalized" to lie within the range 0 to 1.

In section 5.1.3 an experiment using the wire frame scene simulator is described demonstrating looming normalization.

3.2.3.3. CLEARANCE NORMALIZATION

In the situation in which the camera-retina is translating along a straight line, the prediction of obstacles adjacent to this line which it will not "clear" is desirable. The locus of this points for a constant radius of lateral clearance is a cylinder whose axis is collinear with the axis of translation. [ALBUS, RAVIV]

A point at range R will lie on the cylinder whose radius is given by $R \sin\theta$. Hence $v/R \sin\theta$ is constant, and so by 3.2.3.1 we have

$$\frac{v}{R \sin\theta} = \frac{1}{\sin\theta}\frac{\dot{\theta}}{\sin\theta}$$

$$\frac{v}{R \sin\theta}dt = \frac{1}{\sin^2\theta}d\theta$$

$$\int_0^t \frac{v}{R \sin\theta}dt = \int_0^\theta \frac{d\theta}{\sin^2\theta}$$

$$\frac{v}{R \sin\theta}(t - t_0) = -\cot\theta - \cot\theta_0 \qquad 3.2.3.3.1$$

This last equation is the radial displacement which will keep $v/R \sin\theta$ constant. Introducing a fovea of radius F, and requiring that both displacement and the resolution across the boundary of the fovea be equal results in the *clearance normalization* retinal mapping:

$$\hat{\theta}(\theta) = \begin{cases} \dfrac{\theta}{\sin^2 F} & \theta \le F \\ -\cot\theta + \cot F + \dfrac{F}{\sin^2 F} & F \le \theta \le 90^\circ \end{cases} \qquad 3.2.3.3.2$$

Again, it is easily verified that the set of points at a constant clearance, $R \sin\theta$, is given by

$$R \sin\theta = \frac{v \sin^2\theta}{\dot{\theta}} = \frac{v}{\dot{\hat{\theta}}}. \qquad 3.2.3.3.3$$

Figure R4-(d) shows the graph of the depth normalization mapping along with its derivative. Both have been "normalized" to lie within the range 0 to 1.

In section 5.1.4 an experiment using the wire frame scene simulator is described demonstrating clearance normalization.

3.2.4. GENERALIZATION OF RANGE NORMALIZATION

(This subsection may be skipped on first reading as it does not add anything of a substantive nature, and is included here only as a lead for further work.)

In section 3.2.2 range normalization was described for the case of the camera-retina translating parallel to its X axis. In this section this motion is generalized to the case where motion is in the X–Y plane and only the horizontal line $\phi = \cos^{-1} 1$, i.e., $\phi = 0°$ and 180°, of optical flow is considered.

Referring back to equation 3.1.1.4, and letting all motion parameters be zero except for $u \equiv -U$ and $v \equiv -V$, the radial optical flow $\dot{\theta}$ is related to the camera X and Y axis velocity components u and v, range R and eccentricity θ by

$$\dot{\theta} = \frac{u \sin\theta - v \cos\theta \cos\phi}{R}, \quad \text{or,} \quad R = \frac{u \sin\theta - v \cos\theta \cos\phi}{\dot{\theta}}. \qquad 3.2.4.1$$

If $\cos\phi$ is 1, as will be assumed, then we have the following:

Denote by τ the angle, measured in the X–Y plane, between the optical axis of the camera-retina and the direction of camera-retina translation:

$$\tau = \tan^{-1}\frac{v}{u}. \qquad 3.2.4.2$$

This "phase" angle is the amount by which the origin of the range normalization must be shifted in order that it coincide with the "focus of expansion".

The resultant range normalization mapping is then given by

$$\hat{\theta}(\theta) = \begin{cases} \dfrac{\theta}{\sin F} & \theta \le F \\ \ln \tan \dfrac{\theta + \tau}{2} - \ln \tan \dfrac{F}{2} + \dfrac{F}{\sin F} & \theta \ge F \end{cases} \qquad 3.2.4.3$$

This simplifies the interpretation of the optical flow in a manner analogous to the simpler case of motion parallel to the X axis. That this is so can be seen by first noting that

$$\frac{d\hat{\theta}}{d\theta} = \frac{1}{2 \sin \dfrac{\theta + \tau}{2} \cos \dfrac{\theta + \tau}{2}}$$

$$= \frac{1}{\sin (\theta + \tau)}$$

$$= \frac{1}{\dfrac{u}{\sqrt{u^2 + v^2}} \sin \theta - \dfrac{v}{\sqrt{u^2 + v^2}} \cos \theta}, \qquad 3.2.4.4$$

and hence by substituting $\dot{\hat{\theta}}$ for $\dot{\theta}$ in equation 3.2.4.1 we have

$$R = \frac{u \sin \theta - v \cos \theta}{\dot{\theta}}$$

$$= \frac{u \sin \theta - v \cos \theta}{\dfrac{d\hat{\theta}}{dt} / \dfrac{d\hat{\theta}}{d\theta}}$$

$$= \frac{\sqrt{u^2 + v^2}}{\dot{\hat{\theta}}}, \qquad 3.2.4.5$$

which shows that the range calculation is independent of θ.

It is of interest to consider the case where $\tau = 90°$, i.e., when the camera-retina is translating laterally to the optical axis. In this case the "fovea" has shifted from $\hat{\theta} = 0$ to $\hat{\theta} = 90°$ in order to keep optical flow normalized. The derivative in this case is $1/\cos \hat{\theta}$ as opposed to $1/\sin \hat{\theta}$.

The integral

$$\int \frac{d\theta}{w + u \sin\theta + v \cos\theta} \qquad 3.2.4.6$$

has a closed form solution in terms of ln tan θ/2 and we speculate that it may represent an invariant under some more general motion for which the assumption of cos ϕ = 1 is not needed.

3.2.5. NORMALIZATION AND THE GEOMETRIC INTERPRETATION OF OPTICAL FLOW

The proposed solution to the computational problem of interpreting optical flow for the forward translating camera-retina is then a matter of exploiting the logarithmic isometric representation of the normalized optical flow. In particular, the relevant facts concerning this representation are:

(1) Optical flow is independent of θ and ϕ, and hence the interpretation of magnitude is the same throughout the image.

(2) More particularly, two optical flow disparity vectors are equal if and only if they come from the same 1-D parameterization family.

(3) For a static scene, optical flow is a function of $\dot{\hat{\theta}}$ and not of $\dot{\phi}$, and hence disparity extraction and interpretation is a one dimensional numerical extraction along lines of constant ϕ, the same as for a laterally translating camera viewing a static scene as discussed in section 2.3.1.

(4) The logarithmic isometric representation is rotation and scale invariant. This refers to the fact that relative rotation about the optical axis of the camera retina results in a translation along the ϕ axis and translating an object parallel to the axis results in a shift along the $\hat{\theta}$ axis. Thus rotation and scale are just offsets in the logarithmic isometric plane representation.

(Item (4) is included for completeness only, and will not be used in what immediately follows.)

Items (1) through (3) characterize the situation in a manner which make the geometric interpretation of optical flow analogous to that for a laterally translating camera, but with depth replaced by one of the 1-D parameterizations discussed above. The $\hat{\theta}$ and ϕ axes become the y and z axes described there.

Hence the methods described in section 2.3 are applicable. More precisely, the optical flow resulting from the introduction of moving objects into the scene can be treated in an analogous manner as was done there: extracted optical flow normal components for static objects will be shortened by the cosine of the angle they make with the "radial" direction, i. e., lines of constant ϕ. Translating objects of the

scene will have their normal components shifted as described in section 2.3.3., and hence their loci of normal components will shift in the magnitude-orientation plot.

In particular, the algorithm described in section 2.3.4 is applicable. This makes possible the segmentation of the normalized image into regions which are characterized by 1-D parameterized regions moving as distinct rigid bodies viewed from a forward translating camera-retina.

3.3. COMPARISON OF PLANAR PROJECTION AND SPHERICAL PROJECTION

Historically, computer vision research has been carried out using planar projection for both the analytic model and the computational model. We have suggested that for purposes of analysis, spherical projection has certain simplifying advantages for dealing with both binocular disparity and optical flow disparity. In this section planar projection and spherical projection will be related, primarily in terms of the log-polar transform as it compares to the normalization described in this report.

3.3.1. MATHEMATICAL MODEL RELATIONSHIPS BETWEEN PLANAR AND SPHERICAL PROJECTION

The planar projection model and associated notation will be the same as in section 2.1: upper case X, Y, Z will refer to 3-D coordinates and lower case y, z will denote image coordinates related by

$$y = f\frac{Y}{Z}, \qquad \text{and} \qquad z = f\frac{Z}{X}. \tag{3.3.1.1}$$

Note that the optical axis is aligned with the X axis and not the Z axis as is usual.

Spherical projection will be as defined by equations 3.1.1.2 and 3.1.1.3: the primary definitions are

$$\tan\theta_s = \frac{\sqrt{Y^2+Z^2}}{X} \qquad \text{and} \qquad \tan\phi_s = \frac{Z}{Y}. \tag{3.3.1.2}$$

Note that we have used the subscript s to indicate spherical coordinates, and will subscript planar projection variables with p to avoid confusion.

Figure C8 shows how they are related geometrically: the planar projection plane will be tangent to the sphere of projection at $X = f$, where f is the focal length of the planar projection and the radius of the sphere of projection.

Define polar coordinates r_p and θ_p for the planar projection by

$$r_p = \sqrt{y^2+z^2} \qquad \text{and} \qquad \tan\theta_p = \frac{z}{y}. \tag{3.3.1.3}$$

RELATIONSHIP BETWEEN SPHERICAL AND PLANAR PROJECTION

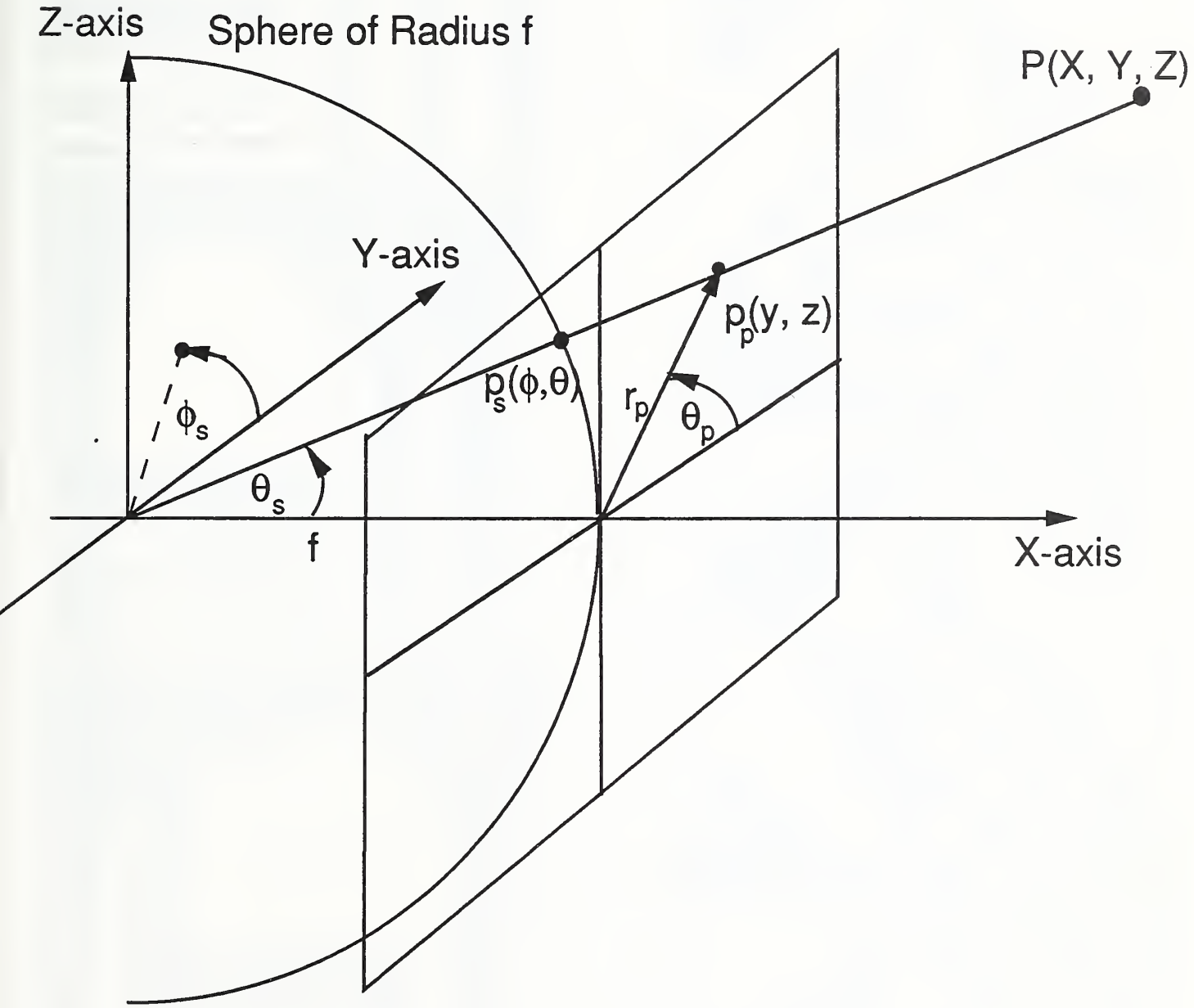


Figure C8: The relationship between a world point $P(X, Y, Z)$ and its projection $p_p(y, z)$ and spherical projection $p_s(\phi, \theta)$ is shown. The relat is $\tan \theta_s = r_p/f$ and $\phi_s = \theta_p$.

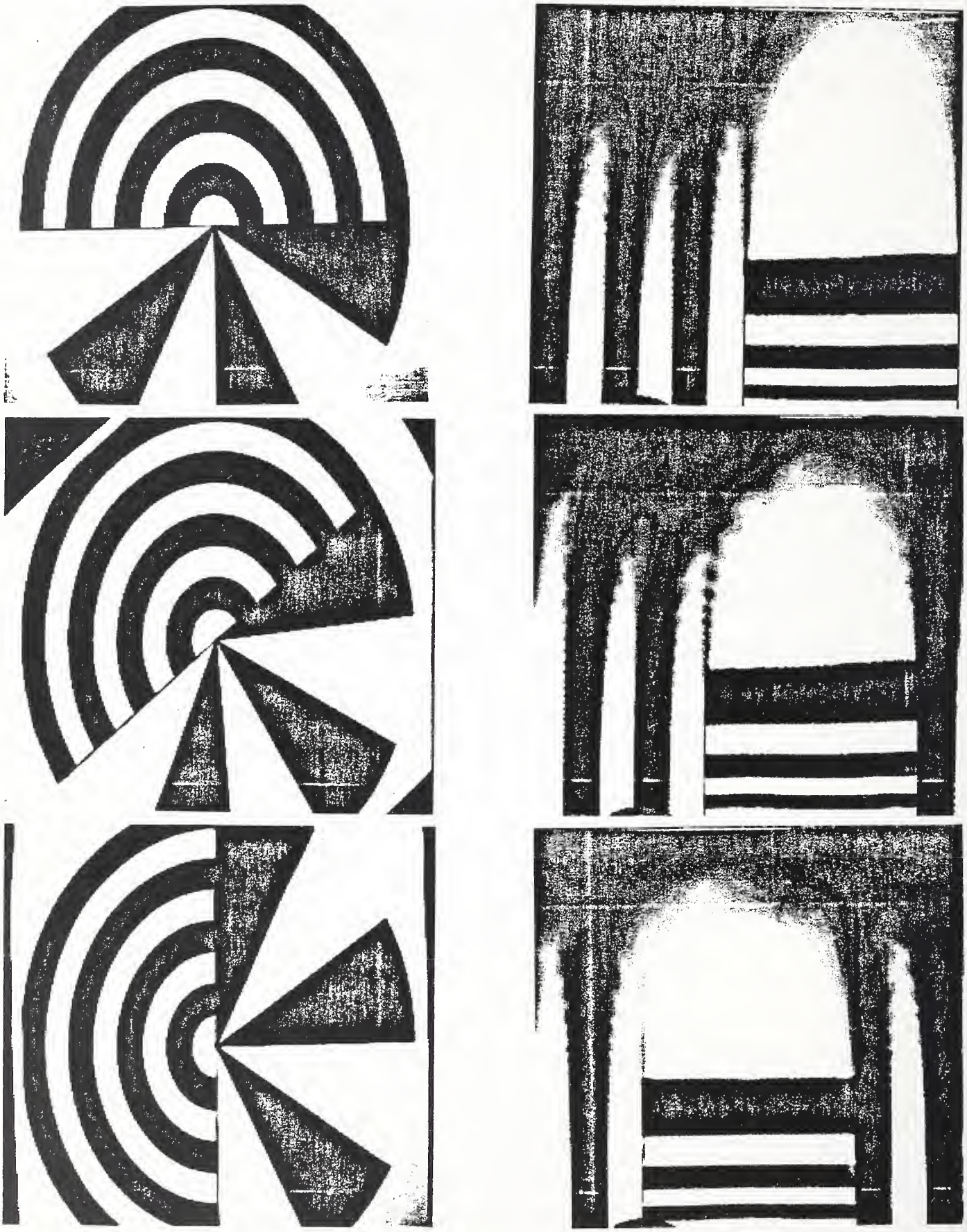


Figure L1: Log-polar Transform of Test Scene - At top left is original image and to its left, its log-polar transform. Pairs below are of the test scene rotated 45° and 90°, respectively, showing how transform is shifted circularly to the left.

Then the basic relationship between spherical and planar projection is

$$\tan \theta_s = \frac{r_p}{f} \qquad \text{and} \qquad \phi_s = \theta_p \qquad 3.3.1.4$$

This is easily seen from figure C8 or by

$$\tan \theta_s = \frac{\sqrt{Y^2 + Z^2}}{X} = \frac{1}{f}\sqrt{(f\frac{Y}{X})^2 + (f\frac{Z}{X})^2} = \frac{1}{f}\sqrt{y^2 + z^2} = \frac{r_p}{f}, \qquad 3.3.1.5$$

and the following:

$$\tan \phi_s = \frac{Z}{X} = \frac{f\frac{Z}{X}}{f\frac{Y}{X}} = \frac{z}{y} = \tan \theta_p \qquad 3.3.1.6$$

We next show how the log-polar transform and the normalizations described in this report are related.

A number of vision researchers, in which [WEIMAN] was one of the first, subsequently followed by many others, have advocated the use of the conformal mapping of planar projection images by the complex logarithm. This has been motivated by both anatomical evidence for the map in primate vision and by the mathematical properties of the transform itself [JAIN, MESSNER]. [FISHER] has reported on programmable hardware for performing this type of remapping in real time.

Properties of the complex logarithm (''log-polar'') transform include ''scale invariance'' and ''rotation invariance''. These refer to the property that concentric circles (points of constant r_p) and radial lines (points of constant θ_p) map to vertical and horizontal lines in the transform space. Hence motions along these lines, e. g., rotation and scale change are translations in the transform space.

Figure L1 demonstrates this invariance. The top left image is a planar projection of a circular disk, the upper half of which contains concentric alternating black and white circles, and the bottom half alternating black and white radial lines.

(Note: the circles are not round due to the video camera pixels not being ''square''; this was compensated for in computing the transform.)

The image to its right is the corresponding log-polar transform. In the transform, the concentric circles map to the horizontal lines on the right, and radial lines map to vertical lines on the left.

The angle ϕ is measured counterclockwise starting with $\phi = 0$ at the ''three o'clock'' position in the original image and goes from left to right in the transform image. The radial r_p of the original maps logarithmically from the top down in the

transformed image, i. e., the highest resolution is at the top.

In the images below the top pair, the circular disk has been turned by 45° and 90° with the corresponding circular shifting to the left along the horizontal in the log-polar transform.

An additional property of the log-polar transform is that it ''linearizes'' optical flow for a forward translating camera with respect to depth in a manner analogous to what has been called depth normalization in this report for a spherical projection camera-retina: The log-polar transform as applied to a planar projection image and the depth normalization mapping as represented by the logarithmic isometric plane result in the same image having the same properties. (Except for differences in sampling resolution between planar and spherical projection).

More precisely, the log-polar transform is defined by

$$\hat{w} = \ln w, \qquad (\ln \equiv \log_e) \tag{3.3.1.7}$$

where w and $\hat{w}$ are complex variables over the original and transformed images respectively, and are of the form ($i \equiv \sqrt{-1}$)

$$w = y + iz = r_p(\cos\theta_p + i\,\sin\theta_p) \qquad \text{and} \qquad \hat{w} = u(y) + iv(z). \tag{3.3.1.8}$$

The functions u and v can be expressed as functions of the modulus r_p and angle θ_p as

$$\begin{cases} u(r_p, \theta_p) = \ln r_p \\ v(r_p, \theta_p) = \theta_p \end{cases} \tag{3.3.1.9}$$

From section 3.2.3.1 the depth normalization mapping is

$$\hat{\theta}(\theta) = \begin{cases} \dfrac{2\theta}{\sin 2F} & \theta \le F \\ \ln\tan\theta - \ln\tan F + \dfrac{2F}{\sin 2F} & F \le \theta \le 90° \end{cases} \tag{3.3.1.10}$$

The critical term is ln tan θ, which by 3.3.1.4 above allows the substitution of r_p/f for tan θ, making the entire mapping outside the fovea

$$\ln\frac{r_p}{f} - \ln\tan F + \frac{2F}{\sin 2F} \tag{3.3.1.11}$$

Note that $\phi_s = \theta_p$ so that the horizontal coordinate in the two representations is the same.

This shows that given the above definition for the log polar transform for planar projection images, the depth normalized image can be generated from a

planar projection image, i. e., values of r_p and θ_p can be used to generate the logarithmic isometric plane representation of spherical projection.

The other direction, showing that depth normalization is a particular log-polar mapping, requires that a fovea be introduced into the definition of log-polar mapping by replacing w by $w+F$ in equation 3.3.1.7. This is easily done and in fact there is some evidence for doing so from the study of cortical topography [SCHWARTZ].

Figure L1 described above was generated from a planar projection image using the depth normalization mapping as it applies to r_p and θ_p. This is discussed in more detail below and in section 5.2.2, where results of a program implementing this transformation is described.

The range normalized spherical projection and the log-polar transform are related through stereographic projection of the first to the second.

More precisely, the critical term of the range normalization mapping is

$$\ln \tan \frac{\theta}{2}. \qquad 3.3.1.12$$

From equation 3.3.1.4 and using the identity $\tan \theta/2 \equiv \sin \theta/(1+\cos \theta)$ we obtain

$$\tan \frac{\theta_s}{2} = \frac{r_p}{f + \sqrt{f^2 + r_p^2}} \qquad 3.3.1.13$$

For the looming and clearance normalizations one finds by similar substitutions that

$$\ln \sin \theta_s = \ln \frac{r_p}{\sqrt{f^2 + r_p^2}} \qquad \text{and} \qquad -cot\ \theta_s = \frac{-f}{r_p}. \qquad 3.3.1.13$$

The relationships between spherical normalization and its planar projection equivalent are summarized in table T2.

The significance of these relations is that they provide the basis for computing the normalizations described in section 3.2.3 for planar projections. The significance of the spherical projection model is that it made the analysis much easier.

In the next section a method for computing the normalizations from planar projection images is briefly discussed. In section 5.2.2 examples using this method on planar projection imagery are shown.

Spherical and Planar Projection Normalization		
Normalization Type	Spherical Projection Coordinates	Planar Projection Coordinates
Range	$\ln \tan \frac{\theta_s}{2}$	$\ln \frac{r_p}{f + \sqrt{f^2 + r_p^2}}$
Depth	$\ln \tan \theta_s$	$\ln \frac{r_p}{f}$
Looming	$\ln \sin \theta_s$	$\ln \frac{r_p}{\sqrt{f^2 + r_p^2}}$
Clearance	$-cot\ \theta_s$	$\frac{-f}{r_p}$

TABLE T2

3.3.2. COMPUTATIONAL MODEL RELATIONSHIPS BETWEEN PLANAR AND SPHERICAL PROJECTION

The computational issue of interest concerns the interchangeability of planar and spherical projection images. By interchangeability is meant that given an image in one form, it may be readily converted to the other form. In particular, we first show here that the ideas developed for normalization can readily be applied to planar projection images.

We first briefly address the technical problem of image transformation.

Assume that a nonlinear mapping function g is to be applied to an input image so as to produce an output image. By nonlinear is meant that the function is not just a translation. The program for performing this must generate values at the regularly spaced predefined locations of the output image. Hence, it must evaluate the inverse of g, g^{-1} at those output image locations, yielding a location in the input image at which that image is to be sampled.

Sampling is performed by fitting a surface to the pixel values in the vicinity of the desired sampling location, followed by evaluating this surface at the desired (subpixel) location. In general, some pixels may be sampled many times while others not at all.

In the next section the inverse of the depth and range normalization mapping is derived. While this is not complex it is "tricky" to think about since things are going in the "reverse" direction. It and the following section, describing the planar projection to spherical projection transform, may be skipped if desired. Experimental

results based on these equations are described in section 5.

3.3.2.1. INVERSION OF NORMALIZATION FOR PLANAR PROJECTION IMAGE SAMPLING

Starting with the depth normalization mapping

$$\hat{\theta}(\theta) = \begin{cases} \dfrac{2\theta}{\sin 2F} & \theta \le F \\ \ln \tan \theta - \ln \tan F + \dfrac{2F}{\sin 2F} & F \le \theta \le 90° \end{cases} \qquad 3.3.2.1$$

solve the peripheral term for $\tan \theta$ obtaining

$$\tan \theta = \tan F \; e^{\hat{\theta} - 2F/\sin 2F}. \qquad 3.3.2.2$$

But $\tan \theta$ is r_p, so that we have, incorporating the fovea,

$$r_p = \begin{cases} \dfrac{\hat{\theta} \sin 2F}{2} & \hat{\theta} \le \dfrac{2F}{\sin 2F} \\ f \tan F \; e^{\hat{\theta} - 2F/\sin 2F} & \dfrac{2F}{\sin 2F} \le \hat{\theta} \le MAXROW \end{cases} \qquad 3.3.2.3$$

But since r_p is the radial distance, y, z must be calculated from (remembering that $\hat{\theta}$ and ϕ are rows and columns of the logarithmic isometric plane representation),

$$\begin{cases} y = r_p \cos \phi \\ z = r_p \sin \phi \end{cases}. \qquad 3.3.2.4$$

This x and y are the coordinates at which the planar projection image is to be sampled.

For range normalization one proceeds as above starting with the range normalization mapping. One obtains for r_p

$$r_p = \frac{-2 \tan \dfrac{F}{2} \; e^{\hat{\theta} - F/\sin F}}{\left(\tan \dfrac{F}{2} \; e^{\hat{\theta} - F/\sin F} - 1\right)\left(\tan \dfrac{F}{2} \; e^{\hat{\theta} - F/\sin F} + 1\right)} \qquad 3.3.2.5$$

In practice additional considerations must be made. For example, the value of the focal length must be given a value reflecting the value for the camera taking the planar projection image in order to not change the scale in a nonlinear way.

In a similar manner to that above, the computational formulas for generating the looming and clearance (from table T2) normalized logarithmic isometric form from a planar projective image may be found.

3.3.2.2. TRANSFORMING PLANAR PROJECTION IMAGES TO SPHERICAL PROJECTION IMAGES

The reflection of a spherical mirror is geometrically invariant under rotation of the sphere, while a reflection from a plane mirror will geometrically change under the same rotation.

One of the apparent advantages of spherical projection is that under saccadic (in analogy with the small involuntary eye rotations in observed in primates) camera-retina rotation, such images are superior due to the nondistorting effect of such rotations. That is, successive images of a spherical camera-retina are easily registered in overlapping regions of the two images. For planar projection images this is not the case. It can be argued that this distortion creates undesirable effects when integrating iconic imagery over saccades. For this reason we have investigated the transforming of planar projection images into the corresponding spherical projection images.

In this section we will indicate the computation needed for this conversion. More precisely, since a spherical projection cannot be mapped to the plane without distortion, we will give the equation which will simultaneously generate the spherical projection and map it to the plane. This later mapping will be the depth normalized spherical projection, but could just as well be the range normalized spherical projection. The polar spherical projection computation is also indicated in what follows.

The computation is nearly identical to the logarithmic plane representation of the depth normalized image. The difference is that f^{-1} is applied to the polar mapping rather than to the isometric mapping.

More precisely let $\hat{y}$ and $\hat{z}$ be the image coordinates of the output image and define $\hat{\theta}$ and ϕ by

$$\begin{cases} \hat{\theta} = \sqrt{\hat{y}^2 + \hat{z}^2} \\ \phi = \tan^{-1} \dfrac{\hat{z}}{\hat{y}} \end{cases} \qquad 3.3.3.1$$

Then r_p is calculated as in the logarithmic isometric plane representation case:

$$r_p = \begin{cases} \dfrac{\hat{\theta} \sin 2F}{2} & \hat{\theta} \le \dfrac{2F}{\sin 2F} \\ f \tan F \; e^{\hat{\theta} - 2F/\sin 2F} & \dfrac{2F}{\sin 2F} \le \hat{\theta} \le \sqrt{y_{max}^2 + z_{max}^2} \end{cases} \qquad 3.3.3.2$$

This then is followed by the conversion to coordinates of the original image:

$$\begin{cases} y = r_p \cos \phi \\ z = r_p \sin \phi \end{cases}. \qquad 3.3.3.3$$

The range normalized analog of this computation is similar. In section 5.2.1 an experimental computation for performing this latter is described.

If in the above equation for r_p, equation 3.3.3.2, r_p is set to equal $\hat{\theta}$, then the polar spherical projection results.

Computationaly converting a sequence of images taken from a rotating planar projection camera and converting them to a sequence of corresponding spherical projection images is computationally expensive. It can be speeded up by precomputing tables of manifold mappings and by representing the image by basis polynomials, e. g., Hermite polynomials, so that spatial interpolation at the subpixel level can be performed efficiently. If for other reasons the image is being sampled in nonlinear ways, such computations as the above can easily be incorporated.

4. BINOCULAR CAMERA-RETINA WIRE-FRAME DYNAMIC SCENE SIMULATOR

As a model of the theoretical work a computer program has been written which provides for the modeling of two camera-retinas viewing a scene in which the user provides parameters for the motion of wire-frame models. Output consists of a variety of graphical depictions of the scene from both a meta perspective and as viewed by the camera-retinas. Included in the camera-retina images are ones in which binocular disparities and optical flow disparities are rendered.

A major purpose of the simulator is that it be suggestive of new relationships and ideas.

4.1. OVERVIEW OF THE SIMULATOR AND ITS GRAPHICAL OUTPUT

The simulator program is designed so that the user may provide one or more rigid objects consisting of point fields and/or wire frame models, together with a description of how the object translates and rotates as a function of time within the viewing area of the simulated camera-retinas, e. g., the translational and angular velocities for each axis.

At each ''time-step'', the user may specify the type of graphical depiction (as itemized below) desired. Graphical output can be in the form of paper hardcopy, color 35 mm slides, or as 8 by 10 transparencies. In addition the program has been designed so that it is feasible to create a movie showing the continuous motion of the optical flow and binocular disparities as a function of time for some user predefined motion.

Graphical renditions available are of three types: meta depictions are those in which the entire scene including the camera-retinas and viewing box are shown, including stereographic projections of the sphere associated with a spherical projection.

The second type consists of what the binocular camera-retinas ''see'', and are plotted as polar spherical projection pairs, azimuthal projection pairs (orthogonally plotted bi-retinal coordinates ψ and δ), and as isometric and logarithmic isometric plane pairs. The logarithmic isometric types indicate in the plot itself the particular normalization function used to map θ to $\hat{\theta}$. In addition, the polar spherical type plot may also plot $\hat{\theta}$ making it a logarithmic polar spherical projection.

The third type consists of any of type two plot, but depicting (1), right and left camera-retina optical flow disparity (temporal disparity), and (2), binocular disparity plots between the right and left retina (spatial disparity). The first are in right and

left pairs, while the latter are singles since they consolidate information from both camera-retinas. These third types will be referred to as *disparity* plots.

Disparity plots may also be cumulative in time: rather than depicting just the optical flow vectors from time 1 to time 2, what is typically plotted are the vectors from time 1 to time 2 to ... time n. These will, for the same 3-D point, lie end to end. Analogously, binocular disparity vectors are plotted for a succession of times, but these move in time at right angles to the spatial disparity vector field.

An enumeration of the three types of graphical output follows, followed by an example simulation in which examples of the most relevant graphical output are shown.

- "Meta" Plots
  - (1) Meta-view of scene in x-y-z Cartesian coordinates.
  - (2) Stereo pair of planar projection images.
  - (3) Meta-view of scene in spherical coordinates.
  - (4) Meta-view of scene depicted on unit sphere retina.
  - (5) Stereo pair of stereographic projection of spherical projection.
- Retina Projections
  - (6) Stereo pair of retina in bi-retinal (azimuth ψ versus elevation δ) coordinates.
  - (7) Stereo pairs of both linear and logarithmic polar-spherical projections of retina (precursor of isometric and logarithmic isometric plane images).
  - (8) Stereo pairs of both linear and logarithmic isometric plane representations of retina.
- Disparity Projection Plots
  - (9) Binocular (spatial) disparities in binocular coordinates λ and δ plotted as an azimuthal projection.
  - (10) Linear and logarithmic polar-spherical projections of binocular (spatial) disparities.
  - (11) Linear and logarithmic isometric plane of binocular (spatial) disparities.
  - (12) Stereo pair of optical flow (temporal) disparities in bi-retinal coordinates.
  - (13) Stereo pairs of both linear and logarithmic polar-spherical projections of optical flow (temporal) disparities.
  - (14) Stereo pairs of both linear and logarithmic isometric plane of optical flow (temporal) disparities.

All the plots which produce both a right and left version are stereo pairs. With a little effort, and without any viewing device, most people can, with a little practice, fuse the two images into a single 3-D perception. Initially, this is most easily done with plots on two separate pages which can be brought nearly together, fused, and then slowly separated, yielding a larger and larger region of stereo perception. When trying this it is very important that the two images remain aligned horizontally.

In the next section we describe and show an example simulation using a box in which the purpose is to familiarize the reader with the graphical output.

4.2. EXAMPLE SIMULATION

In the example shown here unimportant quantitative detail will be avoided. Rather, the goal is to achieve some familiarity with the graphics through the use of a simple figure in a simple motion.

The example consists of stationary camera-retinas with parallel optical axes viewing a rectangular box as the box is translating towards the camera-retinas. The plots shown will consist primarily of right and left pairs at the beginning and end of this motion.

4.2.1. SIMULATOR META GRAPHICS

Figure S1-(a) and (b) show the initial and final location of the translating box within the viewing cube. This latter is not part of the scene, but rather is used to indicate the X axis (labeled X), the Y axis along the horizontal at the left etc. The units are camera-retina radii, and in terms of those units the front of the box will translate from $X = 79$ down to $X = 5$. Figures s1-(c) and (d) show two planar projection stereo meta views of the final position taken from the rear of the viewing cube and slightly behind each of the hemispherical camera-retinas. These are located at ± 3 along the horizontal Y axis. Figure S1-(e) and (f) are the corresponding views of the box projected onto the surface of the of the hemispherical camera-retinas.

4.2.2. SIMULATOR RETINA PROJECTIONS

Figure S2-(a) through (d) are the azimuthal projections of the left and right camera-retinas at the beginning and end of the translation. Note that the left azimuth and right azimuth axes are measured in opposite directions. The azimuthal projection is not a radially symmetric projection: it distorts the image in the corners due to the expansion of the poles at $\delta = \pm 90°$.

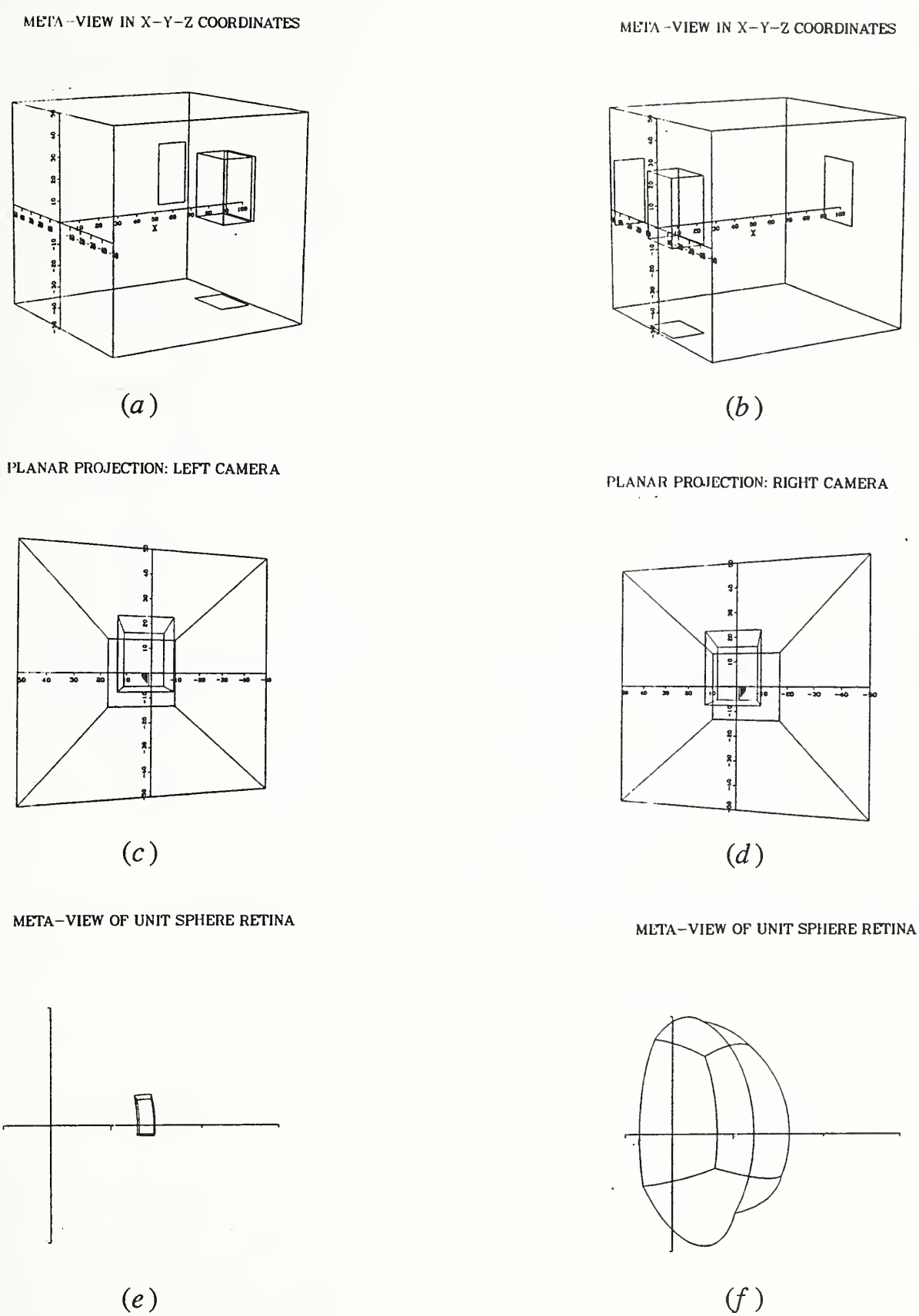


Figure S1: Example simulation for translating box. Meta view plots for the initial and ending position within the viewing cube are shown at the top, a planar projection stereo pair in the middle, and mappings of the box onto the retina at the bottom.

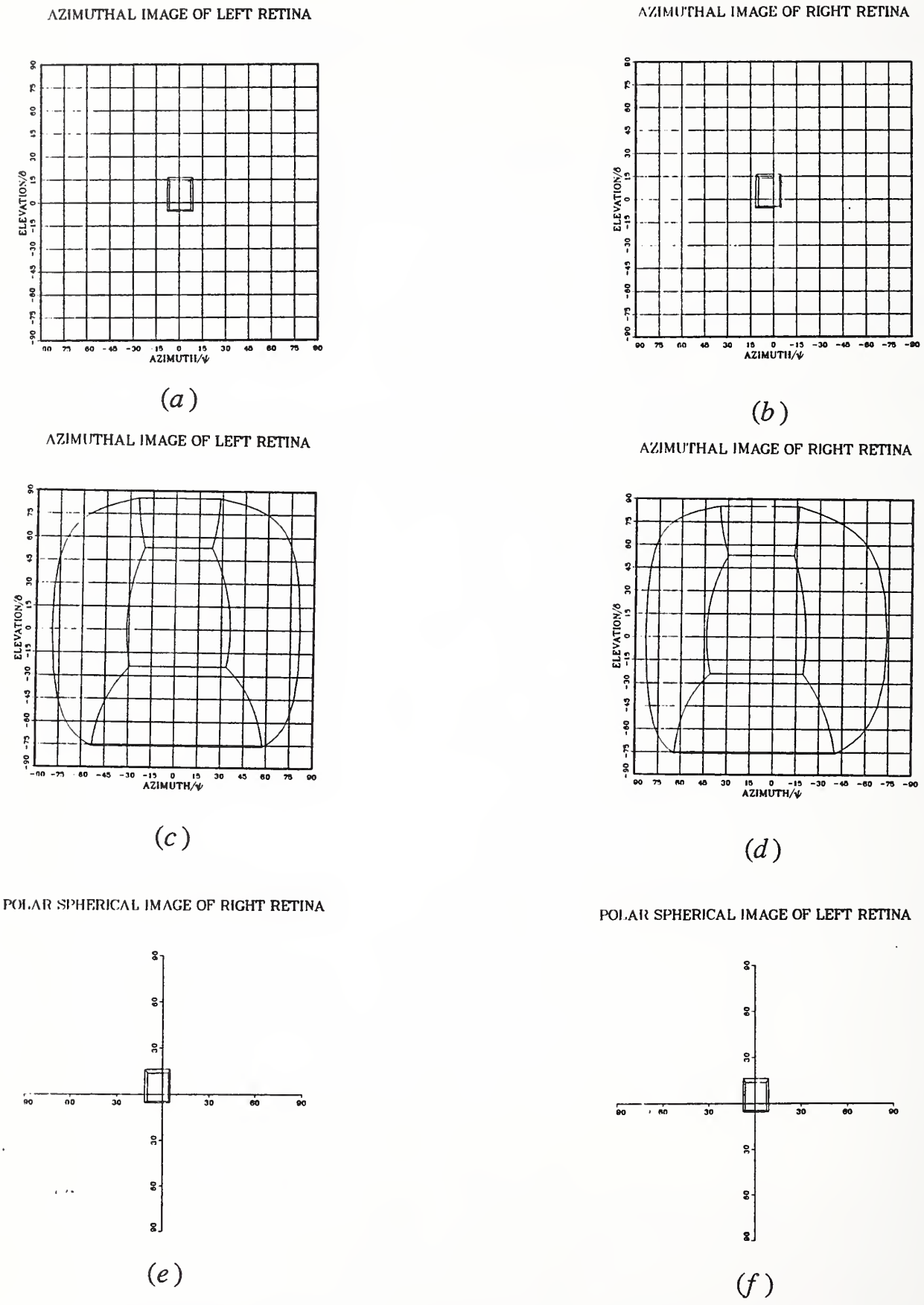
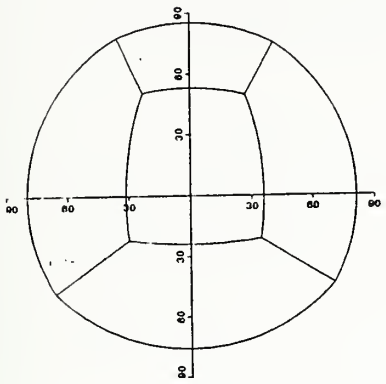


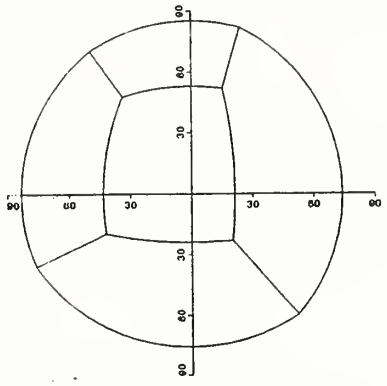
Figure S2: Example simulation for translating box. The right and left retina plotted as azimuthal projections for the box's initial and final position, and below, the right and left polar projections for the initial position.

POLAR SPHERICAL IMAGE OF LEFT RETINA



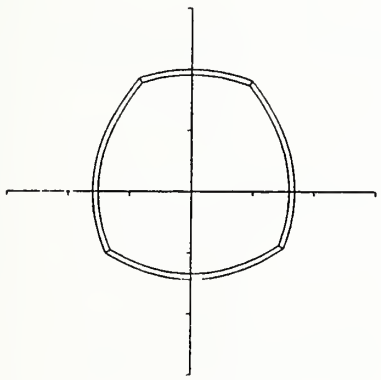
(*a*)

POLAR SPHERICAL IMAGE OF RIGHT RETINA



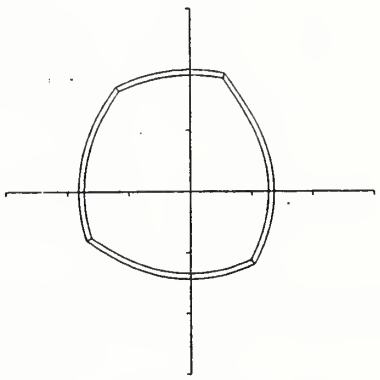
(*b*)

LOGARITHMIC SPHERICAL IMAGE OF LEFT RETINA



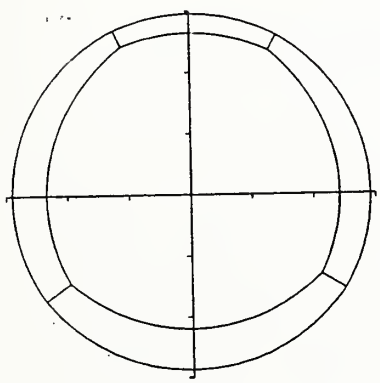
(*c*)

LOGARITHMIC SPHERICAL IMAGE OF RIGHT RETINA



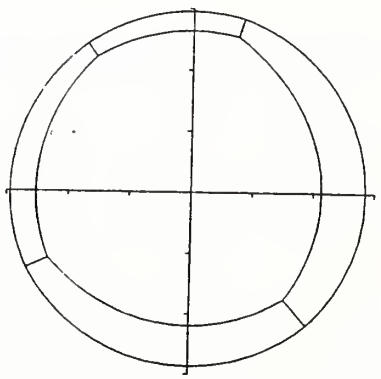
(*d*)

LOGARITHMIC SPHERICAL IMAGE OF LEFT RETINA



(*e*)

LOGARITHMIC SPHERICAL IMAGE OF RIGHT RETINA



(*f*)

Figure S3: Example simulation for translating box. The right and left retina plotted as the right and left polar projections for the final position are at the top. Below are the right and left range normalized, or logarithmic spherical projections, for the beginning and ending box position.

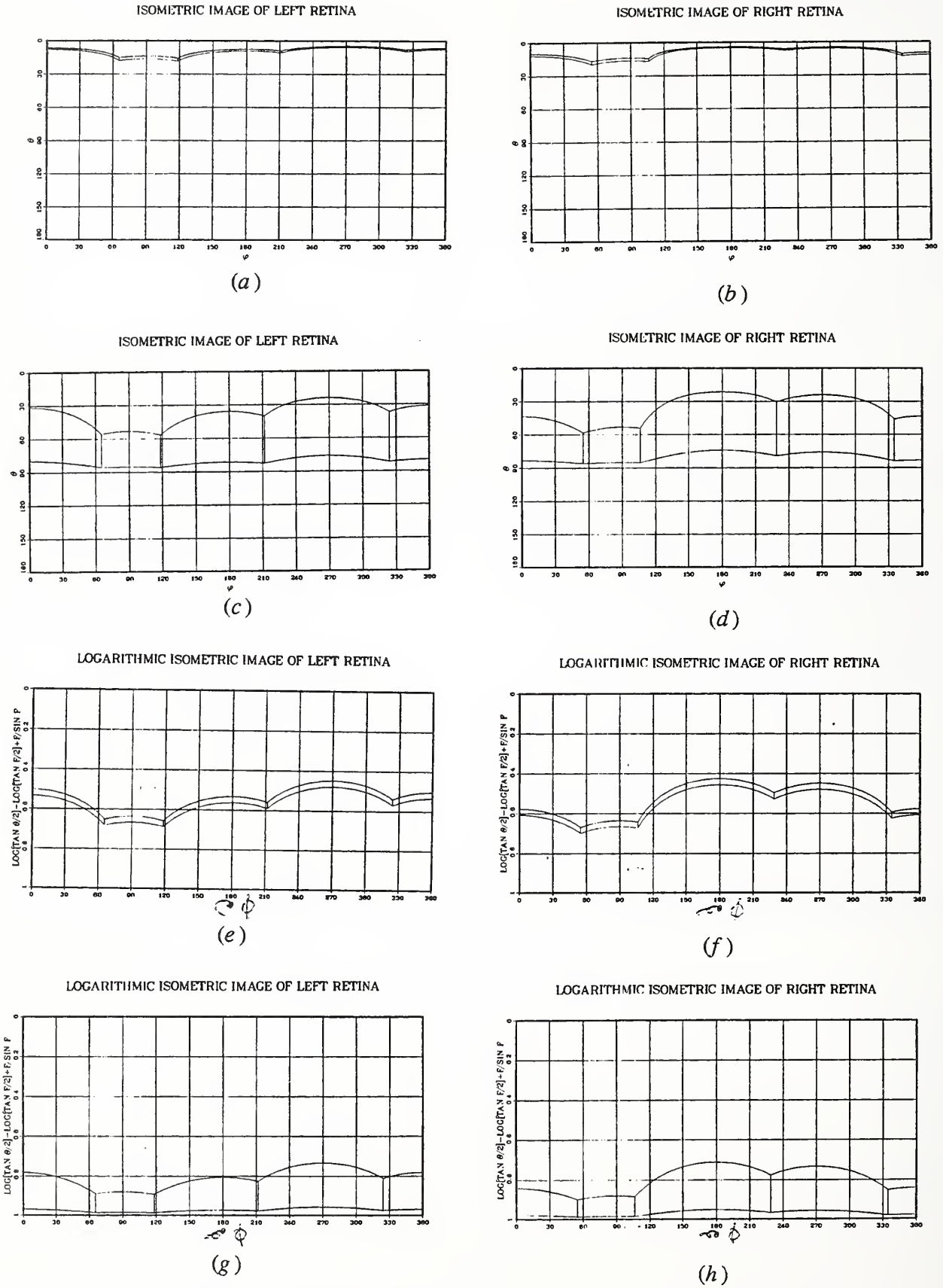


Figure S4: Example simulation for translating box. The right and left retinas represented as the isometric plane for the initial and final box position is shown in the top half. Below is the analogous logarithmic representation normalized for range.

Figure S2-(e) and (f) and S3-(a) and (b) are the left and right polar spherical projections (θ and ϕ plotted as polar coordinates) at the beginning and end of the motion respectively. This projection is radially symmetric, but because the box is not symmetrically placed about the X axis, the image is not symmetric. Here, the straight lines of the box map to portions of great circles on the sphere which in turn are distorted in the polar spherical projection of the sphere.

Figure S3-(c) through (f) are the logarithmic spherical projection analogs of the four previous polar spherical projections, i. e., they are logarithmic polar plots of $\hat{\theta}$ and ϕ in which normalization has been done with respect to range, i. e., the range normalization of section 3.2.2.

The eight plots shown in figure S4-(a) through (h) are the left and right camera-retina projections for the beginning and ending situation as represented by the isometric plane and logarithmic isometric plane.

In these representations the spherical azimuth ϕ is plotted at right angles to the eccentricity θ. For the unnormalized case, eccentricity θ varies from 0° at the top (fovea) to 180° at the bottom, which is to the rear of the camera-retina. In this report, eccentricity angles greater than $\theta = 90°$ will not be used.

For the normalized case, as represented in the logarithmic isometric plane, the eccentricity domain $0° \leq \theta \leq 90°$ is arbitrarily mapped to the normalized range $0 \leq \hat{\theta} \leq 1$ for purposes of plotting.

The normalization in the case of the logarithmic plane representation is with respect to range. (The equation of the $\hat{\theta}$ mapping is indicated on the vertical axis.)

4.2.3. SIMULATOR DISPARITY PROJECTIONS

Both optical flow and binocular disparity is computed by the simulator for each time step. However, they are typically not plotted for each time step, but are accumulated, and plotted as an overlay for some consecutive sequence of time steps.

In the example described here the number of time steps is nine. Figure S5-(a) and (b) are the azimuthal projections of the camera-retina binocular disparities at time 1, and the accumulative binocular disparities at time 9. In these plots, an arrow is from some "feature point" in the right camera-retina projection to the corresponding feature point of the left camera-retina projection. Note that the length of the arrow shaft encodes the vector magnitude. In this example the feature points are just the eight vertices of the box in order to limit the number of arrows generated to a meaningful number. However, they can be made as dense as desired along a line of the figure.

Figure S5-(c) through (f) are the corresponding polar spherical, logarithmic spherical, isometric and logarithmic isometric projections and representations of the binocular disparity accumulated over all nine time steps. The normalization for the logarithmic is again with respect to range.

In figure S6-(a) and (b) the optical flow disparity vectors connecting a feature point at time n with the location of the point at time $n+1$ for $n = 1, \cdots, n-1$ are plotted as azimuthal projections for both camera-retinas. Again, note that the length of the shaft is proportional to the magnitude of the optical flow. In the example, since the translation is parallel to the optical axes, the back four vertices trace out the same spherical meridians as the front four vertices, and hence the arrow traces are superimposed.

Figure S6-(c) through (f) show the corresponding polar spherical and logarithmic spherical projections normalized for range, and figure S7-(a) through (d) are the plots for the corresponding isometric and logarithmic isometric representation.

For plots which are much denser and when viewed as a stereo pair utilizing the binocular disparity, these plots of optical flow provide a very nice 3-D surface of the 3-D motion.

4.3. EXPERIMENTAL VERIFICATION OF THE SIMULATOR

In this section we describe four experiments using the simulator. These experiments consist of using the simulator to translate or rotate a point field for which some aspect of the resulting optical flow disparities and/or binocular disparities is intuitively predictable, e. g., zero or constant or trace out a straight line etc.

These experiments serve the dual purpose of providing in some detail what the simulator generates and also provides some verification that it is working correctly for these simple cases. This hopefully will increase its credibility in cases which are not so intuitively capable of being modeled.

An overview of each of these four verification experiments is given next. This will be followed by a subsection for each experiment in which a detailed description of the experiment, its output and the conclusions which may be drawn from it are given.

V1: *Collapsing Sphere* In this verification experiment a spherical field of points, concentric with a single retina, is located so that the optical axis of the retina passes through the field. The radius of the sphere on which the points lie is then incrementally decreased causing the points to move radially toward the spherical retina center. This should result in a zero magnitude optical flow on

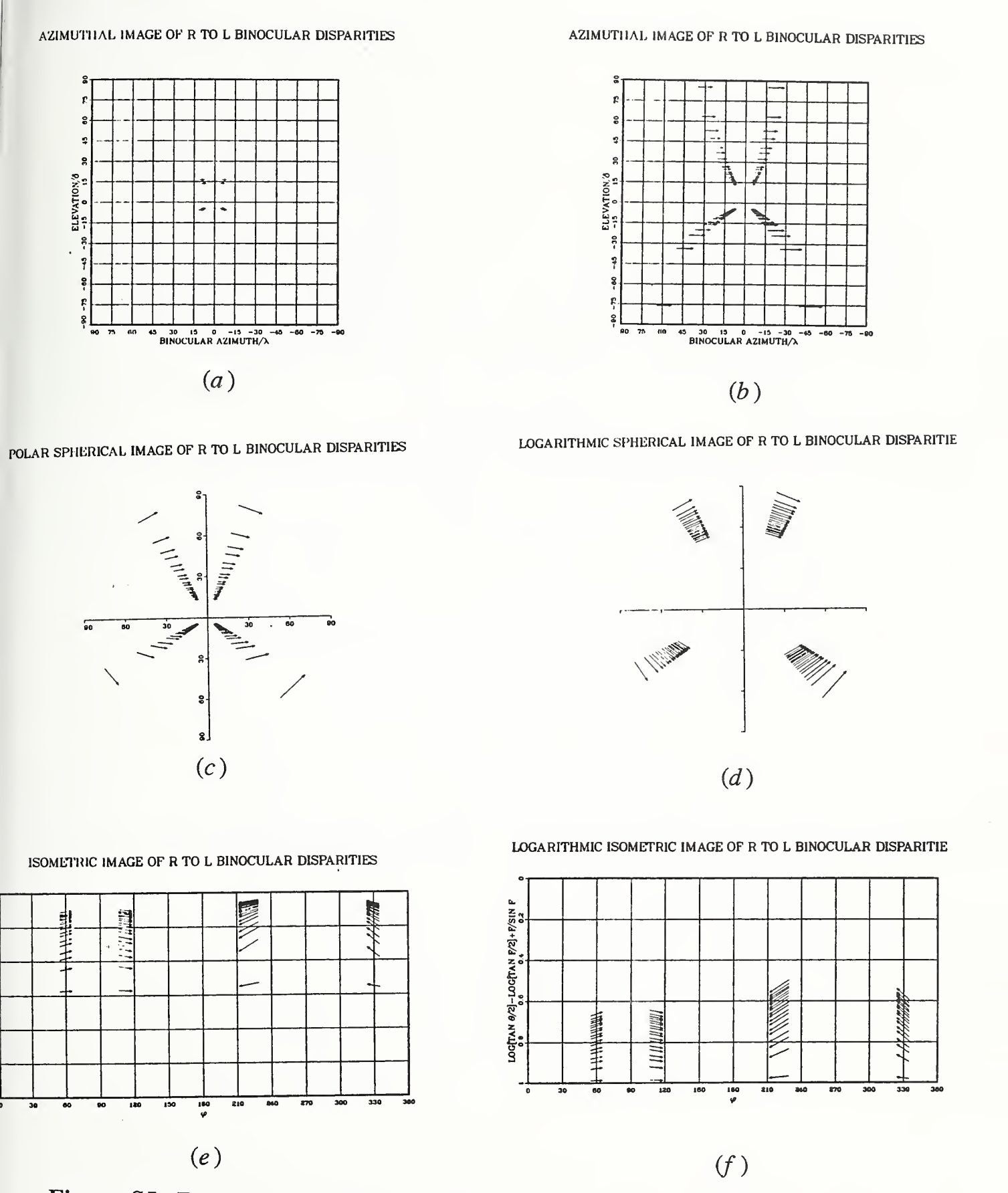
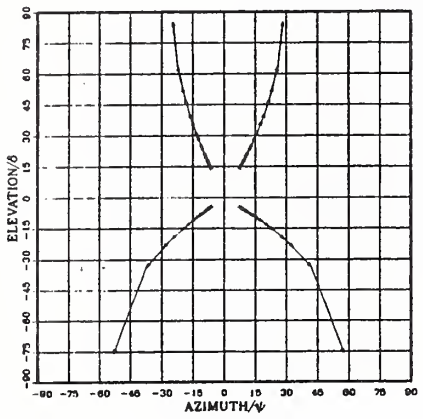


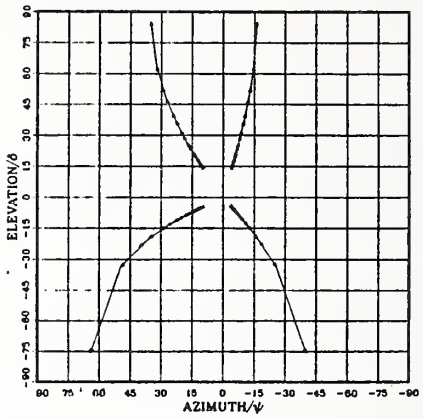
Figure S5: Example simulation for translating box. Binocular disparity, plotted as an azimuthal projection, for the initial position is at top left, and the accumulative over nine positions is at the top right. Below them are the accumulative unnormalized and range normalized polar spherical projections and their representations in the isometric plane.

AZIMUTHAL IMAGE OF LEFT RETINA OPTICAL FLOW



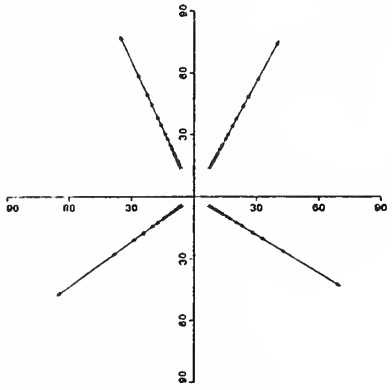
(a)

AZIMUTHAL IMAGE OF RIGHT RETINA OPTICAL FLOW



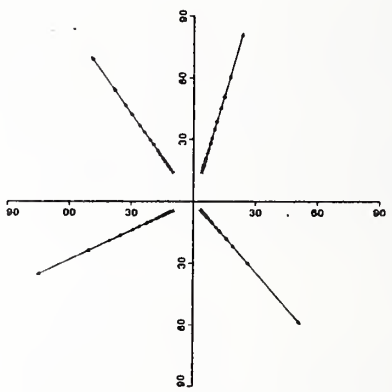
(b)

POLAR SPHERICAL IMAGE OF LEFT RETINA OPTICAL FLOW



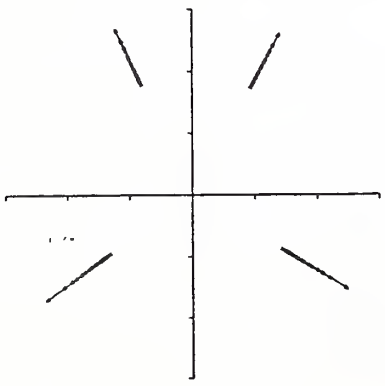
(c)

POLAR SPHERICAL IMAGE OF RIGHT RETINA OPTICAL FLOW



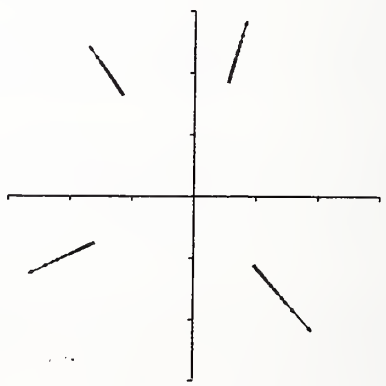
(d)

LOGARITHMIC SPHERICAL IMAGE OF LEFT RETINA OPTICAL FLOW



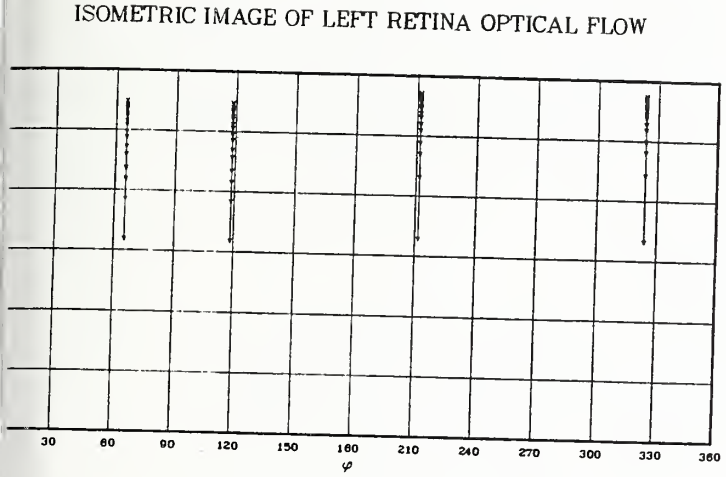
(e)

LOGARITHMIC SPHERICAL IMAGE OF RIGHT RETINA OPTICAL FLOW

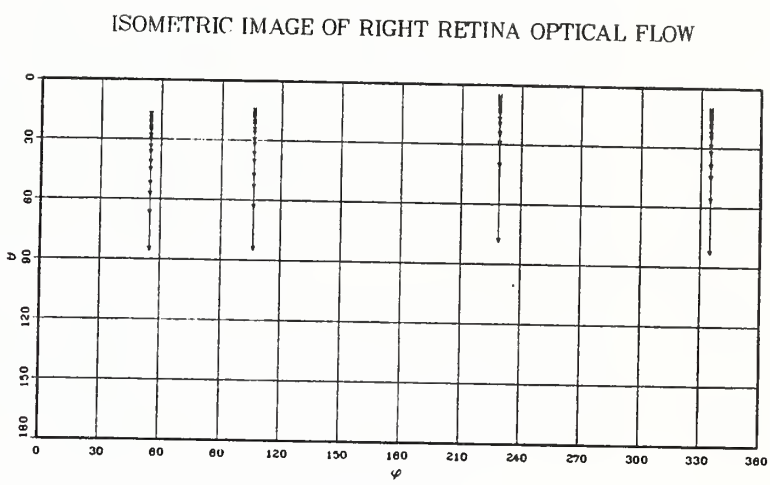


(f)

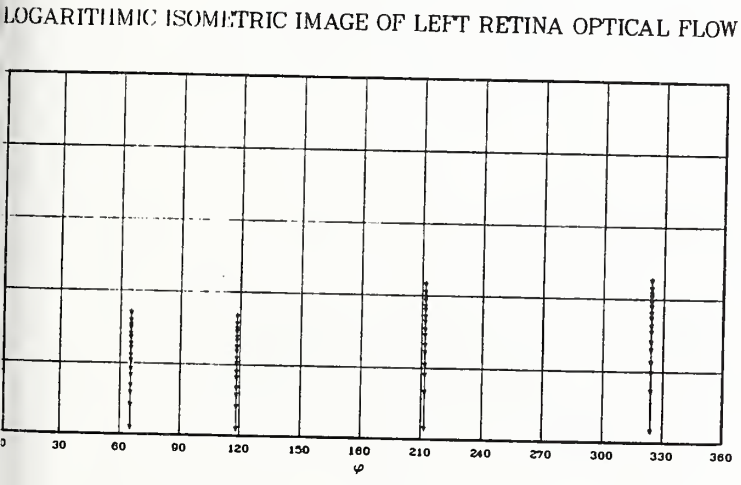
Figure S6: Example simulation for translating box. Accumulative optical flow disparity plotted as right and left azimuthal projections, polar spherical projections, and as range normalized spherical projections.



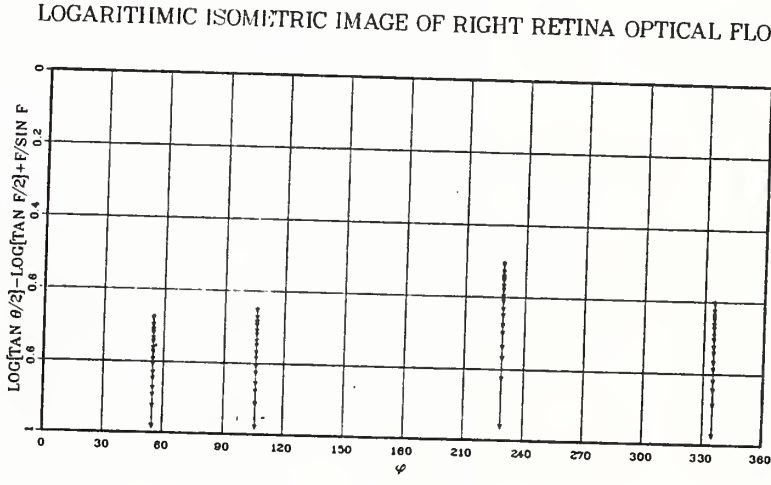
(a)



(b)



(c)



(d)

Figure S7: Example simulation for translating box. Accumulative optical flow disparity plotted in right and left isometric and range normalized isometric representations.

the retina and it is the purpose of this verification experiment to demonstrate this.

V2: *Forward Translating Sphere* In this verification experiment a field of points is again located on a portion of a sphere, which in turn is concentric with a single retina. The field is located so that the optical axis of the retina is made to pass through the center of the field. The points then are made to translate as a single rigid body parallel to the optical axis. The resulting optical flow for a point should have the instantaneous value

$$Optical\ flow = \frac{velocity\ \sin\theta}{range},$$

where θ is the angle of eccentricity for the point. Since both the instantaneous velocity and range are constant, dividing this value by sin θ should result in the constant $\frac{velocity}{range}$ for all points. This purpose of this experiment is to verify this graphically and by tabulating the constant values as a function of eccentricity.

V3: *Constant Binocular Disparity Circles* The purpose of this verification experiment is to verify that points on a Veith-Muller circle have a constant disparity. A Veith-Muller circle is a circle passing through the centers of the two spherical retinas and lying in the plane containing the optical axes. It can be shown that points lying on this circle are imaged to points on the retinas having a constant difference of horizontal angle. Hence all points on this circle may be verged simultaneously, i. e., the optical axes are rotated so as to cause them to intersect at a point on the circle for a zero disparity. Then all other points on the circle will also have a zero disparity.

The experiment consists of graphing and tabulating the disparities for several such circles in which the optical axes are held parallel.

V4: *Rotation of Plane of Elevation* When the camera-retina tilts in order to bring a point into vergence, the bi-retinal azimuths must not change. The purpose of this last verification experiment is to confirm that the simulator is correctly computing these rotations. In particular, it confirms that the bi-retinal azimuth angles are independent of δ, as was indicated in section 3.1.2 in equation 3.1.2.5.

The experiment consists of graphically depicting the results of rotating a horizontal line about the Y axis and noting that the bi-retinal azimuths do not change.

4.3.1. VERIFICATION EXPERIMENT V1: COLLAPSING SPHERE

A point moving directly toward the point of projection is translating along a ray of projection and its optical flow should be zero. This experiment demonstrates this for 180 points uniformly located on the surface of a sphere. Only a single retina is used and the center of that retina, i. e., its point of projection, acts as the center of the sphere(s) on which the points lie.

Figure V1 shows the graphics produced by the simulator. Figures V1-(a) and V1-(b) show the initial location and final location of the 180 points within the viewing box. The radius of the initial spherical field of points is 90, and this is decremented by 10 down to 10 in nine time steps.

The field of points is uniformly spaced in θ–ϕ spherical coordinates, $5° \le \theta \le 45°$, $5° \le \phi \le 355°$, $\Delta\theta = \Delta\phi = 10°$, five rings each containing thirty-six points. The projection of these points should be uniformly located on the isometric plane within these ranges. This in fact is the case as is depicted in figure V1-(f).

In the projected images shown in figure V1-(c) through (g), each depicting the resultant optical flow normally indicated by vectors, we would not expect to see any vectors for this particular motion. However, the simulation program compares the magnitude of the optical flow against the angle 0.00001°, and if the magnitude is less than that value, it plots a small cross at the projection location rather than the normal optical flow disparity vector.

(The angle ω subtended by two points whose coordinates are given in spherical coordinates θ and ϕ is calculated by

$$\cos \omega = \sin \theta_1 \sin \theta_2 + \cos \theta_1 \cos \theta_2 \cos (\phi_2 - \phi_1),$$

and for a unit sphere is equal to the disparity magnitude.)

As evidenced by these figures of the projected optical flow, the motion of the points from their initial location to their final location resulted in crosses indicating a value of less than 0.00001°, (in fact, the largest value was 0.0000068°, a not unreasonable error given the several forward and inverse trigonometric functions involved), and hence we may infer the correctness under these conditions of the simulator's computation of optical flow.

4.4. VERIFICATION EXPERIMENT V2: TRANSLATING SPHERICAL FIELD

The optical flow $[\dot{\theta}, \dot{\phi}]$, generated by a point at range r in the field of view of a spherical projection camera-retina, in which the forward relative translational velocity is $\dot{x}$, is predicted by

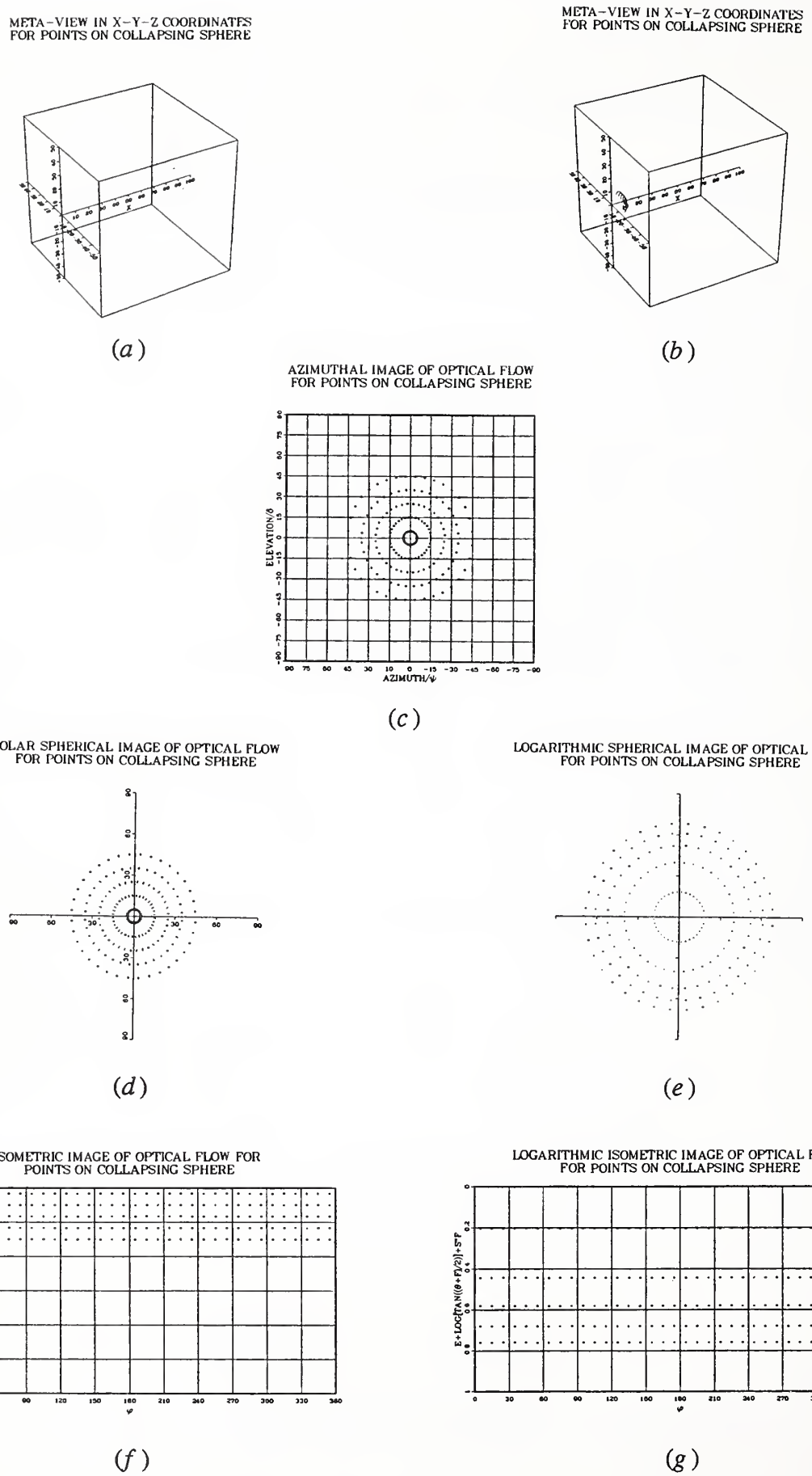
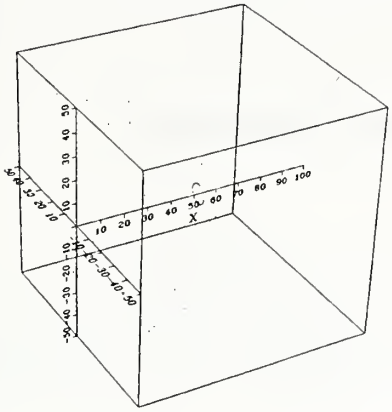
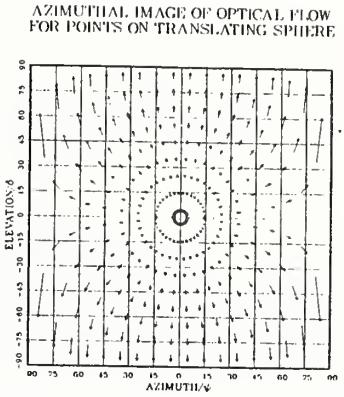


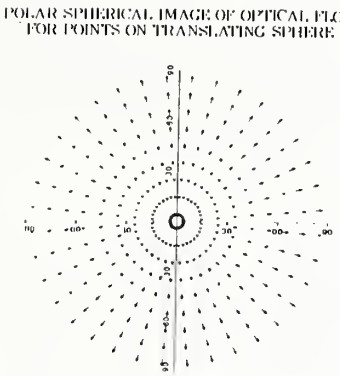
Figure V1: Collapsing Sphere Verification Experiment - Output of Simulator for 180 points collapsing toward retina center indicating zero optical flow.



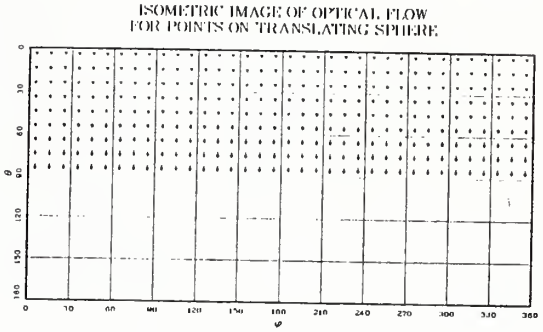
(a)



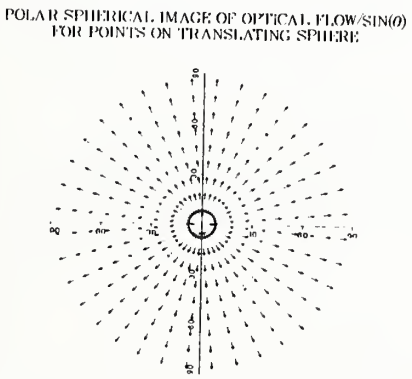
(b)



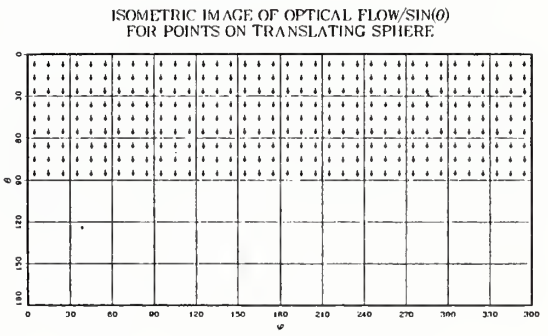
(c)



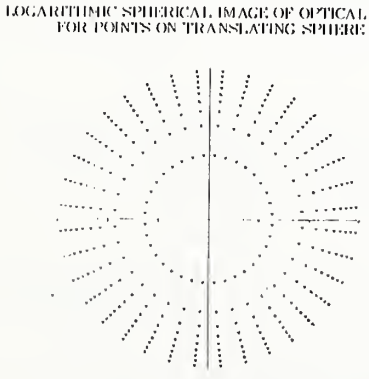
(d)



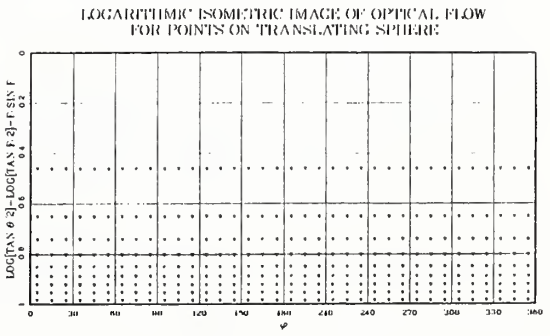
(e)



(f)



(g)



(h)

Figure V2: Forward Translating Sphere Verification Experiment - Output of simulator for 324 points all at the same range translating toward retina. Note that (e) and (f) show *optical flow*/sin θ, a constant.

$$\dot{\theta} = \frac{\dot{x} \sin \theta}{r}, \quad \text{and} \quad \dot{\phi} = 0.$$

This experiment demonstrates this for a field of 324 points uniformly located on a hemisphere of radius 50, oriented so that the optical axis of the single camera-retina pierces it at its center.

Figure V2 shows the graphics produced by the simulator. Figure V2-(a) shows the initial location of the field of of 324 points within the viewing box. (The second position is not shown as it is nearly identical.) The first location is such that the hemisphere intersects the X-axis at $x = 52$, and the second analogous position is at $x = 48$, having translated an amount $\Delta x = -4$ in one time-step.

The field of points is again uniformly spaced in θ–ϕ spherical coordinates, $5° \leq \theta \leq 85°$, $5° \leq \phi \leq 355°$, $\Delta\theta = \Delta\phi = 10°$, nine rings each containing thirty-six points. The projection of these points should be uniformly located on the isometric plane within these ranges. This in fact is the case as is depicted in figure V2-(d).

Figures V2-(b), (c), (d), (g) and (h) show the normal simulator graphical output of the type indicated at the top of each subfigure. The first two clearly show the radially increasing optical flow magnitude, while the last two show the linearizing effect of the log-tan transformation.

Figures (e) and (f) are similar to (c) and (d), respectively, except that they plot the magnitude of the optical flow divided by sin θ. The result, $\frac{\dot{x}}{r}$, is constant for all points and this is indicated qualitatively in these plots. That this is also quantitatively true is tabulated in table TV2.

The table shows for the nine values of θ the optical flow $\Delta\theta°$ in degrees, radians $\Delta\theta_r$, the invariant $\dot{x}/r - \Delta\theta_r / sin(\theta)$, and the deviation of the invariant from being constant. This small numerical error is due to the fact that, while its average distance is 50, the field of points varies in distance between 52 and 48 from the camera-retina. the explained

As evidenced by these figures for the optical flow, the translational motion of the spherical field of points has been as predicted. Hence we may infer the correctness under these conditions of the simulator's computation of optical flow.

4.4.1. VERIFICATION EXPERIMENT V3: CONSTANT BINOCULAR DISPARITY CIRCLES

If the optical axes of the two camera-retinas are held parallel, then a 3-D point feature will project to the camera-retinas at non-corresponding locations. This difference is called the binocular, or spatial, disparity. In turn, this disparity

Optical Flow For Forward Translating Spherical Field of Points				
θ	$\Delta\theta°$	$\Delta\theta_r$	$\Delta\theta_r$ / $sin(\theta)$	Δ
5	0.400	0.006984	4.5836	0.0001
15	1.188	0.02074	4.5835	0.0003
25	1.940	0.03385	4.5832	0.0003
35	2.631	0.04593	4.5829	0.0003
45	3.243	0.05660	4.5824	0.0005
55	3.755	0.06554	4.5820	0.0004
65	4.154	0.07249	4.5817	0.0003
75	4.426	0.07724	4.5814	0.0003
85	4.564	0.07965	4.5812	0.0002

TABLE TV2

determines the amounts by which the two camera-retinas must rotate in order that the disparity be brought to zero, i. e., vergence.

In the plane of zero elevation, the locus of points having constant disparity, or equivalently, constant vergence due to the parallax theorem, is obtained by setting binocular disparity, $\gamma = constant$, (from equation 3.1.3.6) and yields a Veith-Muller circle through the camera retina centers given by

$$(X - d\ cot\ \gamma)^2 + Y^2 = \frac{d^2}{\sin^2 \gamma}.$$

This is a circle located at $[X = d\ cot\ \gamma, Y = 0]$, with radius $d/\sin\gamma$. Note that if the radius R of the circle is given, then its center is located on the X-axis at $d\ cot\ \gamma = \sqrt{R^2 - d^2}$.

The purpose of this verification experiment is to confirm this for thirty-two points located on two Veith-Muller circles of radius 30 and 60. The value of d, the amount the retinas are displaced along the Y-axis is ±3 retina radii.

The graphics produced by the simulator for the case in which the Veith-Muller circle is of radius 60 are shown in figure V3. Figures V3-(a) show the thirty-two points lying in a horizontal plane and passing through the Y axis at ±3 retina radii. Their spherical coordinates, θ_c and ϕ_c with respect to the center of the sphere on which they lie, are given by $0 \le \phi_c \le 180°$, $15° \le \theta_c \le 155°$, $\Delta\phi_c = 180°$ and $\Delta\theta_c = 20°$.

The points will be referred to in the order of left (positive Y-axis) at $\phi_c = 0$, $\theta_c = 155°$ to right at $\phi = 180°$, $\theta_c = 155°$. That is, their projection will pass from the left periphery, through the center of the retina, and back out to the right periphery.

Figures V3-(b) and (f) show the resulting binocular disparities. A vector is drawn from the projection of the feature on the right retina to the corresponding projection point of the left retina, which in this case all lie along the horizontal going from left ($\phi = 0$) to right ($\phi = 180°$).

In the case of the isometric plane, figures V3-(e) and (f), the disparities ascend the vertical line $\phi = 0$, pass through the fovea, and then descend on the vertical line $\phi = 180°$.

It appears from the graphics, i. e., figure V3-(b), that the magnitude of the disparities are all equal. That this is in fact the case is shown in table TV3, where alternate points, i. e., sixteen of the thirty-two, are tabulated. (All angles are in degrees.)

Tabulated are the bi-retinal azimuths ψ_r and ψ_l and the resulting binocular vergence (constant) and azimuth for the points on each circle. Because the points are symmetrically arranged to either side of the X-axis, the top and bottom halves are also symmetric in the bi-retinal azimuths directions, as would be expected. Note also that the disparity for the smaller circle is approximately twice that of the larger, again as is expected.

As evidenced by these numerical results for the binocular disparity, the modeling of the constant binocular disparity circles has been as predicted. Again, we may infer the correctness of the simulator for computing binocular disparity under these conditions.

4.4.2. VERIFICATION EXPERIMENT V4: ROTATION OF PLANE OF ELEVATION

The bi-retinal azimuths $\psi_{r/l}$ are defined by

$$\psi_{r/l} = \frac{Y_{r/l}}{\sqrt{X^2 + Z^2}}.$$

When it is desired that the camera-retina tilt and pan in order to bring a feature point into the "fovea" of the image, it is important that these control coordinates be orthogonal, so that changing the tilt in order to make $\delta = 0$ does not change either $\psi_{r/l}$, which would result in the feature point being lost. This verification experiment demonstrates this for ten collinear horizontal points which are rotated as a

Constant Binocular Disparity Circles									
		Radius = 30				Radius = 60			
ϕ_c	θ_c	ψ_r	ψ_l	$\gamma=\psi_r+\psi_l$	$\lambda=(\psi_r-\psi_l)/2$	ψ_r	ψ_l	$\gamma=\psi_r+\psi_l$	$\lambda=(\psi_r-\psi_l)/2$
0	155	80.4	-74.6	5.7392	77.5	78.9	-76.1	2.8636	77.5
	135	70.4	-64.6		67.5	68.9	-66.1		67.5
	115	60.4	-54.6		57.5	58.9	-56.1		57.5
	95	50.4	-44.6		47.5	48.9	-46.1		47.5
	75	40.4	-34.6		37.5	38.9	-36.1		37.5
	55	30.4	-24.6		27.5	28.9	-26.1		27.5
	35	20.4	-14.6		17.5	18.9	-16.1		17.5
	15	10.4	-4.6		7.5	8.9	-6.1		7.5
180	15	-4.6	10.4		-7.5	-6.1	8.9		-7.5
	35	-14.6	20.4		-17.5	-16.1	18.9		-17.5
	55	-24.6	30.4		-27.5	-26.1	28.9		-27.5
	75	-34.6	40.4		-37.5	-36.1	38.9		-37.5
	95	-44.6	50.4		-47.5	-46.1	48.9		-47.5
	115	-54.6	60.4		-57.5	-56.1	58.9		-57.5
	135	-64.6	70.4		-67.5	-66.1	68.9		-67.5
	155	-74.6	80.4		-77.5	-76.1	78.9		-77.5

TABLE TV3

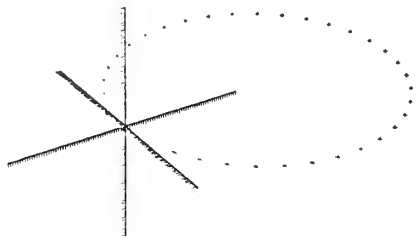
rigid body about the Y axis.

Figure V4 shows the graphics produced by the simulator. Figure V4-(a), (b) and (c) show the starting, mid and ending location of the line, parallel to the Y axis, near the top, center and bottom of the Y–Z plane. Its motion during nine time steps is that of rotation about the Y axis, i. e., points of constant elevation are rotated to a lower plane of elevation.

The ten points are located at constant intervals along the line according to $-45 \leq Y \leq 45$, $\Delta Y = 10$. The nine values of for the plane of elevation δ are given by $85° \geq \delta \geq -85°$, $\Delta\delta = 21.5°$.

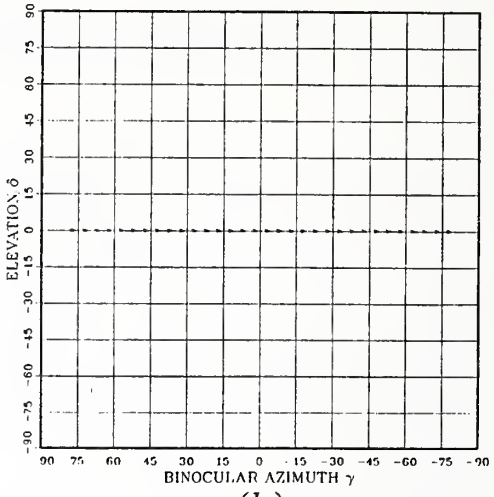
The resulting optical flow plots for ψ_r and ψ_l as a function of δ are shown in figure V4-(d) and (e). They indicate that the values for both bi-retinal azimuths do

META–VIEW IN Φ–Θ–R COORDINATES
FOR POINTS ON CIRCLE OF CONSTANT BINOCULAR DISPARITY



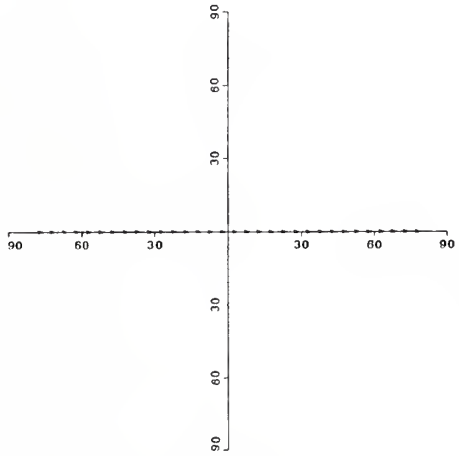
(a)

AZIMUTHAL IMAGE OF R TO L BINOCULAR DISPARITIES
FOR POINTS ON CIRCLE OF CONSTANT BINOCULAR DISPARITY



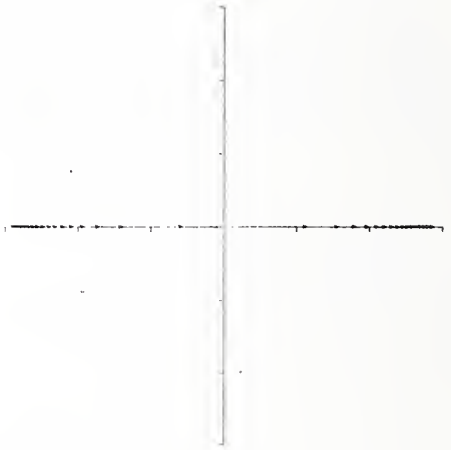
(b)

POLAR SPHERICAL IMAGE OF R TO L BINOCULAR DISPARITIES
FOR POINTS ON CIRCLE OF CONSTANT BINOCULAR DISPARITY



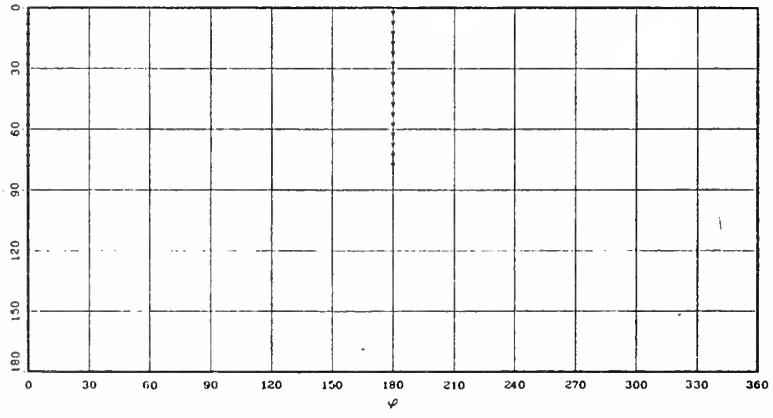
(c)

LOGARITHMIC SPHERICAL IMAGE OF R TO L BINOCULAR DISPARITIE
FOR POINTS ON CIRCLE OF CONSTANT BINOCULAR DISPARITY



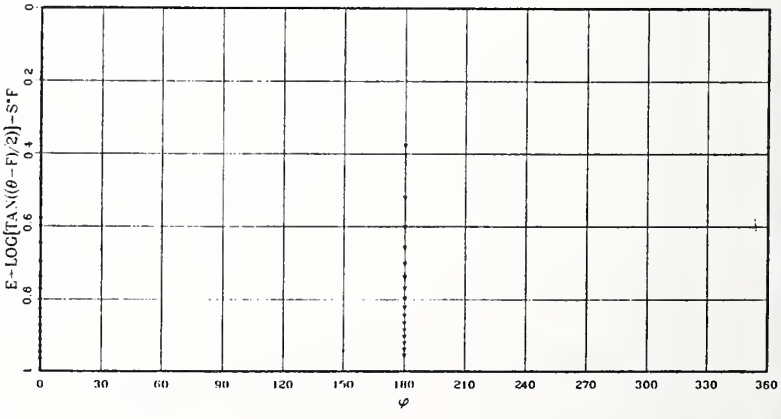
(d)

ISOMETRIC IMAGE OF R TO L BINOCULAR DISPARITIES
FOR POINTS ON CIRCLE OF CONSTANT BINOCULAR DISPARITY



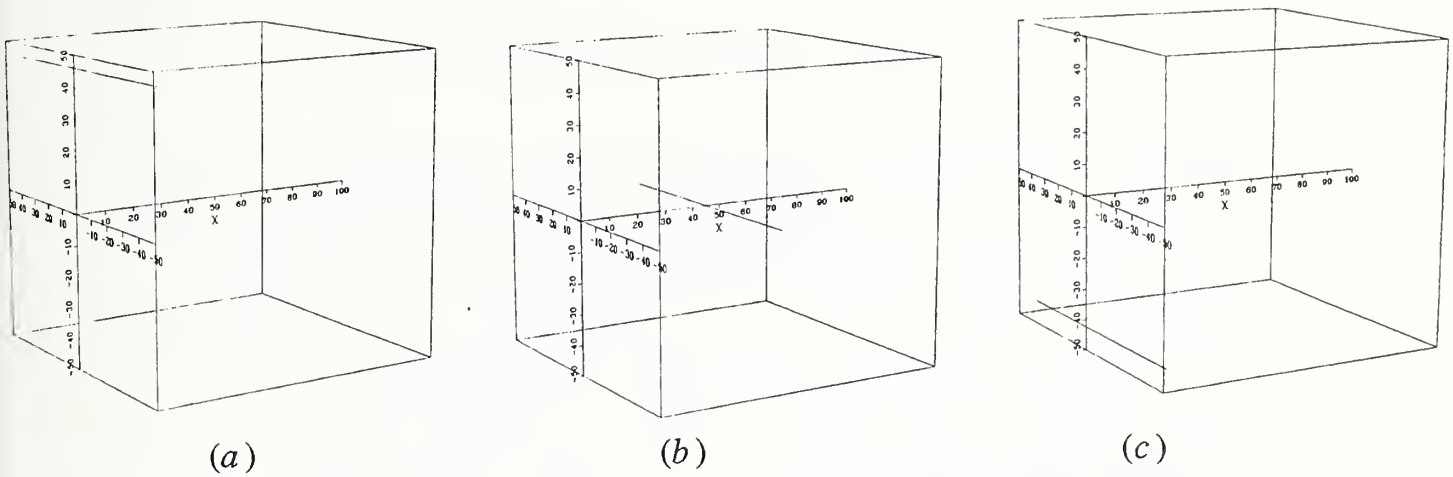
(e)

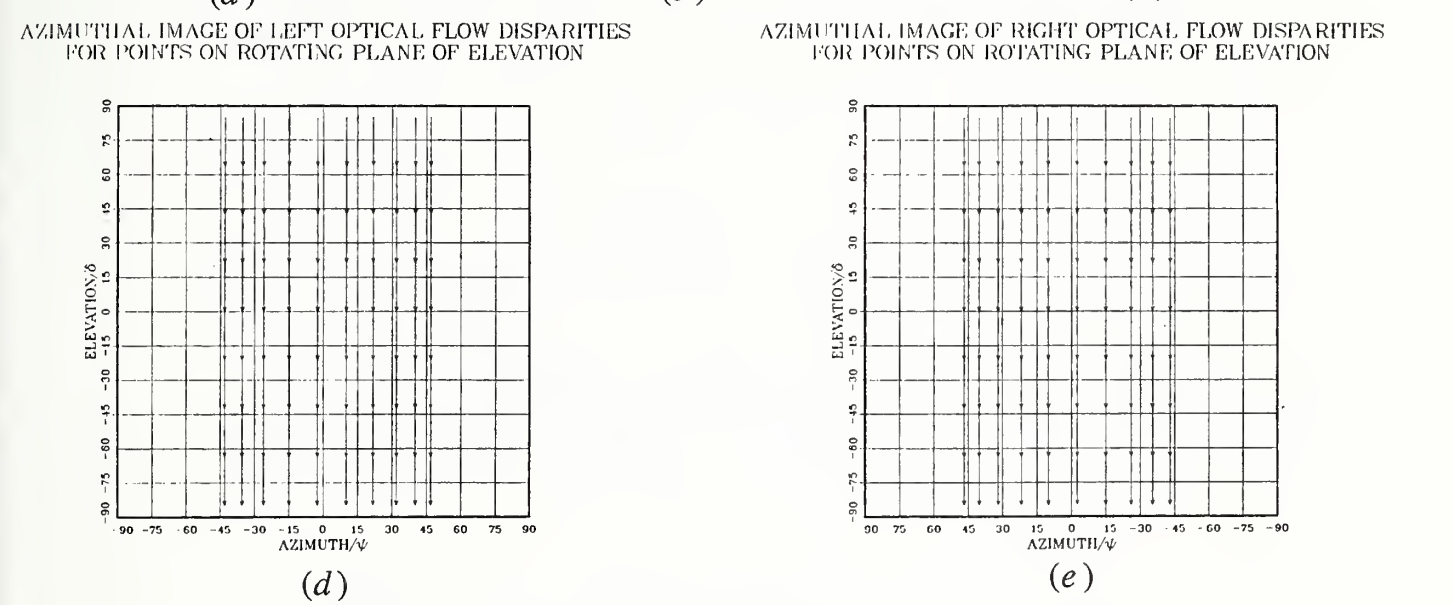
LOGARITHMIC ISOMETRIC IMAGE OF R TO L BINOCULAR DISPARITIE
FOR POINTS ON CIRCLE OF CONSTANT BINOCULAR DISPARITY



(f)

Figure V3: Constant Binocular Disparity Circles Verification Experiment - Output of simulator for thirty-two points on Veith-Muller circle at elevation $\delta = 0$ indicating constant disparity. (See table TV3 for numerical values.)





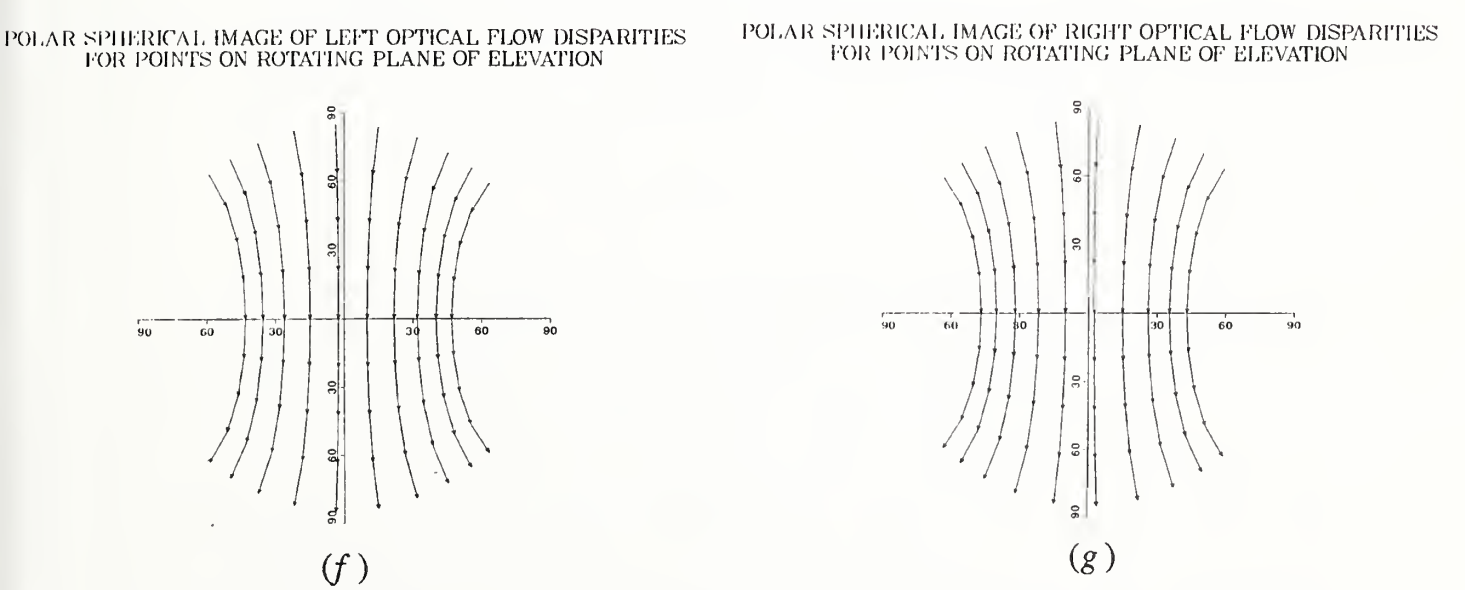


Figure V4: Rotation of Plane of Elevation Verification Experiment - Output of simulator for line rotated about Y axis indicating no change in bi-retinal azimuth. Line starts at top, (a), rotates to zero azimuth, (b), and continues on to bottom, (c).

not change over the interval $85 \leq \delta \leq -85$. This is in contrast to the bi-spherical projection coordinates θ and ϕ plotted in figure V4-(f) and (g) in which the resulting optical flow clearly shows the fact that they are coupled, i. e., non-orthogonality.

As evidenced by these plots of the optical flow for points on a rotating plane of elevation, the bi-retinal azimuth has been shown to remain constant over changes in the angle of elevation as was predicted. Hence we may infer the correctness of the simulator under these conditions for computing the rotation of the plane of elevation and resulting optical flow.

5. EXPERIMENTAL VERIFICATION OF MATHEMATICAL CONCEPTS

This section will describe several experiments in which concepts developed in section 3 are exemplified and verified using the binocular camera-retina wire frame scene simulator and a program for "warping" an image by nonlinear sampling.

5.1. EXPERIMENTAL VERIFICATION OF NORMALIZATION

In section 3.2 the operation of normalization was developed for a forward translating camera-retina. More specifically, four mappings of the spherical azimuth θ to $\hat{\theta}$ were defined in which 1-D parameterizations of 3-D space are projected in a manner making the parameter value for a given optical flow vector inversely related to the optical flow magnitude.

Stated another way, the $\hat{\theta}$-ϕ logarithmic spherical projection and its representation as the logarithmic isometric plane have the property that two optical flow disparity vectors are equal if and only if they are both projections of points having the same 1-D parameter.

The next four subsections exemplify this for four 1-D parameterizations: (1), 3-D space parameterized by range, (2), 3-D space parameterized by depth, (3), 3-D space parameterized by looming, and (4), 3-D space parameterized by lateral clearance.

5.1.1. EXPERIMENTAL VERIFICATION OF RANGE NORMALIZATION

The purpose of this experiment is to demonstrate that the range normalization of section 3.2.2 results in a representation of optical flow which is inversely related to the 1-D parameterization of 3-D space by the parameter range R as given by

$$R = \frac{v}{\dot{\hat{\theta}}} \qquad 5.1.1.1$$

This will be accomplished by placing points all at an arbitrary constant range from a camera-retina and then translating them forward toward the camera-retina. The range normalized logarithmic isometric plane representation of the optical flow which results is constant in both direction and magnitude.

The inference is that, given an optical flow disparity vector in this representation, the x, y location of the image may be immediately associated with a unique range by the calculation $R = v/\dot{\hat{\theta}}$. This is for point features only, and must be modified, via the methods of section 2, for edge optical flow normal components.

Three-hundred twenty-four points all located on the surface of a hemisphere of radius $R = 50$ and centered at $X = 2$ are translated toward the camera-retinas to

$X = -2$ in one time step. The initial configuration is shown in figure E1-(a) and (b). Figure E1-(c) and (e) shows the optical flow disparity vectors as they normally occur: their magnitude increasing radially as sin θ, where θ is the eccentricity. Note that they are regularly spaced in θ-ϕ coordinates in the isometric plane representation.

In figures E1-(d) and (f) the logarithmic mapping for range normalization and its logarithmic isometric plane representation are plotted: here the magnitudes of the disparity vectors are all visibly equal indicating that they have all come from the same range.

In table TE1 are tabulated the minimum and maximum optical flow disparity vectors under several conditions. In the first two rows, the value of d, the distance between the two retinas, in units of retina radii, is zero and the minimum and maximum values of radial optical flow Δθ along lines of constant ϕ are for the case just described. The columns headed $\theta/\hat{\theta}$ and ϕ are the coordinates of the minimum/maximum, and the error is the relative error for the normalized values which should be constant, i.e., the minimum and maximum should be equal.

Range Normalization for Spheres of Constant Range								
Normalized?	d	min Δθ	$\theta/\hat{\theta}$	ϕ	max Δθ	$\theta/\hat{\theta}$	ϕ	error
no	0.0	0.400	5.0	175.	4.56	85.	175.	
yes	0.0	0.01379	0.9849	175.	0.01380	0.4602	165.	0.0007
no	1.5							
yes	1.5	0.01339			0.014214			0.06
no	3.0							
yes	3.0	0.01301			0.01466			0.10

TABLE TE1

In the case of binocular camera-retinas separated by the distance d, the range R is not identical for both retinas and as a result some error is introduced into this demonstration. (However, see remark below.)

Figure E2-(a) and (b) are the logarithmic spherical projection and logarithmic isometric representation of the binocular (spatial) disparity vectors for the case $d = 3$. Below them, figures E2-(c) through (f) are the corresponding right and left retina range normalizations. Clearly, the introduction of binocular disparity has a greater impact on optical flow vector location than on the range normalized

magnitudes themselves.

However, more importantly, in a real use of the normalization technique, normalization would be done separately for each camera-retina followed by its "fusion", using the techniques of section 2.

5.1.2. EXPERIMENTAL VERIFICATION OF DEPTH NORMALIZATION

The purpose of this experiment is to demonstrate that the depth normalization described in section 3.2.3.1 is valid. More particularly, the experiment will demonstrate that this normalization of optical flow results in an inverse relationship between it and the 1-D depth parameterization of 3-D space. This will be accomplished by placing a number of points at an arbitrary but constant depth in front of the camera-retinas, translating them toward the camera-retinas, and observing the normalized results.

Remark: The logarithmic isometric plane representation of a depth normalized image has the same properties as the "log-polar" representation, and in fact is the same image modulo the size of the fovea F.

Again, the inference is that a depth normalized optical flow disparity vector in the logarithmic isometric plane representation can be numerically converted to depth D via the calculation

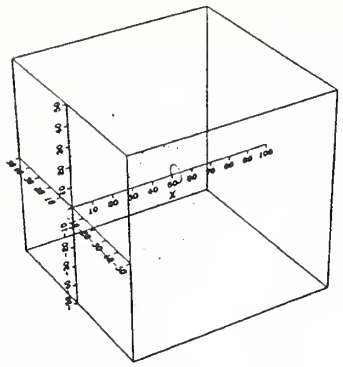
$$D = \frac{v}{\dot{\hat{\theta}}}. \qquad 5.1.2.1$$

This can be done since two optical flow vectors are equal if and only if they project from two 3-D points whose depth is equal.

One-hundred points arranged as a planar 10 by 10 grid are positioned in front of the camera-retinas at a distance $X = 54$ and translated toward the camera-retina to $X = 46$. Figure E3-(a) shows the initial configuration. Since depth is independent of location along the Y axis, there is no difference in depth to the points from the two camera-retinas. However, to emphasize that in the general case normalization must be performed separately for each camera-retina we have included the binocular disparities as figures E3-(b) through (d) to show this. This is because in a real scene, containing edges and contours, the normal components of the optical flow will be different in the two images.

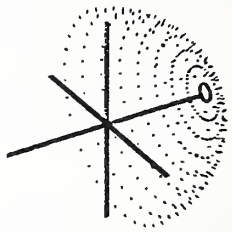
Figures E3-(e) and (f) and E4-(a) through (f) are the left and right camera-retina projections and representations before and after depth normalization. Again, visually it can be seen that the magnitudes of the vectors have been equalized. Numerical results are tabulated in table TE2.

META-VIEW IN X-Y-Z COORDINATES
FOR POINTS ON FORWARD TRANSLATING SPHERE



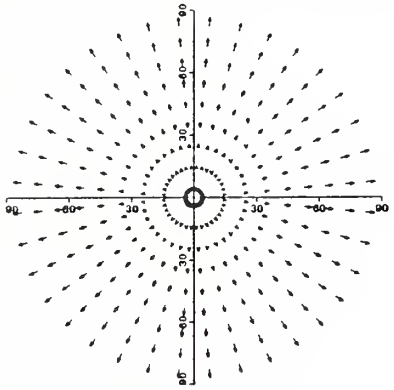
(a)

META-VIEW IN Φ-Θ-R COORDINATES
FOR POINTS ON FORWARD TRANSLATING SPHERE



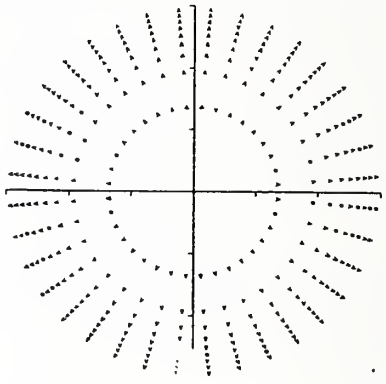
(b)

POLAR SPHERICAL IMAGE OF OPTICAL FLOW
FOR POINTS ON FORWARD TRANSLATING SPHERE



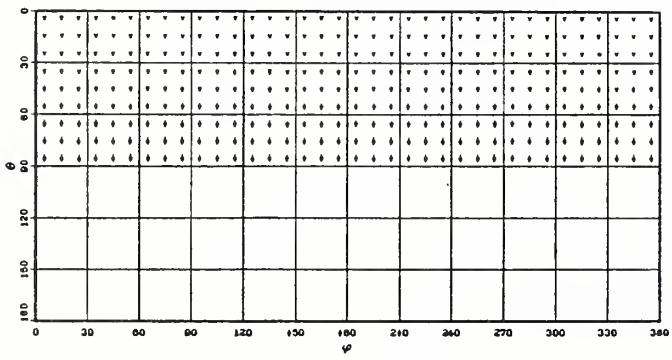
(c)

RANGE NORMALIZATION OF OPTICAL FLOW
FOR POINTS ON FORWARD TRANSLATING SPHERE



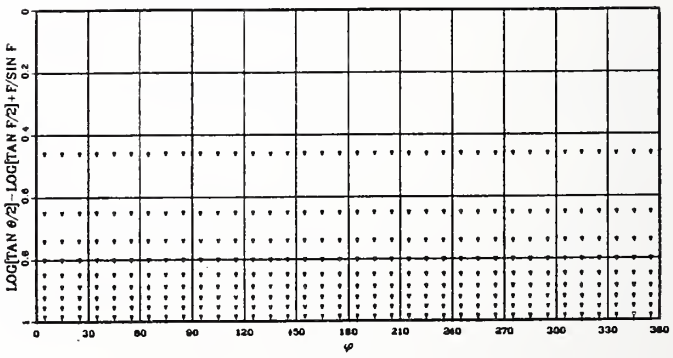
(d)

ISOMETRIC IMAGE OF OPTICAL FLOW
FOR POINTS ON FORWARD TRANSLATING SPHERE



(e)

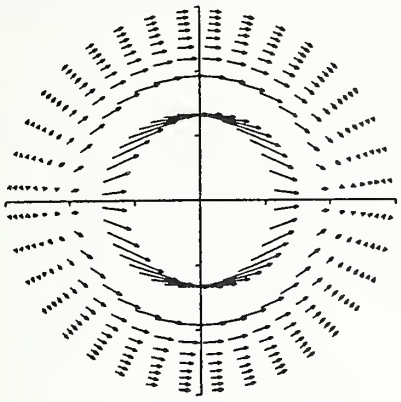
RANGE NORMALIZATION OF OPTICAL FLOW
FOR POINTS ON FORWARD TRANSLATING SPHERE



(f)

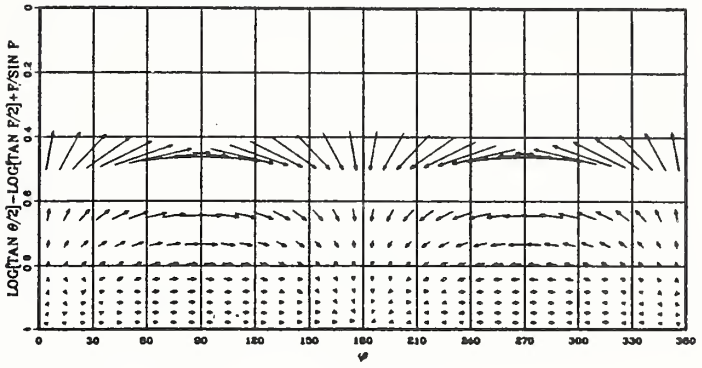
Figure E1: Range Normalization - A spherical field of points, shown at the top, are translated forward resulting in the unnormalized optical flow in center. Range normalized optical flow is at the right, where the optical flow disparity vectors, plotted as a polar spherical projection and as its logarithmic isometric representation, are constant for all points of the sphere.

LOGARITHMIC SPHERICAL IMAGE OF R TO L SPATIAL DISPARITIES
FOR POINTS ON FORWARD TRANSLATING SPHERE



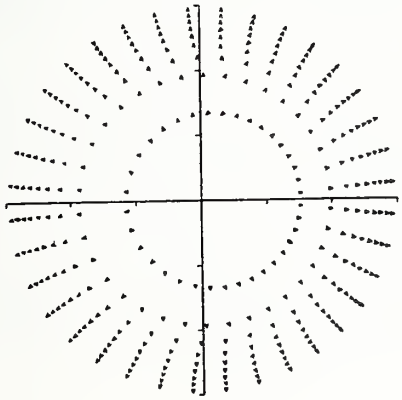
(a)

LOGARITHMIC ISOMETRIC IMAGE OF R TO L SPATIAL DISPARITAES
FOR POINTS ON FORWARD TRANSLATING SPHERE



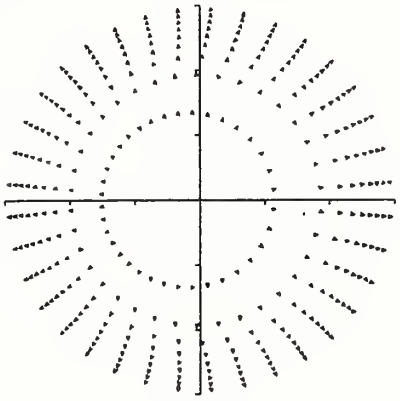
(b)

RANGE NORMALIZATION OF LEFT RETINA OPTICAL FLOW
FOR POINTS ON FORWARD TRANSLATING SPHERE



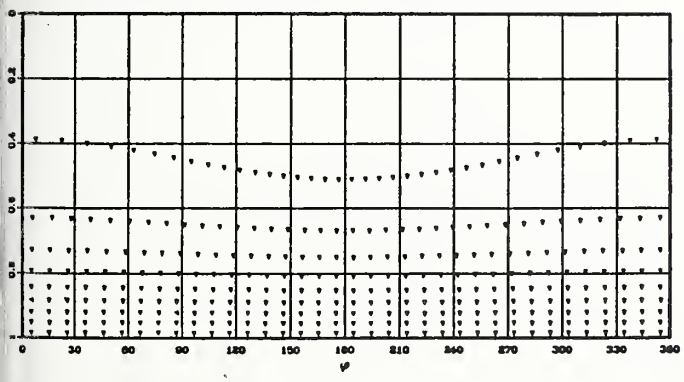
(c)

RANGE NORMALIZATION OF RIGHT RETINA OPTICAL FLOW
FOR POINTS ON FORWARD TRANSLATING SPHERE



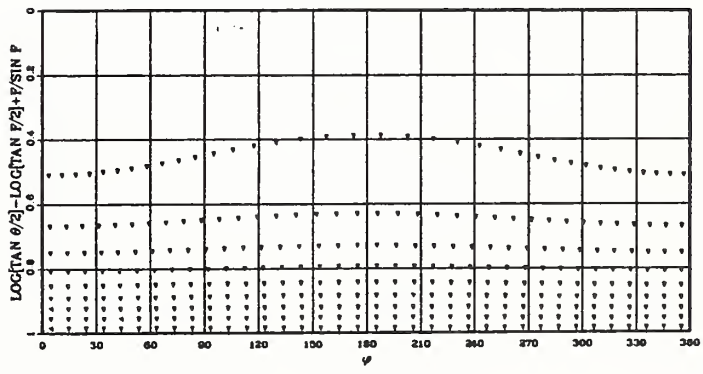
(d)

RANGE NORMALIZATION OF LEFT RETINA OPTICAL FLOW
FOR POINTS ON FORWARD TRANSLATING SPHERE



(e)

RANGE NORMALIZATION OF RIGHT RETINA OPTICAL FLOW
FOR POINTS ON FORWARD TRANSLATING SPHERE



(f)

Figure E2: Range Normalization - Introducing two retinas at a distance $d = 3$ retina radii apart results in the binocular disparity at top. The right and left range normalized logarithmic spherical projections and logarithmic isometric representations, shown below, have a maximum error of 10 per cent.

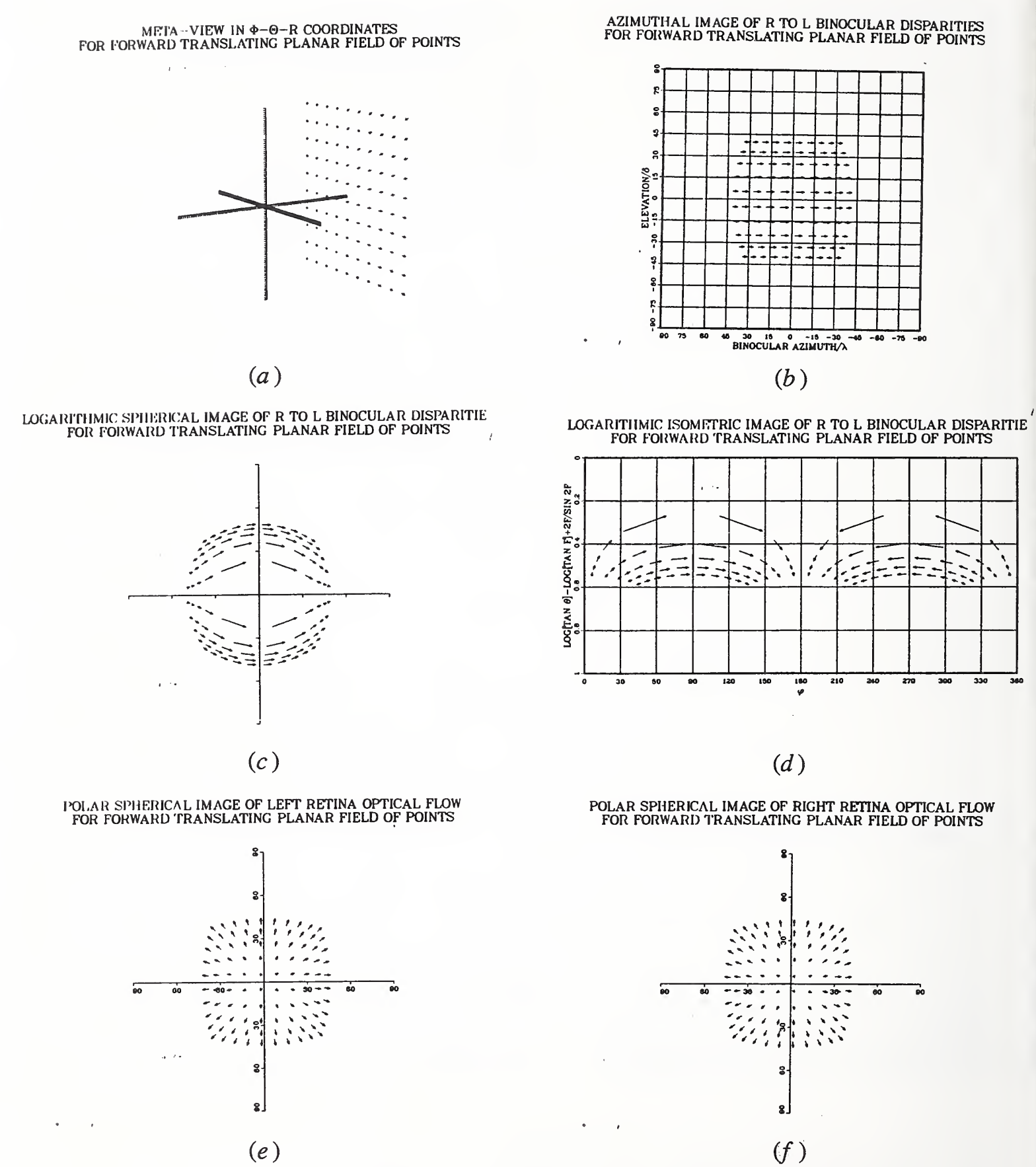


Figure E3: Depth Normalization - A planar field of points at a constant depth are viewed by two retinas at a distance $d = 3$ retina radii apart. The binocular disparities are shown above the unnormalized optical flow.

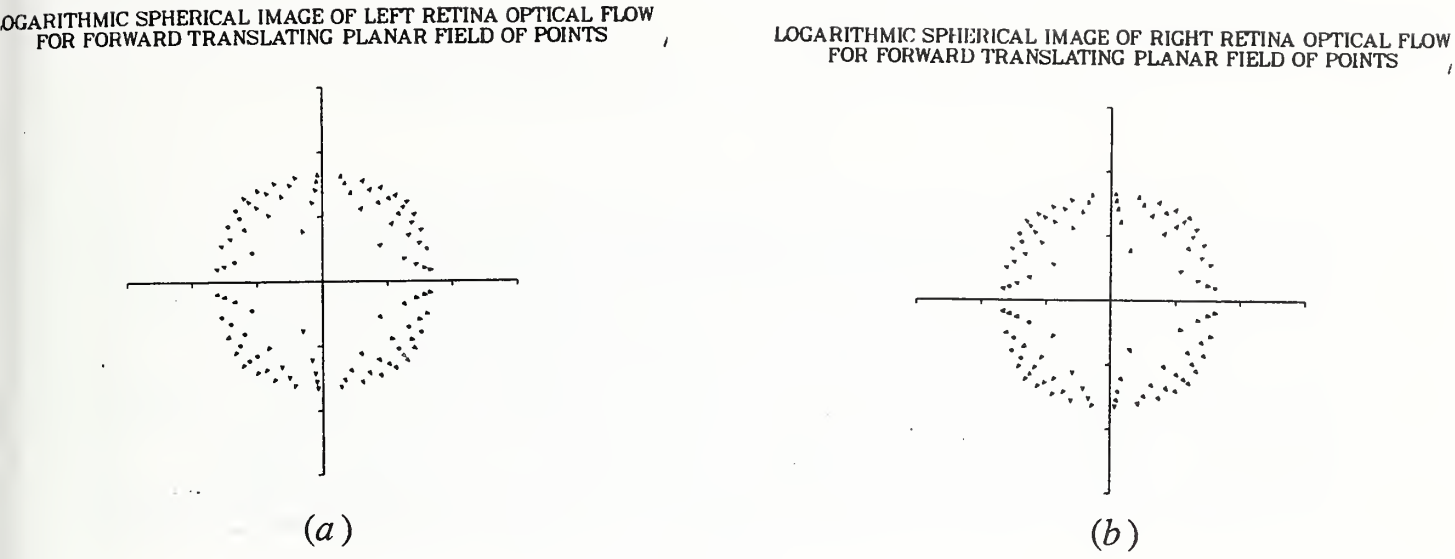


Figure E4: Depth Normalization - The depth normalized logarithmic spherical projections and below, both unnormalized and normalized isometric representations. In this case the binocular disparity results in zero error in normalizing the image. This normalization duplicates the ''log-polar'' transform.

Depth Normalization for Planes of Constant Depth								
Normalized?	d	min $\Delta\theta$	$\theta/\hat{\theta}$	ϕ	max $\Delta\theta$	$\theta/\hat{\theta}$	ϕ	error
no	0.0	0.279	8.1	315.	4.59	44.8	315.	
yes	0.0	0.020489	0.47	348.	0.020489	0.59	45.	0.0000
no	3.0	0.982(?)	6.18	291.8	4.59	45.14	296.	
yes	3.0	0.020489	0.5757	320.	0.020489	0.55	231.	0.0000

TABLE TE2

Here, normalization error in the case of binocular camera-retinas is zero as would be expected.

5.1.3. EXPERIMENTAL VERIFICATION OF LOOMING NORMALIZATION

The purpose of this experiment is to demonstrate that normalization for looming as described in section 3.2.3.2 is valid. More particularly, the experiment will demonstrate that this normalization of optical flow results in a linear relationship between it and the 1-D looming parameterization of 3-D space.

Looming L has been defined [RAVIV] as $L = -\dot{R}/R$ and is identifiable with "obstacle threat". Based on this definition, it can be shown that spheres of constant looming have diameters coincident with the translation vector and passing through the center of a camera-retina.

The demonstration will be accomplished by placing a number of points on an arbitrary but fixed radius sphere of constant looming, translating them toward the camera-retina(s), and observing the normalized representation results.

Again, the inference is that looming normalized optical flow disparity vectors in the logarithmic isometric plane representation can be numerically converted to looming L via the calculation

$$L = \frac{v}{\dot{\hat{\theta}}}. \qquad 5.1.3.1$$

This can be done since two optical flow vectors are equal if and only if they project from two 3-D points whose looming value is equal, and hence if and only if from the same looming sphere.

One-hundred forty-four points, located on the surface of a sphere of radius $R = 50$ and initially centered at $X = 56$, is translated in in one time step to $X = 44$.

Figure E5-(a) shows the initial configuration and E5-(b) the resulting optical flow as an azimuthal projection. (This latter serves only to provide a contrast with the polar spherical projections.) Since each camera-retina has its own distinct sphere of constant looming, the case of a single retina will be discussed first.

Figures E5-(c) through (f) are the resultant unnormalized polar spherical and isometric projection and representation followed by the corresponding normalized versions. Note that the optical flow vectors nearest the center of the fovea are those produced by points at the far end of the sphere, while those at the periphery are those produced by the near end.

Again, visually it can be seen that the magnitudes of the vectors have been equalized. Numerical results are tabulated in table TE3.

Looming Normalization for Spheres of Constant Looming								
Normalized?	d	min $\Delta\theta$	$\theta/\hat{\theta}$	ϕ	max $\Delta\theta$	$\theta/\hat{\theta}$	ϕ	error
no	0.0	0.604	5.0	170.	25.3	75.7	170.	
yes	0.0	0.025919	0.98	170.	0.02723	0.44	150.	0.08
no	3.0	0.604	5.0	170.	25.3	75.7	170.	
yes	3.0	0.021179	0.989	171.	0.032337	0.9837.	11.3	0.53

TABLE TE3

Normalization error in the case of a single camera-retina is larger than expected, though this is the worse case. (Remark: this should be rechecked by running simulation again.)

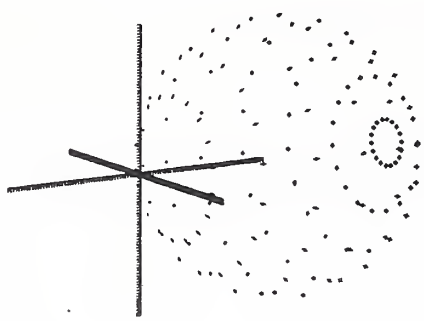
Figures E6-(a) and (b) show the binocular disparity for the case $d = 3$. The resultant normalization is shown in figures E6-(c) through (f).

Again, the error indicated by the difference between the minimum and maximum normalized vector magnitudes is 53%, a number which needs to be rechecked, as again, the average error is much less as a look at the graphics indicate.

5.1.4. EXPERIMENTAL VERIFICATION OF CLEARANCE NORMALIZATION

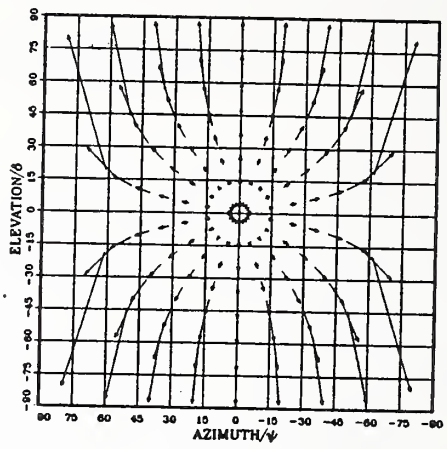
The purpose of this experiment is to demonstrate that normalization for clearance as described in section 3.2.3.3 is valid. More particularly, the experiment will demonstrate that this normalization of optical flow results in a linear relationship between it and the 1-D clearance parameterization of 3-D space.

META-VIEW IN Φ-Θ-R COORDINATES
FOR SPHERICAL POINT FIELD OF CONSTANT LOOMING



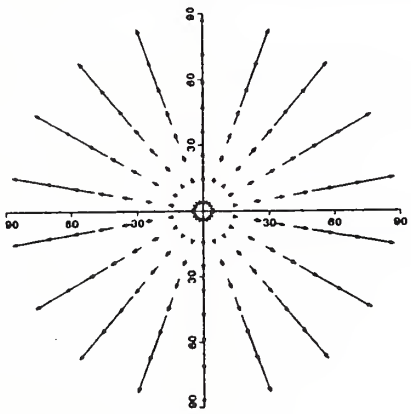
(a)

AZIMUTHAL IMAGE OF OPTICAL FLOW
FOR SPHERICAL POINT FIELD OF CONSTANT LOOMING



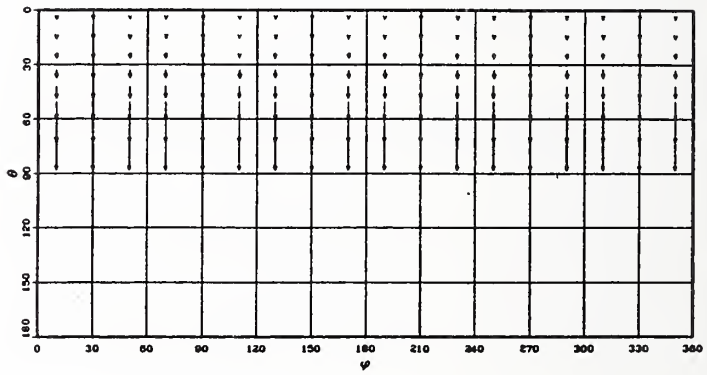
(b)

POLAR SPHERICAL IMAGE OF OPTICAL FLOW
FOR SPHERICAL POINT FIELD OF CONSTANT LOOMING



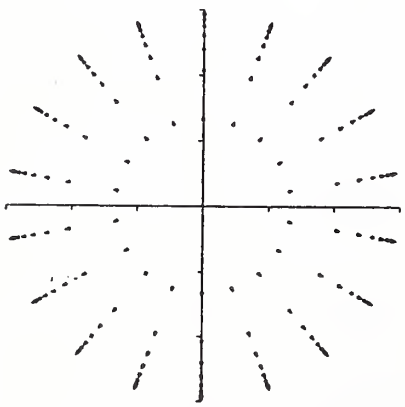
(c)

ISOMETRIC IMAGE OF OPTICAL FLOW
FOR SPHERICAL POINT FIELD OF CONSTANT LOOMING



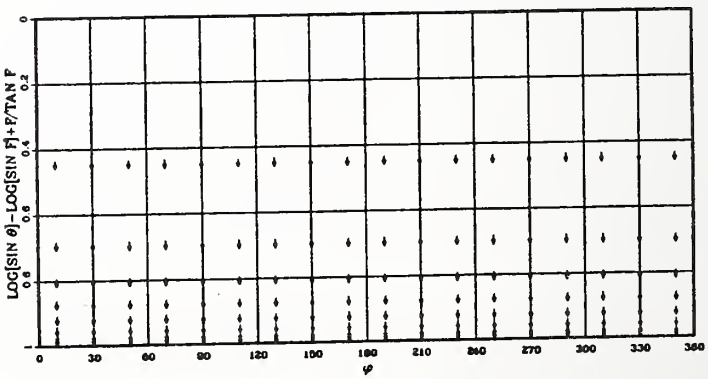
(d)

LOGARITHMIC SPHERICAL IMAGE OF OPTICAL FLOW
FOR SPHERICAL POINT FIELD OF CONSTANT LOOMING



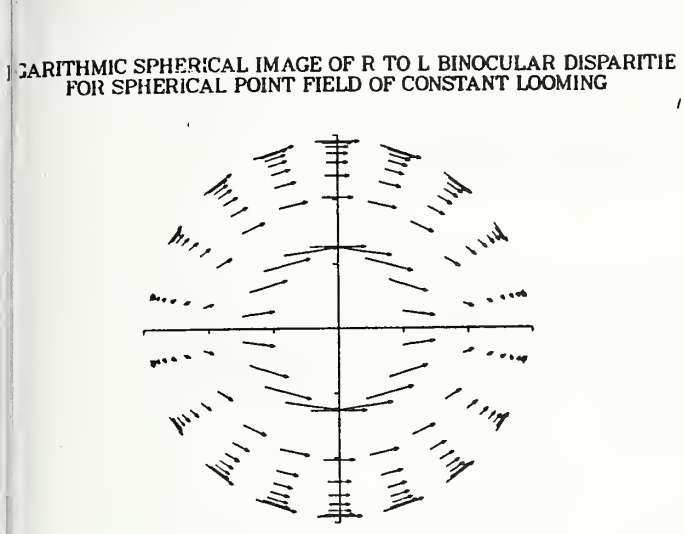
(e)

LOGARITHMIC ISOMETRIC IMAGE OF OPTICAL FLOW
FOR SPHERICAL POINT FIELD OF CONSTANT LOOMING

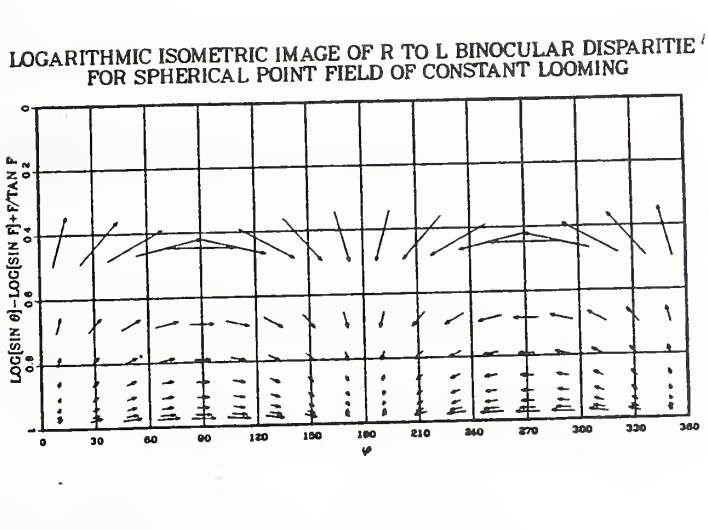


(f)

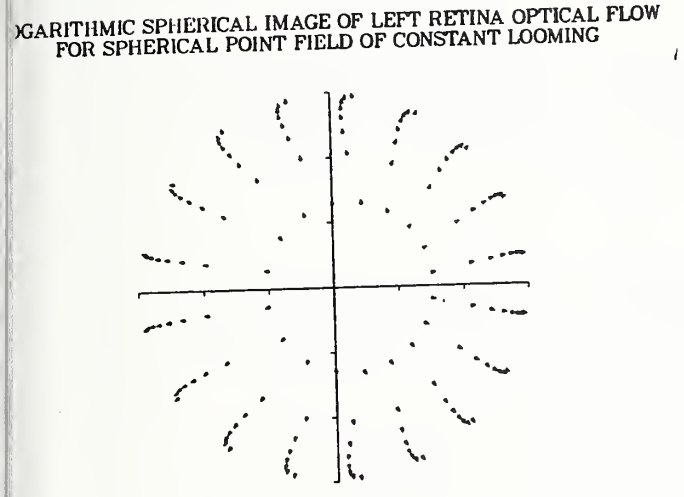
Figure E5: Looming Normalization - A field of points located on a sphere of constant looming, top left, is translated forward. The resulting optical flow is shown as an azimuthal projection, top right, and as unnormalized and normalized polar spherical and isometric representations below.



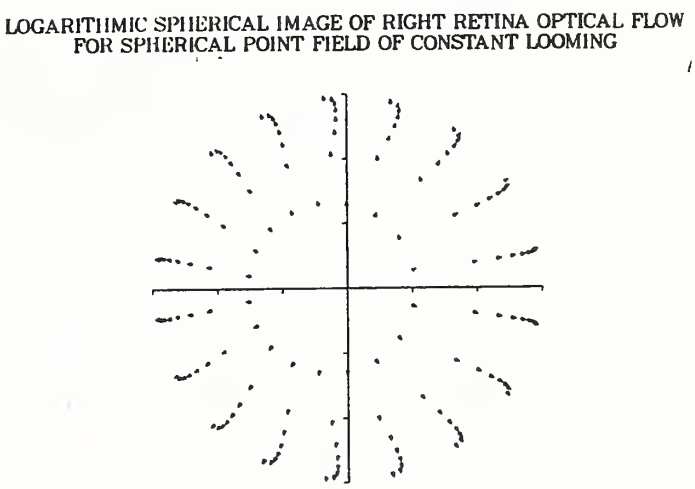
(a)



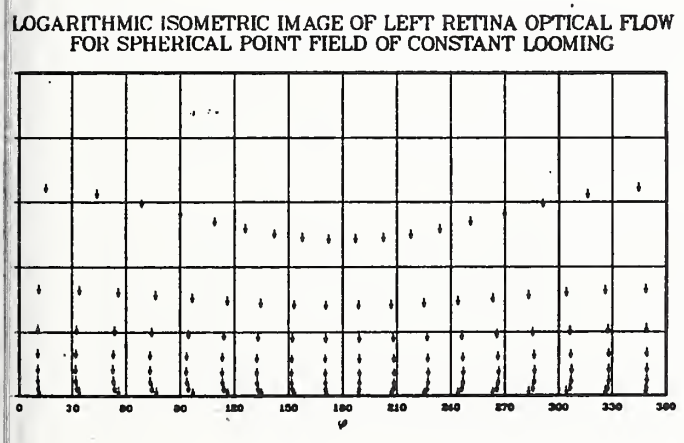
(b)



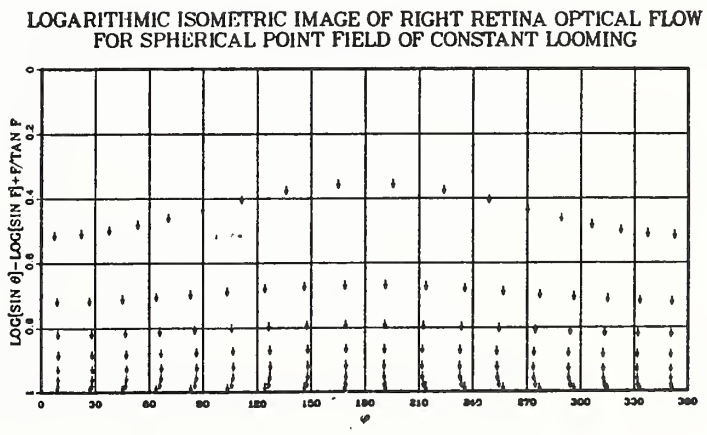
(c)



(d)



(e)



(f)

Figure E6: Looming Normalization - Binocular disparity from two retinas at $d = 3$ apart results in disparity shown at top, and resultant normalized optical shown below. Maximum error is 53 per cent, due to some points being very close.

Clearance here refers to the lateral distance in any direction from the line which represents the extension of the current instantaneous velocity vector v of the camera-retinas. Points of constant clearance C all lie on a cylinder of radius C whose axis is coincident with that extension, and are parameterized by $C = \sqrt{Y^2 + Z^2}$.

The demonstration will consist of six circles, all of the same radius, placed symmetrically about the X axis and at six distinct distances. These circles represent the "tunnel" of equal clearance. The circles are then translated along the X axis toward the camera-retina for one time step thus creating optical flow at six distinct depths.

Again, the inference is that clearance normalized optical flow disparity vectors in the logarithmic isometric plane representation can be numerically converted to clearance via the calculation

$$C = \frac{v}{\dot{\hat{\theta}}}. \qquad 5.1.4.1$$

This can be done since two optical flow vectors are equal if and only if they project from two 3-D points whose clearance value is equal, and hence if and only if from the same clearance cylinder.

Each circle is of radius $C = 6$ and is made up of 31 connected points. The circles are initially centered about the X axis at $X = 96$, 76, 56, 36 and 16, and are translated toward the camera-retina(s) by $\Delta X = -12$. The initial configuration meta view is shown in figure E7-(a).

Figure E7-(c) is the normalized polar spherical projection, and E7-(d) and (f) the corresponding unnormalized and normalized logarithmic isometric plane representations. Again, the magnitudes of the optical flow vectors can be seen to be equal for all circles, whatever their location along the X axis. The numerical results are tabulated in table TE4.

It is of interest to note that in the unnormalized polar spherical projection the circles are projected at nonconstant difference in θ even though they are at a constant difference in range. However, their normalized image locations, shown in figure E7-(b), are at a constant $\Delta\hat{\theta}$ apart. This is a mathematical consequence of the requirement that the normalization equalize optical flow. As a result, radial distance encodes range in a linear manner.

Normalization error in the case of a single camera-retina is zero as is expected. The proximity of the closest circle to the camera-retinas and the subsequent large disparity, as indicated in figure E7-(e), has resulted in a large error for the binocular

Clearance Normalization for Cylinders of Constant Clearance								
Normalized?	d	min $\Delta\theta$	$\theta/\hat{\theta}$	ϕ	max $\Delta\theta$	$\theta/\hat{\theta}$	ϕ	error
no	0.0	0.509	3.8	113.	35.4	38.4	43.5	
yes	0.0	0.032946	0.92	299.	0.032964	0.92	43.	0.00000
no	3.0	0.256	1.9	354.	36.9	45.3	142.	
yes	3.0	0.0220	0.87	181.	0.0657	0.726.	354.	2.1

TABLE TE4

case. Figures E8-(a) and (b) show the polar spherical projection of the optical flow for the right and left retinas. This is followed by the right and left unnormalized and normalized isometric plane representations, where the differences in magnitude for the closest circle are clearly shown.

5.2. EXPERIMENTAL VERIFICATION OF INTERCHANGEABILITY OF PLANAR AND SPHERICAL PROJECTION

In this section several programs for "warping" a planar projection image into spherical projection images and normalized images will be briefly described. The purpose of these experiments is to investigate the feasibility of obtaining planar projection images from standard video cameras and transforming them by nonlinear sampling in order to obtain images with desirable properties described in this report.

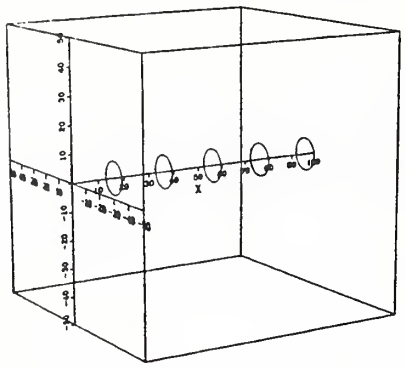
All neighborhood surface modeling required to interpolate the image at subpixel spatial resolutions was done using bilinear interpolation over two by two pixel neighborhoods. Bilinear interpolation was used for reasons of computational speed, but still provides enough detail to confirm that the mapping is being performed.

A more systematic investigation would use at least bi-cubic interpolation and would look at several orders of derivatives in both the original and transformed images. However, there are reasons to represent the image using other sets of orthogonal polynomials, e. g., Hermite polynomials, and in an implementation these would be used to perform the interpolation.

5.2.1. COMPUTING THE SPHERICAL PROJECTION FROM THE PLANAR PROJECTION IMAGE

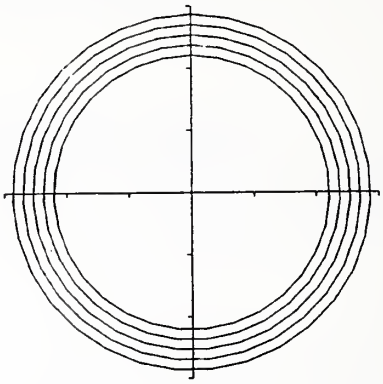
In section 3.3.2.2 equations were described for computing several projections to the plane for a spherical projection. A computer program was written for transforming a planar projection image to a logarithmic spherical projection normalized for

META-VIEW IN X-Y-Z COORDINATES
FOR POINTS ON CONSTANT FLOW CYLINDER



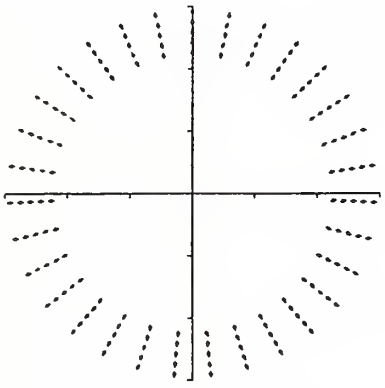
(a)

LOGARITHMIC SPHERICAL IMAGE OF RIGHT RETINA
FOR POINTS ON CONSTANT FLOW CYLINDER



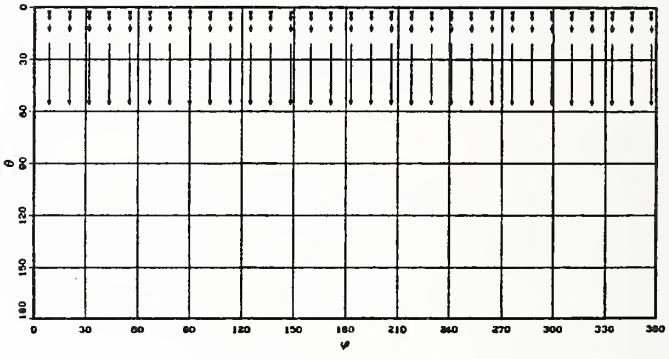
(b)

LOGARITHMIC SPHERICAL IMAGE OF OPTICAL FLOW
FOR POINTS ON CONSTANT FLOW CYLINDER



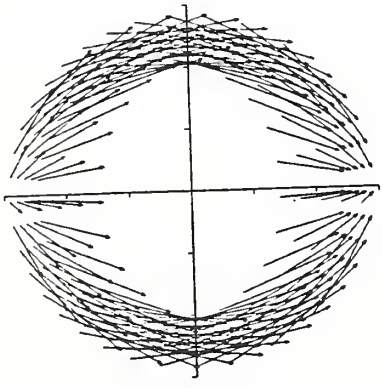
(c)

ISOMETRIC IMAGE OF OPTICAL FLOW
FOR POINTS ON CONSTANT FLOW CYLINDER



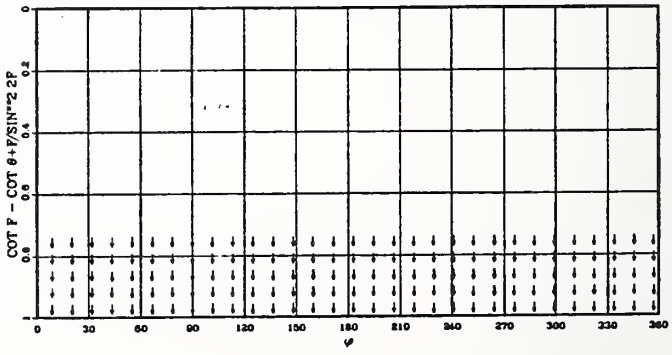
(d)

LOGARITHMIC SPHERICAL IMAGE OF R TO L BINOCULAR DISPARITIE
FOR POINTS ON CONSTANT FLOW CYLINDER



(e)

LOGARITHMIC ISOMETRIC IMAGE OF OPTICAL FLOW
FOR POINTS ON CONSTANT FLOW CYLINDER



(f)

Figure E7: Clearance Normalization - Six circles are placed on the X axes and translated forward are shown at top left. The resultant retinal image of the clearance normalized polar spherical and resultant unnormalized and normalized optical flow is shown below. The introduction of disparity, bottom left results in normalization plots shown in next figure.

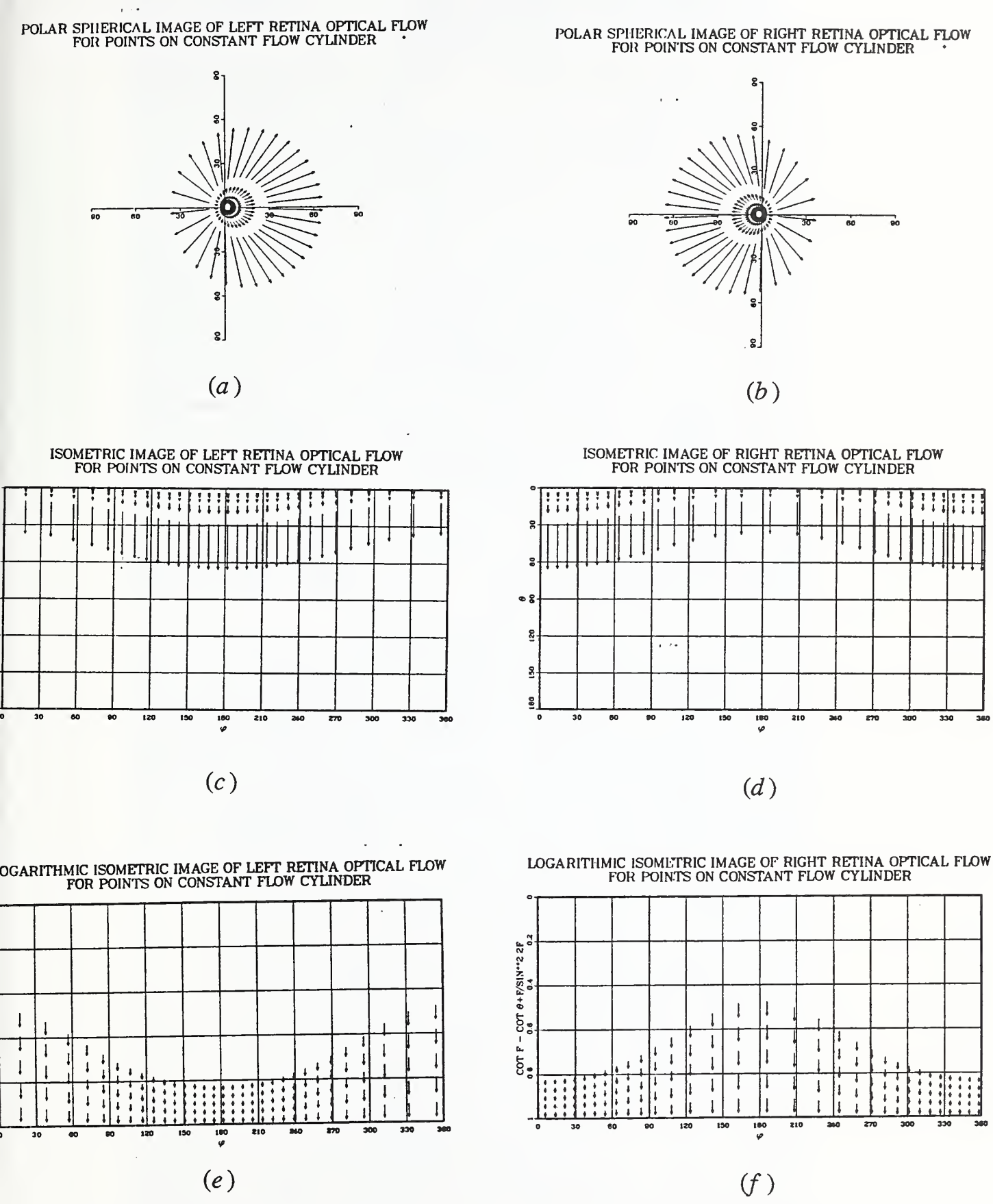


Figure E8: Clearance Normalization - Binocular disparity from two retinas at $d = 3$ apart results in disparity shown in previous figure and at top. Resultant unnormalized and normalized optical flow is shown below. Maximum error in case with disparity is 210 per cent, again due to closest point where disparity is large.

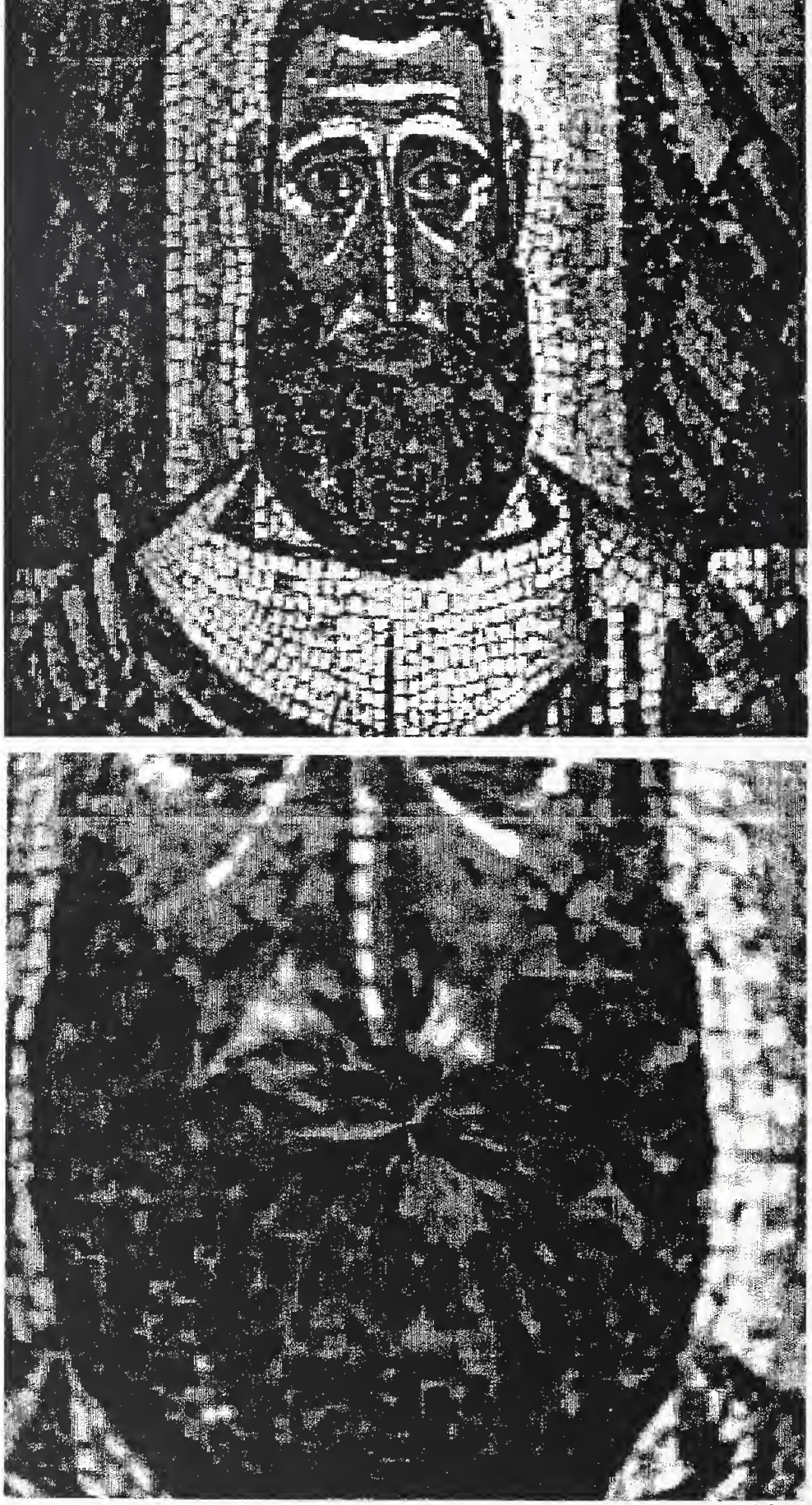


Figure L3: Man With Beard Scene - At top is image made by taking video image of flat picture with planar projection image. Below is the computed spherical projection image, but with longer focal length, i. e., smaller field of view, than was used in planar projection lens.

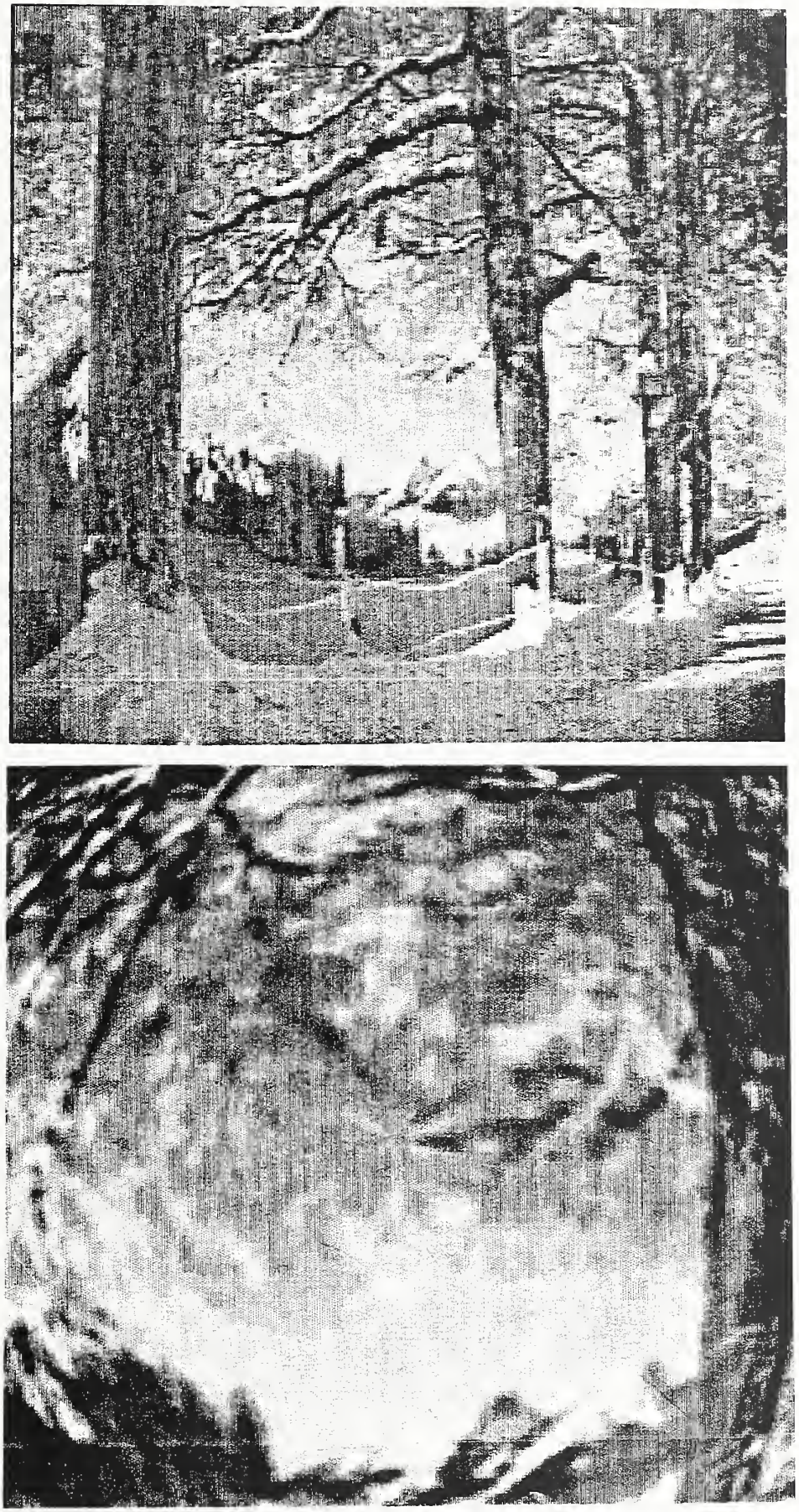


Figure L2: Snow Scene - At top is image made by taking video image of flat picture with planar projection image. Below is the computed spherical projection image.

range.

At the top of figure L2 is an image formed by taking a planar projection image of a flat scene containing a picture of a snow scene. The focal length of the lens was 16mm for a field of view of 30°. (The focal length was converted to image units for the program.)

At the bottom of figure L2 snow scene is the result of transforming it to a range normalized logarithmic spherical projection.

Figure L3 contains a similar pair, the bearded man scene, but with the program working with a longer focal length, i. e., smaller field of view, than was the actual case, and hence the image is much less distorted.

This program and experiment has demonstrated the possibility of using a planar projection camera for the production of spherical projection images.

5.2.2. COMPUTING THE LOGARITHMIC ISOMETRIC PLANE REPRESENTATION FROM THE PLANAR PROJECTION

In section 3.3.2.1 equations were given relating the normalization of spherical projection images from planar projection images. In this section we verify that it is feasible to compute two of these, depth and range normalization, and represent them in the logarithmic isometric plane. This has been done by writing a program which samples the planar projection input image according to those equations.

The log-polar transform was described in section 3.3.1 as being equivalent to depth normalized spherical projection. In that section figure L1 was used as an example of the log-polar transform. The original test pattern images described there were used to debug and calibrate the program for computing the depth normalized logarithmic isometric plane representation from a planar projection image.

Figure L4 shows a larger version of the original test pattern image at top and its transform at the bottom. In this image some experimentation was needed in order to find the center of the scene in order to make it the origin of the transform. Also, as indicated earlier, the nonsquare pixel dimensions had to be taken into account in order to get lines of constant $\hat{\theta}$ to map to horizontal lines in the transform.

All of the original images used here are video images taken of flat pictures, and hence are at a fixed distance from the video camera.

At the top left of figure L5 is the logarithmic isometric plane representation of the bearded man planar projection image shown at the top of figure L3. This clearly shows that pixels in the region of the fovea, at top of transform, result from many samples taken from the same pixel, while at the bottom the reverse is true.

The lower three images are range normalized logarithmic isometric plane representations of the snow scene image shown at the top of figure L2. They are computed with increasing focal lengths, i. e., decreasing fields of view, from top to bottom. The results clearly indicate that the transform is sensitive to this so that in actual use, it must be utilized in a known manner so as not to introduce unknown distortion into the transformed image. This requires more experimentation.

In this section simple experiments were described involving the generation of logarithmic isometric plane representations of iconic imagery from planar projection images. While this does not indicate all the possible ramifications in doing this, it does indicate that it is possible. Again, further work is indicated in which, for example, numerical extraction of optical flow from the logarithmic isometric plane representation is geometrically interpreted as range/depth etc.

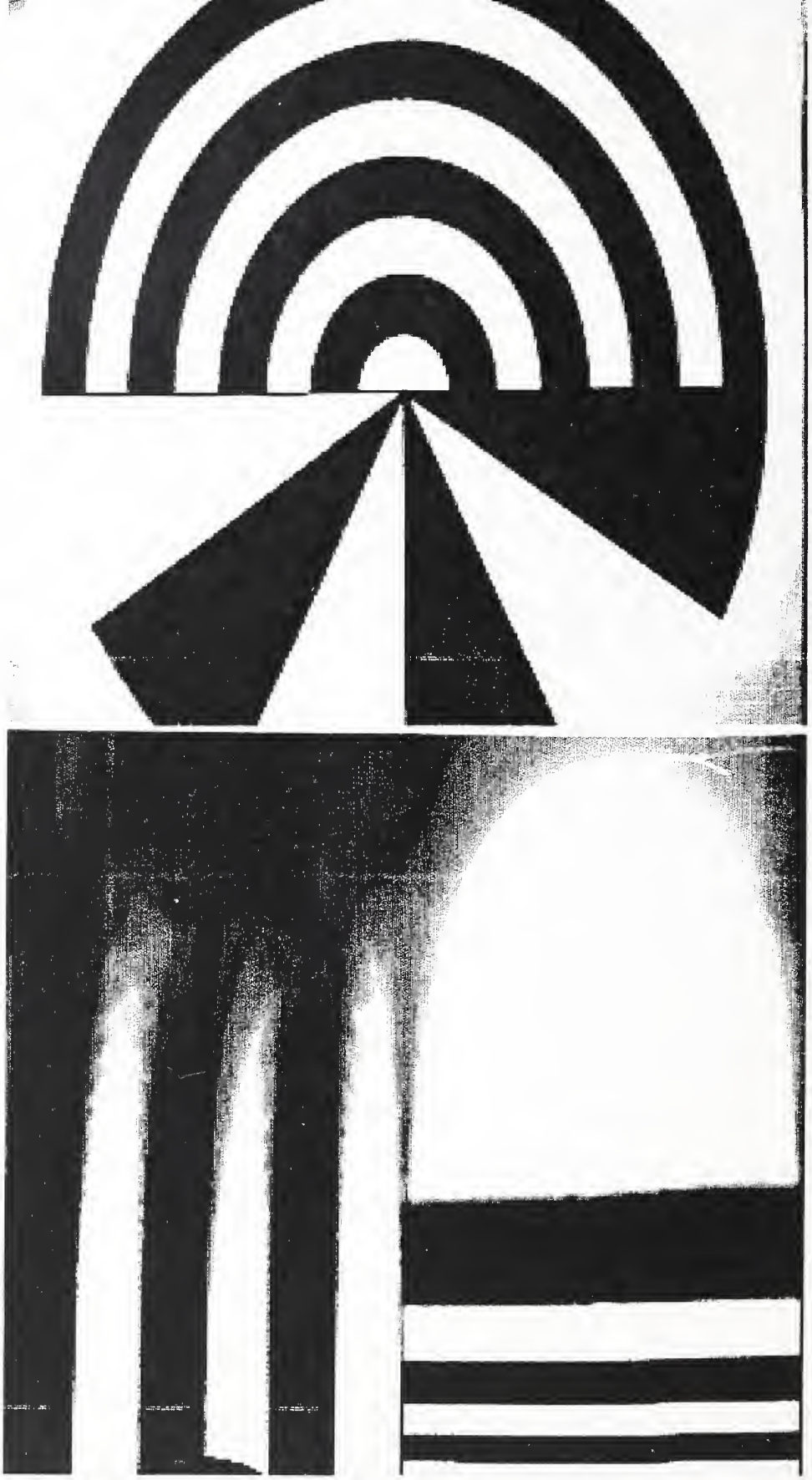
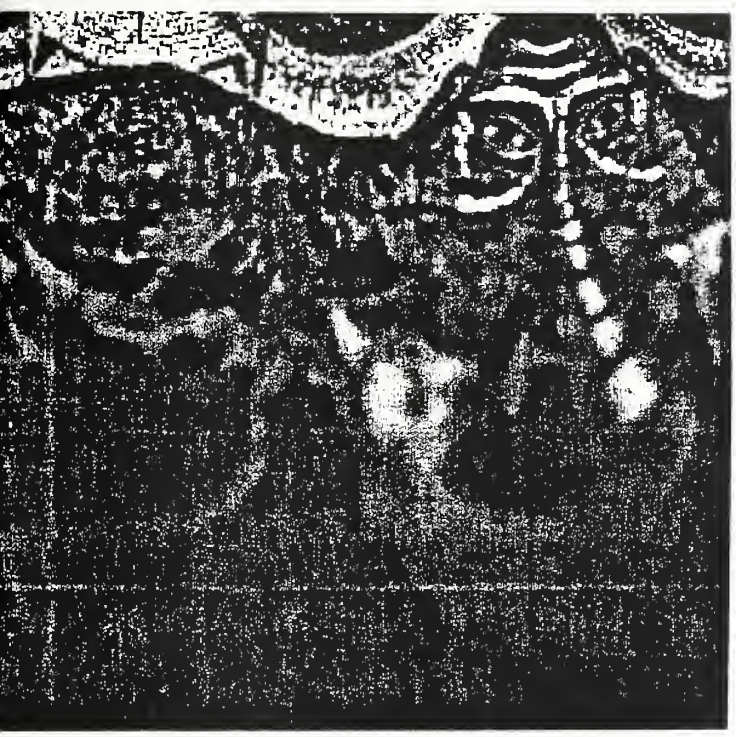


Figure L4: Test Scene Image - At the top is the test image with center nominally coincident with optical axis. Below is the computed depth normalized image, or equivalently, the log-polar transform. The non-square pixel size was taken into account in computing transform, but not in rendering image at top.



(*a*)

(*b*)

(*c*)

(*d*)

Figure L5: Logarithmic Isometric Plane Representation - At top left is the logarithmic isometric plane representation of the bearded man scene of figure L3. The remaining three logarithmic isometric plane representations are of figure L2 range normalized and have been computed with decreasing fields of view.

6. APPLICATIONS AND CONCLUSIONS

The ideas developed and described in this report have had as their motivation two main objectives:

(1) The development of iconic image representations which facilitate the extraction and geometric interpretation of optical flow and binocular disparity, within the context of image sequences as generated by cameras mounted on a forward translating platform, e. g., as would be the case for a self guiding vehicle.

(2) The exploration of spherical projection geometry for two purposes: (a), as a simpler analytic/geometric model of image formation, optical flow and binocular disparity, and (b), as a potentially better computational representation for acquiring imagery and subsequently extracting and interpreting optical flow and binocular disparity.

In the next subsection the first objective is discussed, followed by a second section on the conclusions regarding the utility of using spherical projection over conventional techniques.

6.1. THE PROBLEM OF SEGMENTING IMAGERY FOR A FORWARD TRANSLATING CAMERA

The image analysis problem for a forward moving camera-retina, whether it be biological or artificial, is highly dependent on the task. At one extreme, as eloquently described in [ALBUS], is the problem of obstacle avoidance: an insect moving at a hundred body lengths per second through a random maze of twigs, branches, and leaves making up a bush, all under the highly variable lighting conditions of deep shade broken by sunlight filtering through a moving canopy of tree leaves. In this case little actual object recognition must take place.

At the other extreme, as in the case of a predator looking for prey, object recognition is of extreme importance: a hawk identifying a mouse as distinct from a moving leaf.

In this report we have given key roles to optical flow and binocular disparity in the belief that they play important roles in both extremes. However, this is clearly only part of the story: my view of the world looks about the same from my car whether moving at 60 mph or stopped, with one eye or two. In attacking the problem we must build on what is understood toward what is not.

Most of the self-guiding vehicle research to date is exemplified by [THORPE]: camera imagery is used to build an elaborate internal model of the 3 dimensional world. This may be a valid task, but in large part fails to provide the real-time needs

of navigation involved in getting from point A to point B. This later task is primarily obstacle avoidance while simultaneously selecting the immediate trajectory. Once the point is passed, no more effort need be expended in understanding the scene.

It is from this latter point of view of the navigation task that the work reported on here has application: we wish to know, not *what-thing* is an obstacle, but rather if *any-thing* is an obstacle or has the potential of becoming one. Hence the problem is one of identifying, for the optical flow generated by the visible points of 3-D space, the appropriate category: (a), expected, and hence not an obstacle, e. g., the roadway, (b), not expected, and hence an obstacle, and (c), has the potential of becoming an obstacle. Expected objects are identified with particular events of a ''plan'', while unexpected objects provide the source of new information for incorporation into an updated plan.

It is in this regard that the four parameterizations of space for which normalization has been defined (potentially along with others) seem particularly applicable: range, depth, looming and clearance. Regions in which optical flow of a given magnitude is allowable can be defined by ''carving out'' certain regions of 3-D space, in terms of these 1-dimensional parameterizations.

In this report we have identified a representation for these normalizations of optical flow, namely the logarithmic isometric plane, for which the techniques of optical flow interpretation, as given in section 2.3, are applicable.

In section 2.3 of this report the scenario of a laterally translating camera (translating sideways to the optical axis) viewing a dynamic scene was developed. The linear nature of the geometry of the resulting optical flow facilitated its geometric interpretation. More precisely, in terms of the ideas developed in section 3.2.3, depth is a natural 1-dimensional parameterization of the viewed scene.

The nature of the aperture problem was also analyzed with the intent of demonstrating that edges and contours, at unknown angles, still provide sufficient information to uniquely determine relative motion amongst translating objects. In particular, we outlined an algorithm in that section which provided a method of segmenting an image based on differentiating relative motions in a robust manner, i. e., the Hough transform based algorithm described in section 2.3.4.

In section 3.2.2 the optical flow for a forward translating camera-retina, modeled by spherical projection, was transformed in a manner giving it the same properties as that for a laterally translating camera, but with respect to range, rather than depth. (Range is the Euclidean distance, while depth is used for the distance to the perpendicular plane containing the point.) This range normalized representation

is then amenable to being segmented by the Hough transform algorithm.

In addition, section 3.2.3 developed three additional transformations, called depth, looming and clearance normalization, for which their respective logarithmic isometric plane representations have the linear optical flow properties corresponding to that for range. In particular, the Hough segmentation algorithm is applicable for each of these.

More particularly, optical flow normalized to each of these parameterizations will generate clusters differently in the Hough transform space. A perpendicular flat surface will cluster in the depth normalized transform but not in the others. A point source will maintain a constant range (if the translation velocity is small compared to the range) and hence a saccading sequence of range normalized images will cluster to the same point.

While these ideas may be mathematically correct, more work needs to be done to demonstrate their practicality and usefulness for real imagery.

6.2. CONCLUSIONS REGARDING SPHERICAL PROJECTION

An important aspect of the work described in this report is concerned with whether spherical projection per se is in some way superior to conventional methods of image formation, i. e., planar projection. By spherical projection here is meant either actual projection to a sphere, or as suggested in section 3.2, a simultaneous spherical projection combined with a projection to the plane, e. g., the polar spherical projection. In particular, this question is addressed with respect to the the extraction and interpretation of optical flow and binocular disparity.

Until such time as the principles of vision are well understood, the answer to these questions can only be guessed at. However, within the context of performing vision research, certain conclusions are worth noting. These conclusions have several aspects which are addressed in the following three subsections.

6.2.1. SPHERICAL PROJECTION AND THE ANALYTIC/GEOMETRIC MODEL

The purpose of the analytic/geometric model is to facilitate our reasoning about optical flow and binocular disparity.

For the spherical projection model developed in section 3 of this report, we know a priori that numerical extraction (as opposed to the interpretation, addressed next) of either optical flow or binocular disparity from iconic imagery will not be improved by being on a spherical manifold. The spherical manifold only

complicates the underlying differential geometry, e. g., as mentioned in section 3.2.2 with respect to the visual flow constraint equation. Since numerical extraction is primarily a computational issue it is addressed again below.

With respect to the question of whether the spherical projection model can facilitate the analysis of optical flow and binocular disparity for the purpose of understanding its geometric interpretation, the answer is yes. Section 3 of this report elaborated such a model. In fact it is generally recognized within the vision research community that the analysis of optical flow is facilitated by the spherical projection model [ALBUS, NELSON, RAVIV]. However, for the most part this has been restricted to aspects of analysis only. In particular we make the following additional points:

(1) In sections 3.1.2 and 3.1.3 bi-retinal and binocular coordinates were defined and shown to be related to spherical projection in a simple way. In particular, binocular disparity was shown to be a natural coordinate for parameterizing 3 dimensional space and also for controlling camera pan and tilt for purposes of vergence and bringing a feature point of the scene to the fovea.

These coordinates were defined within the context of the analytic/geometric spherical projection camera-retina imaging model and are an example of its usefulness.

(2) In section 3.2.1 attempts to maintain the Euclidean metric (i. e., the distance between two points is independent of where they are in the image) for spherical projection were abandoned altogether by the introduction of the logarithmic spherical projection as a means of providing variable foveal-peripheral resolution. This was shown, using the analytic/geometric model, to linearize optical flow so as to simplify its geometric interpretation (section 3.2.2).

This report has used the spherical projection analytic/geometric model to relate the geometry of the sphere to a planar manifold. In particular, the simplicity of the spherical model has been combined with a planar surface representation which lends itself to conventional image processing algorithms.

(3) In section 3.2.3 additional mappings were mathematically defined for the analytic/geometric model and shown to provide additional 1-dimensional parameterizations of 3-dimensional space for which optical flow is easily interpreted. In particular, the application of these mappings to iconic imagery, called normalization, led to the logarithmic isometric plane representation.

Again, this analysis was performed within the spherical projection analytic/geometric model and facilitated the understanding of how to interpret

optical flow.

(4) The binocular wire frame scene simulator models the analytic/geometric model. In section 5 experiments verifying the correctness of the analytic/geometric model with respect to the normalization procedures developed in section 3 are described.

We suggest that more development of the spherical projection analytic/geometric model will provide additional benefits. For example, the modeling of saccadic action by camera-retinas in terms of the generators for the Lie group SO(3) of 3-dimensional spherical rotations has been noted by [CHEN], and might take advantage of this latter's high degree of development [KARGER].

Based on these arguments we conclude that for purposes of understanding the geometry of optical flow and binocular disparity, an analytic/geometric model based on spherical projection is simpler and hence more useful than the conventional planar projection models.

6.2.2. SPHERICAL PROJECTION AND THE COMPUTATIONAL MODEL

The computational model consists of iconic image representations, algorithms for acting on those representations, assumptions about the imaging process and the task to be performed etc. It is the environment available to the computation which is to perform the given vision task.

With respect to our conclusions regarding the question of the relative efficiency of spherical projection, the forward translating camera vision task appears not to be of major significance one way or the other.

Rather, it is the incomplete vision research task itself, as performed by any means available, which dictates what representations and strategies seem most appropriate.

Based on our work described in this report it is our conclusion that from a purely computational point of view there are no overriding advantages to using pure spherical projection. This conclusion is in large part the result of the methods conveniently available for storing, i. e., representing, and processing iconic imagery. There are two aspects to this.

At the lowest level, iconic imagery should be stored in a manner which models the underlying manifold geometry of the image itself. This is exemplified by the image formed on the flat chip of a standard video camera: the manifold geometry of the chip tessellation and that of conventional computer memory match.

Given that it were possible to build a spherical analog of a flat chip, and hence obtain a regular sampling of the spherical image, it is still problematic to maintain the spherical geometry.

This is a direct result of the fact that the sphere cannot be mapped to the plane without distorting it, either as a continuous mapping, or as a regular tessellation, e. g., as would be needed to map it to conventional computer memory in a manner allowing systematic access to a pixel and its immediate neighbors in a reasonable way, though this is possible with considerable complications [KAC].

As mentioned above, the solution chosen here is to abandon the Euclidean metric and map the sphere in a way that explicitly facilitates the extraction of optical flow and binocular disparity. In particular, the following point is made for imagery obtained from a forward translating camera:

(1) In the mapping of the sphere to the logarithmic isometric representation, the problem of extraction for optical flow and binocular disparity becomes identical to that for the laterally translating camera, and hence the same techniques for their extraction and interpretation become applicable.

The second aspect concerning the representation of iconic imagery which has caused us to make the above conclusion regarding spherical projection, concerns the use of orthonormal polynomial bases, e. g., Chebyshev, Hermite polynomials, for representing iconic imagery, e. g., the "facet model" [HARALICK2] and their use in computing image gradients, [HASHIMOTO, MEER]. These methods seem attractive to us with a corresponding impact on our conclusions.

The above ideas will not be developed here, but rather the point will be made that once such a representation is engaged upon, the choice of what type of projection to use is of little consequence: whatever projection is desired may be formed out of the original projection through the appropriate (non-linear) spatial sampling of the polynomial representation.

More particularly, in section 3.3, mathematical relationships were developed between the normalization mappings of the sphere to the plane, were put into a computational form, and in section 5.2, computational examples of these nonlinear samplings was given. This provides the basis for what might be called the "computational interchangeability of planar and spherical projection."

In more detail we have demonstrated the following in developing this computational interchangeability:

(2) Based on the mathematical relationship between spherical projection and planar projection, the computation for sampling a planar projection was derived in

section 3.3.2.2 which generates the logarithmic spherical projection normalized for range. In section 5.2.1 an example conversion for doing this is described.

(3) In section 3.3.1 the fundamental relationship between spherical and planar projection was used to derive the normalization sampling needed for each of the 1-D parameterizations, i. e., range, depth, looming and clearance.

(4) In section 3.3.2.1 the computation needed for sampling a planar projection image in a manner which generates the depth and range normalizations as represented by the logarithmic isometric plane are given.

(5) In section 5.2.2 several examples of both depth normalized and range normalized spherical projections, as represented by the logarithmic isometric plane, are generated from planar projections.

Based on these example computations the feasibility of non-linear sampling of images for the purpose of binocular and optical flow disparity interpretation, as well as other purposes, has been demonstrated. These samplings were performed on planar projection images, hence demonstrating that spherical projection images are not required.

Until such time as general artificial vision systems are tailored to perform specific tasks, the question of the relative computational utility of spherical projection should not be a significant one.

6.2.3. SPHERICAL PROJECTION AND PRACTICAL ENGINEERING CONSIDERATIONS

We briefly address some issues related to the question of the practicality of our conclusions regarding the relative utility of spherical projection.

In a research environment, our conclusions seem perfectly appropriate: the giving up of computational efficiency in return for computational versatility. This is particularly true if the research is concerned with the development of a general vision paradigm, and less so for the development of a vision system for a specific task.

One of the characterizing features of spherical projection is its potential very wide field of view. In contrast, planar projection is limited to 180° by its mathematical definition and in practice to much less due to the difficulty of keeping the resulting image distortion free. The economics are such that a planar projection lens approaching 180° will cost something approaching 10,000 dollars.

Equidistant projection (termed polar spherical projection in this report) lenses are available from about 180° (1,200 dollars) up to 220° (14,000 dollars). These

prices are probably more a reflection of the low quantity sold than of the intrinsic difficulty of their manufacture.

The computing of optical flow over large angles of panning does not seem to take place in biological vision. For smaller angles, i. e., eye saccades, the internal maintenance of the image as a spherical projection using polynomials defined on a spherical manifold might provide image stabilization under rotation.

Again, until such time as a general purpose artificial vision methodology is arrived at and subsequently used to design a vision system for a specific task, the relative significance of spherical projection to that engineering design are premature.

ACKNOWLEDGEMENTS

The author would like to thank Dani Raviv of Florida Atlantic University for making his ideas available to him and for many stimulating conversations concerning the nature of both natural and artificial vision. Also to be thanked is Tom Kramer of the Robot Systems Division for suggesting a large number of corrections and improvements to the text, and Jim Albus, Chief of the Robot Systems Division, for providing the necessary support.

REFERENCES

[ALBUS] James S. Albus and Tsai Hong Hong, "Motion Depth, and Image Flow," Proceedings, *IEEE Robotics and Automation Conference,* Cincinnati, Ohio, May 13-18, 1990.

[BRUSS] Anna R. Bruss and Berthold K. P. Horn, "Passive Navigation," *Computer Vision, Graphics, and Image Processing,* Vol. 21, 1983, pp. 3-20.

[BUXT] Buxton, B. F., and Buxton, H., "Computation of Optical Flow from the Motion of Edge Features in Image Sequences," *Image and Vision Computing,* 2, pp. 59-75, May, 1984.

[CHEN] Su-shing Chen, "An Intelligent Computer Vision System," *International Journal of Intelligent Systems,* Vol. 1, 1986, pp. 1-15.

[DOUGHERTY] Edward R. Dougherty and Charles R. Giardina, *Image Processing-Continuous to Discrete: Geometric, Transform, and Statistical Methods,* Prentice-Hall Inc., 1987, p. 236.

[DOW] B. M. Dow, R. G. Vautin and R. Bauer, "The Mapping of Visual Space onto Foveal Striate Cortex in the Macaque Monkey," *The Journal of Neuroscience,* Vol. 5, No. 4., April 1985, pp. 890-902.

[DUDA] Duda, R. O. and Hart, P. E., "Use of the Hough Transformation to detect Lines and curves in Pictures," *Communications of the ACM 15,* no. 1, January 1972, pp. 11-15.

[ESSEN] David C. Essen and Charles H. Anderson, "Information Processing Strategies and Pathways in the Primate Retina and Visual Cortex," *An Introduction to Neural and Electronic Networks,* edited by Steven F. Zornetzer, Joel L. Davis and Clifford Lau, Academic Press, Inc., New York, 1990, pp. 43-72.

[FISHER] T. E. Fisher and R. D. Juday, "A Programmable Video Image Remapper," *Proceedings of SPIE,* Vol. 938, 1988, pp. 122-128.

[GHOSH] Sanjib K. Ghosh, *Analytical Photogrammetry,* 2nd edition, Pergamon Press, New York, 1988.

[GIBSON] James Jerome Gibson, *The Ecological Approach to Visual Perception,* Houghton Mifflen, Boston, 1979.

[GOMBRICH] E. H. Gombrich, "Ambiguities of the Third Dimension," ch. VIII of *Art and Illusion: A Study in the Psychology of Pictorial Representation,* Bollinger Foundation Series, Princeton University Press, Princeton, N. J., 1960, pp. 243-287.

[GROSSO] Enrico Grosso, Giulio Sandini and Massimo Tistarelli, "3-D Object Recognition Using Stereo and Motion," *IEEE Transactions on Systems, Man, and Cybernetics,* Vol. 19, No. 6, November/December 1989, pp. 1465-1476.

[HARALICK1] Robert M. Haralick, "Using Perspective Transforms in Scene Analysis," *Computer Graphics and Image Understanding,* Vol. 13, 1980, pp. 191-221.

[HARALICK2] Robert M. Haralick, "Digital Step Edges from Zero Crossing of Second Directional Derivatives," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol 6, No. 1, January 1984, pp. 58-68.

[HASHIMOTO] M. Hashimoto and J. Sklansky, "Multiple-Order Derivatives for Detecting Local Image Characteristics," *Computer Vision, Graphics, and Image Processing,* Vol. 39, 1987, pp. 28-55.

[HEEGER] David J. Heeger, "Optical Flow Using Spatiotemporal Filters," *International Journal of Computer Vision,* Vol. 1, No. 4., 1987, pp. 279-302.

[HELMHOLTZ] Herman Ludwig Ferdinand Von Helmholtz, *Treatise on Physiological Optics,* Translated and edited from the 3rd German edition by J. P. C. Southall, Volume 3, "The Perceptions of Vision," The Optical Society of America, 1925.

[HERMAN] Martin Herman and Tsai-Hong Hong, "Visual Navigation Using Optical Flow," *Proceedings NATO Defense Research Group Seminar on Robotics in the Battlefield,* Paris, France, 1991.

[HOFFMAN] William C. Hoffman, "The Lie Algebra of Visual Perception," *Journal of Mathematical Psychology,* Vol. 3, 1966, pp. 65-98.

[HORN1] Berthold K.P. Horn and Brian G. Schunk, "Determining Optical Flow," *Artificial Intelligence, Vol. 17, 1981, pp. 185-203.*

[HORN2] Berthold K. P. Horn, *Robot Vision,* The MIT Press, Cambridge, Massachusetts, 1986, pp. 400-416.

[JAIN] Ramesh Jain, Sandra L. Bartlett and Nancy O'Brian, "Motion Stereo Using Ego-Motion Complex Logarithmic Mapping," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 9, No. 3, May 1987, pp. 356-368.

[KAC] C. H. Chen and A. C. Kak, "A Robot Vision System for Recognizing 3D Objects in Low-Order Time," *IEEE Transactions on Systems, Man and Cybernetics,* Vol. 19, No. 6, Nov-Dec, 1989, pp. 1535-1563.

[KARGER] Adolf Karger and Josef Novak, "Motion on the Unit Sphere," *Space Kinematics and Lie Groups,* Translated by Michal Basch, Gordon and Breach

Science Publishers, New York, 1978 and 1985, pp. 94-195.

[KRAKIWSKY] Edward J. Krakiwsky, *Conformal Map Projections in Geodesy,* Department of Surveying Engineering, University of Brunswick, Fredericton, New Brunswick, Canada, 1974.

[KOENDERINK] J. J. Koenderink and A. J. van Doorn, "Local Structure of Movement Parallax of the Plane," *Journal of the Optical Society of America,* Vol. 66, No. 7, July 1976, pp. 717-723.

[KOENDERINK] J. J. Koenderink and A. J. van Doorn, "Exterospecific Component of the Motion Parallax Field," *Journal of the Optical Society of America,* Vol. 71, No. 8, August 1981, 953-957.

[LEE] D. N. Lee, "The Optic Flow Field: the Foundation of Vision," *Phil. Trans. R. Soc. Lond.,* Series B, Vol. 290, 1980, pp. 169-179.

[LUNEBURG] Rudolf K. Luneburg, "Mathematical Analysis of Binocular Vision," Dartmouth Eye Institute, Princeton University Press, Princeton, NJ, 1947.

[MAUNSELL] John H. R. Maunsell and William T. Newsome, "Visual Processing in Monkey Extrastriate Cortex," *Ann. Rev. Neurosci.,* Vol. 10, 1987, pp. 363-401.

[MEER] Peter Meer and Isaac Weiss, *Smoothed Differentiation Filters for Images,* Technical Report CAR-TR-424 / CS-TR-2194, Center for Automation Research, University of Maryland, College Park, Md, February, 1989.

[MESSNER] Richard A. Messner and Harold H. Szu, "An Image Processing Architecture for Real Time Generation of Scale and Rotation Invariant Patterns," *Computer Vision, Graphics and Image Processing,* Vol. 31, 1985, pp. 50-66.

[NAKAYAMA] Ken Nakayama, "Biological Image Motion Processing: A Review," *Vision Research,* Vol. 25, No. 5, 1983, pp. 625-660.

[NELSON] Randal C. Nelson and John Aloimonos, *Finding Motion Parameters from Spherical Flow Fields,* Technical Report CAR-TR-287 / CS-TR-1840, Center for Automation Research, University of Maryland, College Park, Md, April 1987.

[OGLE] Kenneth N. Ogle, *Researches in Binocular Vision,* W. B. Saunders Co., Philadelphia, Pa., 1950.

[PRATT] William K. Pratt, *Digital Image Processing,* John Wiley and Sons, New York, 1978.

[RAVIV1] Daniel Raviv, "A Quantitative Approach to Camera Fixation," National Institute of Standards and Technology Report NISTIR 4324, May 1990.

[RAVIV2] Daniel Raviv, "A Closed Form Massively Parallel Range from Image Flow Algorithm," National Institute of Standards and Technology Report NIS-TIR 4450, October, 1990.

[RAVIV3] Daniel Raviv, "Invariants in Visual Motion," In Preparation, 1991.

[ROJER] Alan S. Rojer and Eric L. Schwartz, "Design Considerations for a Space-Variant Visual Sensor with Complex-Logarithmic Geometry," *Proceedings of the Tenth International Conference on Pattern Recognition,* Atlantic City, New Jersey, June 16-21, 1990, pp. 278-285.

[SCHULTZE] M. Schultze, "Zur Anatomie und Physiologie der Retina," *Archiv Für Mikroscopische Anatomie,* 2, 1866, pp. 175-286.

[SCHUNK] Brian G. Schunk, "The image Flow Constraint Equation," *Computer Vision, Graphics, and Image Processing,* Vol. 35, 1986, pp. 20-46.

[SCHWARTZ1] Eric L. Schwartz, "Spatial Mapping in the Primate Sensory Projection: Analytic Structure and Relevance to Perception," *Biological Cybernetics,* Vol. 25, 1977, pp. 181-194.

[SCHWARTZ2] Eric L. Schwartz, "Computational Anatomy and Functional Architecture of Striate Cortex: A Spatial Mapping Approach to Perceptual Coding," *Vision Research,* Vol. 20, 1980, pp. 645-669.

[SCHWARTZ3] Eric L. Schwartz, Bjorn Merker, Estarose Wolfson and Alan Shaw, "Applications of Computer Graphics and Image processing to 2-D and 3-D Modeling of the Functional Architecture of Visual Cortex," *IEEE Computer Graphics & Applications,* July 1988, pp. 13-23.

[THORPE] Charles E. Thorpe, *Vision and Navigation: The Carnegie Mellon Navlab,* Kluwer Academic Publishers, 1990.

[WAXMAN] Allen M. Waxman and James H. Duncan, "Binocular Image Flows: Steps Toward Stereo-Motion Fusion," *IEEE transactions on Pattern Analysis and Machine Intelligence,* Vol. 8, No. 6, November 1986, pp. 715-729.

[WEIMAN] Carl F. R. Weiman and George Chaikin, "Logarithmic Spiral Grids for Image Processing and Display," *Computer Graphics and Image Processing,* Vol. 11, 1979, pp. 197-226.

[WOOD] Robert W. Wood, "Fish-eye Views," *Phil. Mag.,* August, 1906.

APPENDIX A1: DERIVATION OF OPTICAL FLOW EQUATIONS FOR A PLANAR PROJECTION CAMERA

None of what follows is original, but is available in less detail in for example, [BRUSS, HORN].

1. THE CAMERA MODEL

We imagine the camera to be mounted on a vehicle for which the primary motion is one of translation in the direction of the optical axis of the camera, and further, that the environment through which the vehicle moves moves is static.

The camera may be used to define a coordinate frame X, Y, Z which stays fixed with respect to the camera, in which the positive X-axis is coincident with the forward pointing optical axis of the camera lens. The imaging plane is then parallel to the plane containing the Y and Z axes. We define the origin O_c of this camera coordinate frame to be at the nodal point ("pinhole") for the camera lens. Hence the imaging plane will be at a distance f along the positive X-axis, where f is the focal length of the lens. See figure C1.

We define the origin o_c of the imaging plane coordinate frame y, z to be at $[X = f, Y = 0, Z = 0]$ oriented so that the y and z axes are parallel with the Y and Z axes respectively. A point P, whose three dimensional camera coordinates are $[X_P, Y_P, Z_P]$, maps to a point p on the imaging plane whose coordinates are $[y_p, z_p]$. The mathematical relationship is obtained by noting the two similar triangles O_c, p, o_c and O_c, P and the point on the X-axis containing the perpendicular to it through the point P, i. e.,

$$\frac{y_p}{f} = \frac{Y_P}{X_P}, \text{ and } \frac{z_p}{f} = \frac{Z_P}{X_P}. \qquad 1.1$$

Hence the conversions between the coordinate frames are, (dropping the subscripts for a particular point)

$$y = f\frac{Y}{X}, \text{ and } z = f\frac{Z}{X} \qquad 1.2$$

and inversely, where X becomes in effect a parameter,

$$Y = \frac{y\,X}{f}, \text{ and } Z = \frac{z\,X}{f}. \qquad 1.3$$

Note that equations 1.3 are the parametric equations for the line determined by the points O_c and p. The point P is then constrained to lie on this line.

2. CAMERA MOTION

The camera may be treated as a rigid body for the purpose of describing its motion. It can be shown that any general rigid body motion may be decomposed into a pure translational component **T** and a rotational component ω, each of which is parameterized by three scalars. For our purpose here, we will define the optical axis X to be the axis along and about which this translation and rotation occurs, and hence may characterize **T** by a vector $[U, V, W]$ and ω by an "angular velocity" vector $[A, B, C]$. Letting **r** be the vector $[X, Y, Z]$ representing point P, this decomposition into translational and rotational components relative to P is of the form

$$-\mathbf{T} - \omega \times \mathbf{r}. \tag{2.1}$$

More precisely, what we have described is an instantaneous motion so that the corresponding infinitesimal motion is in fact velocity $\mathbf{V} = [\dot{X} \equiv \frac{dX}{dt}, \dot{Y} \equiv \frac{dY}{dt}, \dot{Z} \equiv \frac{dZ}{dt}]$. Hence we may rewrite 2.1 interpreting it as a velocity:

$$\mathbf{V} = -\mathbf{T} - \omega \times \mathbf{r}, \tag{2.2}$$

or in component form,

$$\begin{bmatrix} \dot{X} \\ \dot{Y} \\ \dot{Z} \end{bmatrix} = -\begin{bmatrix} U \\ V \\ W \end{bmatrix} - \begin{bmatrix} A \\ B \\ C \end{bmatrix} \times \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}$$

$$= \begin{bmatrix} -U - BZ + CY \\ -V - CX + AZ \\ -W - AY + BX \end{bmatrix}. \tag{2.3}$$

In words, this equation allows us to calculate the velocity vector $[\dot{X}, \dot{Y}, \dot{Z}]$ of a point X, Y, Z knowing the parameters of the motion.

3. OPTICAL FLOW

The optical flow at the point $p(x, y)$ in the image plane is a two dimensional vector whose components are $[\dot{y}, \dot{z}]$. Hence differentiating 1.2 with respect to time we obtain

$$\dot{y} = f\left[\frac{\dot{Y}}{X} - \frac{Y\dot{X}}{X^2}\right], \text{ and } \dot{z} = f\left[\frac{\dot{Z}}{X} - \frac{Z\dot{X}}{X^2}\right]. \tag{3.1}$$

Substituting the expressions for $\dot{Y}$ and $\dot{Z}$ obtained in 2.3, we obtain

$$\dot{y} = f\left[\frac{-V - C\,X + A\,Z}{X} - \frac{Y}{X^2}(-U - B\,Z + C\,Y)\right] \qquad 3.2$$

$$\dot{z} = f\left[\frac{-W - A\,Y + B\,X}{X} - \frac{Z}{X^2}(-U - B\,Z + C\,Y)\right], \qquad 3.3$$

which is in terms of the camera motion parameters and three dimensional coordinates of P. This latter may be replaced by depth (i. e., X, as opposed to range $\sqrt{X^2 + Y^2 + Z^2}$), and the corresponding point p expressed in image coordinates y, z by substituting them for Y and Z from 1.2 and simplifying:

$$\dot{y} = f\left[-\frac{V}{X} - C + A\,\frac{z}{f}\right] + y\left[\frac{U}{X} + B\,\frac{z}{f} - C\,\frac{y}{f}\right] \qquad 3.4$$

$$\dot{z} = f\left[-\frac{W}{X} - A\,\frac{y}{f} + B\right] + z\left[\frac{U}{X} + B\,\frac{z}{f} - C\,\frac{y}{f}\right]. \qquad 3.5$$

Each of these expressions may be broken into two components, one of which is a function of the translation parameters and the other of the rotational parameters:

$$\frac{\dot{y}_t}{f} = \left[\frac{-V + y\frac{U}{f}}{X}\right], \quad \text{and} \quad \frac{\dot{y}_r}{f} = \left[B\,y\,z - C\,(y^2 + f) + \frac{A}{f}\,z\right] \qquad 3.6$$

$$\frac{\dot{z}_t}{f} = \left[\frac{-W + z\frac{U}{f}}{X}\right], \quad \text{and} \quad \frac{\dot{z}_r}{f} = \left[B\,(z^2 + f) - C\,y\,z - \frac{A}{f}\,y\right]. \qquad 3.7$$

APPENDIX A2: DERIVATION OF OPTICAL FLOW NUMERICAL EXTRACTION AS RATIO OF TEMPORAL AND SPATIAL GRADIENTS

Starting with the *visual flow constraint equation* [SCHUNK],

$$-\frac{\partial f}{\partial t} = \frac{\partial f}{\partial x}\,u + \frac{\partial f}{\partial y}\,v\,, \qquad 1$$

note that the right hand side can be written as a dot product,

$$-\frac{\partial f}{\partial t} = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y}\right) \cdot \left(\frac{dx}{dt}, \frac{dy}{dt}\right)$$

$$= \nabla f \cdot \left(\frac{dx}{dt}, \frac{dy}{dt}\right)$$

$$= |\nabla f|\,\bar{u}_n \cdot \left(\frac{dx}{dt}, \frac{dy}{dt}\right), \qquad 2$$

where $\bar{u}_n$ is the unit vector parallel to ∇f. Optical flow is identified with the vector $(\frac{dx}{dt}, \frac{dy}{dt})$, but in fact only the component of this flow normal to a moving contour is observable in an image sequence, i. e., the amount observable is $\bar{u}_n \cdot (\frac{dx}{dt}, \frac{dy}{dt})$. The angle ϕ between u_n and $(\frac{dx}{dt}, \frac{dy}{dt})$, i.e., the angle between ∇f and the true direction of contour motion is not uniquely determined and is in general unknown, i.e., the "aperture problem". The magnitude of the dot product of the optical flow and the unit gradient, i.e., the observed normal component is, dividing through by $|\nabla f|$

$$\bar{u}_n \cdot (\frac{dx}{dt}, \frac{dy}{dt}) = \frac{-\frac{\partial f}{\partial t}}{\left[\left[\frac{\partial f}{\partial x}\right]^2 + \left[\frac{\partial f}{\partial y}\right]^2\right]^{\frac{1}{2}}}. \qquad 3$$

In what follows we will denote by $\bar{U}_n$ the observed, or apparent, optical flow, i.e., a vector having magnitude $\bar{u}_n \cdot (\frac{dx}{dt}, \frac{dy}{dt})$ but having the direction of $|\nabla f|$

Denote by Δt, Δx and Δy the calculated values of $\frac{\partial f}{\partial t}$, $\frac{\partial f}{\partial x}$ and $\frac{\partial f}{\partial y}$ at point t, x, and y respectively. Then $\bar{U}_n$ may be calculated from

$$|\bar{U}_n| = \frac{-\Delta t}{\sqrt{(\Delta x)^2 + (\Delta y)^2}} \qquad 4$$

$$\theta = \tan^{-1} \frac{\Delta y}{\Delta x},$$

where θ is the angle $\bar{U}_n$ makes with the positive x–*axis*.

The x and y components of $\bar{U}_n$, (u, v), may be calculated from

$$\begin{aligned} u &= |\bar{U}_n| \cos\theta \\ &= |\bar{U}_n| \frac{\Delta x}{\sqrt{(\Delta x)^2 + (\Delta y)^2}} \\ &= \frac{-\Delta t}{\sqrt{(\Delta x)^2 + (\Delta y)^2}} \frac{\Delta x}{\sqrt{(\Delta x)^2 + (\Delta y)^2}} \\ &= \frac{-\Delta t \Delta x}{(\Delta x)^2 + (\Delta y)^2} \end{aligned} \qquad 5$$

and similarly for $v = |\bar{U}_n| \sin\theta$ for which we obtain

$$v = \frac{-\Delta t \Delta y}{(\Delta x)^2 + (\Delta y)^2}. \qquad 6$$

APPENDIX A3: AN ALGORITHM FOR IMAGE SEGMENTATION BASED ON DIFFERENTIAL OPTICAL FLOW

One application for differential optical flow is to use it to spatially segment imagery into regions which, while indistinguishable from intensity characteristics, are differentiated by reason of their differential optical flow resulting from differences in depth. In this appendix we describe an algorithm which accomplishes the first step of this by classifying edge points in the original image by their depth class. These edge points are in fact the points for which optical flow may be computed.

The central idea in the algorithm is to use the locus of points given by the optical flow normal component theorem to parameterize the family of curves determined by each numerically extracted optical flow vector. This parameterization is an application of a Hough [DUDA] transform in which each data point is used to "vote" for all curves on which it may fall. The result of performing this on many data points is a clustering of votes around particular curves whose parameters then explain the data.

The normal component theorem states that an extracted optical flow vector $|u, v|$ must lie somewhere on the curve

$$|u, v| = |U, V| \cos(\phi - \theta),$$

depending on the (unknown) angle θ the edge makes with the direction of motion and the magnitude of the flow field $|U, V|$ at the point. ($|U, V|$ is nominally $|\dot{x}, \dot{y}|$ but we make a distinction between calculated values and the mathematical model.) However, this equation may also be thought of as characterizing the family of loci all of which contain the data point ($|u, v|, \theta$). That is, we view $|U, V|$ as the dependent variable and ϕ as the independent variable while u, v and θ are constant. The equation

$$|U, V| = \frac{|u, v|}{\cos(\phi - \theta)} \qquad \theta - 90° < \phi < \theta + 90°$$

then generates one point on each possible loci of normal components. A second data point results in a second family of possible loci, and in general will intersect the first family at a point common to both families, i. e., a single locus of normal components on which both data points then fall.

Given many noisy optical flow vector data points the strategy is to find clusters of intersections in the resulting families of loci. This is accomplished by plotting

each family indicated in an accumulative array. A local maxima at location M, Φ in this array then corresponds to a locus of normal components on which falls a number of data points. All these data points support the hypothesis that the corresponding edge points are moving with optical flow magnitude $|U, V| = M$ and direction Φ.

These local maxima in m-ϕ space are rank ordered by some measure of their respective "certainty", e. g., relative height, and the "best" n of them kept. A second pass is then made through the optical flow data points computing for each

$$\min_{k=1}^{n}\left\{M_k - \frac{|u, v|}{\cos(\phi - \theta)}\right\}.$$

Each point is then classed with that hypothesis which yields this minimum, unless the difference is greater than some threshold value, in which case it is placed in residue class $n+1$.

The above described algorithm, while mathematically correct, will work no better than the quality of data supplied to it. However, noise is more likely to result in spurious classification of noise data points than to cause the misassignment of good data points.

A more detailed step by step description of this algorithm follows.

(1) **Initialization:** Set array $MPHI[0{:}MAXMAG; -90°{:}90°] \equiv 0$.

(2) **Generate $m-\phi$ Parameter Space:** For each optical flow normal component $|u, v|$ $\theta = \tan^{-1}\frac{v}{u}$, perform the following:

For $\theta - 90° < \phi < \theta + 90°$ calculate

$$m = \frac{u^2 + v^2}{u \cos\phi + v \sin\phi}$$

$$MPHI[m, \phi] = MPHI[m, \phi] + 1$$

(3) **Locate Clusters:** Locate, rank order and threshold local maxima of MPHI and denote the resulting sequence by

$$[M_1, \Phi_1], [M_2, \Phi_2], \cdots [M_n, \Phi_n].$$

(4) **Classify:** For each optical flow normal component $|u, v|$ $\theta = \tan^{-1}\frac{v}{u}$, with spatial coordinate i, j compute

$$\Delta M_k = \left\{M_k - \frac{u^2 + v^2}{u \cos\Phi_k + v \sin\Phi_k}\right\} \qquad k = 1, \cdots n$$

and set

$$\Delta M_c = \min_{k=1}^{n} \Delta M_k .$$

Then point $I(i,j)$ is in class c with compensated optical flow magnitude M_c moving in direction Φ_c if $\Delta M_c <$ THRESHHOLD, and in residue class $n+1$ otherwise.

In step (2) the entire curve of possible relative motions is generated. An alternative is to plot just the intersections of the curves obtained by solving every pair of optical flow data points for their intersection.

(2′) **Generate m-ϕ point solution Space:** For each distinct pair of optical flow vector data points $m_1 \equiv |u_1, v_1|$, $\theta_1 = \tan^{-1}\frac{v_1}{u_1}$ and $m_2 \equiv |u_2, v_2|$, $\theta_2 = \tan^{-1}\frac{v_2}{u_2}$, calculate

$$m = \frac{\sqrt{(u_1^2 + v_1^2)(u_2^2 + v_2^2)[(u_1 - v_1)^2 + (u_2 - v_2)^2 - 2u_1v_1 - 2u_2v_2]}}{u_2v_1 - u_1v_2}$$

$$\phi = \tan^{-1}\left[\frac{u_2(u_1^2 + v_1^2) - u_1(u_2^2 + v_2^2)}{v_1(u_2^2 + v_2^2) - v_2(u_1^2 + v_1^2)}\right],$$

and set

$$MPHI[m, \phi] = MPHI[m, \phi] + 1$$