

AO-A171 033

CONSTRAINTS ON THE COMPUTATION OF RIGID MOTION
PARAMETERS FROM REEINAL DISPLACEMENTS(U) ROCHESTER UNIV
NY DEPT OF COMPUTER SCIENCE A BANDOPADHAY OCT 85

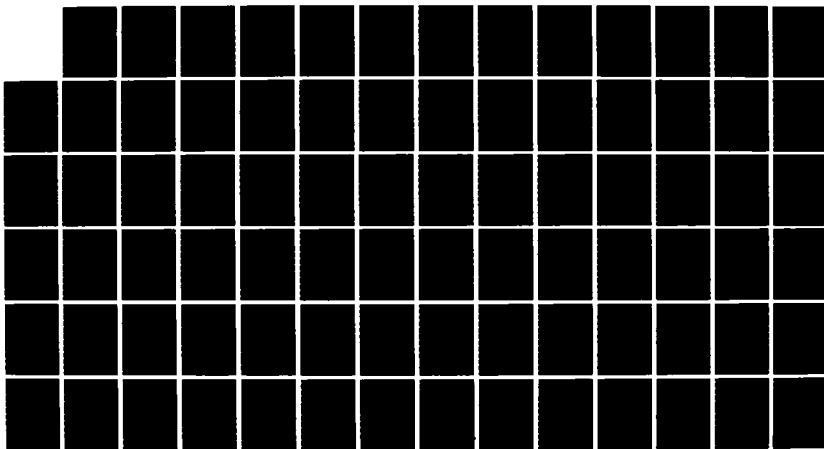
1/1

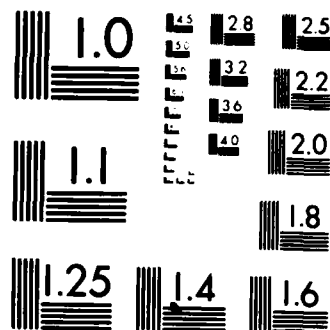
UNCLASSIFIED

TR-168 N00014-82-K-0193

F/G 5/10

NL





MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS 1963-A

AD-A171 033

2

Constraints on the Computation
of Rigid Motion Parameters
from Retinal Displacements

Amit Bandopadhyay
Department of Computer Science
The University of Rochester
Rochester, NY 14627

TR 168
October 1985

DTIC FILE COPY

Rochester
DTIC
ELECTE
AUG 12 1986
E

Department of Computer Science
University of Rochester
Rochester, New York 14627

For information only - not to be used
for any other purpose
10/10/85 10:00 AM

Constraints on the Computation of Rigid Motion Parameters from Retinal Displacements

Amit Bandopadhyay
Department of Computer Science
The University of Rochester
Rochester, NY 14627

TR 168
October 1985

Abstract

A mathematical formulation of the Rigid Motion Perception problem is described. The constraints on the parameters of rigid motion (i.e., three-dimensional velocities) obtained from image motion data (two-dimensional projected velocities) are analyzed. A brief survey of related work shows the lacunae in the existing body of research in this area. Uniqueness results and computational algorithms are presented to compute the rigid motion parameters from retinal velocities. The approximations involved in the velocity representation are stated. Algorithms and constraints to permit cooperative computation of motion and shape are described.

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER TR 168	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Constraints on the Computation of Rigid Motion Parameters from Retinal Displacements		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Amit Bandopadhyay		8. CONTRACT OR GRANT NUMBER(s) N00014-82-K-0193
9. PERFORMING ORGANIZATION NAME AND ADDRESS Computer Science Department University of Rochester Rochester, New York 14627		10. PROGRAM ELEMENT PROJECT TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS DARPA/ 1400 Wilson Blvd. Arlington, VA 22209		12. REPORT DATE October 1985
		13. NUMBER OF PAGES 76
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Office of Naval Research Information Systems Arlington, VA 22217		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Distribution of this document is unlimited		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES None		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number)		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) A mathematical formulation of the Rigid Motion Perception problem is described. The constraints on the parameters of rigid motion (i.e., three-dimensional velocities) obtained from image motion data (two-dimensional projected velocities) are analyzed. A brief survey of related work shows the lacunae in the existing body of research in this area. Uniqueness results and computational algorithms are presented to compute the rigid motion parameters from retinal velocities. The approximations		

20. ABSTRACT (Continued)

involved in the velocity representation are stated. Algorithms and constraints to permit cooperative computation of motion and shape are described.



1. Introduction

Motion is a ubiquitous phenomenon in our everyday life. It is therefore important, in the study of Computer Vision, to understand the retinal motion flux arising both from movement of the observer, as well as the motion of environmental objects. The study of the motion of rigid objects (or surfaces), in particular, is a relevant avenue for investigating motion perception. In general, computing three dimensional motion from monocular two dimensional image motion flux is an underdetermined problem, admitting an infinite number of solutions. The assumption of rigidity makes the problem tractable (see Ullman's paper [33] for a discussion of nonrigid motion perception). Furthermore, most of the moving objects in our environment are rigid. From a practical standpoint, the study of rigid body motion is interesting, since it finds widespread applications in the areas of optical navigation, tracking and recovery of 3D structure of rigid objects.

The motion of a body can be characterized by the rate of change of the positions of various points on its visible surface. Thus, at least instantaneously, this corresponds to a three dimensional velocity field. If the body (or surface) is rigid, then, this velocity field can be described by a vector function of the three dimensional position coordinates and *six global parameters* (see figure 1), which are:

- (i) The three components of the velocity of any point O on the body. These are called the translation parameters.

- (ii) The rotational velocity components of a coordinate frame, with origin O , attached rigidly to the body.

It is a standard result from kinematics and geometry (see [7]) that although the rotational parameters are invariant with respect to the choice of the origin O , of the body frame, the translation parameters are dependent on the choice of O .

When considering motion of rigid bodies, there are two cases of interest, namely, *Egomotion* and *General Motion*. Egomotion or self-motion refers to the movement of the camera or sensor in a static environment. The image flux, or *optical flow*, generated due to such a motion is due to a single relative movement, i.e. between observer and static environment. In contrast, General motion implies that there is more than one object moving with different velocities in the observers field of view. In this case the optical flow field consists of many segments corresponding to the various moving surfaces. Each segment is characterized by the translational and rotational velocities of the associated moving rigid surface inducing the optical flow. These velocities are called the parameters of motion for the rigid surface.

The rigid motion parameters are usually expressed with respect to a frame of reference attached to the moving surface, which is assumed to coincide with the observers frame of reference at the time of observation. The problem is to determine the motion parameters corresponding to a optical flow field segment. If the depth of the scene is unknown then it can be shown that only the rotation - which is depth invariant - can be determined uniquely; whereas the three

translation parameters can only be determined upto a scale factor (this is the depth scaling effect). Thus we can determine five parameters to characterize the motion in this case.

Motion in three dimensions causes the pattern of light falling on the retina (or any two dimensional array of photo-sensitive elements) to vary in time in accordance with the motion. Hence, the input (or stimulus) to any computational process endeavouring to understand the motion, is the two dimensional projection of the three dimensional motion. Since a velocity field is a good representation for the three dimensional motion, it is customary to choose a two dimensional velocity field representation for the image or retinal motion. The latter is called *optical flow*.

The problem addressed in this paper concerns the computation of the parameters of rigid motion and the structure of the moving surface from retinal stimulus such as optical flow.

The optical flow field is a principal source of information about the motion, inducing the "flow", as well as the 3D structure of the moving surface being observed. The optical flow comprises two parts, corresponding to the rotation and the translation, respectively, of the inducing motion. The optical flow due to rigid motion is constrained at every point by the parameters of the motion. However, since the *parameter space* has a *large dimension* and the *constraint is nonlinear* in form, computation of the motion from optical flow (or image displacements) by search techniques is computation intensive.

The optical flow field is mathematically separable into a translational part and a rotational component. It has been long recognized [12] that the motion perception becomes simpler in the instances when the optical flow field can be computationally separated into the translational and rotational parts. A familiar illustration of this is the case of motion parallax observable at depth discontinuities in the retinal field. The effect is to reduce the dimensionality of the space of unknowns. Unfortunately, this seems to be very hard to accomplish, in general. Motion parallax is the basis for an algorithm by Lawton [24]. Other approaches to the problem can be found in [6, 19], involving nonlinear least square techniques or using local constraints involving derivatives of the optical flow.

As stated previously, algorithms for rigid motion perception are difficult to design due to two main reasons:

- (1) The space of parameters is of high dimensionality (e.g. five).
- (2) The Constraint equations obtained by optical flow measurements are non-linear.

There have been some clever implementations of non-linear search algorithms to interpret 3D motion from optical flow data [22,23]. There have also been discrete point tracking algorithms by Tsai and Huang [30] and Fang and Huang [8,9] and Longuet-Higgins [20]. In some of the latter algorithms, the nonlinear motion equations are linearized in terms of synthetic parameters, which are nonlinear combinations of the actual motion parameters. Tsai and Huang, and Fang and Huang, note the cases when such algorithms fail to compute motion

parameters.

In this paper we examine the situations when the optical flow field is capable of being interpreted in more than one way. An instance of such ambiguity is the optical flow field due to motion of a plane [29].

A geometric analysis of the problem of computing 3D motion parameters from 2D image velocities has been done by Longuet-Higgins and Prazdny [19]. The constraint equations that they derive are simple in form, but deal with velocities. To implement a motion analysis algorithm based on these equations, one makes the assumption that the temporal grain of the observations is fine enough to talk meaningfully about the *velocities* or time derivatives of both the image and world positions. Representing motion by velocity parameters entails making a first order approximation of the temporal behaviour associated with the motion. Thus, for example, if the displacement of a particle moving in one dimensional space is Δx in time Δt , then $\frac{\Delta x}{\Delta t}$ is a good approximation for the velocity only when Δt is small enough such that the change in velocity in this time period is small.

An alternative derivation is due to Tsai and Huang [30]. Their approach is to analyze the relation between the projected displacement vectors in the image plane due to an arbitrary rigid displacement of a set of points in 3D. It is known [7] that this type of motion can be characterized by a rotation about an axis passing through the origin of the reference coordinate frame and a translation.

The treatment in this paper assumes the velocity representation for rigid motion. The assumptions underlying the work reported here are:

- (i) The motion being observed, is due to a rigid surface.
- (ii) The time constant (or sampling interval) of the sensor is small enough to make a first order approximation of the temporal behaviour due to the motion being observed.

1.1. Review of previous work

The computation of rigid motion parameters from image displacement vector fields has been studied by a number of researchers. Egomotion has been considered in the literature by Longuet-Higgins and Prazdny [19], Prazdny [22], Waxman and Ullman [34] and Bruss and Horn [6]. Longuet-Higgins and Prazdny examine ways of determining 3D structure and motion parameters from optical flow. Their method depends upon accurate reconstruction of the optical flow field. An interesting result due to them is that for non planar surfaces *local analysis* of the flow field yields a cubic constraint involving the motion parameters. Prazdny ([22]) has devised a five point algorithm to solve for the motion parameters from nonlinear constraint equations. Waxman and Ullman's method depends upon reconstruction of the optical flow field analytically, in local neighbourhoods. Bruss and Horn propose a least square solution to the parameter estimation problem.

Some other computational approaches attempt to segment the optical flow field into translatory and rotatory components, albeit approximately. An example is the method of Reiger and Lawton [24] where the change of rotational flow at

steep depth gradients, is treated as noise. Jain [17,18] computes the *focus of expansion* before computing the image displacements and uses the former to guide the correspondence for finding the latter.

All the above analyses pertain to the computation of motion parameters from optical flow, i.e. continuous or differential image motion. An alternative approach is to consider evaluating the motion parameters and 3D structure from discrete point correspondence. Ullman [32] shows that three views of four non coplanar points is adequate to determine the structure and motion of these points under orthography. Tsai and Huang [30] prove that the motion of seven points not lying on two planes, one of which passes through the origin, nor on a cone passing through the origin, can be uniquely computed, from discrete displacements. Fang and Huang [8,9] prove that structure and motion of nine points not lying on a second order surface passing through the origin is uniquely determined from image displacements. Nagel and Neuman [21] and Roach and Aggarwal [25] have also looked at the problem of determining motion from discrete displacements.

Yet another approach to the problem of motion parameter computation has been to restrict the motion to simplify the analysis. Webb and Aggarwal [35] Hoffman and Flinchbaugh [14] and Hoffman and Bennett [15] analyze rigid motion with the additional assumption of fixed axis of rotation or planarity. An major motivation for this type of analysis is that, it models the locomotion of man and animals.

1.2. Summary of Results reported here

It is evident from the review of the existing body of work in the field of motion perception that, although considerable work has been done, much remains undone. Uniqueness proofs of the type derived by Tsai and Huang and Fang and Huang do not allow us to visualize the situations when the optical flow field is intrinsically ambiguous, admitting more than one interpretation. An analysis of the optical flow field to determine cases of ambiguity will be a major focus of this paper.

When the image formation geometry is modeled by means of the parallel projection model, the constraint equations become simplified. This is also called *Orthographic Projection* model of image formation (see figure IIb). The attendant simplicity in the motion equations can be used to considerable advantage in the preliminary analysis of the motion perception problem. The following results are derived:

1. The component of rotation about the line of sight, the ratio of the other two components of rotational velocity, and the tilt function is uniquely computable from a single optical flow field, for a rigid *non planar* surface.
2. When the surface normals for a rigid surface are known then the motion parameters can be computed uniquely.

The *Perspective Projection model* (see figure IIa) is a more accurate model of image formation by eye or camera. For this model it is proved that:

1. The optical flow field, under the assumptions of rigidity can have at most *three* interpretations.
2. The rigid motion of any surface whose depth from the nodal point of the sensor cannot be expressed by the rational function $\frac{P_1(x,y)}{Q_2(x,y)}$, where P_1 and Q_2 are rational functions of the first and second orders respectively, is uniquely computable from the information in the optical flow field.
3. Two optical flow fields, obtained at different time instants, determine the motion parameters uniquely.
4. The motion parameters are uniquely determined from the optical flow field when the corresponding motion involves rotation only.
5. The optical flow due to planar surfaces is generally ambiguous. However this ambiguity can be resolved either when the flow field is due to more than one plane moving together rigidly, or in the case of a single plane, if its tilt is known.
7. It is feasible to design a cooperative algorithm for computing both shape (e.g. surface normals) and optical flow, under conditions of rigid motion.

2. The Geometry of Rigid Motion

Consider a sensor moving relative to a static scene. The co-ordinate frame (X,Y,Z) is fixed to the sensor (see figure 1). The viewing direction is along the positive z-axis.

A rigid body is defined as a set of points whose relative euclidian distances from all other points in the set are invariants with respect to the transformations of rotation and translation. In addition, since we will generally deal with opaque objects and hence will observe points on a surface (or a manifold) in 3 space. In other words the 3 cartesian coordinates of a point on a rigid body are not independent. Formally,

$$B = (\pi, f)$$

where

$$\pi = \left\{ (X, Y, Z) \mid \text{point on the surface of } B \right\}$$

$$f(X, Y, Z) = 0$$

When the body B is displaced with respect to the frame of reference, we obtain a new representation

$$B' = (\pi', f')$$

The displacement is described by the affine transformation

$$X' = [R]X + T \quad (2-1)$$

Any displacement of a rigid body can be modelled by the above equation, which describes a rotation about an *axis through the origin* and a translation specified by the vector T.

If the rotation angle is small, it can be decomposed into three component rotations about the individual axes separately [16]. In this case R and T are given by

$$R = \begin{bmatrix} 1 & -\omega_z & \omega_y \\ \omega_z & 1 & -\omega_x \\ -\omega_y & \omega_x & 1 \end{bmatrix} \quad T = \begin{bmatrix} t_1 \\ t_2 \\ t_3 \end{bmatrix}$$

Substituting for R and T in equation (2.1) we have,

$$X' = X - \omega_z Y + \omega_y Z + t_1 \quad (2-2.1)$$

$$Y' = Y + \omega_z X - \omega_x Z + t_2 \quad (2-2.2)$$

$$Z' = Z - \omega_y X + \omega_x Y + t_3 \quad (2-2.3)$$

or,

$$\Delta X = t_1 - \omega_z Y + \omega_y Z \quad (2-3.1)$$

$$\Delta Y = t_2 + \omega_z X - \omega_x Z \quad (2-3.2)$$

$$\Delta Z = t_3 - \omega_y X + \omega_x Y \quad (2-3.3)$$

where,

$$\Delta X = X' - X \quad \Delta Y = Y' - Y \quad \Delta Z = Z' - Z$$

We define the parameter vector a for characterizing the motion, where

$$a = (t_1, t_2, t_3, \omega_x, \omega_y, \omega_z)^T$$

Motion perception involves the recovery of the parameters of motion, as well as the structure (or shape) of the moving object. The geometric properties of the three dimensional surfaces and points are related to the geometry of their image. Thus the projective transformation involved in the image formation process must be analyzed. The subsequent analysis considers both the cases of "perspective" as well as "orthographic" projections.

3. Analysis of Motion under Orthography

When the model of image formation involves orthographic or parallel projection, then the mathematical formulation of the problem becomes considerably simpler. It can be argued that this is a valid model of image formation when viewing distant objects, or when the focal length of the camera is large compared to the distance of the viewed surfaces, or when the viewing area is small and centered around the line of sight - as in the case of the field of view corresponding to the *fovea* in the retina.

Under orthography, the projection equation relating the position of a point in three space $P = (X, Y, Z)$ to its image $p = (x, y)$ is:

$$(x, y) = (X, Y)$$

Assuming that after a short while the point moves to a position given by $P' = (X', Y', Z')$ while its image moves to $p' = (x', y')$ the following relations are obtained from equations (2.3):

$$\begin{aligned}\Delta x &= x' - x = \Delta X = t_1 - \omega_2 Y + \omega_3 Z \\ \Delta y &= y' - y = \Delta Y = t_1 + \omega_2 X - \omega_3 Z\end{aligned}\quad (3.1)$$

Optical flow is the time derivative of the image position vector and is denoted by (u, v) where

$$(u, v) = (\dot{x}, \dot{y}) = (\dot{X}, \dot{Y})$$

Alternatively,

$$u = \lim_{\Delta t \rightarrow 0} \frac{\Delta x}{\Delta t} = \frac{dx}{dt} \quad v = \lim_{\Delta t \rightarrow 0} \frac{\Delta y}{\Delta t} = \frac{dy}{dt}$$

The motion parameters are now the translational velocity $V_T = (U, V, W)$ and the rotational velocity $\Omega = (\alpha, \beta, \gamma)$ where:

$$U = \lim_{\Delta t \rightarrow 0} \frac{t_1}{\Delta t} \quad V = \lim_{\Delta t \rightarrow 0} \frac{t_2}{\Delta t} \quad W = \lim_{\Delta t \rightarrow 0} \frac{t_3}{\Delta t}$$

and

$$\alpha = \lim_{\Delta t \rightarrow 0} \frac{\omega_1}{\Delta t} \quad \beta = \lim_{\Delta t \rightarrow 0} \frac{\omega_2}{\Delta t} \quad \gamma = \lim_{\Delta t \rightarrow 0} \frac{\omega_3}{\Delta t}$$

therefore the equations relating image and 3D motion are

$$\begin{aligned} u &= U + \beta Z - \gamma Y \\ v &= V - \alpha Z + \gamma X \end{aligned} \quad (3.2)$$

These equations are exactly identical in form to those obtained under the discrete case (assuming small rotation), i.e. equation (3.1). Strictly speaking, according to the nomenclature adopted before, the motion parameters for the discrete case are $(t_1, t_2, \omega_1, \omega_2, \omega_3)$ and those for the differential case are $(U, V, \alpha, \beta, \gamma)$. However, since equations (3.1) and (3.2) are identical in form, all subsequent analysis is based on the latter equation. Furthermore, the parameters (it will be evident later that only the rotational parameters are of interest here), in both the differential as well as the discrete cases will be referred to by the symbols (α, β, γ) . The treatment of both the cases is identical, the only difference being that derivatives in the differential analysis correspond to differences in the discrete case.

3.1. On the information available in the optical flow field

Observe from equation (3.2) that the image displacement (or image motion field) consists of a translational part and a rotational part. The translational motion parameters are dependent on the origin of reference. In fact the parameters, intrinsic to the motion are those of rotation. Thus relative to a particular point, say the origin (0,0), equation (3.2) becomes:

$$\begin{aligned} u &= \beta Z - \gamma y \\ v &= -\alpha Z + \gamma x \end{aligned} \quad (3.3)$$

where u actually means $u - u(0,0)$, v is $v - v(0,0)$ and Z is $Z - Z(0,0)$. It should be emphasized here that Z denotes depth *relative* to a certain point of reference (in this case it is the origin). If the structure or relative depth is not known then the parameters (α, β, γ) are not completely recoverable. There is an exact analog of equation (3.3) for the discrete case, obtainable from equation (3.1).

Proposition I When the depth function (or structure) is *non planar* the following parameters are uniquely determined from the image displacement field:

- (i) The rotation about the axis aligned with the line of sight, i.e. γ .
- (2) The ratio of the other two parameters, i.e. $\frac{\alpha}{\beta}$.

Proof: The proof is by contradiction. Consider the motion of the *non planar* surface Z_1 , which is described by the parameters $(\alpha_1, \beta_1, \gamma_1)$. The image motion equations (from equation (3.3)) are:

$$\begin{aligned} u &= \beta_1 Z_1 - \gamma_1 y \\ v &= -\alpha_1 Z_1 + \gamma_1 x \end{aligned} \quad (3.4)$$

If possible, let there be another surface Z_2 , whose motion is characterized by the parameters $(\alpha_2, \beta_2, \gamma_2)$, such that the image motion field in both the cases is the same. The motion equation for the second surface is:

$$\begin{aligned} u &= \beta_2 Z_2 - \gamma_2 y \\ v &= -\alpha_2 Z_2 + \gamma_2 x \end{aligned} \quad (3.5)$$

Furthermore, the following relations hold:

$$\begin{aligned}\Delta\gamma &= \gamma_1 - \gamma_2 \neq 0 \\ \frac{\alpha_1}{\beta_1} &\neq \frac{\alpha_2}{\beta_2}\end{aligned}\quad (3.6)$$

From equations (3.4) and (3.5) the following relations are obtained:

$$\begin{aligned}\beta_2 Z_2 - \beta_1 Z_1 - \Delta\gamma x &= 0 \\ -\alpha_2 Z_2 + \alpha_1 Z_1 + \Delta\gamma x &= 0\end{aligned}\quad (3.7)$$

Now since $\alpha_1\beta_2 \neq \alpha_2\beta_1$:

$$Z_1 = \frac{-\alpha_2\Delta\gamma x + \beta_2\Delta\gamma x}{\alpha_2\beta_1 - \alpha_1\beta_2}$$

But this is contrary to the assumption that Z_1 is non planar. Therefore:

$$\frac{\alpha_1}{\beta_1} = \frac{\alpha_2}{\beta_2}$$

Again, this implies (considering equation (3.7)) that

$$\Delta\gamma = 0 \quad \text{or} \quad \gamma_1 = \gamma_2$$

This completes the proof of Proposition I.

Proposition II. The image displacement field generated by a planar surface is linear in the arguments (x, y) . In addition, the parameters $\frac{\alpha}{\beta}$ and γ are uniquely determined by the image displacement field if and only if $\alpha p + \beta q = 0$, where (p, q) is the gradient of the planar surface.

Proof: Consider the equation of the planar surface $Z(x, y)$:

$$Z = \bar{p}x + \bar{q}y + \bar{d}$$

If the motion of the surface is characterized by the parameters $(\bar{\alpha}, \bar{\beta}, \bar{\gamma})$. The image motion (or optical flow) is given by:

$$\begin{aligned}u &= \bar{\beta}(\bar{p}x + \bar{q}y) - \bar{\gamma}x \\ v &= -\bar{\alpha}(\bar{p}x + \bar{q}y) + \bar{\gamma}x\end{aligned}\quad (3.8)$$

The above equation indicates that for planar surfaces the optical flow is linear. It is also true that when the optical flow is linear then the moving surface is planar. Now considering equation (3.3) and substituting for (u, v) from equation (3.8) and rearranging terms:

$$\begin{aligned} 0 &= (\bar{\beta}\bar{p} - \beta p)x + (\bar{\beta}\bar{q} - \beta q - \bar{\gamma} + \gamma)y \\ 0 &= (-\bar{\alpha}\bar{p} + \alpha p + \bar{\gamma} - \gamma)x + (-\bar{\alpha}\bar{q} + \alpha q)y \end{aligned} \quad (3.9)$$

Since the above equations are valid for the entire image we have:

$$\beta p = \bar{\beta}\bar{p} \quad (3.10.1)$$

$$\beta q - \gamma = \bar{\beta}\bar{q} - \bar{\gamma} \quad (3.10.2)$$

$$\alpha p - \gamma = \bar{\alpha}\bar{p} - \bar{\gamma} \quad (3.10.3)$$

$$\alpha q = \bar{\alpha}\bar{q} \quad (3.10.4)$$

Eliminating p, q and γ from the above we get:

$$\beta(\frac{\alpha p}{\alpha}) - \alpha(\frac{\bar{\beta}\bar{p}}{\bar{\beta}}) = (\bar{\beta}\bar{q} - \bar{\alpha}\bar{p})$$

or

$$\mu(\bar{\beta}\bar{p} - \bar{\alpha}\bar{p}) - \bar{\alpha}\bar{q} = 0$$

where $\mu = \frac{\alpha}{\beta}$. The above quadratic equation has a unique solution if and only if:

$$(\bar{\beta}\bar{p} - \bar{\alpha}\bar{p})^2 - 4\bar{\alpha}\bar{\beta}\bar{p}\bar{q} = (\bar{\beta}\bar{q} + \bar{\alpha}\bar{p})^2 = 0$$

Under this condition:

$$\frac{\alpha}{\beta} = \frac{\bar{\alpha}}{\bar{\beta}}, \quad \gamma = \bar{\gamma}, \quad \frac{p}{q} = \frac{\bar{p}}{\bar{q}}$$

Therefore the image motion of planar surfaces uniquely determines the parameters

$(\frac{\alpha}{\beta}, \gamma, \frac{p}{q})$ if and only if $\alpha p + \beta q = 0$.

3.2. Summary

- (1) The analysis under orthographic projection for both differential and discrete motion are nearly identical.
- (2) When the structure of the moving object is known, the motion parameters can be computed uniquely from image motion.
- (3) When the structure is not known then the recoverable parameters are $(\frac{\alpha}{\beta}, \gamma, \frac{p}{q})$. However in this case, the values are unique only when the moving surface is non planar, or a certain condition (see proposition II) holds.

4. Analysis of Rigid Motion for the Perspective Projection Model

Under perspective projection, the "image" is formed by "rays" from points in three space (i.e. world points). These rays are constrained to pass thru a nodal point called the center of perspectivity. The imaging geometry is shown in figure IIa. The nodal point is O, which is also taken as the origin of the frame of reference. An image point $p = (x, y)$ corresponds to the world point $P = (X, Y, Z)$. Here the *focal length* of the imaging system is F. The equation of the ray OP is :

$$\frac{X}{x} = \frac{Y}{y} = \frac{Z}{F}$$

Therefore,

$$x = \frac{FX}{Z}; \quad y = \frac{FY}{Z} \quad (4.1)$$

The above projection is denoted by $(X, Y, Z) \rightarrow (x, y, F)$. Similarly, the projective relation between another world point P' and its image is $(X', Y', Z') \rightarrow (x', y', F)$. Thus from equation (4.1) we have,

$$\Delta x = x' - x = F\left(\frac{x'}{Z'} - \frac{x}{Z}\right)$$

$$\Delta y = y' - y = F\left(\frac{y'}{Z'} - \frac{y}{Z}\right)$$

or,

$$\Delta x = F \frac{Z\Delta x - X\Delta Z}{Z(Z+\Delta Z)} \quad (4-2.1)$$

$$\Delta y = F \frac{Z\Delta y - Y\Delta Z}{Z(Z+\Delta Z)} \quad (4-2.2)$$

Recall that when the 3D rotation angles, characterizing the rigid motion, are "small" then the 3D displacement components are given by the relations:

$$\Delta X = t_1 - \omega_2 Y + \omega_3 Z$$

$$\Delta Y = t_2 + \omega_2 X - \omega_3 Z$$

$$\Delta Z = t_3 - \omega_1 X + \omega_1 Y$$

Thus, substituting for ΔX , ΔY and ΔZ in the equation (4.2) we have:

$$\Delta x = F \frac{Z(t_1 + \omega_3 Z - \omega_2 Y) - X(t_2 + \omega_1 Y - \omega_3 X)}{Z^2 + Z(t_2 + \omega_1 Y - \omega_3 X)}$$

or,

$$\Delta x = \frac{(F t_1 - X t_2) Z / Z + F \omega_3 - \omega_2 Y - \omega_1 \frac{X^2}{F} + \omega_1 \frac{X^2}{F}}{1 + \frac{t_2}{Z} + \omega_1 \frac{Y}{F} - \omega_3 \frac{X}{F}}$$

similarly, we obtain an expression for the other component of the retinal displacement,

$$\Delta y = \frac{(F t_2 - Y t_2) Z / Z - F \omega_1 + \omega_2 X - \omega_3 \frac{Y^2}{F} + \omega_3 \frac{Y^2}{F}}{1 + \frac{t_2}{Z} + \omega_1 \frac{Y}{F} - \omega_3 \frac{X}{F}}$$

The above equations express the the retinal displacement vector $(\Delta x, \Delta y)$ at an image point $P = (x, y)$ in terms of the parameter vector $\vec{\sigma}$ and the "depth"

coordinate Z for corresponding world point $p = (X, Y, Z)$. Another form of the above equations is,

$$\Delta x = \frac{(x_0 - x) \frac{l_z}{Z}}{1 + \frac{l_z}{Z} + \omega_x \frac{y}{F} - \omega_y \frac{x}{F}} + \frac{F \omega_y - \omega_z y - \omega_x \frac{xy}{F} + \omega_y \frac{x^2}{F}}{1 + \frac{l_z}{Z} + \omega_x \frac{y}{F} - \omega_y \frac{x}{F}} \quad (4-3.1)$$

$$\Delta y = \frac{(y_0 - y) \frac{l_z}{Z}}{1 + \frac{l_z}{Z} + \omega_x \frac{y}{F} - \omega_y \frac{x}{F}} + \frac{-F \omega_x + \omega_z x - \omega_x \frac{y^2}{F} + \omega_y \frac{xy}{F}}{1 + \frac{l_z}{Z} + \omega_x \frac{y}{F} - \omega_y \frac{x}{F}} \quad (4-3.2)$$

where,

$$(x_0, y_0) = \left(\frac{Fl_x}{Z}, \frac{fl_y}{Z} \right)$$

Note here that, when the displacement is purely translational

$$\frac{\Delta y}{\Delta x} = \frac{(y_0 - y)}{(x_0 - x)} \quad (4-4)$$

This means that when the rotational component of the displacement is zero, the image displacement vectors meet at one point (x_0, y_0) . That is to say, the retinal displacement field converges to or diverges from a single point in the image plane. This point is called the *focus of contraction* (FOC) or the *focus of expansion* (FOE), depending on whether the translational motion is directed away from or towards the image plane (figure III).

If we can measure the retinal displacement field due to a particular motion, then it is possible to estimate the parameters characterizing the motion. In addition, if the temporal sampling rate of our imaging process is high - meaning

that the components of the displacement for a single time interval is small and the following condition holds.

$$\frac{l_z}{Z} + \omega_x \frac{y}{F} - \omega_y \frac{x}{F} \ll 1 \quad (\text{A})$$

It is possible to derive the equations relating image motion to the motion parameters in the differential case. This is obtained by dividing equation (4.3) by a small time interval, Δt , and taking the limit as $\Delta t \rightarrow 0$. The image displacement then becomes image velocity, and is called *optical flow*. The optical flow is denoted by the vector (u, v) where:

$$u = \lim_{\Delta t \rightarrow 0} \frac{\Delta x}{\Delta t} = \frac{dx}{dt} \quad v = \lim_{\Delta t \rightarrow 0} \frac{\Delta y}{\Delta t} = \frac{dy}{dt}$$

Similarly the motion parameters are now the translational velocity $T = (U, V, W)$ and the rotational velocity $\Omega = (\alpha, \beta, \gamma)$ where:

$$U = \lim_{\Delta t \rightarrow 0} \frac{l_x}{\Delta t} \quad V = \lim_{\Delta t \rightarrow 0} \frac{l_y}{\Delta t} \quad W = \lim_{\Delta t \rightarrow 0} \frac{l_z}{\Delta t}$$

and

$$\alpha = \lim_{\Delta t \rightarrow 0} \frac{\omega_x}{\Delta t} \quad \beta = \lim_{\Delta t \rightarrow 0} \frac{\omega_y}{\Delta t} \quad \gamma = \lim_{\Delta t \rightarrow 0} \frac{\omega_z}{\Delta t}$$

Equation (2.6) now becomes:

$$u = (x_0 - x) \frac{W}{Z} + F\beta - \gamma y - \alpha \frac{xy}{F} + \beta \frac{x^2}{F} \quad (4-5.1)$$

$$v = (y_0 - y) \frac{W}{Z} - F\alpha + \gamma x - \alpha \frac{y^2}{F} + \beta \frac{xy}{F} \quad (4-5.2)$$

where the 3D motion is now characterized by a translational velocity (U, V, W) and a rotational velocity (α, β, γ) . Furthermore the FOE is now given by

$$(x_0, y_0) = \left(-\frac{FU}{W}, -\frac{FV}{W} \right).$$

Motion perception involves the computation of the parameters of motion from the image displacement field. The latter, becomes in the limiting case, a field of velocities, called optical flow. The relation that optical flow has with the motion parameters, is embodied in equations (4.5). These motion equations involve velocities, both in 3D as well as in the retina. However, in a practical vision system, the retinal measurements that are actually made involve displacements over a small time interval. This means the above velocity equations, are not strictly applicable. Under certain conditions, the penalty paid for doing this may not be too severe. This happens when the error introduced by the velocity approximation is within some predetermined bounds.

There are two separate approximations embodied in the usage of the equations (4.5) to express the constraints on image motion due to the 3D motion parameters:

- (1) *The three dimensional velocity approximation* - The velocity of a point $\rho = (X, Y, Z)$ on a rigid body, moving with a translational velocity $T = (U, V, W)$, and a rotational velocity $\Omega = (\alpha, \beta, \gamma)$ is given by

$$\frac{d\rho}{dt} = T + \Omega \times \rho$$

Integrating the above with respect to time we have

$$\int_0^{\Delta t} \frac{d\rho}{dt} dt = \int_0^{\Delta t} (T + \Omega \times \rho) dt$$

Here \times denotes the vector cross product. The three dimensional velocity approximation implies that, for small Δt , the image displacement can be expressed as:

$$\Delta \rho = (\Delta X, \Delta Y, \Delta Z) \equiv T \Delta t + (\Omega \Delta t) X \cdot \rho$$

- (2) *The retinal velocity approximation* - This enables us to treat retinal displacements as retinal velocities and is valid so long as $\frac{\Delta Z}{Z} \ll 1$. This can also be written as relation (A) stated previously.

When both the translational velocity T as well as the depth function Z is multiplied by the same constant, the latter cancels out leaving the equations (4.5) unchanged. The same applies to the equations (4.3). This means that scaling the translation by a constant factor, and at the same time, causing a depth dilation by the same factor leaves the image displacement field unchanged. Thus, from the information available in the image displacement field, the translation vector is obtainable, only upto a scale factor.

In equation (4.3) the depth variable Z is an unknown. An equation relating image displacement to the motion parameters is obtained by eliminating $\frac{t}{Z}$ from equations (4.3):

$$\frac{\Delta x(1 + \omega_1 \frac{1}{f} - \omega_2 \frac{1}{f}) + (f\omega_1 + \omega_2 x - \omega_1 \frac{x^2}{f} + \omega_2 \frac{x^2}{f})}{\Delta y(1 + \omega_1 \frac{1}{f} - \omega_2 \frac{1}{f}) + (f\omega_1 + \omega_2 y - \omega_1 \frac{y^2}{f} + \omega_2 \frac{y^2}{f})} = \frac{x_0 - x - \Delta x}{y_0 - y - \Delta y}$$

or,

$$\frac{\omega_1 q_1 + \omega_2 q_2 + \omega_2 y f + \Delta x}{\omega_1 q_1 + \omega_2 q_2 + \omega_2 y f + \Delta y} = \frac{x_0 - x - \Delta x}{y_0 - y - \Delta y} \quad (4-6)$$

where

$$\begin{aligned}\varphi_1 &= y \Delta x + x_1 & \varphi_2 &= F^2 + x \Delta x + x^2 \\ \varphi_3 &= F^2 + y \Delta y + y^2 & \varphi_4 &= x \Delta y + x_1\end{aligned}$$

The above equation relates the motion parameters to the image displacements, which are observables. This is a bilinear equation in the unknown motion parameters. A similar relation is obtained for the differential motion case, by eliminating $\frac{W}{Z}$ from equation (4.5):

$$\frac{u - (1 - \beta - \gamma) - \alpha \frac{x_1}{F} + \beta \frac{x^2}{F}}{v - (-1 - \alpha + \gamma)x - \alpha \frac{y^2}{F} + \beta \frac{xy}{F}} = \frac{x_0 - x}{y_0 - y} \quad (4-7)$$

In the above analysis, the relations between image motion and 3D motion has been derived by assuming general displacement of a rigid constellation of points in space. This relation is given by equation (4-5). From this, by taking the limiting case, for infinitesimal displacement, the "continuous" or differential motion case is obtained. The latter relation can also be obtained directly from the kinematic equations of rigid motion (see Appendix I or [19] for details).

4.1. The Information available in the image displacement field

The foregoing analysis illustrates the dependence of the optical flow field on the motion parameters. In other words 3D motion constrains image motion. The magnitude of the translation parameter vector cannot be computed from the optical flow field. The rigid motion parameters observable from monocular retinal optical flow measurements are given by the parameter vector \vec{d}

$$\vec{d} = (x_0, y_0, \omega_1, \omega_2, \omega_3)$$

Now, we examine the motion equations to see whether the displacement field

uniquely determines a .

The question to be answered, before attempting the design of algorithms to compute the motion parameters from optical flow is whether such computation is feasible. This means that given an optical flow field, when can we say that it could be produced by a unique set of motion parameters. The following theorem answers this question, by giving a sufficient condition for uniqueness.

Theorem I: *The optical flow field is uniquely determined by the rigid motion parameters when the moving surface cannot be expressed as a rational function of the form $\frac{P_1(x,y)}{Q_2(x,y)}$, where P_1 and Q_2 are polynomials of the first and second orders respectively, and (x,y) are image coordinates.*

Proof: Let a rigid surface Z' , moving with translational and rotational velocities (U',V',W') and (α',β',γ') respectively, generate the optical flow field (u,v) given by

$$\begin{aligned} u &= \frac{U' - xW'}{Z'} + F\beta' - \gamma'y - \alpha'\frac{xy}{F} + \beta'\frac{x^2}{F} \\ v &= \frac{V' - yW'}{Z'} - F\alpha' + \gamma'x - \alpha'\frac{y^2}{F} + \beta'\frac{xy}{F} \end{aligned} \quad (4-8)$$

where the translation parameter vector is (U',V',W') and the rotational velocity is (α',β',γ') .

Assume that there is another surface $Z(x,y)$ moving with a different set of motion parameters but giving rise to the same optical flow field (u,v) , or

$$\begin{aligned} u' &= \frac{U - xW}{Z} + F\beta - \gamma'y - \alpha'\frac{xy}{F} + \beta'\frac{x^2}{F} \\ v' &= \frac{V - yW}{Z} - F\alpha + \gamma'x - \alpha'\frac{y^2}{F} + \beta'\frac{xy}{F} \end{aligned} \quad (4-9)$$

where the 3D motion is now due to a translational velocity (U,V,W) and a

rotational velocity (α, β, γ) .

Since $u - u' = 0$ and $v - v' = 0$ everywhere in the image, we have from equations (4.8) and (4.9):

$$\frac{U - xW}{Z} - \frac{U' - xW'}{Z'} + F\Delta\beta - y\Delta\gamma - \frac{xy}{F}\Delta\alpha + \frac{x^2}{F}\Delta\beta = 0 \quad (4-10.1)$$

$$\frac{V - yW}{Z} - \frac{V' - yW'}{Z'} - F\Delta\alpha + x\Delta\gamma - \frac{y^2}{F}\Delta\alpha + \frac{xy}{F}\Delta\beta = 0 \quad (4-10.2)$$

where, $\Delta\alpha = \alpha - \alpha'$, $\Delta\beta = \beta - \beta'$, and $\Delta\gamma = \gamma - \gamma'$.

Considering the above set of equations and solving for the variable Z' we have:
(assuming the focal length F to be unity)

$$Z' = \frac{P_1(x, y)}{Q_1(x, y)}$$

where

$$P_1(x, y) = (U'V - U'V') + x(V'W'' - V''W') + y(U''W' - U'W'') \quad (4.11)$$

and,

$$Q_1 = (\Delta\beta V + \Delta\alpha U) - x(\Delta\alpha W + \Delta\gamma U) - y(\Delta\beta W + \Delta\gamma U) - xy(\Delta\alpha V + \Delta\beta U) + x^2(\Delta\beta V + \Delta\gamma W) + y^2(\Delta\alpha U + \Delta\gamma W) \quad (4.12)$$

The above implies that the surface Z' that originally generated the optical flow must be a rational function of the form $\frac{P_1}{Q_1}$, to permit ambiguous interpretation of its rigid motion. This is contrary to the the statement of the theorem. This proves the theorem.

Corollary I: *When the motion of a surface is purely rotational, the optical flow field is uniquely determined by the motion.*

Proof: In equation (4.10) making the substitutions $U'' = V'' = W'' = 0$ we get:

$$- \frac{U - xW}{Z} + F\Delta\beta - y\Delta\gamma - \frac{xy}{F}\Delta\alpha + \frac{x^2}{F}\Delta\beta = 0$$

$$- \frac{V - yW}{Z} - F\Delta\alpha + x\Delta\gamma - \frac{y^2}{F}\Delta\alpha + \frac{xy}{F}\Delta\beta = 0$$

Now, eliminating Z from the above equations and setting focal length 'F' to unity, we obtain:

$$\begin{aligned} &(\Delta\beta V + \Delta\alpha U) - x(\Delta\alpha W + \Delta\gamma U) - y(\Delta\beta W + \Delta\gamma V) \\ &- xy(\Delta\alpha U + \Delta\beta U) + x^2(\Delta\beta V + \Delta\gamma W) + y^2(\Delta\alpha U + \Delta\gamma W) = 0 \end{aligned}$$

From the above equation we have a set of six equations:

$$\Delta\alpha U + \Delta\beta V = 0$$

$$\Delta\alpha V + \Delta\beta U = 0$$

$$\Delta\beta V + \Delta\gamma W = 0$$

$$\Delta\alpha U + \Delta\gamma W = 0$$

$$\Delta\alpha W + \Delta\gamma U = 0$$

$$\Delta\beta W + \Delta\gamma V = 0$$

The above equations imply either $U = V = W = 0$ or $\Delta\alpha = \Delta\beta = \Delta\gamma = 0$.

Both these conditions mean that the optical flow field due to a pure rotational motion has a unique interpretation. This proves the corollary.

Corollary II: *It is possible for a flow field generated by pure translatory motion to be identical to one generated by another flow field due to both translation and rotation. In other words convergence of the flow vectors directly onto a point on the image plane does not imply purely translatory motion.*

The truth of the above corollary will be demonstrated by a numerical example. Consider two flow fields generated by different surfaces undergoing different motions:

In the first case the motion is due to a planar surface given by the equation :

$$Z = \frac{1}{2}X - \frac{5}{6}Y + 1$$

The motion is rigid and is specified by

$$(x_0 = -\frac{7}{2}, y_0 = \frac{35}{6}, \alpha = 5, \beta = 3, \gamma = 0)$$

Assume the translation in depth to be unity. Then, from equation (4.8) we have,

$$u = (x - \frac{7}{2})(1 - \frac{1}{2}x + \frac{5}{6}y) - 3 + 5xy - 3x^2$$

$$u = -\frac{3}{4}x + \frac{35}{12}y + \frac{35}{6}xy - \frac{7}{2}x^2 + \frac{1}{2}$$

$$v = \frac{35}{12}x - \frac{139}{36}y + \frac{35}{6}y^2 - \frac{7}{2}xy - \frac{5}{6}$$

In the second case the motion is due to the planar surface given by the equation :

$$Z = \frac{7}{2}X - \frac{35}{6}Y + 1$$

and the motion is specified by the parameter vector

$$(x_0 = -\frac{1}{2}, y_0 = \frac{5}{6}, \alpha = 0, \beta = 0, \gamma = 0)$$

The optical flow field in both the examples are identical.

The question of multiple interpretations of the same flow field, has received some attention in the literature. The foregoing example illustrates the fact that motion of planes can be potentially open to more than one interpretation. It is known (see [27-29,34]) that the motion of planes have dual interpretations. Uniqueness of interpretation for planes requires three views of four points, or two views of seven points which uniquely define two planes neither of which pass through the origin. In another study Fang and Huang [9] showed that nine points not lying on a second order surface passing through the origin can be used to

determine the motion parameters uniquely. Another significant theoretical result is due to Longuet-Higgins [20], and Tsai and Huang [30], where eight points are used to solve for the motion parameters from a set of linear equations. The important question as yet unanswered are, under what conditions the optical flow field is *inherently* ambiguous and, what is the degree of the ambiguity possible in optical flow fields. The following analysis answers these questions.

Theorem II. *Under the assumption of rigidity, an optical flow field is amenable to at most three interpretations.*

Proof: Theorem I shows that the optical flow field is enough to determine the rigid motion parameters uniquely for most surfaces. It was seen however that in case of certain rational functions there is potential ambiguity in the interpretation of motion. These are the rational functions belonging to the class R_3^2 , and written as

$$Z = \frac{ax + by + c}{t + \dots + fx + gxy + hx^2 + y^2} \quad (4.13)$$

Planar surfaces belong to the above class of surfaces. It has been mentioned previously that planar surfaces can have at most two interpretations. When a surface is non planar, to have multiple interpretations of its motion, it must be of the type given by equation (4.13) with the added property that there is no common factor between the numerator and the denominator.

Let such a surface be undergoing rigid motion $(u, v, w, \alpha, \beta, \gamma)$. Let there be another motion $(U, V, W, \alpha + \Delta\alpha, \beta + \Delta\beta, \gamma + \Delta\gamma)$ that produces an identical flow field. Then from equation (4.11) we have

$$\begin{aligned}
vW - Uv &= ak \\
Uw - uW &= bk \\
uU - Uv &= ck
\end{aligned}
\tag{4.14}$$

where k is some constant factor. Since by definition of the class R_2^1 at least one of a , b and c must be non zero, therefore $k \neq 0$. This is because if k is zero then from the above set of three equations we get the result that the translations (u, v, w) and (U, V, W) are identical upto a scale factor. Hence by Lemma I of Appendix I, the motion is not ambiguous.

Multiplying the first equation by u , the second by v , and the third by w and adding the three equations we have

$$(ua + bv + cw)k = 0$$

This means that the motion can only be ambiguous when

$$ua + bv + cw = 0 \tag{4.15}$$

Similarly it can be shown that

$$uU + bV + cW = 0 \tag{4.16}$$

Again comparing the denominator of the rational function with equation (4.12),

and combining the constant $-$ with the translation parameter (U, V, W) :

$$\Delta\beta U + \Delta\alpha U = d \tag{4.17}$$

$$\Delta\alpha W + \Delta\gamma U = -e \tag{4.18}$$

$$\Delta\beta W + \Delta\gamma V = -f \tag{4.19}$$

$$\Delta\alpha U + \Delta\beta U = -g \tag{4.20}$$

$$\Delta\beta U + \Delta\gamma W = h \tag{4.21}$$

$$\Delta\alpha U + \Delta\gamma W = i \tag{4.22}$$

From equations (4.17), (4.21) and (4.24) we get:

$$2\Delta\alpha U = q \quad (4.23)$$

$$2\Delta\beta U = r \quad (4.24)$$

$$2\Delta\gamma W = s \quad (4.25)$$

where $q = d + i - h$, $r = d - i + h$, $s = -d + i + h$. Substituting from the above equations into equations (4.18), (4.19) and (4.20):

$$qU^2 + rU^2 + 2gUW = 0 \quad (4.26)$$

$$rW^2 + sU^2 + 2fUW = 0 \quad (4.27)$$

$$qU^2 + sW^2 + 2eUW = 0 \quad (4.28)$$

Equations (4.26), (4.27) and (4.28), together with equation (4.16) can admit no more than two solutions. This is because at least one of (q, r, s, e, f, g) must be nonzero. Therefore, since there can be at the most two spurious solutions (recall that the veridical solution corresponds to $k = 0$), the implication is that:

When the optical flow field has more than one interpretation, the number of globally consistent solutions for the motion parameters can be at most three.

This completes the proof of the theorem.

It will be shown that there exist surfaces whose rigid motion induces optical flow that is compatible with *three* distinct interpretations. This fact explains why Longuet-Higgins and Prazdny [19] noted, that from *local* optical flow constraints and their derivatives three interpretations of the motion are possible since the constraint equations were cubic.

An example of 2D motion field with three distinct rigid motion interpretations:

The equation of the moving surface is

$$Z = \frac{1}{gxy}$$

the motion parameters are $(U^*, V^*, 0, \alpha, \beta, \gamma)$ the expression for optical flow is therefore

$$\begin{aligned} u &= U^* gxy - \alpha xy + \beta(x^2 + 1) - \gamma y \\ v &= V^* gxy - \alpha(y^2 + 1) + \beta xy + \gamma x \end{aligned}$$

Alternative interpretation I:

$$\frac{1}{Z} = \frac{g}{U^*} [U^* xy - V^*(x^2 + 1)]$$

where the motion parameters are $(U, 0, 0, \alpha, \beta + gU^*, \gamma)$. The optical flow field is given by

$$\begin{aligned} u_1 &= U \frac{g}{U^*} [U^* xy - V^*(x^2 + 1)] - \alpha xy + (\beta + gU^*)(x^2 + 1) - \gamma y \\ v_1 &= -\alpha(y^2 + 1) + (\beta + gU^*)xy + \gamma x \end{aligned}$$

Alternative interpretation II:

$$\frac{1}{Z} = \frac{g}{V^*} [V^* xy - U^*(y^2 + 1)]$$

The motion parameters are $(0, V, 0, \alpha - gU^*, \beta, \gamma)$. The optical flow field is

$$\begin{aligned} u_2 &= -(\alpha - gU^*)xy + \beta(x^2 + 1) - \gamma y \\ v_2 &= V \frac{g}{V^*} [V^* xy - U^*(y^2 + 1)] - (\alpha - gU^*)(y^2 + 1) + \beta xy + \gamma x \end{aligned}$$

It is easily verified that $u = u_1 = u_2$ and $v = v_1 = v_2$.

Theorem I states that under certain cases the optical flow field may not indicate the motion parameters uniquely. The next theorem shows how unambiguous determination of the motion parameters can be achieved from optical flow data.

Theorem III: *Given the optical flow values at three non collinear retinal locations, where the temporal derivative (or time difference) of the flow is nonzero, the motion parameters are uniquely determined.*

Proof: The essential fact on which the proof is based is that the rotational component of optical flow is not dependent on time. Thus if during a short observation period the parameters of motion remain fixed then the temporal derivative of the flow is only dependent upon the change in the translational segment of the flow. Although the following proof uses temporal derivatives, differences also lead to the same result.

Differentiating equation (4.8) with respect to time we have,

$$\frac{\partial u}{\partial t} = - (x_0 - x) \frac{W}{Z^2} \frac{\partial Z}{\partial t} \quad (4.29a)$$

$$\frac{\partial v}{\partial t} = - (y_0 - y) \frac{W}{Z^2} \frac{\partial Z}{\partial t} \quad (4.29b)$$

Now we assume that $\frac{\partial Z}{\partial t} \neq 0$. Furthermore, the case where the translational motion is in the frontal plane is easily dealt with and will not be considered any further. From the above equations we have,

$$\frac{\frac{\partial v}{\partial t}}{\frac{\partial u}{\partial t}} = \frac{(y_0 - y)}{(x_0 - x)} \quad (4.30)$$

The above equation provides a linear constraint for locating the FOE. It is therefore evident that we can determine the FOE uniquely if we have two such constraint lines that are independent. This is guaranteed when we have three

retinal locations where the temporal derivatives of the flow are non zero. Once the FOE is determined, the rotational velocity can also be uniquely computed. (Note that, instead of temporal derivatives, differences can also lead to the same result.)

Another way of resolving the ambiguity in the optical flow is by using shape information. There is a strong relationship between the parameters of motion, the optical flow field and the structure of a moving surface. The following propositions makes this concept clear.

Proposition I. *When the parameters (i.e. $x_0, y_0, \alpha, \beta, \gamma$) describing the motion of a rigid surface are known then the structure of the surface is uniquely determined from the optical flow field.*

Proof: The proof is evident from equation (4.5). Note that we can obtain the depth function upto a constant dilation factor W . In other words the ratio of depths at any two image points can be computed.

Proposition II. *When the structure of a surface is known then the parameters describing its rigid motion are uniquely obtained from the optical flow generated by the motion.*

Proof: See Appendix II.

Even the partial specification of shape can lead to a correct perception of rigid motion. A illustration of the fact that shape information can disambiguate between alternative motion interpretations comes from the next theorem.

Theorem IV: *The motion of a planar surface whose direction of translation does not lie in the plane of its surface normal and the line of sight, can be interpreted correctly from the optical flow generated, when the tilt of the plane is known.*

Proof: Let the equation of the planar surface be

$$z = \frac{d}{1 - px - qy}$$

where (p,q) is the orientation of the depth plane and 'd' is the distance from the origin along the z axis (e.g. line of sight). Substituting the above into equation (4.5) and observing that we can ignore multiplication of the translational parameters by a constant (such as d) since we can compute the former upto a scale factor anyway, we have

$$\begin{aligned} u^2 &= l_2 x^2 - l_3 x + l_4 xy + l_5 y^2 \\ v^2 &= l_6 x^2 - l_7 x - l_8 xy + l_4 y^2 \end{aligned} \quad (4.31)$$

where the unknowns $\{ a_i \}$ are given by

$$U + \beta = l_1 \quad (4.32.1)$$

$$Up + W = l_2 \quad (4.32.2)$$

$$Uq + \gamma = l_3 \quad (4.32.3)$$

$$Wq - \alpha = l_4 \quad (4.32.4)$$

$$Wp + \beta = l_5 \quad (4.32.5)$$

$$V - \alpha = l_6 \quad (4.32.6)$$

$$\gamma - Up = l_7 \quad (4.32.7)$$

$$Vq + W = l_8 \quad (4.32.8)$$

Note that (4.32) are linear homogeneous equations in eight unknowns. Thus if we can solve for the *synthetic parameters* $\{ l_i \}$ by making measurements at four

suitable points, and *in addition* can measure the *tilt* of the depth plane, i.e.

$$\frac{E}{q} = \tau \quad (4.33)$$

Then from (4-32.7) and (4-32.8) and (4-33) we have:

$$\gamma + \tau W = l_1 + \tau l_3 \quad (4.34.1)$$

From (4-32.2), (4-32.3) and (4-33) we have :

$$\tau \gamma - W = \tau l_3 - l_2 \quad (4.34.2)$$

Therefore, since $\tau^2 + 1 \neq 0$ we have:

$$\gamma = \frac{l_1 + \tau(l_3 - l_2) + \tau^2 l_3}{\tau^2 + 1} \quad (4.34.3.1)$$

$$W = \frac{l_2 + \tau(l_1 - l_3) + \tau^2 l_3}{\tau^2 + 1} \quad (4.34.3.2)$$

Now if $W \neq l_3$ (i.e. $q \neq 0$) we have from (4-32.8) and (4-32.3):

$$\frac{l_1}{l_3} = \frac{U}{V} = \frac{l_3 - \gamma}{l_3 - W} = k \quad (4.34.4)$$

otherwise if $l_1 \neq 0$ (i.e. $p \neq 0$) we have from equations (4-32.7) and (4-32.2):

$$\frac{l_1}{l_2} = \frac{U}{V} = \frac{l_2 - W}{\gamma - l_1} = k \quad (4-34.4)$$

(if both p and q are zero then the parameters are easily solved for)

Now from (4-34.4), (4-32.6) and (4-32.1) we have:

$$k\alpha + \beta = 1 - kl_6 \quad (4.34.5)$$

Also from (4-32.5) and (4-32.4) we have:

$$\tau\alpha + \beta = l_5 - \tau l_4 \quad (4.34.6)$$

Therefore, since $\tau \neq k$, from the assumption made in the statement of the theorem, then equations (4-34.5) and (4-34.6) are independent, and we have:

$$\alpha = \frac{(1 - kl_6) - (l_5 - \tau l_4)}{k - \tau} \quad (4.34.7.1)$$

$$\beta = \frac{k(l_5 - \tau l_4) - \tau(1 - kl_6)}{k - \tau} \quad (4.34.7.2)$$

Now U and V can be determined from equations (4-32.6) and (4-32.1). Thus we have determined the motion parameters uniquely from the optical flow and tilt information.

At this point it may be mentioned in passing that it is possible to obtain the motion parameters uniquely from the optical flow generated by two planes moving together rigidly. In this case the optical flow is locally second order. If the eight synthetic parameters are now measured at two different regions of the flow field then

$$\begin{aligned} U \Delta \frac{1}{d} &= \Delta l_1 \\ U \Delta \frac{p}{d} + W \Delta \frac{1}{d} &= \Delta l_2 \\ U \Delta \frac{q}{d} &= \Delta l_3 \\ W \Delta \frac{q}{d} &= \Delta l_4 \\ W \Delta \frac{p}{d} &= \Delta l_5 \\ V \Delta \frac{1}{d} &= \Delta l_6 \\ -V \Delta \frac{p}{d} &= \Delta l_7 \\ V \Delta \frac{q}{d} + W \Delta \frac{1}{d} &= \Delta l_8 \end{aligned} \quad (4.35)$$

where the two planes involved in the motion are given by $z = \frac{d}{px + qy + 1}$ and

$z = \frac{d'}{p'x + q'y + 1}$. The Δ operator in front of any quantity denotes the difference

of the corresponding parameters for the two planes, e.g. $\Delta p = \frac{p}{d} - \frac{p'}{d'}$.

The above equations imply that when at least one of, Δp or Δq or $\Delta \frac{1}{d}$ is non zero the translational parameters are uniquely determined. Hence in such a case the rigid motion parameters are determined uniquely from the optical flow field (see Appendix I). Therefore

When two planes, neither of which pass through the origin, move rigidly together, their motion is uniquely determinable from the optical flow field generated.

4.2. Summary and Discussions

The analysis presented here leads to considerable insight into the 3D motion interpretation problem. Previous results (e.g. [9, 30]) by Huang and his colleagues presented *sufficient* conditions for uniqueness of three dimensional motion interpretation, since, they were concerned with specific algorithms. The work, reported here, deals with necessary conditions for unique interpretation of 3D motion from the optical flow field.

While the surface denoted by equation (4.13) does mean second order surfaces containing the nodal point of the camera, it is certainly true that all such surfaces do not admit ambiguous interpretations of their 3D motions. Multiple interpretations require, in addition, that the constraints given by (4.16), (4.26), (4.27) and (4.28) all be satisfied.

Thus consider, an algorithm, such as Prazdny's [22], where nonlinear (and independent) flow constraints at five retinal locations are used to obtain a 3D

motion interpretation. It is now possible to answer the question as to whether the solution obtained is the only one possible. Since now a set of motion parameters is known, from equation (4.5) the relative depth $\frac{Z}{W}$ can be obtained at the five retinal locations. The latter, when substituted into equation (4.13), generates five linear equations in the surface parameters $a, b, c, d, e, f, g, h, i$. These together with the four constraints (4.16), (4.26), (4.27) and (4.28) constitute nine linear homogeneous equations in the nine surface parameters. Therefore uniqueness of interpretation is possible if the determinant of the above system is zero. Which in turn implies, that all the surface parameters must be zero. This makes it impossible to construct any other interpretation from measurements at the five retinal locations, guaranteeing that the solution obtained is the only one possible.

5. Computational Techniques for obtaining the Rigid Motion Parameters

The main difficulty in computing the 3D Rigid Motion parameters is that the equation constraining the image motion to the 3D motion is nonlinear. Another complication arises from the high dimensionality of the parameter space. If it were possible to separate the component of the image displacement due to translation from that due to rotation we could have efficient algorithms for the computation of the 3D motion.

The constraint equations developed by Longuet-Higgins and Prazdny [19] are used by Bruss and Horn [6] to arrive at the parameter set that minimizes the square of the error between the measured optical flow and the flow computed

from the parameter constraint. In general such a technique will give rise to a system of non-linear equations from which the parameters must be computed using some suitable iteration scheme. Longuet-Higgins and Prazdny mention the possibility of using motion parallax to simplify the computation of the global motion parameters. Lawton and Rieger [24] uses a similar idea to factor out the rotational component of the optical flow at depth discontinuities or regions where the depth gradient is large. This method is not reliable since it hinges upon the ability to compute flow vectors reasonably accurately at discontinuities. Since almost all algorithms, to date, for computing optical flow face problems at regions where the field is sharply discontinuous.

5.1. Computing Rigid Motion Parameters From Optical Flow

Attempts at segmenting the parameter space of rigid motion into translational and rotational components can be termed marginally successful, at best. A simple way to estimate the motion parameters from the bilinear flow constraint equation (2.10) is by means of the hough transform technique [2.5]. There are two problems that are immediately apparent, namely, the nonlinearity of the constraint, and the large dimension (e.g. five) of the parameter space. Another method is to linearize the constraint equation by writing (2.10) as a linear equation in *eight* parameters. Obviously these eight parameters are each functions of the values of the five actual parameters. This implies that linear least square methods are not applicable here, since the eight *synthetic* parameters are not independent of one another. Finally it is shown that the information in the variation in the

optical flow field, i.e. the spatio temporal derivatives of the flow field facilitate the computation of the motion parameters.

5.2. The Analysis of General Motion

Here the situation is complicated by the fact that we have to determine several sets of parameters, corresponding to the several bodies in motion. This problem, is obviously, quite hard and is still open. It has been studied in restricted domains by Fennema & Thompson [11] The Hough transform technique proposed earlier in this paper still works. The only difference is that we have to look for multiple peaks in the parameter space after houghing. These then would, of course, correspond to the parameter set of the various bodies in motion with respect to the sensor.

Methods involving the spatial derivatives of the optical flow can again be applied. There is no known technique for obtaining the optical flow in all types of imaging situations. Also the computed flow field is noisy, to say the least. This difficulty is compounded when we consider the case where several bodies are in motion with respect to the sensor. Thus obtaining spatial derivatives of the flow may not be practically possible over large portions of the image frame.

Recently a way of determining motion parameters from 3D flow has been suggested [3]. This method is amenable to adaptation to the general motion case. It is not clear as to how difficult it is to compute the 3D flow in this case. However, it can be shown that in case a depth map can be obtained (by some stereo matching technique), the 3D map can be calculated.

5.3. Algorithms for motion perception

Computer algorithms for determining the parameters of rigid motion will now be discussed in the light of the various ideas put forth in earlier sections. The treatment will consider both *orthographic* and *perspective* projections, as well as *differential* and *discrete* motions. In some of the cases the steps of the algorithms will be described with a fair amount of detail. In others details will be omitted, particularly when the algorithm in question has a structure which is similar to one already described. In all of the algorithms the *Hough Transform* technique (see [2] for details) is used to compute the desired global parameters from sets of constraint equations obtained at different image (or retinal) locations. It should be noted that least square error minimization techniques are also applicable in most cases.

For the sake of simplicity the motion of a single rigid body is considered. To extend the following methods to the motion of several moving bodies, either the image motion field has to be segmented, or, when hough transform is used, multiple modes have to be detected in the parameter "voting" distribution.

Recall that for the case of differential motion, optical flow is denoted by (u, v) , the translation parameters (velocity) by (U, V, W) or $(x_0 = \frac{U}{W}, y_0 = \frac{V}{W})$ and the rotational parameters by (α, β, γ) .

5.4. Differential motion under Orthography

This case has been analyzed by Hoffman, Sugihara previously [13,26]. Hoffman's shows that motion parameters are not uniquely determinable from local analysis of optical flow. However, this is not the case for global analysis techniques. It has been previously shown that, for non planar surfaces, global analysis will give rise to unambiguous results. Sugihara computed structure from two optical flow frames. Another interesting result was obtained by Aloimonos [1] where it is shown when absolute depth can be recovered under pure rotation under orthography when shape is known. Under orthography the translational part of the optical flow field is constant and hence the translational parameters are not computable. Hence motion parameters here, always refer to the rotational velocity parameters (α, β, γ) .

The relevant equations are

$$\begin{aligned}\Delta u &= \beta \Delta z - \gamma \Delta y \\ \Delta v &= -\alpha \Delta z + \gamma \Delta x\end{aligned}\tag{5.1}$$

where the Δ symbol denotes that the following quantity is a difference obtained from measurements made at two different retinal locations. The relation between the surface gradients and the optical flow derivatives are:

$$\frac{\partial u}{\partial x} = \beta \frac{\partial Z}{\partial x}\tag{5.2.1}$$

$$\frac{\partial v}{\partial x} = -\alpha \frac{\partial Z}{\partial x} + \gamma\tag{5.2.2}$$

$$\frac{\partial u}{\partial y} = \beta \frac{\partial Z}{\partial y} - \gamma\tag{5.2.3}$$

$$\frac{\partial v}{\partial y} = -\alpha \frac{\partial Z}{\partial y} \quad (5.2.4)$$

Algorithm 1: *Motion parameters from image motion and structure information.*

The simplest instance is when the structure of the moving object is known. In the discrete case the relative depth function, $\Delta Z(x,y)$, values are enough to compute the parameters (α, β, γ) uniquely from the linear equation (5.1). For the differential case structure or shape can be represented by the surface normals $(\frac{\partial Z}{\partial x}, \frac{\partial Z}{\partial y})$. If the surface normals are known everywhere, then we can integrate the surface normals to obtain the depth upto a constant additive term. In other words $\Delta Z(x,y)$ is computable. In this case measurement of optical at three non collinear points is enough to compute the rotational parameters. However, if the surface normals are only known at sparse locations, but the optical flow field is locally known at these locations then we can use equation (5.2) for computing the rotation parameters. In this case we are relying on the fact that the first derivatives of the flow can be reliably computed. This is possible when, in the neighbourhood of the points of interest, the optical flow values have been measured at enough locations so as to allow analytic reconstruction of the optical flow function. Finally note that, if the motion parameters are known then the structure can be obtained from the image motion for both the discrete and the differential cases. The steps in the algorithm are:

1. Set up a three dimensional accumulator array for the rotation parameters:

$$h[\alpha, \beta, \gamma] := 0.$$

2. For every point in the image where optical flow *and* surface normals are known, select the constraint equation (5.1) if the estimated measurement error in the surface normal function is less than that estimated for the optical flow function; otherwise select equation (5.2).

For all values of (α, β, γ) :

If (α, β, γ) satisfies the constraint equation selected

$$h[\alpha, \beta, \gamma] := h[\alpha, \beta, \gamma] + 1$$

3. Obtain the maximum value in the accumulator array. The corresponding indices are the desired values for the rotation parameters.

Algorithm II: *Motion parameters and structure from image motion.* When the structure is not known then, considering the differential case and eliminating $(\frac{\partial Z}{\partial x}, \frac{\partial Z}{\partial y})$ from equations (5.2) :

$$\frac{\partial v}{\partial x} = -\frac{\alpha}{\beta} \frac{\partial u}{\partial x} + \gamma \quad (5.3.1)$$

$$\frac{\partial u}{\partial y} = -\frac{\beta}{\alpha} \frac{\partial v}{\partial y} - \gamma \quad (5.3.2)$$

Similarly, eliminating ΔZ from equation (5.1):

$$\mu u - \gamma x + \mu \gamma y + v = 0 \quad (5.4)$$

where $\mu = \frac{\alpha}{\beta}$.

It is easy to obtain quadratic equations in either γ or $\frac{\alpha}{\beta}$ from the equations (5.3).

This means that in general, at every image location, from the measurement of the spatial derivatives of the optical flow at most two sets of values of the parameters

$(\frac{\alpha}{\beta}, \gamma, \frac{U}{q})$ may be obtained. However, if some global assimilation technique, like the hough transform (see [2]) is used, then, as shown previously, if the moving surface is non planar, only one set of parameters will be globally consistent. An exactly similar method, but using differences of image displacements, can be devised for the discrete case starting from equation (5.4).

5.5. Differential motion under Perspective

The relation between the optical flow and the motion parameters is given by the equation:

$$\begin{aligned} u &= \frac{U - xW}{Z} - \alpha xy + \beta(x^2 + 1) - \gamma y \\ v &= \frac{V - yW}{Z} - \alpha(y^2 + 1) + \beta xy + \gamma x \end{aligned} \quad (5.5)$$

From the above we obtain, by eliminating Z :

$$\frac{u + \alpha xy - \beta(x^2 + 1) + \gamma y}{v + \alpha(y^2 + 1) - \beta xy + \gamma x} = \frac{U - xW}{V - yW} \quad (5.6)$$

Observe from the right hand side of the above equation, that its value is unchanged when the translational parameters are multiplied by some constant. Hence we can determine the translational parameters only upto a scale factor. If we assume that $W \neq 0$ then the previous equation can be written as:

$$\frac{u + \alpha xy - \beta(x^2 + 1) + \gamma y}{v + \alpha(y^2 + 1) - \beta xy + \gamma x} = \frac{x_0 - x}{y_0 - y} \quad (5.7)$$

If $W = 0$ then (5.6) reduces to:

$$\frac{u + \alpha xy - \beta(x^2 + 1) + \gamma y}{v + \alpha(y^2 + 1) - \beta xy + \gamma x} = \frac{U}{V} \quad (5.8)$$

Equations (5.6), (5.7) and (5.8) are bilinear in the translation and the rotation

parameters. This nonlinearity makes it difficult to combine constraints from different image locations to compute the motion parameters. To summarize, the problems with computation of motion parameters are:

1. The constraint equations are nonlinear.
2. The parameter space is of high (e.g. five) dimensionality.

Algorithm III: *Hough transform in 5D parameter space.* This type of algorithm can be easily realised by simple parallel neuronal hardware (see [10]). The parameters that are to be determined are the polar angles (or direction cosines) representing the directions of translation and rotation, and the magnitude of the rotation vector. This representation for the rigid motion parameters is convenient since the parameter subspaces representing directions in space become easy to quantize by means such as *geodesic tessellation* of the gaussian sphere. The steps in the algorithm are:

1. Select a coarseness scale for the parameter subspaces. For instance, how many distinct directions in space, the range of values estimated for the rotation magnitude and the sampling interval in this range. Initialize the parameter units belonging to the *hough transform space* (this is the five dimensional accumulator array where the "votes" for every parameter vector is tallied).
2. For all retinal locations where optical flow has been measured do step 3:
3. For all possible parameter values (i.e. values of the parameter quintuple) admitted in step 1, do:

(i) If the direction of the translational velocity is not parallel to the image plane select equation (5.7) else select equation (5.8).

(ii) If the parameter values satisfy the chosen constraint equation vote for the corresponding parameter vector.

4. Find the parameter quintuple that has received the maximum number of votes.
5. Restrict the parameter space to a neighbourhood of the selected parameter quintuple. Repeat the steps from 2 to 4 after choosing a finer parameter space quantization.
6. If the error due to the parameter quantization is acceptable then stop and return the parameter values computed. Otherwise repeat step 5.

Some Remarks:

- (i) The space and time required by the algorithm is reduced by periodically examining the parameter accumulator units and purging those that have collected only a few "votes" compared to the top contenders. This is possible, since it is assumed that the noise in the optical flow data is uniformly distributed in retinal space.
- (ii) The confidence of the computed parameter quintuple is the ratio of the votes it received to the maximum votes possible.
- (iii) If in step 4 instead of a clear winner, a number of contenders are found then step 5 might have to be repeated for each of these for finer resolutions. Then

the winner is the parameter quintuple that comes thru with the highest confidence.

- (iv) If it is estimated that $\rho\%$ of the optical flow values is corrupted by noise, then the acceptable confidence of the result is $(100 - \rho)\%$ with a tolerance of, say 10%.

Algorithm 1, performs well when the quantization of the parameter space is not "too coarse". This is mainly due to the nonlinearity of the constraint equation used. This problem can be alleviated by linearizing the constraint equation. Although in this case the price we pay is that the dimensionality of the parameter space increases. In the following discussion it is assumed that the not all the translational velocity components are zero. This is a valid assumption since it has been shown in a previous section that the motion parameters for pure rotational motion are uniquely detectable.

From equation (5.6) we have:

$$(yu - xv)W' + vU' - uV' - x(\alpha W' + \gamma U') - y(\beta W' + \gamma V') - z(\alpha U' + \beta V') + x^2(\beta V' + \gamma W') + y^2(\alpha U' + \gamma W') \quad (5.9)$$

Now we state and prove a lemma regarding the feasibility of computing the motion parameters using the constraint given above.

Lemma 1: *The optical flow components can be expressed as an implicit polynomial equation $F(u, v, x, y; p_i, i = 1, \dots, 8) = 0$ involving the image coordinates (x, y) and eight linearly independent parameters p_i unless the depth function is a rational function $\frac{P_1(x, y)}{Q_2(x, y)}$, where P_1 and Q_2 are polynomials of first and second orders*

respectively.

Proof: Equation (5.9) is homogeneous in the motion parameters. Assume that the parameter $W \neq 0$ (The case where $W = 0$ but either U or $V \neq 0$ can be worked out in an analogous manner). Dividing the above equation by W yields:

$$(yu - xv) + p_1v - p_2u + p_3 - p_4x - p_5y + p_6x^2 + p_7y^2 + p_8xy = 0 \quad (5.10)$$

where

$$p_1 = x_0 \quad (5.11a)$$

$$p_2 = y_0 \quad (5.11b)$$

$$p_3 = \alpha x_0 + \beta y_0 \quad (5.11c)$$

$$p_4 = \alpha + \gamma x_0 \quad (5.11d)$$

$$p_5 = \beta + \gamma y_0 \quad (5.11e)$$

$$p_6 = \gamma + \beta x_0 \quad (5.11f)$$

$$p_7 = \gamma + \alpha x_0 \quad (5.11g)$$

$$p_8 = \beta x_0 + \alpha y_0 \quad (5.11h)$$

The parameters p_i 's are linearly dependent iff

$$k_1v - k_2u + k_3 - k_4x - k_5y + k_6x^2 + k_7y^2 + k_8xy = 0 \quad (5.12)$$

where the k_i 's are constants not all of which are zero. Let the optical flow be due to a rigid surface Z moving with velocity $(\bar{U}, \bar{V}, \bar{W}, \bar{\alpha}, \bar{\beta}, \bar{\gamma})$. In this case:

$$\begin{aligned} u &= \frac{\bar{U} - x\bar{W}}{\gamma} - \bar{\alpha}xy + \bar{\beta}(x^2 + 1) - \bar{\gamma}y \\ v &= \frac{\bar{V} - y\bar{W}}{\gamma} - \bar{\alpha}(y^2 + 1) + \bar{\beta}xy + \bar{\gamma}x \end{aligned} \quad (5.13)$$

Assume that the parameters p_i are linearly dependent. This implies that in equation (5.12) there must be at least one k_i that is not equal to zero. However, if both k_1 and k_2 are zero, then, all the k_i 's must be zero. Hence, if the parameters

p_i are linearly dependent, then at least one of k_1 and k_2 must be nonzero.

Substituting for \bar{u} and \bar{v} in equation (5.12) from equation (5.13) we obtain:

$$k_1 \left(\frac{\bar{U} - x\bar{W}}{Z} - \bar{\alpha}(x) + \bar{\beta}(x^2 + 1) - \bar{\gamma}(x) \right) - k_2 \left(\frac{\bar{V} - y\bar{W}}{Z} - \bar{\alpha}(y^2 + 1) + \bar{\beta}(xy) + \bar{\gamma}(x) \right) + k_3 - k_4x - k_5y + k_6x^2 + k_7y^2 - k_8xy = 0$$

Since both k_1 and k_2 are not zero, we obtain Z as a rational function of the form

$$\frac{P_1(x,y)}{Q_2(x,y)}. \text{ This proves the lemma.}$$

Lemma II: *The five parameters of rigid motion are uniquely determined by the parameters p_i .*

Algorithm IV: Equation (5.10) is the basis of a hough transform scheme to recover the motion parameters. The advantage of this scheme is that the constraint equation is linear in the "synthetic" parameters p_i . Once these parameters are computed the five rigid motion parameters are uniquely determined.

Algorithm V: Differentiating equation (5.10) with respect to the retinal space coordinates we have two independent equations:

$$(yu_x - v - xv_x) + p_1v_x - p_2u_x - p_4 + 2p_6x - p_8y = 0 \quad (5.14)$$

$$(u + yu_y - xv_y) + p_1v_y - p_2u_y - p_5 + 2p_7y - p_8x = 0 \quad (5.15)$$

The parameters in equations (5.14) and (5.15) are linearly independent when the depth function is not of the form given in lemma I. Selecting five suitable points we obtain two alternative sets of simultaneous equations in five unknowns. These can then be solved for the five motion parameters. Note, however, that when $p_1 = x_0 = 0$ then equation (5.14) alone cannot be used for the computation.

This is because the parameters $(p_1, p_2, p_4, p_6, p_8)$ cannot then be used to solve for the five motion parameters. A similar restriction holds for equation (5.15) when $p_2 = y_0 = 0$.

Algorithm VI: It has been shown that when two optical flow fields obtained at two different time instants is available then the motion parameters are uniquely determined from measurements at three non collinear points on the retina. The assumption here is that the motion parameters are stable during the measurement period. This can be used as a basis for the motion estimation algorithm.

Algorithm VII: *Motion parameters from structure and optical flow.*

When the structure of the moving surface is known, its motion is unambiguous. This method also reduces the dimensionality of the parameter space by isolating the rotational parameters. Two alternative constraint equations can be used here. In the first form spatial derivatives of the optical flow function are needed. This implies local analytic reconstruction of the flow function. In the alternative form of the constraint depth ratios are needed, implying reliable (and dense) measurement of surface normals.

From eq. (5.5) the expressions for the spatial derivatives of the optical flow (u, v) are obtained as:

$$u_x = -\frac{W'}{Z} - (\lambda_0 - \lambda) \frac{W'}{Z^2} \frac{\partial Z}{\partial x} - \alpha_x + 2\beta_x \quad (5.16.1)$$

$$u_x = -(\lambda_0 - \lambda) \frac{W'}{Z^2} \frac{\partial Z}{\partial x} - \alpha_x - \gamma \quad (5.16.2)$$

$$u_1 = - (x_0 - x) \frac{W}{Z^2} \frac{\partial Z}{\partial x} + \beta y + \gamma \quad (5.16.3)$$

$$v_1 = - \frac{W}{Z} - (x_0 - y) \frac{W}{Z^2} \frac{\partial Z}{\partial y} - 2\alpha y + \beta x \quad (5.16.4)$$

Substituting $(x_0 - x) \frac{W}{Z}$ and $(x_0 - y) \frac{W}{Z}$ in the above equations from equation (5.5)

we get:

$$u_1 - v_1 = (-u - \alpha xy + \beta(x^2 + 1) - \gamma y)\psi + (x - \alpha(y^2 + 1) - \beta xy - \gamma x)\rho - \alpha y + \beta x \quad (5.17.1)$$

$$u_1 = (-u - \alpha xy + \beta(x^2 + 1) - \gamma y)\rho - \alpha x - \gamma \quad (5.17.2)$$

$$v_1 = (-x - \alpha(y^2 + 1) + \beta xy + \gamma x)\psi + \beta y + \gamma \quad (5.17.3)$$

where $\psi = \frac{\partial Z}{\partial x}$ and $\rho = \frac{\partial Z}{\partial y}$

Thus at every image location (x, y) , a set of three linear independent equations involving the rotation parameters can be obtained. The functions $\psi(x, y)$ and $\rho(x, y)$ are computable from the surface orientation values $\frac{\partial Z}{\partial x}|_{x,y}$ and $\frac{\partial Z}{\partial y}|_{x,y}$ (see Appendix II).

When it is not possible to measure derivatives of the optical flow, but the ratio of depths at any two image locations can be estimated, an alternative linear constraint equation can be derived involving only the rotation parameters. Consider two image points (x_1, y_1) and (x_2, y_2) with depths z_1 and z_2 respectively. The optical flow values at these points are (u_1, v_1) and (u_2, v_2) . The motion parameters are $(U, V, W, \alpha, \beta, \gamma)$. Using equation (5.5) we have the following equations

$$u_1 z_1 - u_2 z_2 = (x_2 - x_1)W + z_1(-\alpha x_1 y_1 + \beta(x_1^2 + 1) - \gamma y_1) - z_2(-\alpha x_2 y_2 + \beta(x_2^2 + 1) - \gamma y_2)$$

$$v_1 z_1 - v_2 z_2 = (y_2 - y_1) + z_1(-\alpha(y_1^2 + 1) + \beta x_1 y_1 + \gamma x_1) - z_2(-\alpha(y_2^2 + 1) + \beta x_2 y_2 + \gamma x_2)$$

Eliminating W from the above equations we have

$$l_{12}\alpha + m_{12}\beta + r_{12}\gamma + s_{12} = 0 \quad (5.18)$$

where

$$l_{12} = x_1 y_1 y_2 - x_2 y_1^2 + x_1 - x_2 + \frac{z_2}{z_1}(x_2 y_1 y_2 - x_1 y_2^2 - x_1 + x_2)$$

$$m_{12} = x_1 x_2 y_1 - x_2^2 y_2 + y_1 - y_2 + \frac{z_2}{z_1}(x_1 x_2 y_2 - x_2^2 y_1 - y_1 + y_2)$$

$$r_{12} = x_1 x_2 + y_1 y_2 - x_1^2 - y_1^2 + \frac{z_2}{z_1}(-x_2^2 - y_2^2 + x_1 x_2 + y_1 y_2)$$

$$s_{12} = u_1(y_2 - y_1) - u_2(x_2 - x_1) + \frac{z_2}{z_1}(-u_2(y_2 - y_1) + v_2(x_2 - x_1))$$

If the surface normal values are available everywhere in a region enclosing two image points, then the depth ratio, $\frac{z_2}{z_1}$, (corresponding to those locations) can be estimated (of course, mathematically, it is possible to compute this ratio if the surface normal values are known along a path from the one image location to the other). Consequently, each pair of image points gives rise to a linear constraint in the rotation parameters. Thus by a suitable choice of three pairs of image points we can uniquely solve for the rotation parameters and subsequently the translation parameters $(\frac{U}{W}, \frac{V}{W})$ (see Appendix I).

The novel feature of the above algorithm is that it can combine shape and motion information under two different conditions:

- (1) In the first case the optical flow field has been measured sufficiently 'densely' to enable local reconstruction of the flow field. This enables the first order spatial derivatives of the flow field to be estimated. Then at all retinal points where the surface normals are known, we can locally solve for the rotation parameters by means of a set of three linear constraint equations.
- (2) Alternatively, if the flow measurements are not dense, but the shape measurements allow reconstruction of the depth function (upto a constant scale factor), then again locally we obtain linear constraints in the rotation parameters (e.g. equation (5.18)).

This means that in any image neighbourhood, full reconstruction of either shape or 2D motion, helps to recover both structure and motion. The schematic diagram of the algorithm is given in figure IV.

Remarks:

- (i) Note the similarity between algorithms I and VII. In both, the local analytic reconstructability of either the optical flow function or the surface normal function, determines the selection of the constraint equation that is to be used.
- (ii) From equations (5.17), ψ and ρ can be eliminated to obtain a cubic polynomial equation in the three rotation parameters. Thus if the optical flow and its first spatial derivatives are measured we can use the cubic constraint to estimate the rotation parameters by the hough transform technique. So,

although the nonlinearity remains, the dimension of the parameter space is reduced, which reduces the size of the search space.

5.6. Discrete motion under orthography

This case is of interest to researchers in the field of Visual Cognitive Modeling [31]. The reason for this, is that psychological experiments by Ullman [32] to explain human capabilities in the perception of structure from motion, agree more with the orthographic projection (actually an extension of orthography, termed *polar parallel projection* [32]) than with the perspective projection models.

For the case of biological motion a plethora of proposals have been put forward by several researchers in the area, and many potentially powerful algorithms have been proposed [14] [15], [4] [35]. The research reported here, however does not cover this case of motion analysis.

5.7. Discrete motion under Perspective

This is the most involved among all the motion types. To simplify the analysis, Ullman [31] assumed the rotation axis to be along the z axis. The constraint he obtained was an equation of the fourth degree in the sine of the rotation angle. Another simplification is due to Fang and Huang, whose "small rotation" assumption makes their analysis similar to the differential case. The most extensive work done in this particular area is due to Tsai and Huang [30]. Their work is innovative and based on elegant mathematical formalisms. However a general unambiguous solution to the motion perception problem in case of

Smith, Reid G. & Randall Davis (1981) A Framework for Cooperation in Distributed Problem Solving, *IEEE Transactions on Systems, Man, and Cybernetics*, SMC-11, no. 1, Jan.

Stefik, Mark (1980) *Planning with Constraints*, Report STAN-CS-80-784, Department of Computer Science, Stanford University.

Wilensky, Robert (1982) Talking to UNIX in English: An Overview of U.C., *Proceedings of AAAI-82*, Carnegie-Mellon University, Pittsburgh.

Wilensky, Robert (1983) *Planning and Understanding*, Addison-Wesley, Reading, MA.

discrete perspective is still unavailable.

6. Conclusion

The problem of interpretation of a moving retinal image has been studied, for both "short range" and "long range" motion. Our findings, indicate that motion information available in optical flow (differential case) is less than that in the discrete displacements field (long range motion).

We saw that three temporally contiguous image frames contain enough information to uniquely recover 3-D motion and structure under perspective projection. Since the optical flow field (two temporally proximal frames) is, in general, ambiguous, two frames can recover structure when the moving surface satisfies the conditions of Theorem 1.

We proved that structure and 3-D motion parameters are equivalent - the one constrains the other uniquely - and both problems (determination of structure and 3-D motion parameters from retinal displacements) are better tackled this way.

We believe that our work forms an important extension to Ullman's and Huang's theories, and, in conjunction with interpretation schemes for recovering structure in the case of biological motion (using the planarity assumption), constitutes a significant advance towards the solution of the problem of Motion Perception.

7. Acknowledgements

This paper would not be written without the invaluable help of Dana Ballard. Jerry Feldman, Chris Brown and Peter Lennie provided helpful critiques. Thanks

are also due to the Rochester vision group, especially John Aloimonos and Paul Chou, for the many fruitful discussions. This work was supported in part by the Defense Advanced Research Projects Agency under grant number N00014-82-K-0193 and in part by the National Science Foundation under grant number DCR-8405720.

8. References

1. J. Aloimonos, One eye suffices : A computational model of monocular depth perception, Tech. Rep. 160, Dept. of Computer Science, Univ. Rochester, 1985.
2. D. H. Ballard and C. Brown, *Computer Vision*, Prentice Hall, 1982.
3. D. H. Ballard and O. Kimball, Rigid Body Motion From Depth and Optical Flow, *CVGIP* 22, (1983), 95-115.
4. B. M. Bennett and D. D. Hoffman, The Computation of Structure from Fixed - Axis Motion: Nonrigid Structures, *Biological Cybernetics* 51, (1985), 293-300.
5. C. M. Brown, M. B. Curtiss and D. B. Sher, Advanced Hough Transform Implementations, *Proceedings of the Eighth International Joint Conference on Artificial Intelligence*, , 1983, 1081.
6. A. Bruss and B. K. P. Horn, Passive Navigation, *CVGIP* 21, (1983), 3-20.
7. H. S. M. Coxeter, *Introduction to Geometry*, John Wiley & Sons.

8. J. Q. Fang and T. S. Huang, Some Experiments on Estimating the 3D Motion Parameters of a Rigid body from two Consecutive Image Frames, *IEEE Trans. PAMI* 6(5), (1984), 545-554.
9. J. Q. Fang and T. S. Huang, Solving three dimensional small- rotation motion equations: Uniqueness, algorithms, and numerical results, *CVGIP* 26, (1984), 183-206.
10. J. A. Feldman and D. H. Ballard, Connectionist Models and Their Properties, *Cognitive Science* 6, 205 - 254.
11. C. L. Fennema and W. B. Thompson, Velocity Determination in Scenes Containing Several Moving Objects, *CVGIP* 9, (1979), 301-315.
12. H. L. F. V. Helmholtz, *Treatise of Physiological Optics*, NY Dover Publications, 1925.
13. D. D. Hoffman, Inferring local surface orientation from motion fields, *J. Opt. Soc. Am.* 72(7), (July 1982), 888-892.
14. D. D. Hoffman and B. F. Flinchbaugh, The Interpretation of Biological Motion, *Biol. Cybernetics* 42, (1982), 195-204.
15. D. D. Hoffman and B. M. Bennett, Inferring the relative three dimensional positions of two moving points, *J. Opt. Soc. Am. A* 2(2), (February 1985), 350-353.
16. T. S. Huang and R. Y. Tsai, Image Sequence Analysis: Motion Estimation, in *Image Sequence Analysis*, T. S. Huang (ed.), Springer-Verlag, 1981, ch. 1.

17. R. Jain, An approach for the Direct Computation of the Focus of Expansion. *PRIP*, , 1982, 262-268.
18. R. Jain, Direct Computation of the Focus of Expansion, *IEEE Trans. PAMI* 5(1), (January 1983), 58-63.
19. H. C. Longuet-Higgins and K. Prazdny, The interpretation of a moving retinal image, *Proc. Royal Soc. London B* 208, (1980), 385 - 397.
20. H. C. Longuet-Higgins, A Computer Algorithm for Reconstructing a Scene from Two Projections. *Nature* 293, (September 1981), 133-135.
21. H. H. Nagel and B. Neumann, On 3-D Reconstruction from Two Perspective Views, *IJCAI*, , 1981, 661-663.
22. K. Prazdny, Egomotion and Relative Depth Map from Optical Flow, *Biol. Cybernetics* 36, (1980), 87-102.
23. K. Prazdny, Determining the instantaneous direction of motion from optical flow generated by curvilinearly moving observer, *CVGIP* 17, (1981), 94 - 97.
24. J. H. Rieger and D. T. Lawton, Determining the Instantaneous Axis of Translation from Optic Flow Generated by Arbitrary Sensor Motion, *COINS tech. report 83-1*, , January 1983.
25. J. W. Roach and J. K. Aggarwal, Determining the Movement of Objects from a Sequence of Images, *IEEE Trans. PAMI* 2 (6), (November 1980), 54-562.

26. K. Sugihara and N. Sugie, Recovery of Rigid Structure from Orthographically Projected optical flow, *CVGIP* 27, (1984), 309-320.
27. R. Y. Tsai and T. S. Huang, Estimating Three-Dimensional Motion Parameters of a Rigid Planar Patch, *IEEE Trans. A.S.S.P. ASSP-29*, (December 1981), 1147-1152.
28. R. Y. Tsai and T. S. Huang, Estimating Three-Dimensional Motion Parameters of a Rigid Planar Patch III: Finite Point Correspondences and the Three-View Problem, *Proc. IEEE Conf. ASSP, Paris* . . May 1982.
29. R. Y. Tsai, T. S. Huang and W. Zhu, Estimating Three-Dimensional Motion Parameters of a Rigid Planar Patch II : Singular Value Decomposition, *IEEE Trans. A.S.S.P. ASSP-30*, (August 1982), 525-534.
30. R. Y. Tsai and T. S. Huang, Uniqueness and Estimation of Three Dimensional Motion Parameters of Rigid Objects with Curved Surfaces, *IEEE, Trans. PAMI* 6, (January 1984), 13-27.
31. S. Ullman, The Interpretation of Visual Motion, *Ph.D. Thesis* . , 1977.
32. S. Ullman, The Interpretation of Structure from Motion, *Proc. R. Soc. Lond. (B)* B 203, (1979), 405-426.
33. S. Ullman, Maximising rigidity: The incremental recovery of 3D structure from rigid and nonrigid motion, *Perception* 13, (1984), 255-274.
34. A. M. Waxman and S. Ullman, Surface structure and 3D Motion from Image Flow: A Kinematic Analysis, CAR-Tech. Rep.-24, Center for

Automation Research, University of Maryland.. October 1983.

35. J. Webb and J. K. Aggarwal, Structure from motion of rigid and jointed objects, *Artificial Intelligence* 19, (1982), 107-130.

APPENDIX I

Uniqueness of Motion Parameters computed from Optical flow under Perspective Projection

Consider a point P in space whose coordinates are (X, Y, Z) with respect to a fixed inertial frame XYZ . The image of this point is $p = (x, y)$ whose coordinates are given with respect to a xy frame located on the image plane. The relation between the world point P and the image point p is given by

$$(x, y) = \left(\frac{FX}{Z}, \frac{FY}{Z} \right) \quad (i)$$

where ' F ' is the *focal length* of the imaging system. This is assumed to be unity in the following analysis.

Now if a rigid surface moves with a translational velocity $V_T = (U, V, W)$ and a rotational velocity $\Omega = (\alpha, \beta, \gamma)$. Then, from kinematics, the three dimensional velocity of any point on the surface can be written as

$$\left(\frac{dX}{dt}, \frac{dY}{dt}, \frac{dZ}{dt} \right) = V_T + \Omega \times (X, Y, Z) \quad (ii)$$

where ' t ' is the time variable and ' \times ' denotes vector product.

In differential motion case the image motion or optical flow is denoted by

$(u, v) = \left(\frac{dx}{dt}, \frac{dy}{dt} \right)$. Differentiating equation (1) and substituting from equation (2) we

have the following relations

$$u = \frac{U - \alpha W}{Z} - \alpha x + \beta(x^2 + 1) - \gamma y \quad (iii.a)$$

$$v = \frac{1 - xW}{Z} = \alpha(y^2 + 1) + \beta xy + \gamma x \quad (\text{iii.b})$$

Eliminating the unknown depth variable from the above we get

$$\frac{u + \alpha xy - \beta(x^2 + 1) + \gamma x}{x + \alpha(y^2 + 1) - \beta xy - \gamma x} = \frac{U - xW}{V - yW} \quad (\text{iv})$$

The above equation describes the constraint imposed by the measured value of optical flow (u, v) , at an image point (x, y) , on the six motion parameters $(U, V, W, \alpha, \beta, \gamma)$.

Proposition I. *Given the rotation parameters the translation parameters can be uniquely determined from the optical flow field*

Proof: First we define the function $\mu(x, y)$ where,

$$\mu = \frac{u + \alpha xy - \beta(x^2 + 1) + \gamma x}{x + \alpha(y^2 + 1) - \beta xy - \gamma x} \quad (\text{iv})$$

Now we analyse the following cases:

Case 1: If $\mu = \text{constant}$ then from equation (iv) we have $W = 0$. In this case we can only obtain the ratio $\frac{U}{V}$ from the optical flow field.

Case 2: If $\mu \neq \text{constant}$ then there are two image points where μ is different. In which case we can solve the resultant set of two linear equations, obtained from (iv), to get $x_0 = \frac{U}{W}$ and $y_0 = \frac{V}{W}$.

Proposition II. *Given the translation parameters the rotation parameters can be uniquely determined from optical flow.*

Proof: Here the values of x_0 and y_0 are known. The expression for optical flow is,

$$u = (x_0 - x)\varphi - \alpha xy + \beta(x^2 + 1) - \gamma y$$

$$v = (y_0 - y)\varphi - \alpha(y^2 + 1) + \beta xy + \gamma x$$

Where (α, β, γ) are the rotation parameters and $\varphi = \frac{W}{Z}$ is the reciprocal of the scaled depth function. If possible let there be another surface moving with the same translation but different rotation parameters, but generating the same optical flow. Thus we have,

$$u = (x_0 - x)\varphi' - \alpha' xy + \beta'(x^2 + 1) - \gamma' y$$

$$v = (y_0 - y)\varphi' - \alpha'(y^2 + 1) + \beta' xy + \gamma' x$$

Now from the above sets of equations by subtracting appropriately we get,

$$0 = (x_0 - x)(\varphi - \varphi') - \Delta\alpha xy + \Delta\beta(x^2 + 1) - \Delta\gamma y \quad (v.a)$$

$$0 = (y_0 - y)(\varphi - \varphi') - \Delta\alpha(y^2 + 1) + \Delta\beta xy + \Delta\gamma x \quad (v.b)$$

where $\Delta\alpha = \alpha - \alpha'$, $\Delta\beta = \beta - \beta'$ and $\Delta\gamma = \gamma - \gamma'$. Eliminating $(\varphi - \varphi')$ from the above we have,

$$\begin{aligned} (\Delta\alpha x_0 + \Delta\beta y_0) - x(\Delta\gamma + \Delta\alpha) - y(\Delta\gamma y_0 + \Delta\beta) + x^2(\Delta\beta y_0 + \Delta\gamma) \\ + y^2(\Delta\alpha y_0 + \Delta\gamma) - xy(\Delta\beta x_0 + \Delta\alpha y_0) = 0 \end{aligned} \quad (vi)$$

Since the above equation is valid everywhere in the image,

$$\begin{array}{ll} \Delta\alpha x_0 + \Delta\beta y_0 = 0 & \Delta\alpha y_0 + \Delta\beta x_0 = 0 \\ \Delta\gamma x_0 + \Delta\alpha = 0 & \Delta\beta y_0 + \Delta\gamma = 0 \\ \Delta\gamma y_0 + \Delta\beta = 0 & \Delta\alpha x_0 + \Delta\gamma = 0 \end{array}$$

From the above we obtain,

$$\Delta\alpha = 0 \quad \Delta\beta = 0 \quad \Delta\gamma = 0$$

This means that $\alpha = \alpha'$, $\beta = \beta'$ and $\gamma = \gamma'$ and therefore, *the rotation parameters are*

uniquely determined when the translation parameters are known.

Proposition III *If the structure of a Rigidly moving surface is known, then the parameters describing its motion is uniquely determined.*

Proof: Knowing structure means that we have the depth values available upto some scale factor. Thus in equation (iii) the value 'Z' is no longer an unknown. The unknown scale factor is lumped with the translation parameters. Now proceeding in a manner analogous to the previous proof we have,

$$\frac{1}{Z}(\Delta U' - x \Delta W') = \Delta \alpha x y - \Delta \beta (x^2 + 1) + \Delta \gamma y \quad (\text{vii.a})$$

$$\frac{1}{Z}(\Delta V' - y \Delta W') = \Delta \alpha (y^2 + 1) - \Delta \beta x y - \Delta \gamma x \quad (\text{vii.b})$$

Eliminating $\frac{1}{Z}$ we have,

$$\begin{aligned} &(\Delta \alpha \Delta U' + \Delta \beta \Delta V') - x(\Delta \gamma \Delta U' + \Delta \alpha \Delta W') - y(\Delta \beta \Delta W' + \Delta \gamma \Delta V') \\ &+ x^2(\Delta \gamma \Delta W' + \Delta \beta \Delta V') + y^2(\Delta \gamma \Delta W' + \Delta \alpha \Delta U') - xy(\Delta \alpha \Delta V' + \Delta \beta \Delta U') \end{aligned}$$

Since the above equation must be valid all over the image plane, the following relations hold:

$$\Delta \alpha \Delta U' + \Delta \beta \Delta V' = 0 \quad \Delta \alpha \Delta W' + \Delta \gamma \Delta U' = 0 \quad \Delta \beta \Delta W' + \Delta \gamma \Delta V' = 0$$

$$\Delta \alpha \Delta V' + \Delta \beta \Delta U' = 0 \quad \Delta \beta \Delta V' + \Delta \gamma \Delta W' = 0 \quad \Delta \alpha \Delta U' + \Delta \gamma \Delta W' = 0$$

From eqn. (vii) and the above relations we have,

$$\Delta U' = \Delta V' = \Delta W' = \Delta \alpha = \Delta \beta = \Delta \gamma = 0$$

Therefore, once the structure is known for a rigidly moving surface, its translation (upto a scale factor) and its rotation is determined uniquely from the optical flow generated by the motion.

APPENDIX II

Representations of surface orientation and their properties

In computer vision, the terms *surface orientation map* and *shape* are sometimes used interchangeably. The following is an attempt to explain the basis of this usage. The cases of *Perspective* as well as *Orthographic* projections are considered. Shape information obtainable from a surface orientation map in image coordinates is also explored.

Representations for surface orientation

A direction in three space is specified by two independent parameters.

- A. (Latitude, Longitude): The coordinates are denoted by (θ, φ) where $0 \leq \theta < \pi$, $0 \leq \varphi < \pi$.
- B. Coordinates on the *gaussian* (or unit radius) sphere. If the coordinates are (l, m, n) then $l^2 + m^2 + n^2 = 1$.
- C. (slant , tilt): Slant is the tangent of their latitude angle (or $\tan \theta$) while tilt is the longitude angle. The symbolic notation is (σ, τ) .
- D. (Gradient): If the depth is expressed in the form $Z = f(X, Y)$, then it is the level surface $F(X, Y, Z) = 0$, where

$$F(X, Y, Z) = f(X, Y) - Z$$

The gradient of $F(\cdot)$, i.e. $(\frac{\partial f}{\partial X}, \frac{\partial f}{\partial Y}, -1)$ gives the orientation of the surface (in the direction of increasing $F(\cdot)$). The gradient notation is written as (p, q) .

where $(p, q) = (\frac{\partial f}{\partial X}, \frac{\partial f}{\partial Y})$.

Relationship among the surface normal representations:

$$\sqrt{p^2 + q^2} = \tan \theta = \sigma$$

$$\frac{q}{p} = \tan \phi = \tan \tau$$

$$(l, m, n) = (\frac{p}{g}, \frac{q}{g}, \frac{-1}{g}), \quad g = \sqrt{p^2 + q^2 + 1}$$

Shape under Perspective Projection

In the case of perspective projection the relation between a world point (X, Y, Z) and its projection (x, y) in the image plane is given by

$$(x, y) = (\frac{FX}{Z}, \frac{FY}{Z}) \quad (i)$$

where F is the focal length of the imaging system.

The surface is represented in the world frame by the functional form $Z(X, Y)$. It is assumed that the surface can also be represented (at least locally) by the function $z(x, y)$ in image coordinates. Here the relation between the surface normals $(\frac{\partial Z}{\partial X}, \frac{\partial Z}{\partial Y})$ corresponding to an image point (x, y) and the partial derivatives of $z(x, y)$ are sought.

A. Relationship between surface gradients in image and world coordinates. Now a small displacement $(\delta x, \delta y)$ in the image plane corresponds to a displacement $(\delta X, \delta Y, \delta Z)$ in the world frame, along the surface $Z(X, Y)$. From equation (i) we get the relation

$$\delta X = \frac{\delta_x Z + x \delta Z}{F} \quad (\text{ii.a})$$

$$\delta Y = \frac{\delta_y Z + y \delta Z}{F} \quad (\text{ii.b})$$

Furthermore the following identity holds

$$Z(X + \delta X, Y + \delta Y) = z(x + \delta x, y + \delta y) \quad (\text{iii})$$

Using the Taylor series expansion of the above

$$Z(X + \delta X, Y + \delta Y) = Z(X, Y) + \delta X \frac{\partial Z}{\partial X} + \delta Y \frac{\partial Z}{\partial Y} + (\text{higher order terms}) \quad (\text{iv.a})$$

$$z(x + \delta x, y + \delta y) = Z(x, y) + \delta x \frac{\partial Z}{\partial x} + \delta y \frac{\partial Z}{\partial y} + (\text{higher order terms}) \quad (\text{iv.b})$$

Neglecting the higher order terms in equation (iv) and substituting for δX and δY from equation (ii) in equation (iv.a)

$$Z(X + \delta X, Y + \delta Y) - Z(X, Y) = \delta Z = \frac{1}{F}(\delta_x Z + x \delta Z) \frac{\partial Z}{\partial X} + \frac{1}{F}(\delta_y Z + y \delta Z) \frac{\partial Z}{\partial Y}$$

or

$$\delta Z \left(F - x \frac{\partial Z}{\partial X} - y \frac{\partial Z}{\partial Y} \right) = Z \delta x \frac{\partial Z}{\partial X} + Z \delta y \frac{\partial Z}{\partial Y} \quad (\text{v})$$

Recall now that

$$Z(X + \delta X, Y + \delta Y) - Z(X, Y) = z(x + \delta x, y + \delta y) - z(x, y)$$

Therefore combining equations (iii), (iv) and (v)

$$\delta x \frac{Z}{F - x \frac{\partial Z}{\partial X} - y \frac{\partial Z}{\partial Y}} \frac{\partial Z}{\partial X} + \delta y \frac{Z}{F - x \frac{\partial Z}{\partial X} - y \frac{\partial Z}{\partial Y}} \frac{\partial Z}{\partial Y} = \delta x \frac{\partial z}{\partial x} + \delta y \frac{\partial z}{\partial y} \quad (\text{vi})$$

Since δx and δy are independent of each other we have

$$\frac{\partial z}{\partial x} = \frac{Z \frac{\partial Z}{\partial X}}{F - x \frac{\partial Z}{\partial X} - y \frac{\partial Z}{\partial Y}} \quad (\text{vii.a})$$

$$\frac{\partial z}{\partial y} = \frac{Z \frac{\partial Z}{\partial Y}}{F - x \frac{\partial Z}{\partial X} - \frac{\partial Z}{\partial Y}} \quad (\text{vii.b})$$

B. What Shape means Consider the shape information available from the field of surface normals indexed by the image coordinates. Making the appropriate substitutions from equations (vii) in equation (iv.b) we have:

$$\frac{z(x + \delta x, y + \delta y) - 1}{Z(x, y)} = 1 + \delta x \frac{\frac{\partial Z}{\partial X}}{F - x \frac{\partial Z}{\partial X} - \frac{\partial Z}{\partial Y}} + \delta y \frac{\frac{\partial Z}{\partial Y}}{F - x \frac{\partial Z}{\partial X} - \frac{\partial Z}{\partial Y}}$$

Thus the following statement can be made:

Under perspective projection, when the field of surface normals is available, indexed by image coordinates, then the image centered depth function can be computed upto a dilation factor.

Lemma 1. If the surface Z is represented by an algebraic function $Z(X, Y)$ and furthermore if the function $z(x, y)$ denotes the same surface in terms of the image coordinates (x, y) , then the tilt function $\tau(x, y)$ is given by

$$\tau = \frac{\frac{\partial Z}{\partial X}}{\frac{\partial Z}{\partial Y}} = \frac{\frac{\partial z}{\partial x}}{\frac{\partial z}{\partial y}}$$

Proof: Since $Z(X, Y)$ is an algebraic function, by definition it can be expressed implicitly by the polynomial equation $F(X, Y, Z) = 0$. We can write $F(\cdot)$ as

$$\sum_{i=0}^L \sum_{j=0}^M \sum_{k=0}^N c_{ijk} X^i Y^j Z^k = 0 \quad (\text{viii})$$

where the c_{ijk} 's are real constants and L, M, N are finite positive integers. By using

the implicit function theorem we get

$$\tau = \frac{\frac{\partial Z}{\partial Y}}{\frac{\partial Z}{\partial X}} = \frac{-\frac{F_Y}{F_Z}}{-\frac{F_X}{F_Z}} = \frac{F_Y}{F_X}$$

where F_X, F_Y, F_Z denote the partial derivative of $F(\cdot)$ with respect to X, Y and Z .

Therefore we have from equation (viii):

$$\tau = \frac{\sum_{i=0}^I \sum_{j=1}^M \sum_{k=0}^N j c_{ijk} X^i Y^{j-1} Z^k}{\sum_{i=1}^I \sum_{j=0}^M \sum_{k=0}^N i c_{ijk} X^{i-1} Y^j Z^k} \quad (\text{ix})$$

Observe now that we can obtain an implicit representation for the depth in terms of the image coordinates (x, y) from equation (viii) by substituting for X and Y in accordance with $x = \frac{X}{Z}$ and $y = \frac{Y}{Z}$ (where the focal length is assumed to be 1).

Thus we obtain the representation $G(x, y, z) = 0$ or

$$\sum_{i=0}^I \sum_{j=0}^M \sum_{k=0}^N c_{ijk} x^i y^j z^{i+j+k} = 0 \quad (\text{x})$$

Again by the implicit function theorem we have

$$\frac{\frac{\partial z}{\partial y}}{\frac{\partial z}{\partial x}} = \frac{-\frac{G_y}{G_z}}{-\frac{G_x}{G_z}} = \frac{G_y}{G_x} = \frac{\sum_{i=0}^I \sum_{j=1}^M \sum_{k=0}^N j c_{ijk} x^i y^{j-1} z^{i+j+k}}{\sum_{i=1}^I \sum_{j=0}^M \sum_{k=0}^N i c_{ijk} x^{i-1} y^j z^{i+j+k}}$$

or

$$\frac{\frac{\partial z}{\partial y}}{\frac{\partial z}{\partial x}} = \frac{\sum_{i=0}^I \sum_{j=1}^M \sum_{k=0}^N j c_{ijk} x^i y^{j-1} z^{i+j+k-1}}{\sum_{i=1}^I \sum_{j=0}^M \sum_{k=0}^N i c_{ijk} x^{i-1} y^j z^{i+j+k-1}} \quad (\text{xi})$$

Consider now, equation (ix) and substitute $x = xz$ and $y = yz$

$$\frac{\frac{\partial Z}{\partial Y}}{\frac{\partial Z}{\partial X}} = \frac{\sum_{i=0}^L \sum_{j=1}^M \sum_{k=0}^N i c_{ijk} x^i y^{j-1} z^{i+j+k-1}}{\sum_{i=1}^L \sum_{j=0}^M \sum_{k=0}^N i c_{ijk} x^{i-1} y^j z^{i+j+k-1}} \quad (\text{xii})$$

But the right hand sides of the equations (xi) and (xii) are identical. This means,

$$\tau = \frac{\frac{\partial Z}{\partial Y}}{\frac{\partial Z}{\partial X}} = \frac{\frac{\partial z}{\partial y}}{\frac{\partial z}{\partial x}}$$

which concludes the proof of the lemma.

Shape under Orthographic Projection:

Under orthography the image coordinates of a point are equal to the corresponding three dimensional coordinates, or

$$(x, y) = (X, Y)$$

Thus

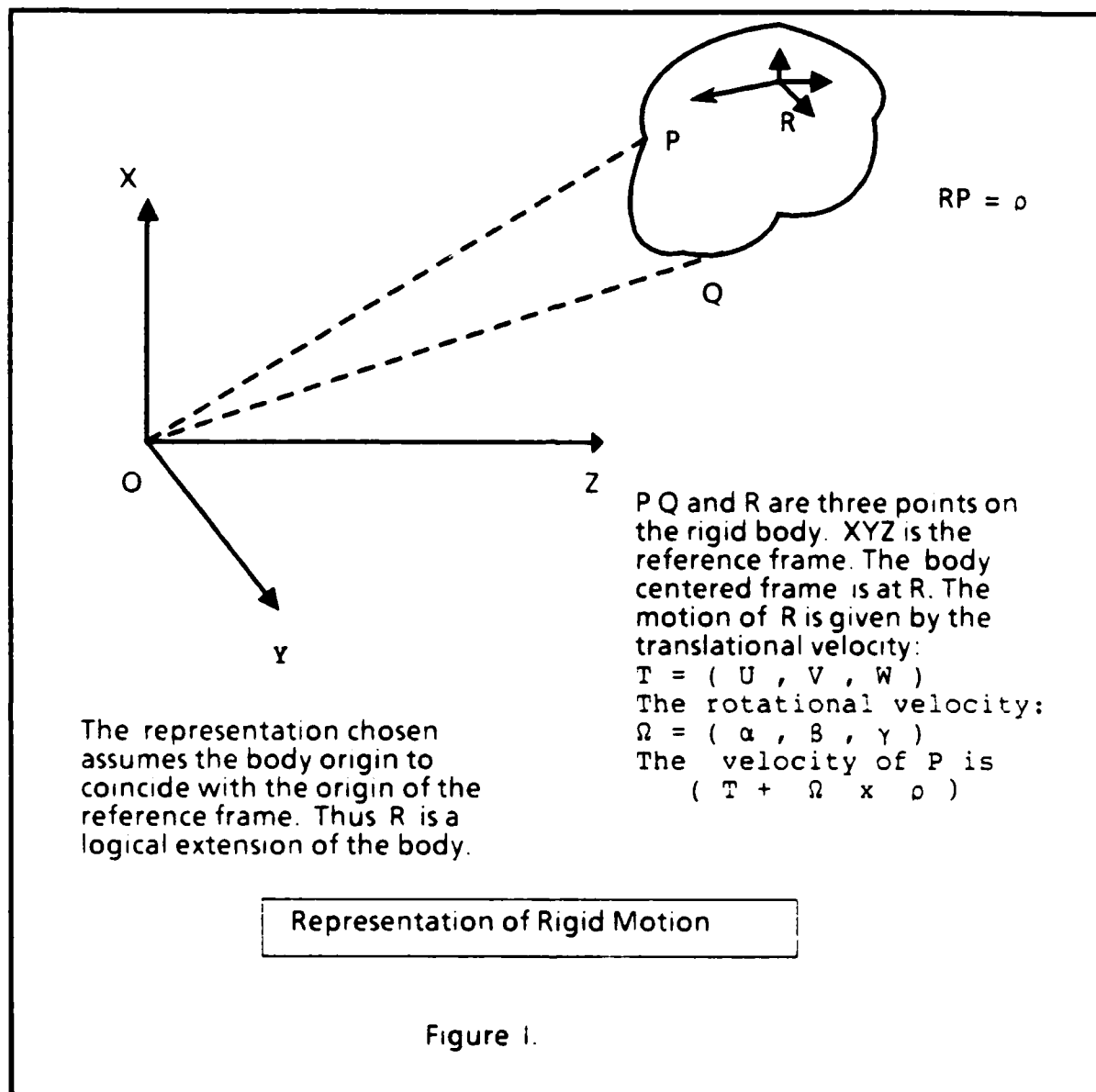
$$\left(\frac{\partial Z}{\partial x}, \frac{\partial Z}{\partial y} \right) = \left(\frac{\partial Z}{\partial X}, \frac{\partial Z}{\partial Y} \right)$$

Now observe from equation (iv.a) that when the surface normals are known at an image point (x, y) , then the depth difference between this point and neighbouring image points are known:

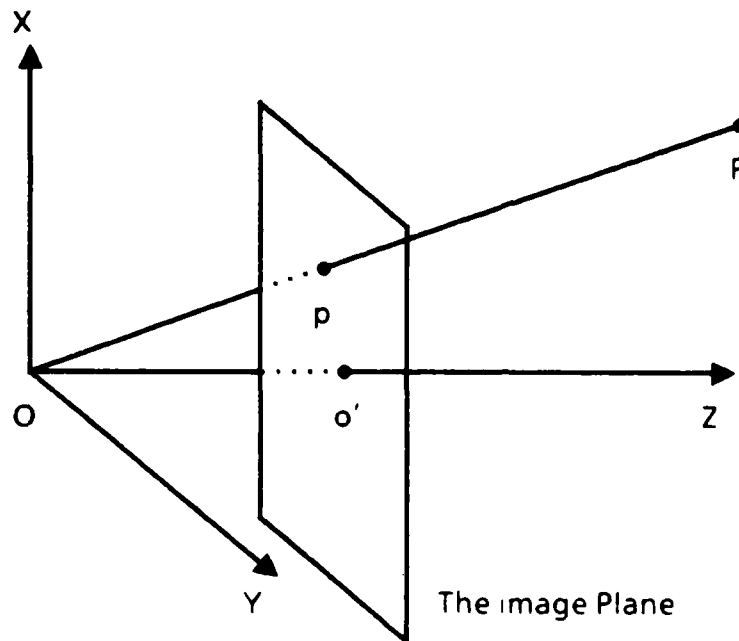
$$Z(X + \delta X, Y + \delta Y) - Z(X, Y) = \delta X \frac{\partial Z}{\partial X} + \delta Y \frac{\partial Z}{\partial Y} + (\text{higher order terms})$$

Thus we can state the following:

When a map of surface normals is available under orthography, the depth function can be computed upto a constant additive term.



The Perspective Projection Model



The image $p = (x, y)$ of the world point $P = (X, Y, Z)$ is projected by the ray OP . The focal length of the system is 'F'. The equation of the image plane is:

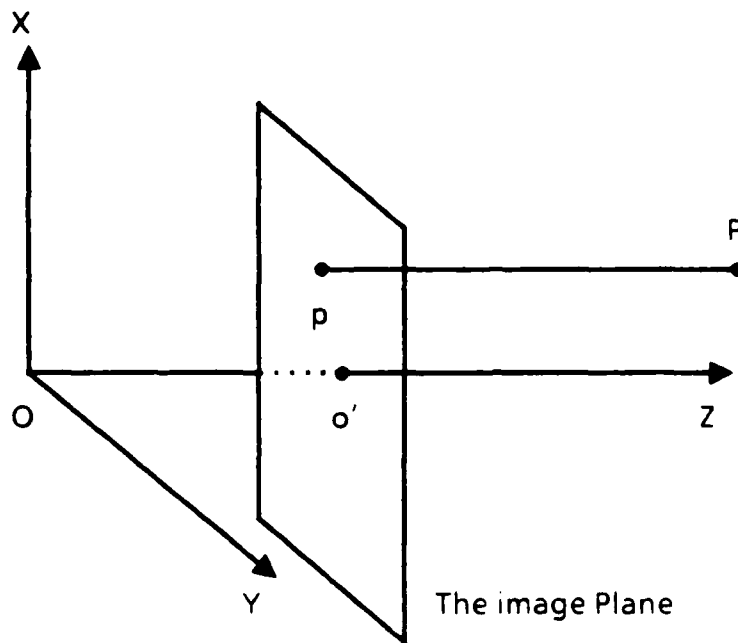
$$Z = F$$

The relation between image and world coordinates is:

$$x = FX/Z \quad \text{and} \quad y = FY/Z$$

Figure II a.

The Orthographic Projection Model

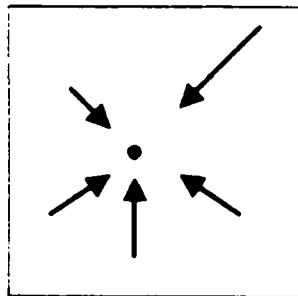


The image $p = (x, y)$ of the world point $P = (X, Y, Z)$ is projected by a ray parallel to the line of sight (Z axis). The relations are

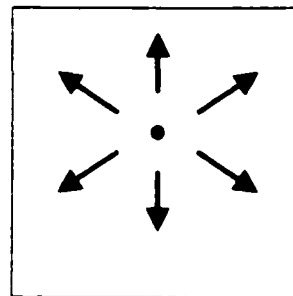
$$x = X \quad \text{and} \quad y = Y$$

Note that all depth information is lost.

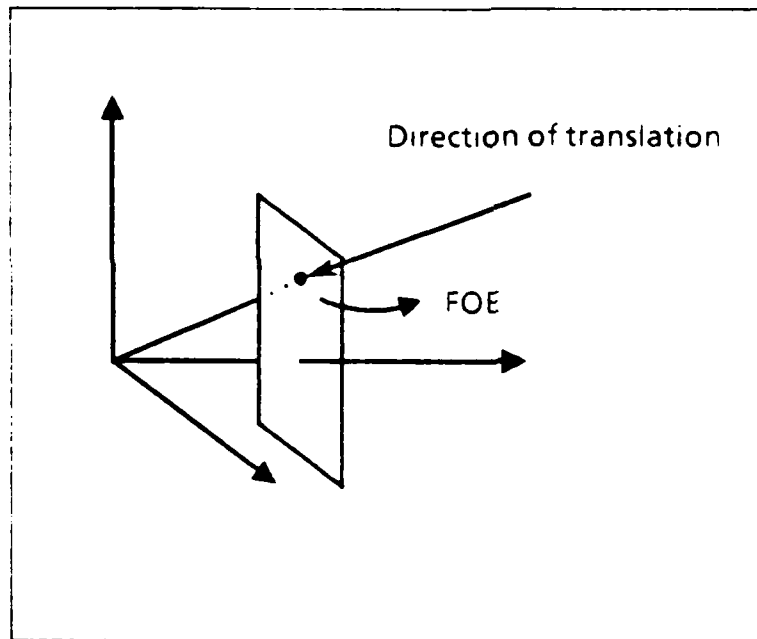
Figure II b.



FOC



FOE



The structure of the translational flow field

Figure III

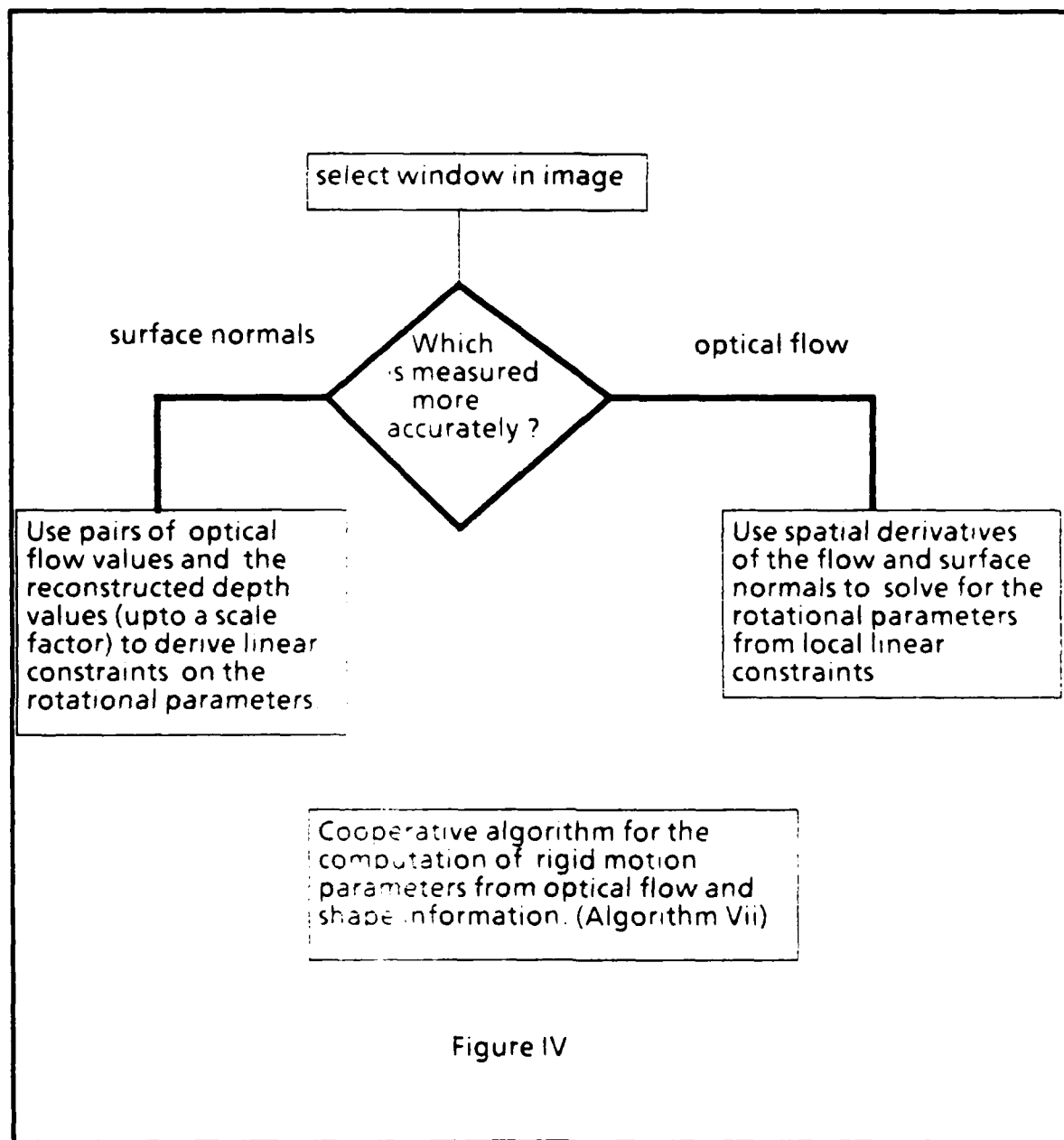


Figure IV

END

DTIC

9-86