MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY

A.I. Memo No. 592 December, 1980

Inferring Shape From Motion Fields

D.D. Hoffman

ABSTRACT

The human visual system has the ability to utilize motion information to infer the shapes of surfaces. More specifically, we are able to derive descriptions of rigidly rotating smooth surfaces entirely from the orthographic projection of the motions of their surface markings.

A computational analysis of this ability is proposed based on a "shape from motion" proposition. This proposition states that given the first spatial derivatives of the orthographically projected velocity and acceleration fields of a rigidly rotating regular surface, then the angular velocity and the surface normal at each visible point on that surface are uniquely determined up to a reflection.

The computational analysis proceeds in three main steps. First it is shown that surface tilt and one component of the angular velocity may be obtained entirely from the first spatial derivatives of the velocity field. Second it is shown that surface slant and the remaining two components of the angular velocity are computable if the first spatial derivatives of the acceleration field are also given. Finally the problem of constructing a velocity field from the temporally changing optic array is briefly discussed.

ACKNOWLEDGEMENT: This report describes research done at the Artificial Intelligence Laboratory and Psychology Department of the Massachusetts Institute of Technology. Support for this research is provided in part by the Advanced Research Projects Agency of the Department of Defense under Office of Naval Research contract N00014-80-C-0505 and by NSF and AFOSR under grant 7923110-MCS. The author was also supported by a Hughes Aircraft company graduate fellowship.

1. Introduction

Visual motion provides a powerful base for inferences about the layout of the immediate environment and the motions of the various constituents of that environment. The focus of this paper is one inference that the human visual system does appear to perform routinely based on visual motion alone. In particular, the human visual system has a remarkable ability to utilize motion information to infer the three dimensional shapes of surfaces. More specifically, we are able to derive correct descriptions of rigidly rotating smooth surfaces entirely from the orthographic projection of the motions of their surface markings.

A demonstration of this ability, similar to that of Ullman (1979), is illustrated in figure 1. Dots are randomly placed on a sphere in the memory of a computer. Successive snapshots of this random dot sphere are generated at five degree intervals and orthographically projected in quick temporal succession (using an ISI of 20 msec and a presentation time per frame of 20 msec) on a computer driven crt. Figure 1a shows three successive frames as they would appear statically on the crt. As is obvious from the figure each individual frame gives no impression of being a sphere.[1] Rather it just looks like a somewhat circular array of random dots. However, when the frames are presented in quick temporal succession one obtains a compelling perception of a smooth sphere in rotation (see figure 1b).

It is important to note that the perception is of a *smooth* spherical surface, not, for example, of invisible wires connecting the individual dots as in Johannson's biological motion (Johannson, 1973). One has the feeling that there is an almost tangible smooth black pearl with little lights attached to its surface. The importance of noting this is that it indicates the type of description that appears to be built by the visual system. It is a description whose primitives relate to *surfaces* rather than to positions of isolated points.[2]

That this visual ability is a nontrivial feat becomes apparent when it is realized that the mapping from the environment onto the retina is many to one. The information available to the visual system underdetermines the surface which is the source of the motion observed, so that any conclusions drawn about that surface are *in principle* nondemonstrative. Yet, surprisingly, our perception is, in general, of a unique surface in rotation. More surprisingly, it is more often than not correct. Clearly the visual system must be utilizing generally valid constraints about the nature of surfaces and objects in our world in order to obtain this unique solution. One constraint of central importance in obtaining a unique surface is the rigidity constraint; the environment is usually, though not always, composed of rigid objects (Ullman, 1979; Johansson, 1964 & 1975; Hay, 1966; Green 1961; Wallach & O'Connell, 1953; Gibson & Gibson, 1957). Later this constraint will be given a precise mathematical formulation and its utility in arriving at a unique interpretation clearly illustrated.

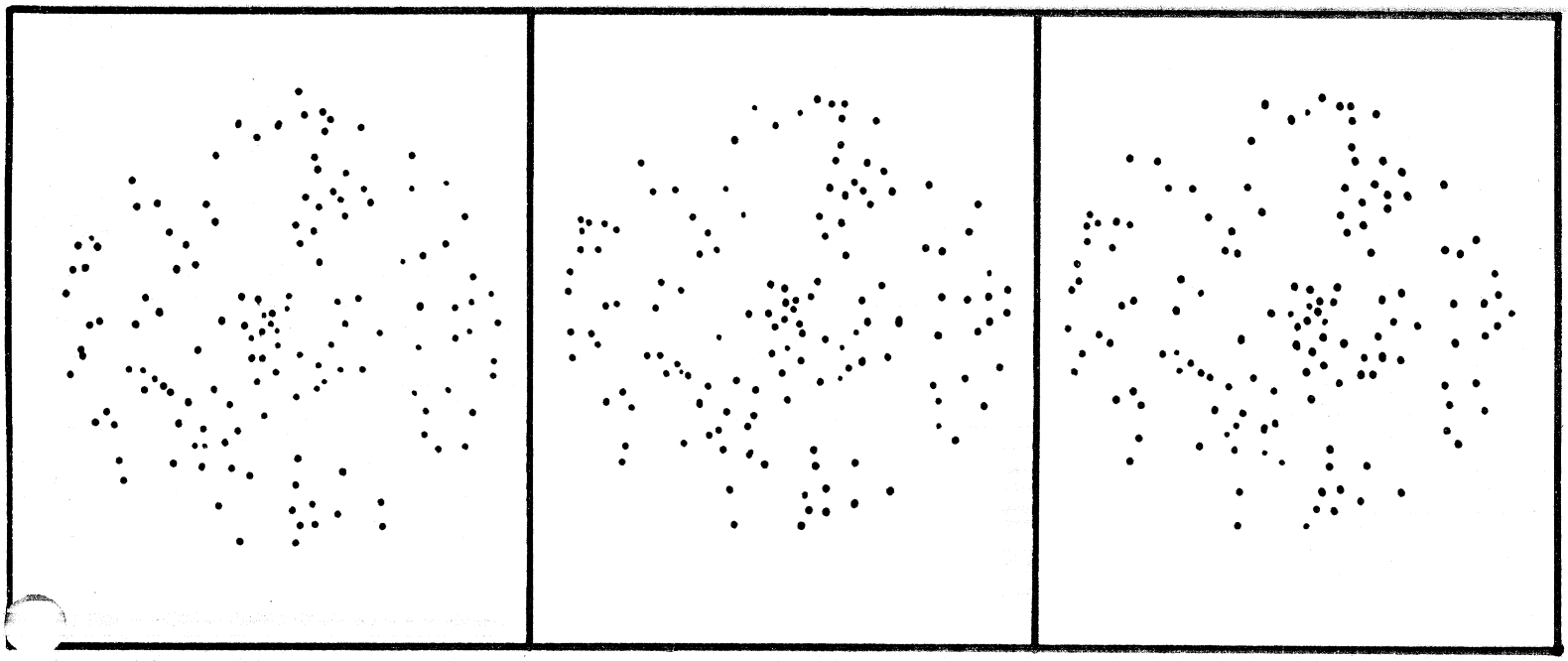
The goal of this paper is to provide a description of this perceptual ability at a level which Marr and Poggio have called a computational theory (Marr & Poggio, 1977). The computational analysis proposed is based on a "shape from motion" proposition[3] which states that given the first spatial

[1]This eliminates single frame information such as texture gradients from being a plausible explanation of this ability.

[2]This does not discount the possibility, of course, that positions of points might be computed first and smooth surfaces fitted through them afterwards. In fact, just such a scheme appears to be utilized in stereo vision (Grimson, 1980).

[3]The term "proposition" is not intended to imply any hubristic claims regarding the complexity of this result or its derivation. Rather it is intended to emphasize that the present inquiry is a *computational* analysis.

(a)



(b)

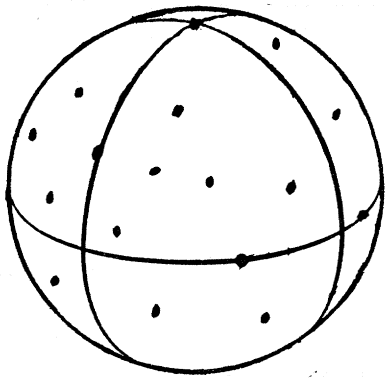


Figure 1. (a) Three successive frames of a rotating random dot sphere. Each frame is rotated five degrees to the right about the vertical axis with respect to the previous frame. (b) The resulting spherical percept when the frames are presented in quick succession.

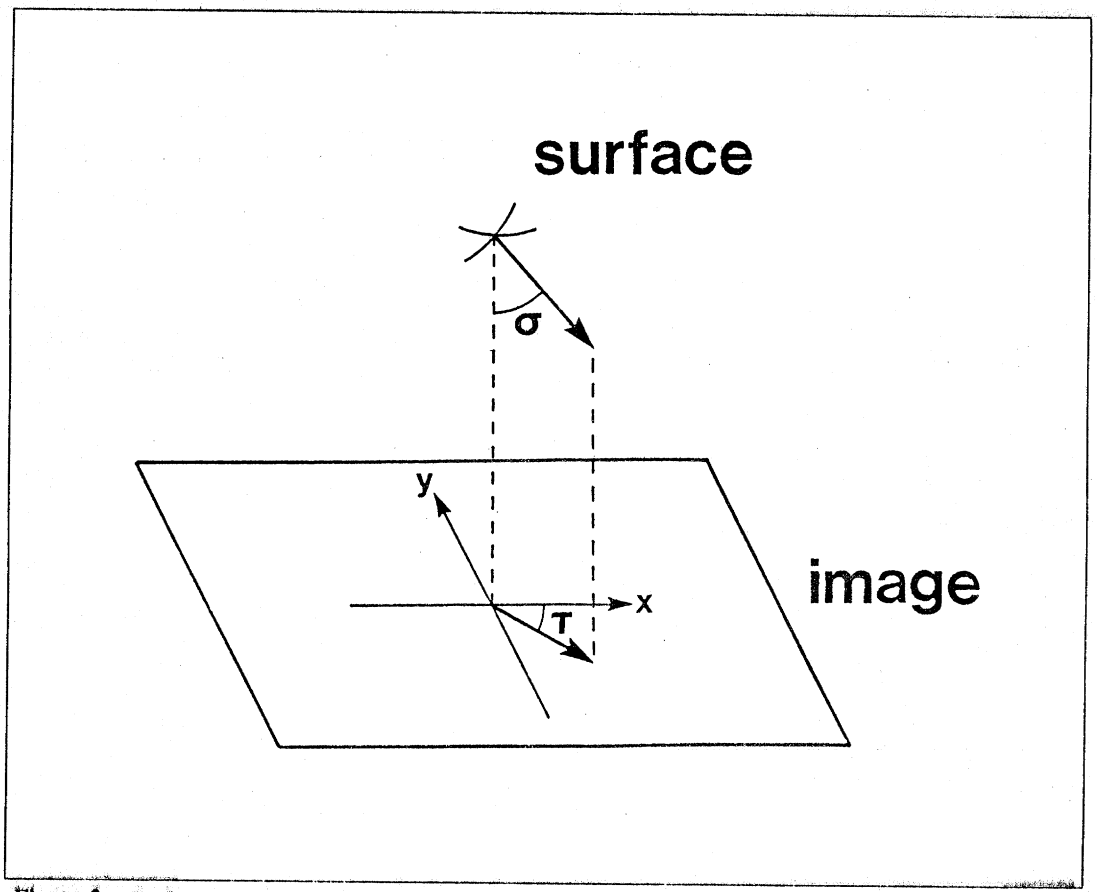


Figure 2. Surface representations using slant, σ, and tilt, τ. Rather than representing the surface normal at a point in terms of surface gradients (z_x and z_y) it is convenient to adopt the slant and tilt convention proposed by both Stevens (1980) and Attneave (1972). Briefly, tilt indicates in which direction a surface is rotated from the observer's frontal plane and slant indicates how much it is rotated away from the frontal plane in that direction. Whereas surface gradients tend to infinity at occluding contours, tilt ranges only between 0 and 180 degrees and slant ranges from 0 to 90 degrees. The equations of transformation are $\sigma = \tan^{-1}\sqrt{z_x^2 + z_y^2}$ and $\tau = \tan^{-1}\sqrt{z_y/z_x}$.

derivatives of the orthographically projected velocity and acceleration fields of a rigidly rotating regular surface, its angular velocity and the surface normal at each visible point on the surface are uniquely determined up to a reflection about the image plane.

For clarity the computational analysis is presented in three main steps. First it is shown that surface tilt (see figure 2) and one component of the angular velocity may be obtained from the first spatial derivatives of the velocity field. Then it is shown that surface slant and the remaining two components of the angular velocity are computable if the first spatial derivatives of the acceleration field are also given. Finally, since the computational analysis assumes as one of its givens a velocity field, the problem of constructing a velocity field from the temporally changing optic array is discussed briefly.

2. Two Previous Computational Analyses

The ability of the human visual system to infer the correct three dimensional description of an object from its projected motion alone has been investigated computationally several times before. Two of these previous analyses will be briefly discussed to illustrate the two basic types of computational approaches that can be taken to this problem and the two basic types of resulting descriptions.

Ullman (1979) took what may be called the "discrete approach" to the problem. The givens for his computational analysis are three successive snapshots of isolated points moving in a rigid configuration. The resulting description he builds is essentially a set of triples giving the three dimensional positions of the points in relation to each other. Fundamental to Ullman's elegant analysis is his "structure from motion" theorem which states that the structure of four non-coplanar points in a rigid configuration is recoverable from three orthographic projections.

An example of the "continuous approach" to the problem can be found in Longuet-Higgins and Prazdny (1980).[4] Rather than utilizing discrete orthographic projections as input, they assume a velocity field arising from a perspective projection. The resulting description computed involves surfaces instead of sets of triples. In short they prove that given the perspective projection and first and second spatial derivatives of the velocity field presented to a moving observer it is in principle possible to compute both the observer's motion and the surface gradients at each point in the visual field.

The present analysis falls into the continuous category. Flow fields are assumed as the input and a description of the surface of interest in terms of the surface normal (slant, tilt) at each visible point is the desired result. Where the current analysis differs from that of Longuet-Higgins and Prazdny and other previous work within the continuous approach is that here orthographic projection is assumed instead of perspective projection. Consequently in this analysis it proves impossible to derive both the observer's motion and a complete surface description merely from the velocity field and its spatial derivatives. The relations of these various approaches is summarized in figure 3.

3. Why Use Orthographic Projection?

Why bother performing a computational analysis of the problem assuming orthographic projection? After all it will be shown that less information about local surface properties can be computed from the velocity field in orthographic projection than in perspective. Specifically, surface slant computation requires the temporal derivative of the velocity field. There are several motivations.

First, as Ullman (1979) points out, perspective effects are often rather noisy and unreliable. To utilize them locally would require very careful measurements by the visual system.

Second, orthographic projection provides a good local approximation to the actual retinal projection. A theorem from differential topology allows us to conclude that whatever the true retinal

[4]Several other researchers have examined aspects of this problem from a continuous point of view (Koenderink & Van Doorn, 1976; Nakayama & Loomis, 1974; Gibson, 1950).

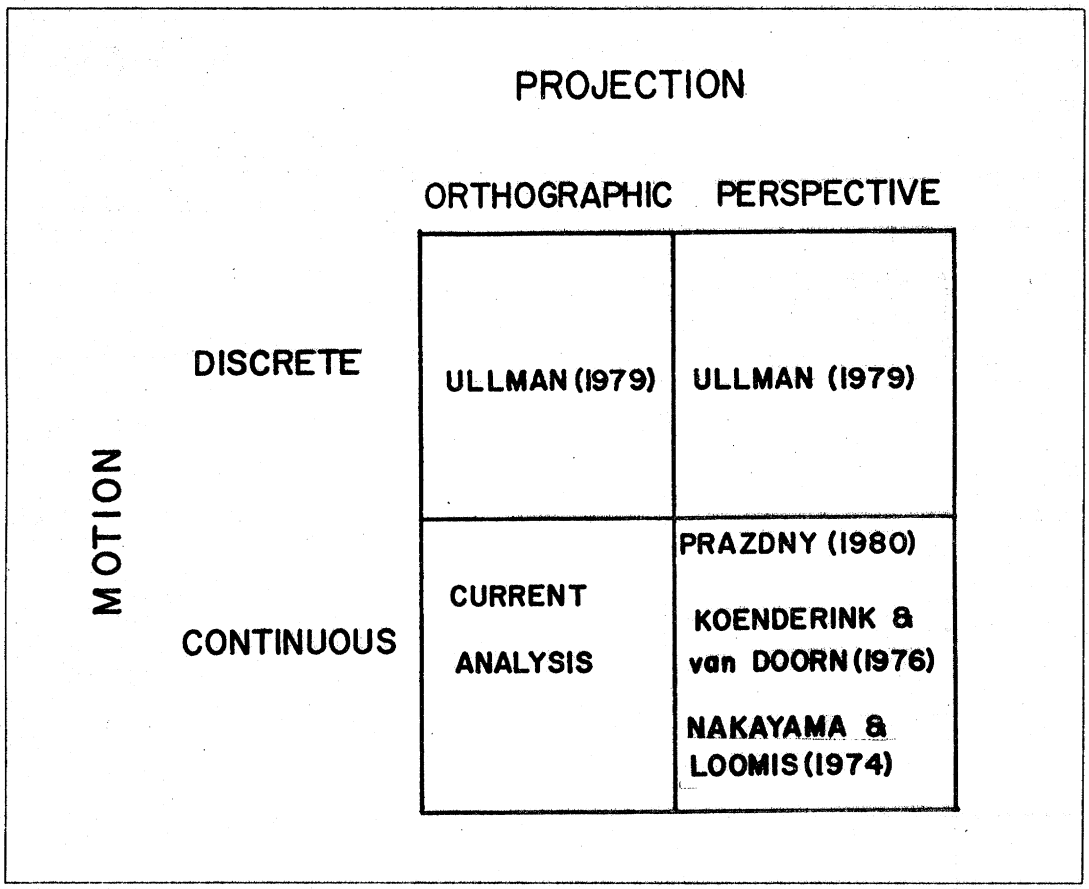


Figure 3. A categorization of the various computational approaches to the problem of deriving shape from motion. The categorization scheme is given by crossing projection type (orthographic or perspective) with motion type (discrete frames versus optical flow).

projection is, it is locally equivalent to orthographic projection.[5]

A third motivation is provided by the results of some psychophysical tests done by Ullman. Using a cylinder composed of random dots he showed that observers can recover the correct structure entirely from the orthographic projection of the motion of the dots when the cylinder is rotated about its axis. However observers cannot recover the structure under perspective projection when the object is alternately receding and approaching without any rotation. This is significant because a computational analysis shows that if perspective effects are taken into account the structure can in principle be recovered from receding and approaching motion alone. These results tend to support the psychological reality of a computational theory based on a locally orthographic projection for the recovery of shape from motion.

Alternate computational analyses provide clear candidate hypotheses that may be tested for their psychological reality and that each lend different insights into the subject of study. For example it will be shown later that the tilt component of the surface normal is much more easily recovered than the slant component, both in the nature of the motion information required and the computations in-

[5]The theorem is called the **Local Submersion Theorem** (see, for example, Guillemin & Pollack (1974)). It states, "Suppose that $f: X \mapsto Y$ is a submersion at x, and $y = f(x)$. Then there exists local coordinates around x and y such that for $k \geq l$, $f(x_1, \ldots, x_k) = (x_1, \ldots, x_l)$. That is, f is locally equivalent to the canonical submersion near x."

volved. This is an interesting result and one that could provide a basis for psychophysical examination of the psychological reality of this analysis.

Finally, the equations for surface orientation and motion derived using orthographic projection are much simpler than those derived under perspective projection. Not only are the the equations simpler, they do not require measurements of the *second* spatial derivatives of the velocity field as is typical in the perspective case.

4. Geometrical Model

The idealized geometry underlying the following computational analysis is illustrated in figure 4. A rigid patch of surface, S, is considered to be an open set of points each of which has an associated position vector **R**. The position vector for a point on S with respect to the x, y, z coordinate system is given by:

$$\mathbf{R} = x\mathrm{i} + y\mathrm{j} + z(x, y)\mathrm{k} \tag{1}$$

where i, j, k are unit vectors along the x, y, z axes respectively. The surface, S, has an angular velocity Ω given by:

$$\Omega = \omega_1\mathrm{i} + \omega_2\mathrm{j} + \omega_3\mathrm{k} \tag{2}$$

with respect to the x, y, z coordinate system.

Note that Ω may either result from rotary motions of the surface or from movement of the image plane, I, with respect to S or both. As long as Ω is not zero it doesn't matter whether the surface is rotating and the observer remains stationary or whether the surface is stationary and the observer's motion with respect to the surface includes an angular component.

Associated with S is a velocity vector field, **V**, which at any point $p \in S$ is given by:

$$\mathbf{V} = \Omega \times \mathbf{R} + \mathbf{T} \tag{3}$$

where **T** is any net translation between the observer and the surface.[6]

The velocity field available to the observer is an orthographic (parallel) projection of the velocity field, **V**, associated with S onto the image plane, I.

Now this is clearly an idealization. The real observer is definitely not given a velocity field but must *construct* such a field from the temporally changing optic array. This problem will be discussed briefly later. For the analysis of the present problem of inferring the shape of S, the orthographically projected velocity field is assumed as a given.

With this simple geometrical model as background the computational analysis of the problem of inferring shape from orthographically projected motion is now presented as the proof to the following proposition.

[6]Actually **T** is any net translation between the observer and the *axis of rotation* of the surface. However, the translation term is of no consequence for the present analysis since it will drop out when the spatial derivatives of (3) are taken.

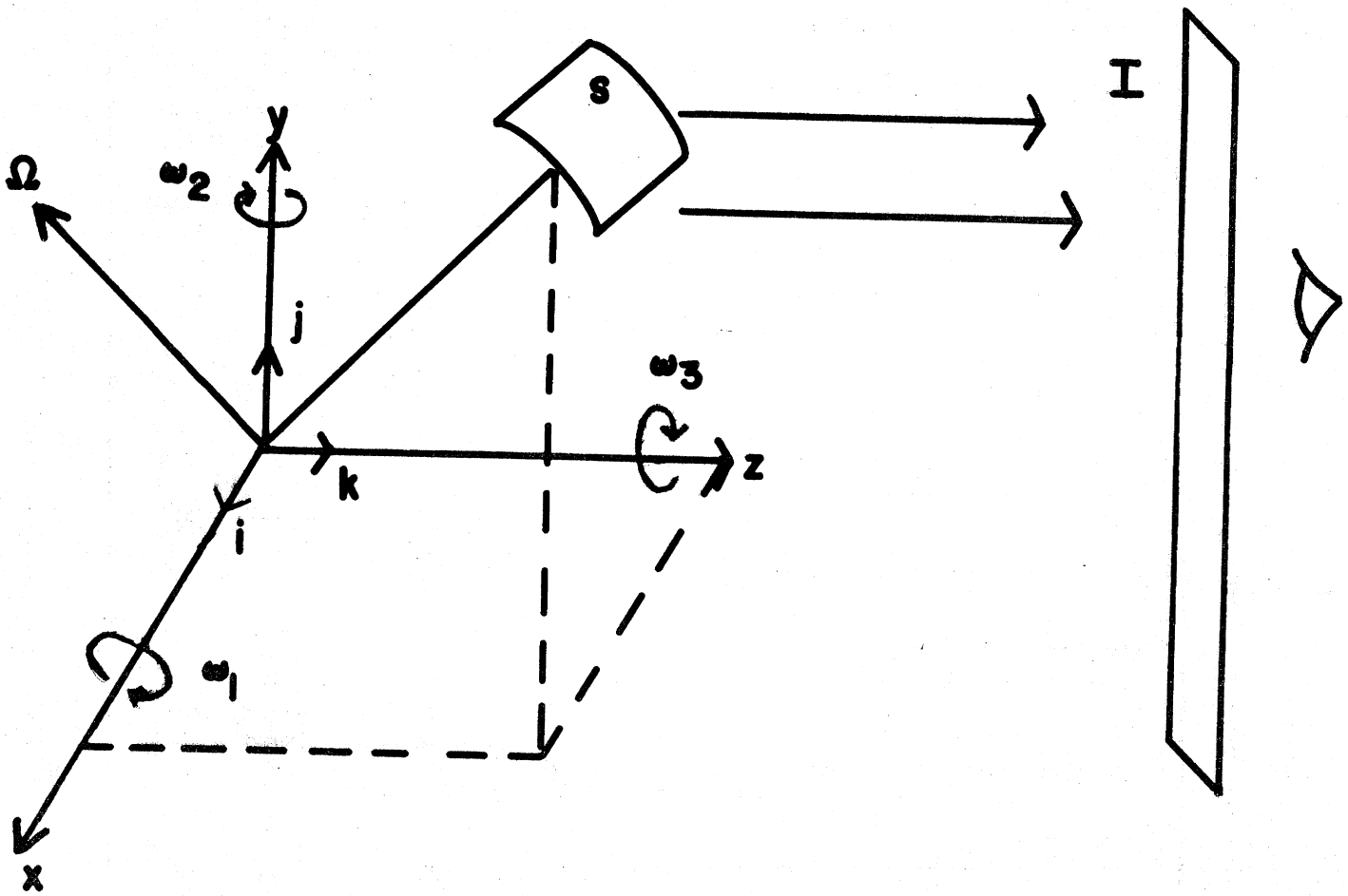


Figure 4. Geometrical model underlying the computational analysis. S is some surface with angular velocity Ω. The resulting velocity field $\mathbf{V}$ is orthographically projected onto the image plane I.

5. The Shape From Motion Proposition

Given the first spatial derivatives of the orthographically projected velocity and acceleration fields of a rigidly rotating regular surface, the angular velocity and surface normal at each visible point on the surface are uniquely determined up to a reflection about the image plane.

Proof. The proof of this proposition involves deriving equations for the two components (σ, τ) of the surface normal, N, at each visible point and for the three components of the angular velocity $(\omega_1, \omega_2, \omega_3)$. For clarity of presentation the proof is divided into two lemmas. In the first lemma equations for the tilt, τ, and one component of the angular velocity, ω_3, are derived and discussed. In the second lemma the same is done for the slant, σ, and the remaining two components of the angular velocity, ω_1 and ω_2.

5.1 Lemma 1.

Both the tilt, τ, at each visible point on S and the component of angular velocity about the axis orthogonal to the image plane, ω_3, are computable given only the first spatial derivatives of the orthographically projected velocity field.

To make the claims of this lemma clearer figure 5 illustrates the tilt fields associated with two simple surfaces and figure 4 illustrates with which axis the angular velocity component, ω_3, is associated.

Proof of Lemma 1. Since the projection plane, I, is orthogonal to the unit vector, k, the orthographic projection of the velocity field, $\mathbf{V}^*$, is given by:[7]

$$\mathbf{V}^* = \mathbf{V} - (\mathbf{V} \cdot \mathbf{k})\mathbf{k} \tag{4}$$

What this essentially means is that the components of the velocity field along the x and y axes survive orthographic projection unaltered, whereas the component along the z axis (i.e., along the observer's line of sight) is eliminated completely. Consequently the only spatial derivatives of the velocity field that need be computed are along the x and y directions. Denoting spatial partial derivatives by subscripts, the first spatial derivatives of the velocity field (equation 3) along the x and y axes are:

$$\mathbf{V}_x = \Omega_x \times \mathbf{R} + \Omega \times \mathbf{R}_x \tag{5}$$

$$\mathbf{V}_y = \Omega_y \times \mathbf{R} + \Omega \times \mathbf{R}_y \tag{6}$$

Before investigating equations (5) and (6) further it is helpful to introduce a mathematical expression for the rigidity constraint that will allow these equations to be simplified. The motivation for the particular mathematical expression to be used here is simple. One consequence of surface rigidity is that the entire surface can have only one angular velocity, Ω. Regardless of which neighborhood of

[7] This characterization of the orthographic projection of a vector is borrowed from Witkin (1980).

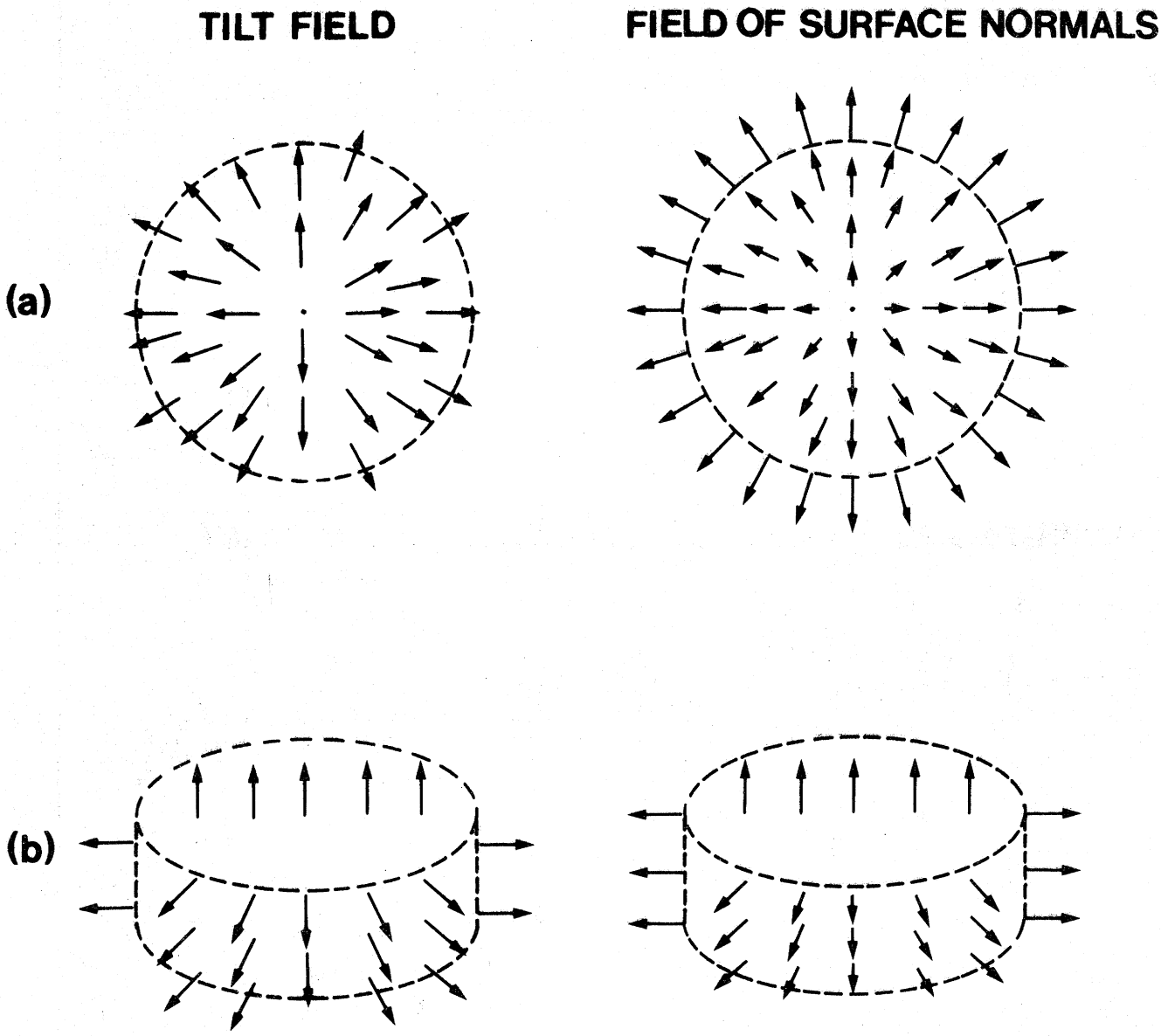


Figure 5. Tilt fields compared with fields of surface normals for two surfaces. According to lemma 1 one can obtain the correct tilt field $(1, \tau)$ from the velocity field but, unfortunately, not the field of surface normals (σ, τ). A tilt field is an example of a *field of directions* (Do Carmo, 1976, p.178). Since no magnitude information is known, only the direction of tilt, the tilt fields in (a) and (b) are indicated by constant length vectors pointing in the direction of tilt. The surfaces are (a) a sphere and (b) a cylinder.

5.4 Remarks on the Shape From Motion Proposition

It has been shown that the angular velocity and the surface normal at each visible point of a rigidly rotating regular surface are uniquely determined up to a reflection if one is given the orthographic projection and first spatial derivatives of the associated velocity and acceleration fields. This proposition and its proof are proposed as the basis for a computational theory of the human visual ability to perceive the shape of a smooth moving surface from its motion alone.

Some disclaimers are in order. First, only arguments for the sufficiency of this approach, not its necessity, have been suggested. Alternative computational theories are available, some of which were discussed earlier. It is a matter for empirical investigation to determine which, if any, of the current theories is to some extent psychologically real.

Two pieces of psychophysical evidence may be adduced to *suggest* the greater psychological reality of the present approach over previous ones. First are Ullman's (1979) experiments, mentioned before, which indicate that only orthographic, not perspective, information seems to be utilized by the visual system in recovering surface shapes. The second is that the resulting perceptual effect (illustrated in figure 1) is of a smooth surface as opposed to isolated points connected by invisible wires. This suggests greater psychological reality for an approach which builds a description whose primitives relate to surfaces.

A second disclaimer must be mentioned. The visual system may utilize additional generally valid constraints for the interpretation of surface shapes from motion. For example, shortcuts in computing the slant, σ, might be based on noting that σ must be 90 degrees at external occluding contours and must vary smoothly between them. Another potentially powerful constraint is that the tilt field must be locally orthogonal to the image of its occluding contour (for smooth surfaces). Thus further investigation of valid means to reduce the computational complexity of this approach is warranted before serious claims for its psychological reality can be sustained.

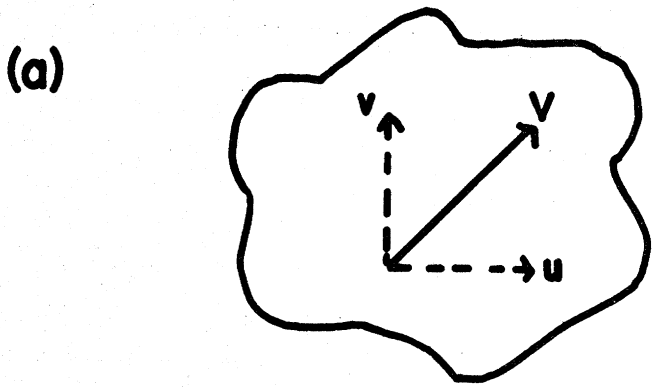
6. Computing Velocity Fields

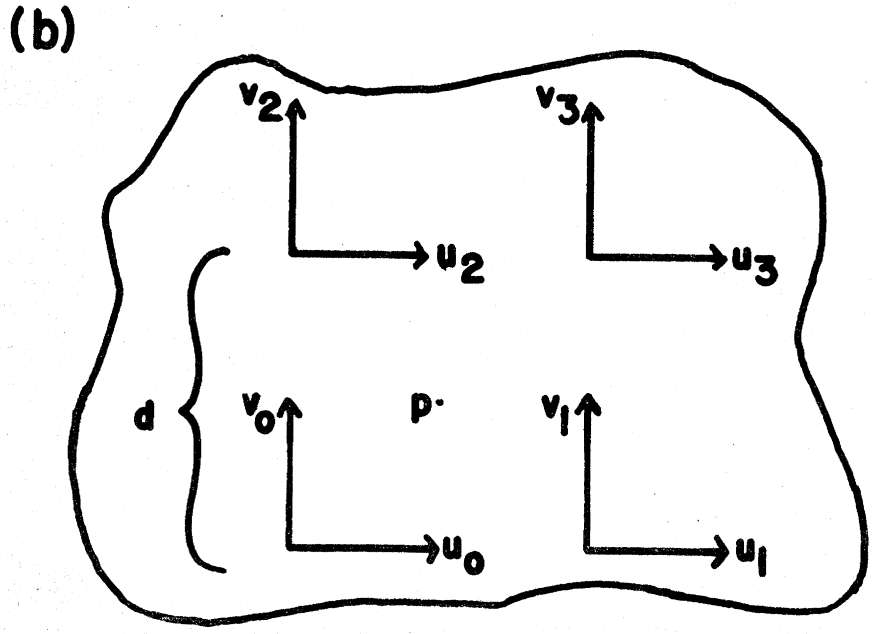
To this point the analysis has assumed the velocity field and its first spatial derivatives are given.[13] Clearly this is not the case for a real observer. The real observer is presented with a temporally changing optic array. If a velocity field is required it must be *constructed* from the changing optic array.

The problem of computing a velocity field has remained nontrivial despite much recent research. One can show that the motion information available at any single point in a changing optic array is insufficient to uniquely determine the velocity field at that point. Consequently much of the research in the field of optical flow has been devoted to discovering valid means of integrating motion information from local neighborhoods to uniquely determine the flow at each point in the neighborhood.

A detailed analysis of the problem of determining optical flow is presented in Horn and Schunck (1980), which also includes a representative list of references on the topic.

[13]Actually only the first spatial derivatives have been used.





$$\partial u/\partial x \simeq [(u_3 - u_2) + (u_1 - u_0)]/2d \qquad \partial u/\partial y \simeq [(u_2 - u_0) + (u_3 - u_1)]/2d$$

$$\partial v/\partial x \simeq [(v_3 - v_2) + (v_1 - v_0)]/2d \qquad \partial v/\partial y \simeq [(v_2 - v_0) + (v_3 - v_1)]/2d$$

Figure 6. The orthographically projected velocity field and its first spatial derivatives. (a) illustrates the decomposition of a velocity vector at a point in the field into its x component, u, and its y component, v. (b) illustrates how the spatial derivatives u_x, u_y, v_x and v_y can be approximated at some point, p, from the local velocity field.

5.2 Remarks on lemma 1.

An important problem for the computational investigation of early vision is the initial carving up of the visual array into tentative objects and a background. This is important because it is a fundamental contention of the bottom up computational approach that there exist autonomous low level visual processes capable of providing a useful initial segregation of the visual world *independent* of higher level cognitive influences. For example, it is a primary goal of the primal sketch and 2-$\frac{1}{2}$D sketch, early visual representations proposed by Marr (1976) and Marr and Nishihara (1978), to make explicit exactly that information in a visual image which is required to build useful descriptions of the image in terms of objects and their relations. The processes proposed both to build and to operate upon these representations are invariably bottom up. If this endeavor fails, so too does much of the computational approach to vision. Therefore a high priority activity of computational research in vision is to provide convincing existence proofs (e.g., running computer programs) for this contention.

Visual motion seems a likely candidate base for tentative structuring of the visual array via autonomous processes. This has been suggested many times before. Ullman (1979, p. 76) proposes that a primitive motion correspondence process might be causally related to the child's acquisition of object constancy over changing views of an object. Marr and Ullman (1979) have suggested that retinal velocity fields may be used to segregate the visual world by exploiting the "principle of continuous flow". This principle states that "the velocity field of motion within the image of a rigid object varies continuously almost everywhere."

The results of lemma 1 suggest four further motion based segregation methods. The first two arise again from the fact that a rigid body can have but one angular velocity at any instant. Since lemma 1 provides methods to compute ω_3 and ω_1/ω_2 locally, it is possible to segregate the field into regions of constant ω_3 and constant ω_1/ω_2. In fact, the segregations obtained by the two methods should agree, providing the necessary redundancy to check for gross errors.

The third method is based on noting that the discriminant of equation (16) for ω_3 remains real over regions in the image which are the projections of smooth rigid surfaces.[9] Therefore points where ω_3 becomes complex indicate regions in the image where the rigidity constraint is violated (or where the surface has a discontinuity from the current viewpoint).

Finally, we can utilize constraints on tilt fields. For smooth rigid surfaces a "principle of continuous tilt" analogous to that proposed by Marr and Ullman for optical flow may be invoked to segregate the visual array. This principle states that "the tilt field within the image of a rigid object varies continuously almost everywhere."

These four segregation techniques are not isomorphic to Marr and Ullman's principle of continuous flow. The methods suggested here segregate the image into regions which are the projections of *rigid* objects. The principle of continuous flow cannot. Since it does not explicitly incorporate a rigidity constraint,[10] the principle of continuous flow cannot be used to distinguish regions of smooth

[9]This is easily proved by substituting from equations (12)-(15) into the appropriate terms of the discriminant of (16). Simplifying gives $(\omega_2 z_y + \omega_1 z_x)^2$ which is always greater than zero. Implicit in equations (12)-(15) is the rigidity assumption.

[10]The word rigid does appear in the statement of their principle, but it is equally true that "the velocity field of motion within the image of a *bending* object varies continuously almost everywhere."

flow in the image which arise from rigid objects from those which arise from bending or otherwise non-rigid substances. Consequently the segregations provided by the different methods are not identical but are useful for different purposes.

5.3 Lemma 2.

The surface slant, σ, and the remaining two components of the angular velocity, ω_1 and ω_2, are computable given the spatial derivatives of the orthographically projected acceleration field in addition to those of the velocity field.

Proof of Lemma 2. The acceleration field associated with a smooth rigid surface is found by taking the time derivative of equation (3). Indicating temporal derivatives by primes we have:

$$\mathbf{V}' = \Omega' \times \mathbf{R} + \Omega \times \mathbf{R}' + \mathbf{T}' \tag{19}$$

where

$$\mathbf{R}' = \Omega \times \mathbf{R} = \mathbf{V} \tag{20}$$

$$\Omega' = \omega_1' \mathbf{i} + \omega_2' \mathbf{j} + \omega_3' \mathbf{k} \tag{21}$$

If we take the first spatial derivatives of (19), simplify the results using the rigidity constraint of (7), and expand the indicated cross products as before, we obtain the four equations:

$$u_x' = \omega_2' z_x - \omega_2^2 - \omega_3^2 + \omega_3 \omega_1 z_x \tag{22}$$

$$u_y' = \omega_2' z_y - \omega_3' + \omega_1 \omega_2 + \omega_1 \omega_3 z_y \tag{23}$$

$$v_x' = \omega_3' - \omega_1' z_x + \omega_2 \omega_3 z_x + \omega_1 \omega_2 \tag{24}$$

$$v_y' = -\omega_1' z_y + \omega_2 \omega_3 z_y - \omega_1^2 - \omega_3^2 \tag{25}$$

Equations (12)–(15) and (22)–(25) relate eight quantities measurable in principle from the image, $(u_x, u_y, v_x, v_y, u_x', u_y', v_x', v_y')$, to the eight unknowns of interest: the local surface normal, (σ, τ), the three components of the angular velocity, $(\omega_1, \omega_2, \omega_3)$, and the three components of the angular acceleration, $(\omega_1', \omega_2', \omega_3')$. The simple fact that we have eight equations in eight unknowns does not necessarily imply that this system has but a finite number of solutions. To ascertain if there are a finite number of solutions we apply the inverse function theorem.[11] This theorem allows us to conclude

[11] For an informal discussion of the utility of the inverse function theorem, Bezout's theorem, and Sard's theorem for problems involving systems of nonlinear equations see Richards, Rubin, and Hoffman (1981).

that wherever the Jacobian of these equations is nonsingular the mapping defined by the equations is locally one to one and onto (ie, a local diffeomorphism). Consequently any roots at points where the Jacobian is nonsingular are isolated and not part of a continuum of solutions. The determinant of the Jacobian of (12)–(15) and (22)–(25) is:

$$\begin{vmatrix}
\omega_2 & 0 & 0 & z_x & 0 & 0 & 0 & 0 \\
-\omega_1 & 0 & -z_x & 0 & 1 & 0 & 0 & 0 \\
0 & \omega_2 & 0 & z_y & -1 & 0 & 0 & 0 \\
0 & -\omega_1 & -z_y & 0 & 0 & 0 & 0 & 0 \\
\omega_1\omega_3 + \omega_2' & 0 & \omega_3 z_x & -2\omega_2 & \omega_1 z_x - 2\omega_3 & 0 & z_x & 0 \\
0 & \omega_1\omega_3 + \omega_2' & \omega_3 z_y + \omega_2 & \omega_1 & \omega_1 z_y & 0 & z_y & -1 \\
\omega_2\omega_3 - \omega_1' & 0 & \omega_2 & \omega_3 z_x + \omega_1 & \omega_2 z_x & -z_x & 0 & 1 \\
0 & \omega_2\omega_3 - \omega_1' & -2\omega_1 & \omega_3 z_y & \omega_2 z_y - 2\omega_3 & -z_y & 0 & 0
\end{vmatrix}$$

This Jacobian has rank eight. Consequently the system of equations has but a finite set of solutions in general.[12] By Bezout's theorem[11] we know that the sum of the multiplicities of the solutions does not exceed the product of the degrees of the equations.

We have shown that there are but a finite number of solutions given the spatial derivatives of the velocity and acceleration fields *at one point*. In fact (12)–(15) and (22)–(25) can be solved uniquely (up to a reflection) for σ, ω_1, and ω_2 in terms of ω_3':

$$\sigma = \tan^{-1}\sqrt{\frac{\beta+\gamma}{\alpha}} \tag{26}$$

$$\omega_1 = \pm(v_x - \omega_3)\sqrt{\frac{\alpha}{\beta}} = \pm v_y\sqrt{\frac{\alpha}{\gamma}} \tag{27}$$

$$\omega_2 = \mp(u_y + \omega_3)\sqrt{\frac{\alpha}{\gamma}} = \mp u_x\sqrt{\frac{\alpha}{\beta}} \tag{28}$$

where

$$\alpha = (\omega_3' + u_y')(v_x' - \omega_3') - (u_x' + \omega_3^2)(\omega_3^2 + v_y') \tag{29}$$

$$\beta = (\omega_3 - v_x)^2(\omega_3^2 + u_x') + u_x(v_x' + u_y')(\omega_3 - v_x) + u_x^2(\omega_3^2 + v_y') \tag{30}$$

$$\gamma = (\omega_3^2 + v_y')(\omega_3 + u_y)^2 - v_y(u_y' + v_x')(\omega_3 + u_y) + v_y^2(\omega_3^2 + u_x') \tag{31}$$

This concludes the proof of lemma 2 and of the shape from motion proposition.

[12]Degenerate conditions can be found by determining when this determinant is zero.

5.4 Remarks on the Shape From Motion Proposition

It has been shown that the angular velocity and the surface normal at each visible point of a rigidly rotating regular surface are uniquely determined up to a reflection if one is given the orthographic projection and first spatial derivatives of the associated velocity and acceleration fields. This proposition and its proof are proposed as the basis for a computational theory of the human visual ability to perceive the shape of a smooth moving surface from its motion alone.

Some disclaimers are in order. First, only arguments for the sufficiency of this approach, not its necessity, have been suggested. Alternative computational theories are available, some of which were discussed earlier. It is a matter for empirical investigation to determine which, if any, of the current theories is to some extent psychologically real.

Two pieces of psychophysical evidence may be adduced to *suggest* the greater psychological reality of the present approach over previous ones. First are Ullman's (1979) experiments, mentioned before, which indicate that only orthographic, not perspective, information seems to be utilized by the visual system in recovering surface shapes. The second is that the resulting perceptual effect (illustrated in figure 1) is of a smooth surface as opposed to isolated points connected by invisible wires. This suggests greater psychological reality for an approach which builds a description whose primitives relate to surfaces.

A second disclaimer must be mentioned. The visual system may utilize additional generally valid constraints for the interpretation of surface shapes from motion. For example, shortcuts in computing the slant, σ, might be based on noting that σ must be 90 degrees at external occluding contours and must vary smoothly between them. Another potentially powerful constraint is that the tilt field must be locally orthogonal to the image of its occluding contour (for smooth surfaces). Thus further investigation of valid means to reduce the computational complexity of this approach is warranted before serious claims for its psychological reality can be sustained.

6. Computing Velocity Fields

To this point the analysis has assumed the velocity field and its first spatial derivatives are given.[13] Clearly this is not the case for a real observer. The real observer is presented with a temporally changing optic array. If a velocity field is required it must be *constructed* from the changing optic array.

The problem of computing a velocity field has remained nontrivial despite much recent research. One can show that the motion information available at any single point in a changing optic array is insufficient to uniquely determine the velocity field at that point. Consequently much of the research in the field of optical flow has been devoted to discovering valid means of integrating motion information from local neighborhoods to uniquely determine the flow at each point in the neighborhood.

A detailed analysis of the problem of determining optical flow is presented in Horn and Schunck (1980), which also includes a representative list of references on the topic.

[13]Actually only the first spatial derivatives have been used.

7. Summary

A computational analysis of the human visual ability to infer surface shapes entirely from their motion has been presented. The analysis proceeded in three main steps. First it was shown that surface tilt, τ, and the component of angular velocity orthogonal to the image plane, ω_3, may be derived from just the spatial derivatives of the velocity field (assuming orthographic projection). Then it was shown that surface slant, σ, and the two components of angular velocity lying parallel to the image plane, ω_1 and ω_2, are computable if the first spatial derivatives of the acceleration field are also available. Finally the problem of computing velocity fields from changing optic arrays was discussed briefly.

8. Acknowledgement

I thank A. Witkin, W. Richards, S. Ullman, and D. Marr for many valuable discussions. C. Papineau kindly drew the figures.

REFERENCES

Attneave, F., "Representation of physical space," *Coding processes in human memory*, Melton, A.W. and Martin, E. eds. New York: John Wiley (1972).

Do Carmo, M.P., *Differential Geometry of Curves and Surfaces*, Prentice-Hall, New Jersey, 1976.

Gibson, J.J., *The Perception of the Visual World*, Houghton Mifflin, Boston, 1950.

Gibson, J.J. & Gibson, E.J., "Continuous perspective transformations and the perception of rigid motion," *Journal of Experimental Psychology* **54**, 2 (1957), 129–138.

Green, B.F., "Figure coherence in the kinetic depth effect," *Journal of Experimental Psychology* **62**, 3 (1961), 272–282.

Grimson, W.E.L., "Implementation of a Theory of Stereo Vision," *MIT PhD Thesis* (1980).

Guillemin, V. & Pollack, A., *Differential Topology*, Prentice-Hall, Inc., New Jersey, 1974.

Hay, C.J., "Optical motions and space perception – an extension of Gibson's analysis," *Psychological Review* 73 (1966), 550–565.

Horn, B.K.P & Schunck, B., "Determining Optical Flow," *MIT AI Memo 572* (1980).

Johansson, G., "Perception of motion and changing form," *Scandinavian Journal of Psychology* **5** (1964), 181–208.

Johannson, G., "Visual Perception of Biological Motion and a Model for its Analysis," *Perception & Psychophysics* **14**, 2 (1973), 201–211.

Johannson, G., "Visual Motion Perception," *Scientific American* **232**, 6 (1975), 76–88.

Koenderink, J.J & van Doorn, A.J., "Local Structure of Movement Parallax of the Plane," *J. Opt. Soc. Am.* **66**, 7 (1976), 717–723.

Longuet-Higgins, H.C. & Prazdny, K., "The interpretation of a moving retinal image," *Proc. R. Soc. London B.* **208** (1980), 385–397.

Marr, D., "Early Processing of Visual Information," *Philosophical Transactions of the Royal Society of London* 275, 942 (1976), 483–534.

Marr, D. & Poggio, T., "From Understanding Computation to Understanding Neural Circuitry," *Neuroscience Research Program Bulletin* **15**, 3 (1977), 470–488.

Marr, D. & Nishihara, H.K., "Representation and Recognition of the Spatial Organization of Three-dimensional Shapes," *Proc. Roy. Soc. London B* **200** (1978), 269–294.

Marr, D. & Ullman, S., "Directional Selectivity and its Use in Early Visual Processing," *MIT AI Memo 524* (1979).

Nakayama, K. & Loomis, J.M., "Optical Velocity Patterns, Velocity-sensitive Neurons, and Space Perception: a Hypothesis," *Perception* **3** (1974), 63–80.

Richards, W.A., Rubin, J.M., and Hoffman, D.D., "Application of the Jacobian test and Bezout's theorem to problems in natural computation," *MIT AI Memo 614* (1981).

Stevens, K., "Surface Perception from Local Analysis of Texture and Contour," *MIT PhD Thesis (AI-TR-512)* (1980).

Ullman, S., *The Interpretation of Visual Motion,* MIT Press, Cambridge, 1979.

Wallach, H. & O'Connell, D.N., "The Kinetic Depth Effect," *Journal of Experimental Psychology* **45**, 4 (1953), 205–217.

Witkin, A.P., "Shape From Contour," *MIT PhD Thesis* (1980).