MASSACHUSETTS INSTITUTE OF TECHNOLOGY
ARTIFICIAL INTELLIGENCE LABORATORY

A.I. Memo No. 663 October, 1981

THE IMPLICIT CONSTRAINTS OF THE PRIMAL SKETCH

W.E.L. Grimson

ABSTRACT: Computational theories of structure-from-motion and stereo vision only specify the computation of three-dimensional surface information at points in the image at which the irradiance changes — for example, the *zero-crossings* of a $\nabla^2 G$ operator applied to the image. Yet, the visual perception is clearly of complete surfaces, and this perception is consistent for different observers. Since mathematically the class of surfaces which could pass through the known boundary points provided by the stereo system is infinite and contains widely varying surfaces, the visual system must incorporate some additional constraints besides the known points in order to compute the complete surface.

Using the image irradiance equation, we derive the *surface consistency constraint*, informally referred to as *no news is good news.* The constraint implies that the surface must agree with the information from stereo or motion correspondence, and *not* vary radically between these points. An explicit form of this surface consistency constraint is derived, by relating the probability of a zero-crossing in a region of the image to the variation in the local surface orientation of the surface, provided that the surface albedo and the illumination are roughly constant.

The surface consistency constraint can be used to derive an algorithm for reconstructing the surface that "best" fits the surface information provided by stereo or motion correspondence.

This report describes research done at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for the laboratory's artificial intelligence 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 in part by National Science Foundation Grant 79-23110MCS.

1. Introduction

Although our world has three spatial dimensions, the projection of light rays onto the retina presents our visual system with an image of the world that is inherently two-dimensional. We must use such images to physically interact with this three-dimensional world, even in situations new to us, or with objects unknown to us. That we easily do so implies that one of the functions of the human visual system is to reconstruct a three-dimensional representation of the world from its two-dimensional projection onto our eyes.

Methods that could be used to effect this three-dimensional reconstruction include stereo vision [e.g. Wheatstone, 1838; Helmholtz, 1925; Julesz, 1971] and structure from motion [for example, Miles, 1931; Wallach and O'Connell, 1953; Johannson, 1964]. Both of these methods may be considered as correspondence techniques, since they rely on establishing a correspondence between identical items in different images, and using the difference in projection of these items to determine surface shape.

Most of the current computational theories of these processes [Marr and Poggio, 1979; Grimson, 1980, 1981a; Mayhew and Frisby, 1981; Ullman, 1979; Longuet-Higgins and Prazdny, 1980] argue that the correspondence process cannot take place at all points in an image. Rather, the first stage of the correspondence process is to derive a symbolic description of points in the image at which the irradiance undergoes a significant change [Marr and Hildreth, 1980]. This symbolic representation (called the *primal sketch* [Marr, 1976; Marr and Hildreth, 1980]) forms the input to the second stage of the process in which the actual correspondence is computed. As a consequence of the form of the input, the correspondence process can compute explicit surface information only at scattered points in the image. Yet our perception is clearly of complete surfaces. For example, in Figure 1, a sparse random dot stereogram yields the vivid perception of a square floating in space above a background plane, rather than a collection of dots suspended in space. This suggests that some type of filling-in, or interpolation of surface information, is taking place in the visual system.

This poses an interesting problem, since mathematically the class of surfaces that could pass through the known boundary points provided by the stereo algorithm is infinite and contains widely varying surfaces. The implication of our perception of complete surfaces, which is consistent for different viewers, is that the visual system must incorporate some additional constraints, besides the known points, in order to compute the complete surface.

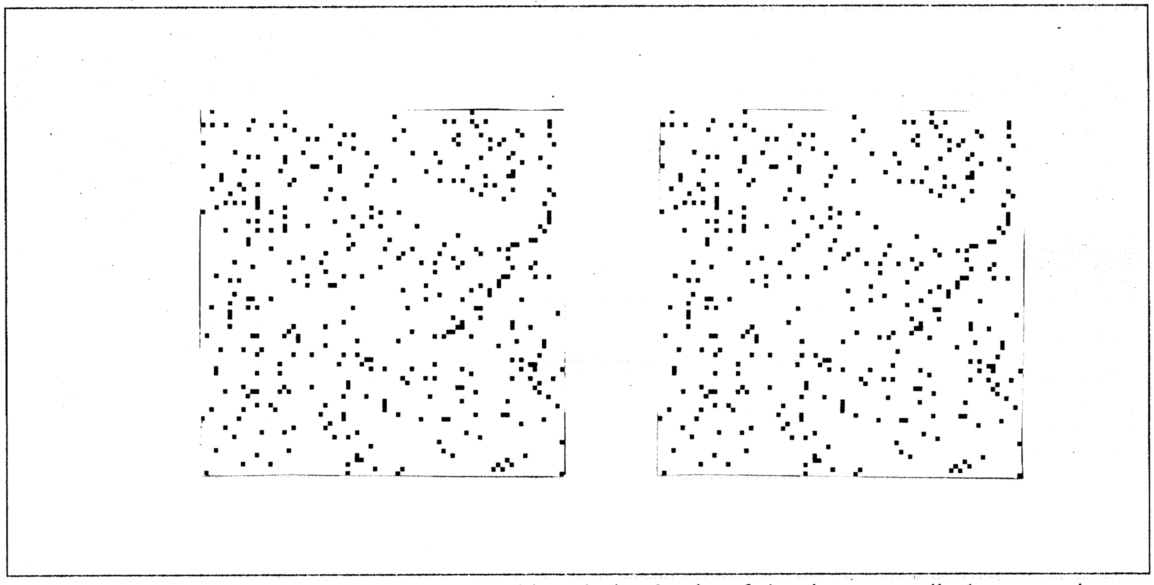

Figure 1. A Sparse Random Dot Pattern. Although the density of dots is very small, the perception obtained upon fusing this pattern is one of two disjoint planes, rather than dots isolated in depth.

In this paper, we will use the image irradiance equation [Horn, 1977] to derive the *surface consistency constraint*, informally referred to as *no news is good news*. The constraint implies that the surface must agree with the information from stereo or motion correspondence, and not vary radically between these points. We will derive an explicit form of this surface consistency constraint, by relating the probability of an abrupt irradiance change in a region of the image to the variation in the local surface orientation of the corresponding surface. We will then indicate how this constraint can be used to derive an algorithm for reconstructing the surface that "best" fits the surface information provided by stereo or motion correspondence [Grimson, 1981b, 1982].

2. The Computational Constraint

The input representation used by the stereo algorithm [Marr and Poggio, 1979; Grimson, 1981a, 1981b] consists of a symbolic representation of points at which the image irradiance undergoes a significant change. These points are identified by the zero-crossings of a Laplacian of a Gaussian ($\nabla^2 G$) operator applied to the image; that is, the image is convolved with a $\nabla^2 G$ operator and the points at which the resulting convolution changes sign are located [Marr and Hildreth, 1980; Hildreth, 1980]. As a consequence of the correspondence process, explicit three-dimensional information about the surface shape will be computed only at these zero-crossings.

Suppose one were to attempt to construct a complete surface description based only on the surface information known along the zero-crossings (where by complete, we mean that an explicit surface value is assigned to each point on a rectangular grid). An infinite number of surfaces would consistently fit the boundary conditions provided by these surface values. Yet there must be some way of deciding which surface, or at least which small family of surfaces, could give rise to the zero-crossing descriptions. This means that there must be some additional information available from the visual process which, when taken into account, will identify a class of nearly indistinguishable surfaces that represent the visible surfaces of a scene.

In order to determine what information is available from the visual system, one must first carefully consider the process by which the zero-crossing contours are generated. For instance, sudden changes in the reflectance of a surface, caused by surface scratches or texture markings, will give rise to zero-crossings in the convolved image [Marr and Hildreth, 1980; Hildreth, 1980]. Of more interest here is the fact that sudden or sharp changes in the shape of the surface will under most circumstances also give rise to zero-crossings. This can be used to constrain the possible shapes of surfaces that could produce particular surface values along the zero-crossing contours

We illustrate the basic argument with an example. Suppose one is given a closed zero-crossing contour, within which there are no other zero-crossings. An example would be a circular contour, along which the depth is constant. There are many surfaces which could fit this set of boundary conditions. One such surface is a flat disk. One could, however, also fit other smooth surfaces to this set of boundary conditions. For example, the highly convoluted surface formed by $\sin(x^2+y^2)^{\frac{1}{2}}$ would be consistent with the known disparity values (see Figure 2). Yet in principle, such a rapidly varying surface should give rise to other zero-crossings. This follows from the observation that if the surface orientation undergoes a periodic variation, then it is likely that the irradiance values will also undergo such a variation. Since the only zero-crossings lie at the borders of the object, this implies that the surface $\sin(x^2+y^2)^{\frac{1}{2}}$ is a highly improbable representative surface for this set of boundary conditions.

Hence, the hypothesis, which will be evaluated in the following sections, is that the set of zero-crossing contours contains implicit information about the surface as well as explicit information. If

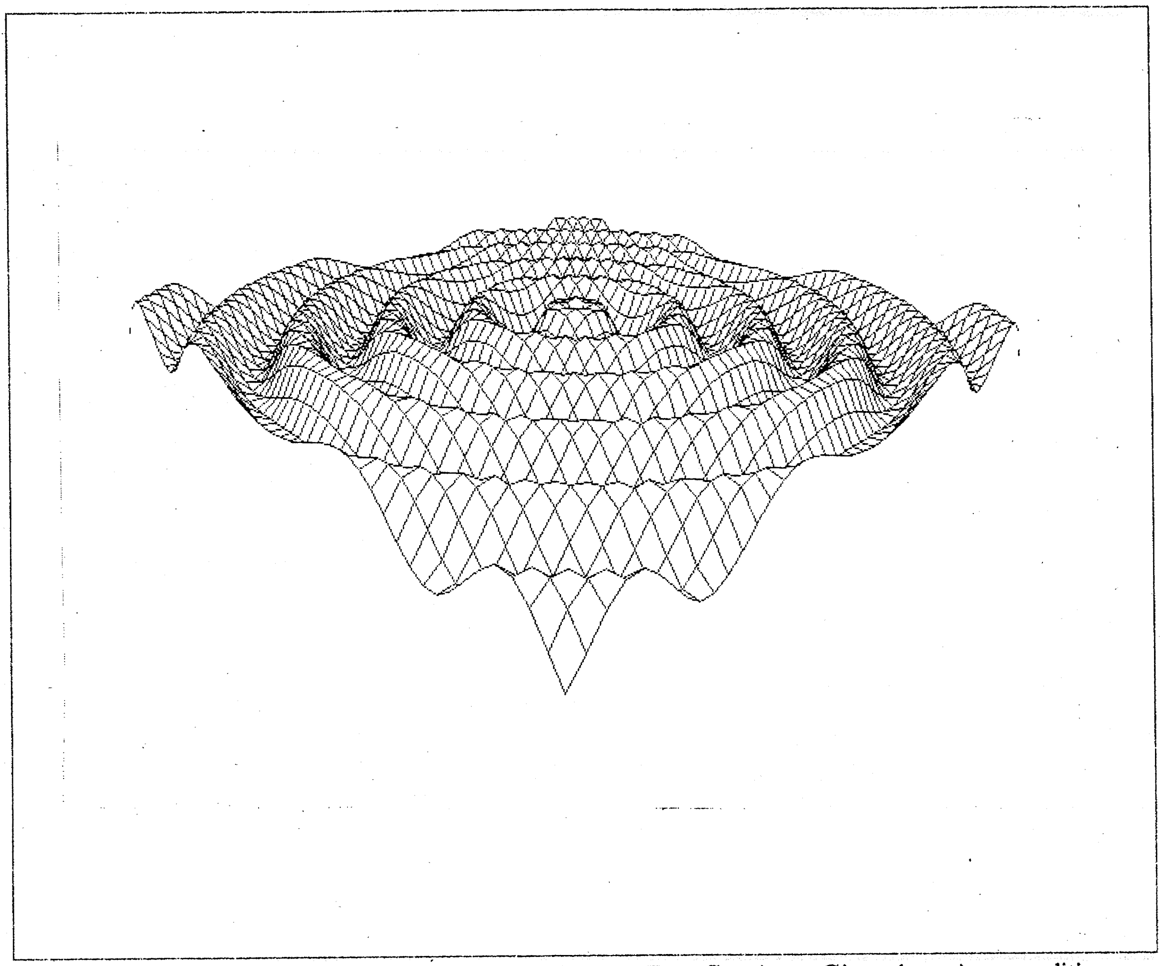

Figure 2. Possible Surfaces Fitting Depth Values at Zero-Crossings. Given boundary conditions of a circular zero-crossing contour, along which the depth is constant, there are many possible surfaces which could fit the known depth points. Two examples are a flat disk, and the highly convoluted surface formed by $\sin\left(x^2 + y^2\right)^{\frac{1}{2}}$, shown here. (From Grimson [1981b].)

one can determine a set of conditions on the surface shape that cause inflections in the irradiance values, then one may be able to determine a likely surface structure, given a set of boundary conditions along the zero-crossing contours.

2.1 No News is Good News

In general, any one of a multitude of widely varying surfaces could fit the boundary conditions imposed along the zero-crossings. Our intention is to show that to be completely consistent with the imaging process, such surfaces must meet both explicit conditions and implicit conditions. The explicit conditions are given by the depth values along the zero-crossing contours. The implicit conditions are that the surface must not impose any zero-crossing contours other than those which appear in the convolved image. This implicit condition leads to the *surface consistency constraint* [Grimson, 1981b], namely:

The absence of zero-crossings constrains the possible surface shapes.

Just as the presence of a zero-crossing tells us that some physical property is changing at a given location, the absence of a zero-crossing tells us the opposite, that no physical property is changing, and in particular that the surface topography is not changing in a radical manner. We informally refer to this constraint as *no news is good news*, since it says that in general the surface cannot change radically without informing us of this fact by means of zero-crossings. Note that in practice, the convolved image may not signal the entire extent of a discontinuity in the surface topography, but at least some portions of the discontinuity will be evidenced by a zero-crossing. In order to complete such a partial discontinuity, a subjective contour must be inserted between the known portions [Brady and Grimson, 1981].

In order to make explicit any constraints on the shape of the surface for locations in the image not associated with a zero-crossing, we will carefully examine the factors which influence irradiance and hence image intensity. (Note that while irradiance refers to the illuminant flux emitted from the surface of an object and intensity refers to the brightness recorded by a sensor, we will use the terms interchangably, since they are proportional [Horn and Sjoberg, 1979].) These factors will be expressed in the image irradiance equation, which describes the manner in which a particular irradiance is formed at a point in the image. Two goals are kept in mind. The first is to determine what surface conditions will cause a local change in irradiance, and the second is to combine this constraint with the input from the visual processes, such as stereo or structure-from-motion, in order to design an algorithm for interpolating surface information.

The basic result, corresponding to the intuitive argument given above, is that the probability of a zero-crossing occurring, in regions where the illumination is roughly constant and the surface material does not change, is a monotonic function of the variation in the orientation of the surface normal. This means that the probability of a zero-crossing increases as the variation in surface orientation increases. By inverting this argument, we will show that the best surface to fit the known data is that which minimizes the variation in surface orientation since this surface is most consistent with the zero-crossings in the convolved image.

3. Image Formation

In order to examine the process of zero-crossing formation, a review of the factors involved in the formation of an image will be presented. One factor is the geometry of the projection (either orthographic or perspective) from the scene to the image. The second factor, which we will review briefly below, is the manner in which image irradiance values at a point are formed. A detailed analysis of these processes has been undertaken by several investigators [Horn, 1970, 1975, 1977; Nicodemus, et al., 1977; Woodham, 1978; Horn and Sjoberg, 1979].

3.1 Grey-Level Formation

In this section, we will outline the factors involved in the creation of image irradiances in order to illustrate the relationship between changes in irradiance and zero-crossings in the convolved image, (most of the discussion is based on Horn and Sjoberg, [1979]). To do this, we need to determine what irradiance value will be associated with a particular image location.

The apparent "brightness" recorded by an imaging device is a measurement of *image irradiance* E, the radiant flux striking a unit area of the receptive field. The flux reaching a small portion of the receptive field will be exclusively a function of a corresponding small surface element on some object in the scene, provided the imaging system is properly focused and the lens is small relative to its distance from the object [Horn and Sjoberg, 1979]. We want to specify the factors that determine the image irradiance at a point in the image, as a function of the corresponding surface element.

In general, there are three factors governing the irradiance recorded at any position in the image:

(1) the amount of radiant flux striking a surface, a function of the distance of the surface from the source r, the orientation of the surface relative to the source, and the intensity I of the source itself;

(2) the percentage of incident flux reflected by the surface (as opposed to the percentage absorbed or transmitted), a function of the surface material and usually described by the albedo ρ; and

(3) the distribution of that reflected flux as a function of direction, usually described by a reflectance function.

We shall show that a sharp change in each of these factors will give rise to a zero-crossing in the convolved image. Further, in regions where both the incident illumination and absorption characteristics of the surface material are roughly constant, an equation relating the recorded irradiances to the shape of the surface will be derived. This image irradiance equation will be used in later sections to verify our surface consistency constraint.

Most surfaces have the property that the reflectance is not changed by rotating a surface element about its normal (exceptions are diffraction gratings, irridescent plumage and "tiger's eye"). Such surfaces are referred to as *isotropic*. If a surface possesses this property, only three angles are needed to determine reflectance, as shown in Figure 3. The angle between the local surface normal and the incident ray is called the incident angle and is denoted by i. The angle between the local surface normal and the emitted ray is called the view angle and is denoted by e. The angle between the incident and emitted rays is called the phase angle and is denoted by g. In the following, we shall assume that the surfaces are isotropic. This reduces the image irradiance to a function of I, r, i, e, g and ρ, where I is the intensity of the source, r is the distance of the source from the surface, i, e, and g are the angles specified above, and ρ is the albedo of the surface material.

We know that the incident flux density follows an inverse-square reduction as a function of distance. If two surfaces lie at very different depths, and all other factors are roughly equal, this inverse-square dependence will cause a noticeable difference in the irradiances associated with the different surfaces. If the projections of the two surfaces are adjacent in the image, the different irradiances will give rise to a zero-crossing in the convolved image. This is to be expected, since the Marr-Hildreth theory of edge detection was based on the requirement of detecting changes in the image corresponding to changes in some physical property of the surfaces. In this case, the changing physical property is object continuity.

Given that these situations will cause a zero-crossing, we can restrict our attention to regions of the image between the borders of the objects. We will further assume that the source is distant relative to the surfaces. This has two consequences. The first is that each surface receives roughly uniform illumination. This implies that the I/r^2 term is roughly constant and may be ignored in what follows.

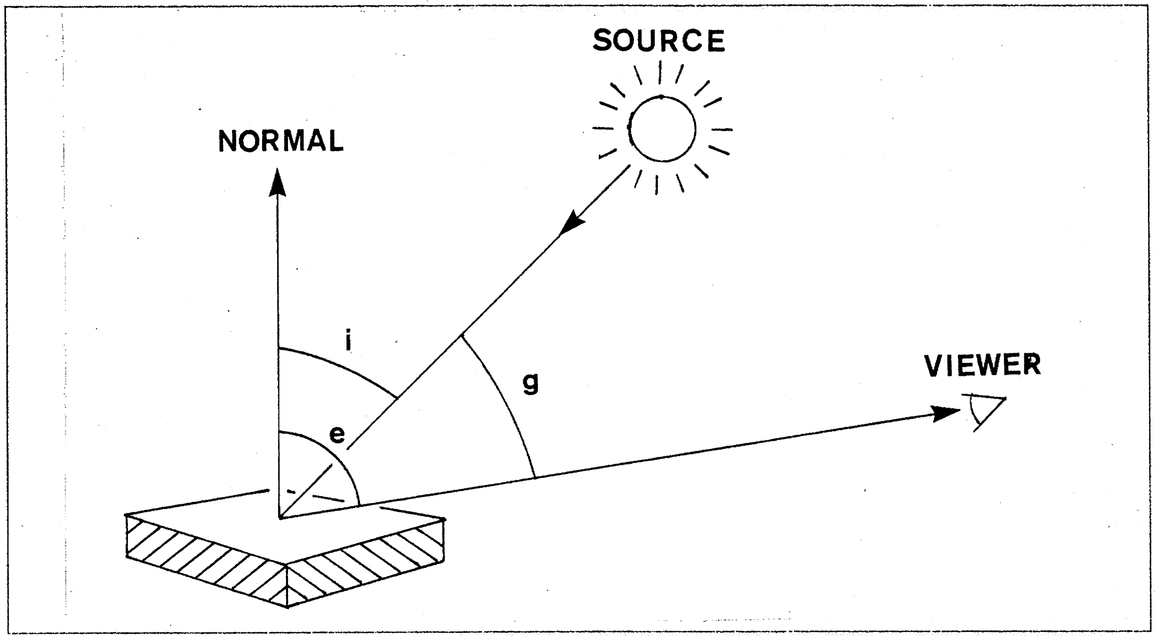

Figure 3. The Imaging Geometry. The incident angle i is the angle between the incident ray and surface normal. The view angle e is the angle between the emergent ray and the surface normal. The phase angle g is the angle between the incident and emergent rays. (Redrawn from Woodham, [1978]).

The second is that the angle between the viewer direction and the source direction (the phase angle g) will be roughly constant over each surface.

If the surface material changes, there will frequently be an associated change in the albedo. If all other factors are roughly equal, this will cause a change in the irradiances and there will be a corresponding zero-crossing in the convolved image. This is again expected from the Marr-Hildreth theory of edge detection: a changing physical property of the surface should correspond to a zero-crossing in the convolved image. In general we shall restrict our attention to regions of the surfaces between such material changes, so that the albedo ρ is roughly constant. In these regions, we can consider the image irradiances to be a function of i, e and ρ (ignoring scale constants).

If we assume that the source is distant enough and small enough to be treated as a point source, then we can simplify the geometry further by reducing the angles i and e to a specification of the local surface normal, in the following manner. Consider a Cartesian coordinate system with z-axis aligned with the line of sight, and x-axis some arbitrary direction in the plane normal to the line of sight. Suppose the surface is specified in the coordinate system by $z = f(x, y)$. Elementary calculus states that the vectors

$$\left\{1, 0, \frac{\partial z}{\partial x}\right\} \qquad \left\{0, 1, \frac{\partial z}{\partial y}\right\}$$

are tangent to the surface at a given point. The cross product

$$\left\{\frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}, -1\right\}$$

is normal to the tangent plane, and hence to the surface. Defining

$$p = \frac{\partial z}{\partial x} \qquad q = \frac{\partial z}{\partial y}$$

the normal vector becomes $\{p, q, -1\}$. The quantity (p, q) is usually referred to as the *gradient* and the space of all points (p, q) is referred to as *gradient space* [Huffman, 1971; Mackworth, 1973].

In this case, the angles i and e can be straightforwardly transformed into the normal components p and q [Horn, 1977]. Since we are considering the case of a distant point source, its direction relative to the surface can be represented by a direction vector $\{p_s, q_s, -1\}$. A straightforward calculation of the vector dot products shows that the angles i and e are given by:

$$\cos(i) = \frac{1 + pp_s + qq_s}{\sqrt{1 + p^2 + q^2}\sqrt{1 + p_s^2 + q_s^2}}$$

$$\cos(e) = \frac{1}{\sqrt{1 + p^2 + q^2}}.$$

Extended sources (sources whose size relative to their separation from the scene cannot be treated as a point source, for example, fluorescent lights in a normal room) can be frequently treated by superposition of point sources, or as being equivalent to an appropriately chosen point source [Brooks, 1978; Silver, 1980]. Thus, the irradiance reduces to a function of p, q and ρ.

Finally, if we assume that the objects are distant relative to the viewer, the image projection may be treated as orthographic. In this case, the coordinate system of the image and the coordinate system of the scene can be treated as identical so that we obtain the *image irradiance equation:*

$$E(x, y) = \rho(x, y) R\big(p(x, y),\ q(x, y)\big)$$

where $E(x, y)$ is the image irradiance recorded at a point in the image, ρ is the albedo associated with the surface intersecting the ray from (x, y) and $R(p, q)$ is the reflectance map [Horn, 1970, 1975, Horn and Sjoberg, 1979, Horn and Bachman, 1977]. Note that the reflectance map R differs from the reflectance function for a surface [Nicodemus, et al., 1977]. The reflectance function describes the amount of flux reflected in a particular direction, given an incident flux from a second given direction. The reflectance map combines the effects of the reflectance function in the case of uniform illumination and isotropic surface material with the illumination geometry and the image sensor's view point. In the case of constant albedo the image irradiance equation reduces to the partial differential equation [Horn, 1970, 1975; Horn and Sjoberg, 1979]:

$$E(x, y) = R\big(p(x, y),\ q(x, y)\big).$$

These two equations describe the manner in which an image irradiance (or grey-level) is obtained at a particular point in the image.

4. The Interpolation Theorem

We have seen that zero-crossings can arise from many factors. We shall restrict our attention to regions in which illumination is constant, albedo is roughly constant and surface material is isotropic. Our intention is to show that in such cases, if the surface topography changes radically, the image irradiances must also change radically. We will then be able to use the contrapositive statement — if the image irradiances do not change radically, then the surface topography also does not change. Note that if the albedo or the illumination is not constant over the region, we cannot make this statement. That is, there could be situations in which the surface topography does change radically without causing a corresponding change in the image irradiances, because one of the other factors, such as the albedo, is also changing in such a manner as to mask out the effects of the change in topography. These situations are fortunately very rare, since they require a precise meshing of the effects of the changing albedo with the changing surface topography and in general are highly dependent on the viewer position and illuminant direction. Thus, a small change in the viewer position will usually decouple the conflicting effects of albedo and surface shape, and the change in topography will give rise to a zero-crossing. (In other words, it is generally difficult to paint a radically curved surface so as to appear flat from all viewpoints. The opposite is not true, of course. One can easily paint a flat surface to appear curved. In other words, we prefer to see topographic changes, as evidenced by the informal observation that photographs with smoothly changing albedo are usually interpreted as having smoothly changing topography.)

The concern here is whether there can be zero-crossings corresponding to topographic rather than photometric changes in an object, for example, in the radial sine surface of Figure 2. Indeed, if the albedo is roughly constant, the form of the equations indicates that there may be topographic effects that could also cause sharp changes in irradiance. The surface conditions under which such a change in irradiance can occur are important, since the absence of zero-crossings in a region would then imply the absence of such surface conditions for that region. It is precisely these restrictions on surface shape which will allow us to determine a surface consistent with the depth values along the zero-crossing contours. The basic problem is, under what conditions does bending of the surface force an inflection in the irradiance array? This question will be answered in the following sections by considering specific cases.

There are two points to note, before beginning the mathematical details. The first concerns the role of the Gaussian filter in constructing the primal sketch representations upon which the stereo correspondence is computed. The Gaussian performs a smoothing of the image, thereby isolating irradiance changes at a particular scale. At the same time, it removes some of the noise problems which arise from using a discrete grid to represent the image. In what follows, we shall concentrate on the differential operator (∇^2) and its role in the creation of the zero-crossings. The effect of the Gaussian will not be considered in any detail. Hence, the surface reconstruction will not account for minor surface fluctuations on the scale of the grid resolution. This is not a major problem. (Equivalently, the problem may be considered as one of applying the Laplacian ∇^2 to the image $G * E$. The reconstruction will be based on this image information.)

The second point concerns the use of the Laplacian operator ∇^2 as opposed to directional derivatives. The mathematical arguments which follow are based on the consideration of zero-crossings of the convolution of a directional second-order differential operator with the image, $(\mathbf{v} \cdot \nabla) * E$. We have already seen that the Marr-Hildreth theory of edge detection is based on the use of zero-crossings of the image convolved with the Laplacian, $\nabla^2 * E$. This difference is not critical, for the following two reasons. Marr and Hildreth show that under some simple assumptions, the zero-crossings obtained by either operator are identical. In particular, the *condition of linear variation* [Marr and Hildreth, 1980] states that: *the irradiance variation near and parallel to a line of zero-*

crossings should locally be linear. If this condition is assumed to be true then it can be shown [Marr and Hildreth, 1980, Appendix A] that the zero-crossings obtained with the Laplacian are identical to those obtained using directional derivatives. Thus, in those situations for which the condition of linear variation is valid, no difference is obtained. In the case where the condition of linear variation does not hold, the difference in the zero-crossings of the operators lies in their position, not in the existence of such zero-crossings. For example, at a corner (which frequently carries strong visual information), both $\nabla^2 * E$ and $(\mathbf{v} \cdot \nabla) * E$ will give rise to zero-crossings. The only difference will be in the exact position of the zero-crossings. We shall see that such variation in position is of negligible consequence to the reconstruction process. Hence, it will be assumed that the arguments developed in the next section, based on the zero-crossings of a directional second derivative will also generally apply to the zero-crossings of the Laplacian.

Finally, in what follows, the functions ρ, R and f are assumed to have continuous second order partial derivatives. Throughout, it is assumed that the albedo effects can be ignored relative to the topographic ones, that is, the albedo factor may be considered roughly constant and will therefore not affect any study of derivatives. Without loss of generality, one may assume that $\rho = 1$ and ignore the albedo in what follows.

4.1 General Argument

The basic hypothesis is that in order for a surface to be consistent with a given set of zero-crossings, not only must it give rise to a zero-crossing in the convolved irradiances at those points, but it must also *not* give rise to zero-crossings anywhere else. Under most situations, this restriction would require that the surface not change in a radical manner between zero-crossing contours. (For example, the surface shown in Figure 2 is not consistent with the boundary conditions of a circular zero-crossing contour of uniform depth). It is difficult to prove this assertion in general, since the image formation equation includes terms dependent on the imaging process and on the light source geometry, as well as factors dependent on the photometric properties of the surface itself. However, under some fairly weak assumptions concerning the relative strengths of photometric and topographic changes, and the form of the reflectance function, we will prove the *surface consistency theorem*, describing the probability of a zero-crossing as a function of the shape of the surface. The importance of this theorem is that it leads to a method for measuring the probability of a particular surface being

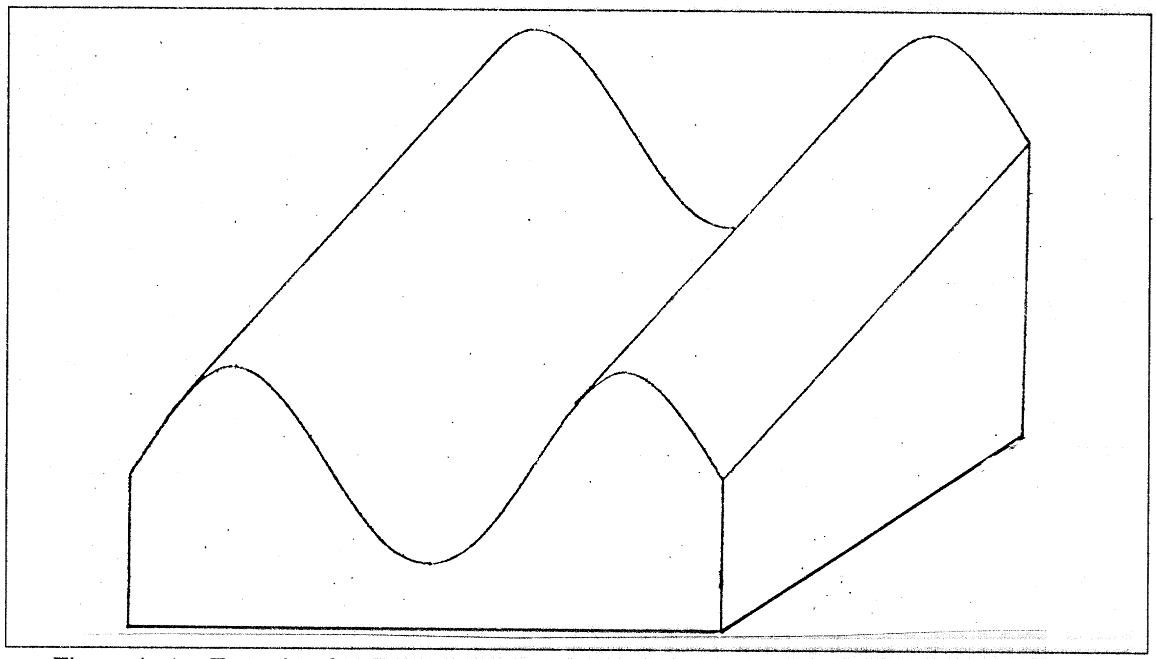

Figure 4. An Example of a Developable Surface. The component of the surface orientation in the y direction is constant for this region of the surface, so that the only variations in surface orientation take place in the x direction.

inconsistent with the zero-crossing information. This in turn suggests that it will be possible to derive a method for determining the best possible surface to fit the known information [Grimson, 1982].

4.2 A One-Dimensional Example

To illustrate the scope of the surface consistency theorem, we shall consider first the one-dimensional case of a developable surface, before proving the general theorem concerning arbitrary surfaces. Note that the Laplacian is orientation independent, so that without loss of generality, one may rotate the coordinate system of the image to suit our needs. One may assume that the surface has the form $f(x, y)$ such that $f_y(x, y) = q(x, y) = c$, in the local region under consideration. Hence, the partial derivatives of q vanish, as do any partial derivatives of p involving y. A sample surface is shown in Figure 4.

Suppose that a one-dimensional slice in the x direction of the surface contains at least two inflection points. Figure 5 indicates a sample surface and its derivatives. Since q is assumed constant, the derivatives of the image irradiance equation are given by:

$$\nabla^2 E = \left(\nabla^2 \rho\right) R + 2\rho_x R_p p_x + \rho\left(R_{pp} p_x^2 + R_p p_{xx}\right).$$

(Note that except where explicitly stated otherwise, we shall use a subscript to denote a partial derivative, so that, for example, $p_x = \partial p / \partial x$.) In this situation of a developable surface, we can prove the following result.

Theorem 0: Consider a portion of a second differentiable, developable surface oriented along the y axis such that $f_y(x, y) = q = c$, for some constant c. If the following conditions are true:

1. *The surface portion contains exactly two inflection points in the x direction, at x_1 and x_2,*
2. *At the points x_1 and x_2, normalized changes in albedo are dominated by normalized changes in reflectance,*

$$\left|\frac{\nabla^2 \rho}{\rho}\right| < \left|\frac{R_p p_{xx}}{R}\right|,$$

3. *The reflectance map R does not pass through an extremum in this region of the surface,*
4. *The reflectance map R is not constant over this region of the surface,*
5. *The albedo ρ is non-zero,*

then there exists a point $x_1 < x' < x_2$ such that $\nabla^2 E(x') = 0$. That is, there exists a zero-crossing here.

Proof: The signum function is defined by:

$$sgn(x) = \begin{cases} 1 & \text{if } x > 0 \\ -1 & \text{if } x < 0 \\ 0 & \text{if } x = 0 \end{cases}$$

From the derivatives of the image equation,

$$\nabla^2 E = (\nabla^2 \rho) R + 2\rho_x R_p p_x + \rho(R_{pp} p_x^2 + R_p p_{xx}).$$

At the inflection points x_1 and x_2, $p_x(x_i) = 0$. Hence, evaluation of the equation yields

$$\nabla^2 E(x_i) = \nabla^2 \rho(x_i) R(p(x_i)) + \rho(x_i) R_p(p(x_i)) p_{xx}(x_i) \quad \text{for } i = 1, 2.$$

Condition (2) implies that the albedo changes are negligible in this region, so that the first term may be ignored,

$$sgn(\nabla^2 E(x_i)) = sgn(\rho(x_i) R_p(p(x_i)) p_{xx}(x_i)).$$

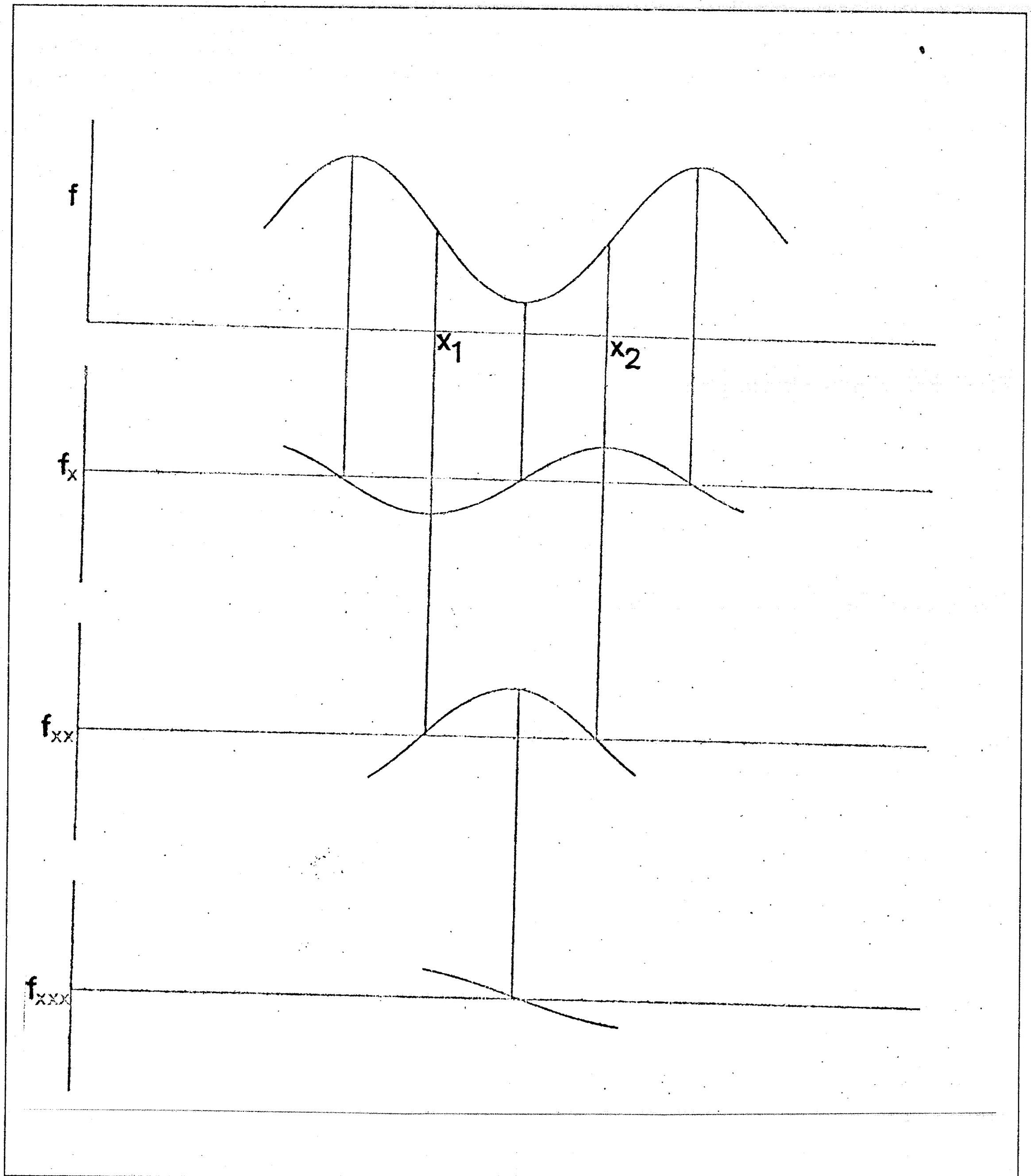

Figure 5. One-Dimensional Example. The top figure illustrates a slice of the surface, containing two inflection points. The second figure illustrates the first derivative of the surface function. The two inflection points of the surface correspond to extrema in the first derivative. The third figure illustrates the second derivative of the surface function. The two inflections in the surface correspond to zero-crossings in the second derivative. The bottom figure illustrates the third derivative of the surface function. Between the points corresponding to the two inflection points of the surface, the third derivative contains a zero-crossing. (From Grimson [1981b]).

Condition (5) implies that $sgn(\rho) = 1$. Observing that

$$sgn(xy) = sgn(x)\, sgn(y),$$

the sign of the convolved intensity function at the surface inflection points is given by

$$sgn(\nabla^2 E(x_i)) = sgn(R_p(p(x_i))p_{xx}(x_i)).$$

Condition (3) implies that R_p does not change sign in this region of the surface and hence

$$sgn(R_p(p(x_1))) = sgn(R_p(p(x_2))).$$

Note that $p_x = 0$ at x_1, x_2. The fact that there are exactly two inflections on the surface implies that $p_x \not\equiv 0$ over the neighborhood (x_1, x_2). Thus,

$$sgn(p_{xx}(x_1)) \neq sgn(p_{xx}(x_2)),$$

and thus

$$sgn(\nabla^2 E(x_1)) \neq sgn(\nabla^2 E(x_2)).$$

The second differentiability of the surface implies that there exists a point $x' \in (x_1, x_2)$ such that $\nabla^2 I(x') = 0$. ▮

The contrapositive of this theorem states something important about the possible surfaces which can fit the known depth information. Specifically, given the conditions of the theorem, if there is a set of known depth points to which a surface is to be fit, there cannot be two or more inflections in the surface between any two zero-crossing points. If there were, there would have to be an additional zero-crossing there as well.

Corollary 0.1: Suppose one is given a set of known depth points at a set of zero-crossings, along a developable surface. If the albedo ρ is non-zero and the reflectance map R is not constant for this region of the surface, then the surface cannot contain two or more inflection points between any pair of adjacent zero-crossings. ▮

There are many ways of extending this result. Some are indicated in the appendix. The most important extension is to consider general surfaces, rather than developable ones. We make this extension in the following sections.

4.3 The Analytic Argument

Take a planar slice of some surface f, normal to the image plane, along some direction (given for example by the angle α between the x axis and the direction of the slice). This is illustrated in Figure 6. At each point along the resulting curve C, one may associate a surface orientation, or gradient, given by the pair of partial derivatives, $p(x, y) = \partial f(x, y)/\partial x$, and $q(x, y) = \partial f(x, y)/\partial y$. Thus, one may construct a two-dimensional space spanned by a coordinate system with axes given by p and q, the gradient space introduced by Huffman [1971] and used by Mackworth [1973], and by Horn [1977] to relate the geometry of image projections to the radiometry of image formation. The curve obtained by the planar intersection of the surface transforms into a new, parametric curve in gradient space:

$$\mathbf{N}_f(t) = \{p(t), q(t), -1\}$$

In fact, this curve corresponds to the mapping of the normal to the surface, as one moves along the planar slice; the subscript f is used to indicate that this is the normal to the surface f.

Because the albedo is roughly constant, the image irradiances are determined by the reflectance map $R(p, q)$. Thus, the irradiances may be related to Horn's reflectance map, in which one considers the surface defined by $R(p, q)$ in gradient space. Hence, the curve C on the original surface will map onto a curve N_f in gradient space and this may be projected onto the surface $R(p, q)$ to obtain a new curve $C'(t)$. At each point t along the parametric curve, the corresponding irradiance is given by $R(p(t), q(t))$, (scaled by the constant ρ, which we will ignore without loss of generality).

We are interested in the conditions on the original surface and the reflectance surface that will cause a zero-crossing in the second directional derivative of the image irradiances. (By reflectance surface, we mean the reflectance map $R(p, q)$ considered as a surface in gradient space.) Let $\mathbf{v}$ be a vector in three-space with direction cosines $\cos\alpha_x$, $\cos\alpha_y$, and $\cos\alpha_z$ (the direction cosines refer to the cosine of the angle made by the vector and the coordinate axes of the space). Thus, a unit vector in the direction of $\mathbf{v}$ is denoted by

$$\mathbf{v} = \cos\alpha_x\,\mathbf{i} + \cos\alpha_y\,\mathbf{j} + \cos\alpha_z\,\mathbf{k}$$

where $\mathbf{i}, \mathbf{j}, \mathbf{k}$ are unit vectors along the coordinate axes. The directional derivative of $\Phi(\mathbf{r})$, a function of a vector $\mathbf{r}$, is the rate of change of Φ with distance s along the direction $\mathbf{v}$ and is denoted by

$$(\mathbf{v} \cdot \nabla)\Phi = \cos\alpha_x \frac{\partial\Phi}{\partial x} + \cos\alpha_y \frac{\partial\Phi}{\partial y} + \cos\alpha_z \frac{\partial\Phi}{\partial z}.$$

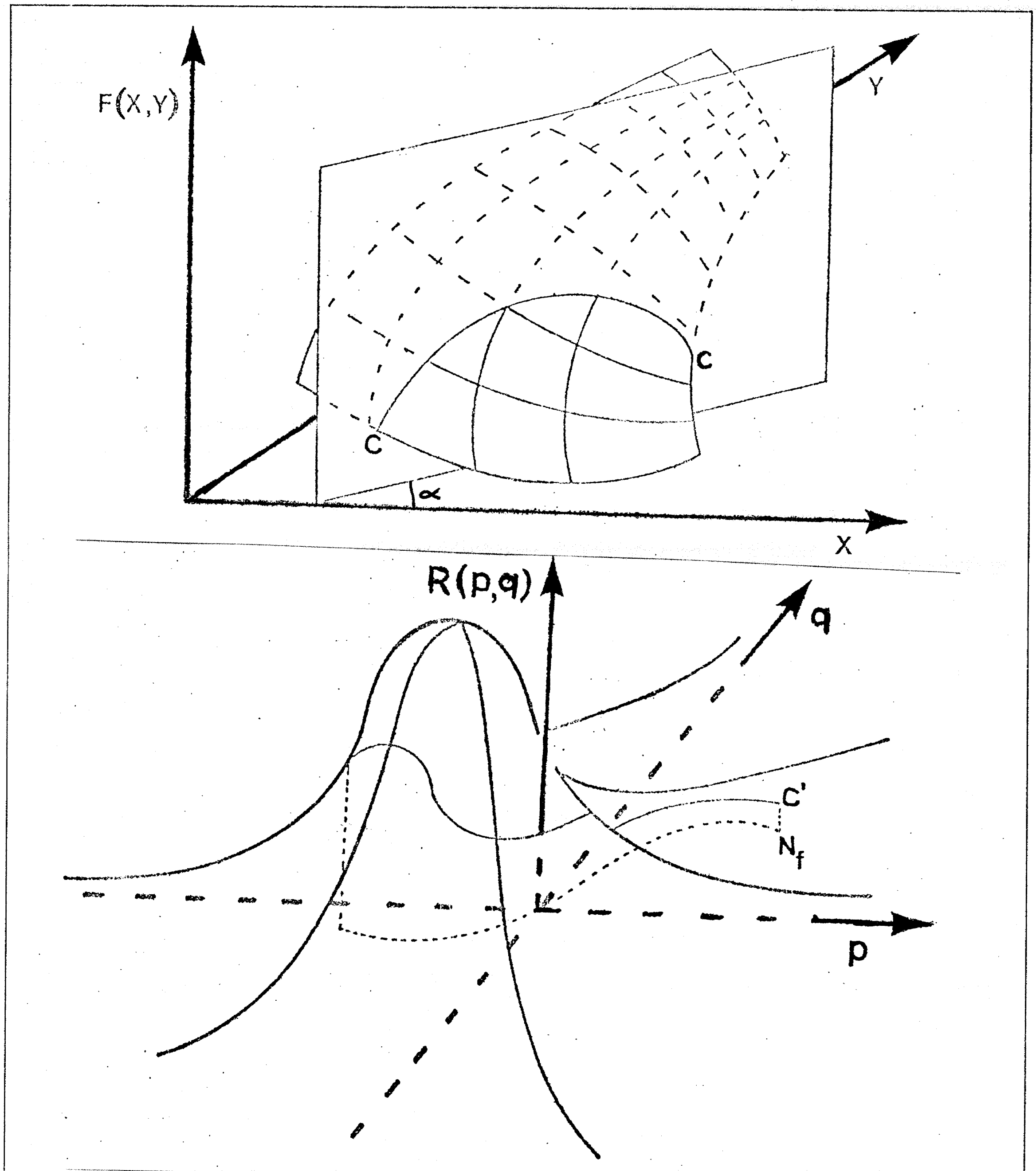

Figure 6. The Generation of Image Irradiance. The curve C is generated by taking the intersection of a plane normal to the image plane with the surface. The direction of the plane is given by the angle α it makes with the x-axis. By taking the surface normal at each point, this curve C can be mapped into gradient space, resulting in a parametric curve N_f. Furthermore, this curve can be projected onto the reflectance map R (which defines a surface on p-q space) to obtain a parametric curve C' along which the irradiance is given by the corresponding value of R. (From Grimson [1981b]).

If the direction vector is constrained to lie in the x-y plane, then $\cos\alpha_z = 0$ and $\cos\alpha_y = \sin\alpha_x$.

Now consider a planar slice normal to the image along the direction $\mathbf{v}$ and construct the parametric normal curve $\mathbf{N}_f$ in $p - q$ space. Then the directional derivative of irradiance along this slice is given by

$$(\mathbf{v}\cdot\nabla)E = R_p((\mathbf{v}\cdot\nabla)p) + R_q((\mathbf{v}\cdot\nabla)q).$$

This can be rewritten as

$$(\mathbf{v}\cdot\nabla)E = \nabla R\cdot(\mathbf{v}\cdot\nabla)\mathbf{N}_f$$

showing that the directional derivative along the slice is given by the projection of the gradient of the reflectance surface R onto the tangent of the parametric p-q curve $\mathbf{N}_f$. Now, if there are two points t_1, t_2 such that this dot product vanishes at those points, then Rolle's theorem, and the assumption that the surface f is twice continuously differentiable would imply the existence of a zero-crossing in the second directional derivative at some point between t_1 and t_2. Thus, to evaluate the probability of a zero-crossing in some region of the image, we need to consider the probability of the vectors ∇R and $(\mathbf{v}\cdot\nabla)\mathbf{N}_f$ being orthogonal.

There are two cases to consider. The first is when the gradient of R vanishes at some point, corresponding to an extremum in the reflectance surface. If we assume that over all possible viewpoints, ∇R is uniformly distributed, then the probability of $\nabla R = 0$ for some point $(p(t), q(t))$ on the parametric curve $\mathbf{N}_f(t)$ is given by the arclength of the curve

$$\int\left[(p_t)^2 + (q_t)^2\right]^{\frac{1}{2}} dt,$$

where the subscript t indicates partial differentiation with respect to the parameter t.

The second case is when the gradient of R is non-zero and the two vectors are non-trivially orthogonal. Consider first the one-dimensional case in which two inflections in a developable surface imply the existence of a zero-crossing (as shown in Theorem 0). Since in this case $q = c$, the parametric curve $\mathbf{N}_f(t)$ in p-q space is parallel to the p-axis, and reverses direction at each of the inflection points. At these points, the curvature of the parametric curve is essentially infinite, and the tangent can be considered to sweep through all possible directions. As a consequence, the tangent must be orthogonal to ∇R for one of those directions, and hence $(\mathbf{v}\cdot\nabla)E$ is zero at each point of infinite curvature. Thus, Rolle's theorem guarantees a zero-crossing along the slice of the image.

Now suppose that $q \neq c$ but changes slowly. In this case, the points of infinite curvature become points of high curvature. Here, the tangent to the parametric curve sweeps through a large range of directions. If we assume that the second order partial derivatives of the reflectance surface R are small, then over a small region of p-q space, ∇R is nearly constant. Hence, in the region of the points of high curvature, the direction of ∇R is nearly constant, while the direction of the tangent to N_f sweeps through a large range of directions. Thus, at points of high curvature, the probability of $(\mathbf{v} \cdot \nabla)E = 0$ is high. If, on the other hand, the parametric curve is locally a straight line, the tangent is constant and the probability of the directional derivative vanishing is small. Thus, along a slice of the image, the probability that $(\mathbf{v} \cdot \nabla)N_f$ and ∇R are non-trivially orthogonal, under the assumption that R_{pp}, R_{pq} and R_{qq} are small, is monotonically related to the integral of square curvature

$$\int \frac{(p_t q_{tt} - q_t p_{tt})^2}{(p_t^2 + q_t^2)^{\frac{5}{2}}},$$

where again the subscript t indicates partial differentiation with respect to the parameter t.

Thus, we can combine the above discussion into the following result.

Theorem 1 (Surface Consistency Theorem): Consider some region of a surface

$$\left\{ f(x, y): \quad (x, y) \in \mathfrak{R} \subset E^2 \right\}.$$

If the reflectance surface R is isotropic, its normal ∇R is uniformly distributed, and the second partials R_{pp}, R_{pq}, R_{qq} are small, then the probability of a zero-crossing in $(\mathbf{v} \cdot \nabla)^2 E$ for some direction **v**, *with associated directional angle α, is a monotonic function of*

$$\int \int_{\mathfrak{R}} \int_t \left[(p_t)^2 + (q_t)^2 \right]^{\frac{1}{2}}$$

and of

$$\int \int_{\mathfrak{R}} \int_t \frac{(p_t q_{tt} - q_t p_{tt})^2}{(p_t^2 + q_t^2)^{\frac{5}{2}}}$$

where the integral with respect to t is taken along the parametric curve obtained by slicing the image in some direction, and the integral with respect to $\mathfrak{R}$ is taken over all such slices of the image region $\mathfrak{R}$. ∎

This theorem has the immediate corollary of specifying the "best" surface to fit through a set of known points.

Corollary 1.1: The surface f most consistent with a set of known points provided by the visual system (e.g. from stereo) is that which is least likely to introduce zero-crossings not present in the primal sketch. By the Surface Consistency Theorem, this surface minimizes arclength of the parametric normal to the surface along each direction and curvature of the parametric normal along each direction, subject to passing through the known points.∎

Finally, we would like to obtain a simple expression for this surface. Note that the curvature of a curve at a point is simple the rate of change of its tangent at that point. Thus, the curve with minimum integral square curvature is one with a tangent as nearly constant as possible, relative to the constraints imposed by the known points. Next, to minimize arclenth, we require the curve of constant tangent to have as small a tangent as possible. This can be done by find the curve which minimzies

$$\int_t p_t^2 + q_t^2 \, dt$$

which is equivalent to

$$\int (p_x \cos\alpha + p_y \sin\alpha)^2 + (q_x \cos\alpha + q_y \sin\alpha)^2 \, dt.$$

Integrating this form over all directions α, and using the fact that $p_y = q_x$ (since f is assumed to be twice continuously differentiable), we obtain

$$f_{xx}^2 + 2f_{xy}^2 + f_{yy}^2.$$

As a consequence, we obtain the following result.

Corollary 1.2: The surface f which "best" fits a set of known data in a region $\mathfrak{R}$ *of the image is that which passes through the known points and minimizes*

$$\int\int_{\mathfrak{R}} f_{xx}^2 + 2f_{xy}^2 + f_{yy}^2 \, dx dy.$$

It is interesting to note that this expression for "best" surface interpolation as the surface which minimizes the integral of quadratic variation has been derived elsewhere from an analysis of fitting this elastic plates to the known points [Grimson, 1981b, 1982].

Three assumptions were made in the statement of the surface consistency theorem. One was that the reflectance surface R is isotropic, a second, that its normal ∇R is uniformly distributed, and a third, that the second order partial derivatives of R are small. What do these assumptions imply about the imaging situation? The assumption of an isotropic reflectance function is minor. The assumption of a uniform distribution of the reflectance normal is stronger. Although a stronger version of the surface consistency theorem can be proved by weakening this assumption, we can also justify the assumption by an appeal to the "general position" principle of viewer geometry and illuminant geometry. The notion of general position requires that all properties of the observed image irradiances be roughly independent of the specific viewer position or illuminant position. In other words, a small movement of the viewer or the illuminant should not grossly affect the properties of the image irradiances. In terms of the reflectance surface, a slight alteration of the viewer position or illuminant position will generally result in a translation (and possibly a scaling) of the reflectance surface R relative to the p-q coordinate system. Within the proof of the theorem, the assumption of uniform distribution is used to argue that at any point in the image, the reflectance normal ∇R is equally likely to be found in any direction. By applying the principle of general position, we can loosely argue that the illuminant is equally likely to be in any of a range of positions and that the viewer is equally likely to be in any of a range of positions. As a consequence, the reflectance normal corresponding to a point in the image is equally likely to be found in any direction, and hence is uniformly distributed. The assumption of small R_{pp}, R_{pq} and R_{qq} indicates that the theorem may not be valid for highly specular surfaces, but the assumption is reasonable for surfaces that are generally matte in nature.

4.4 The Importance of the Surface Consistency Theorem

We began this paper by discussing the problem of creating complete surface representations from sparse data such as that provided by stereo, and argued that additional constraints were needed to effect such a surface reconstruction. The above corollary suggests that one such constraint has been identified. In particular, given a class of surfaces that pass through a set of known points, the surface which is most consistent with the imaging information is that which is least likely to introduce additional zero-crossings. By the surface consistency theorem, we see that this surface is defined as the surface that minimizes the functional of quadratic variation, or at least minimizes some measure which is monotonically related to this functional. In [Grimson, 1982], we combine this fact with additional constraints, in order to develop a computational theory of visual surface interpolation.

5. Examples

To illustrate these results, we show how inflections in the surface can cause zero-crossings in the convolved image, for different light source positions. Figures 7, 8, 9 and 10 show examples of a one-dimensional slice of a developable surface, with a Lambertian reflectance function (a perfectly diffuse surface which reflects flux uniformly in all directions), and the irradiance values obtained for different positions of a point light source. Figure 7 indicates the sample surfaces and the rough positions of the light sources for the different examples. Figures 8, 9 and 10 indicate sample surfaces and the corresponding irradiance profiles for different positions of the light source, as indicated in Figure 7. The positions in the irradiance function which would give rise to a zero-crossing in the convolved image are indicated.

6. Summary

In this paper, we have seen the need for interpolating the results of the stereo algorithm to create complete surface descriptions. Determining which surface to fit to the known stereo data is at first sight difficult, since any one of a widely varying family of surfaces could be used. We saw, however, that most of these surfaces can be discarded as possible interpolations, since they contain rapid fluctuations in surface orientation and these fluctuations should give rise to additional zero-crossings, not contained in the stereo data. Such zero-crossings would, of course, be inconsistent with the imaging data, and hence such a surface is considered to be unacceptable as a possible interpolation of the known data. Thus, we proved that the *surface consistency constraint* restricts the problem:

The absence of zero-crossings constrains the possible surface shapes

To make this constraint precise, we outlined the physics of image formation and outlined a derivation of Horn's image irradiance equation. We then showed that this expression accounts for the zero-crossings of the Marr-Hildreth edge detection algorithm in the case of surface boundaries, surface markings, texture changes, and so forth. The equations further suggest that even in regions where the physical factors causing the previous types of zero-crossings are constant, there may still be zero-crossings, due to a change in surface shape. We used this fact to prove the *surface consistency theorem* which relates the probability of a zero-crossing in the second derivative of an image to the

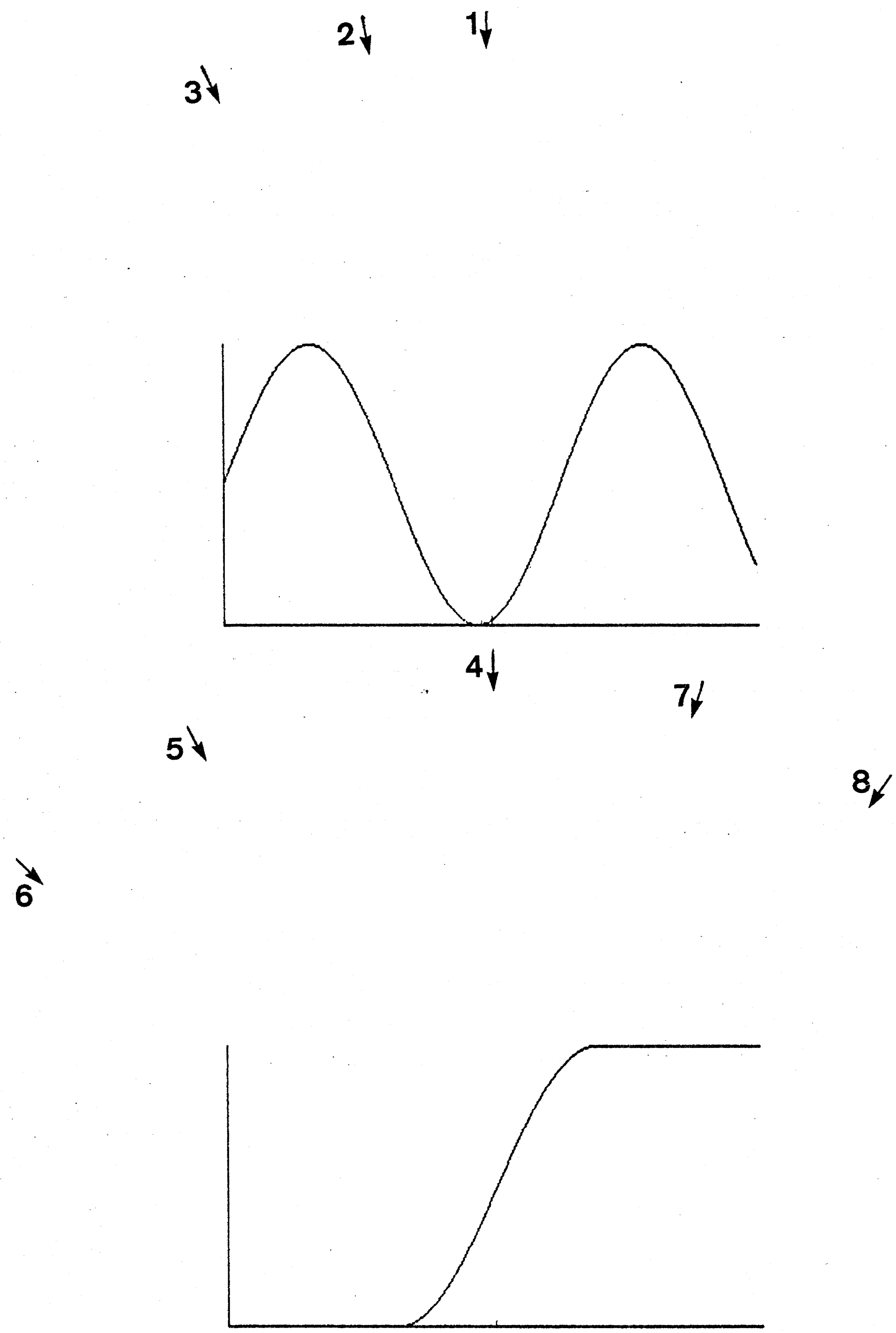

Figure 7. Examples of One-Dimensional Surfaces. The top figure shows one surface and the arrows indicate the rough orientations of the light source. The numbers refer to the irradiance profiles in Figure 8. The bottom figure shows a second surface, with a set of rough orientations of the light source. The numbers refer to the profiles of Figures 9 and 10. (From Grimson [1981b]).

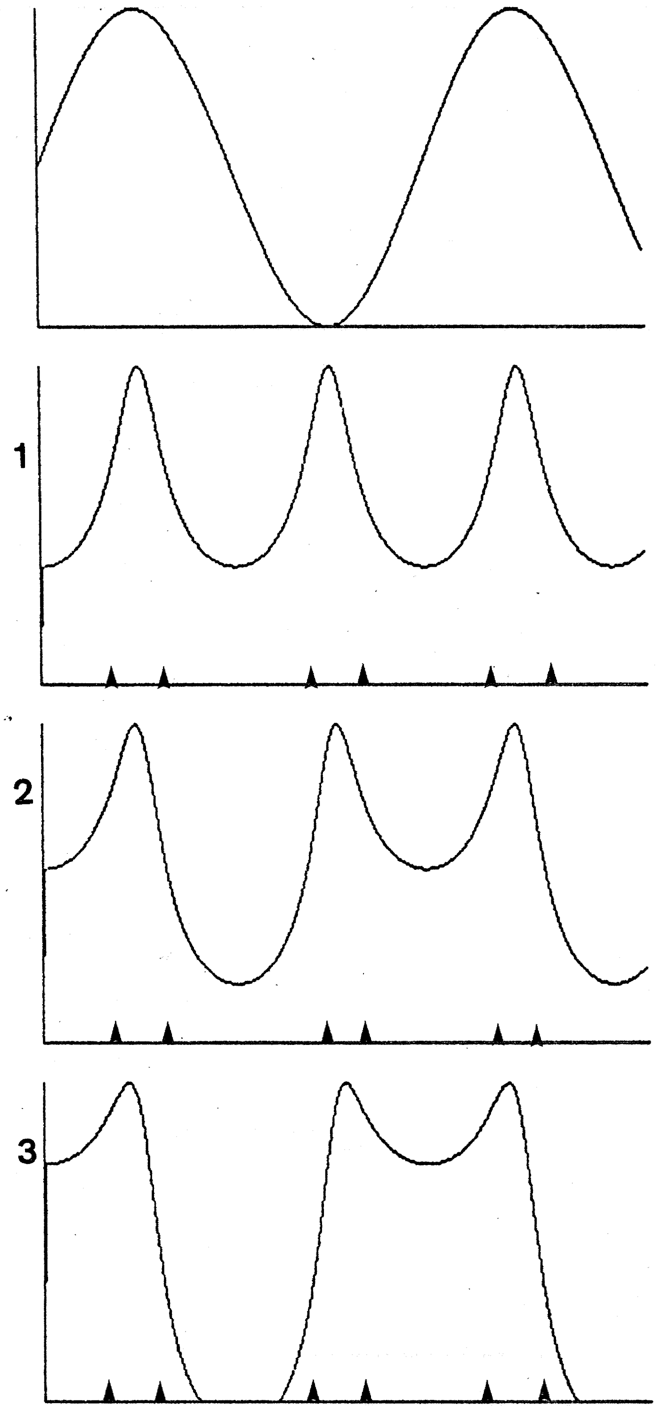

Figure 8. Examples of a Surface with Two Inflections. The top figure shows a slice of a surface. The bottom three figures indicate irradiance profiles for different positions of the light source. Note that in all cases, there are six irradiance inflections. In case 3, the irradiances also undergo a self-shadowing, where the irradiance value is zero. The positions in the irradiance function which would give rise to a zero-crossing in the convolved image are indicated. (From Grimson [1981b]).

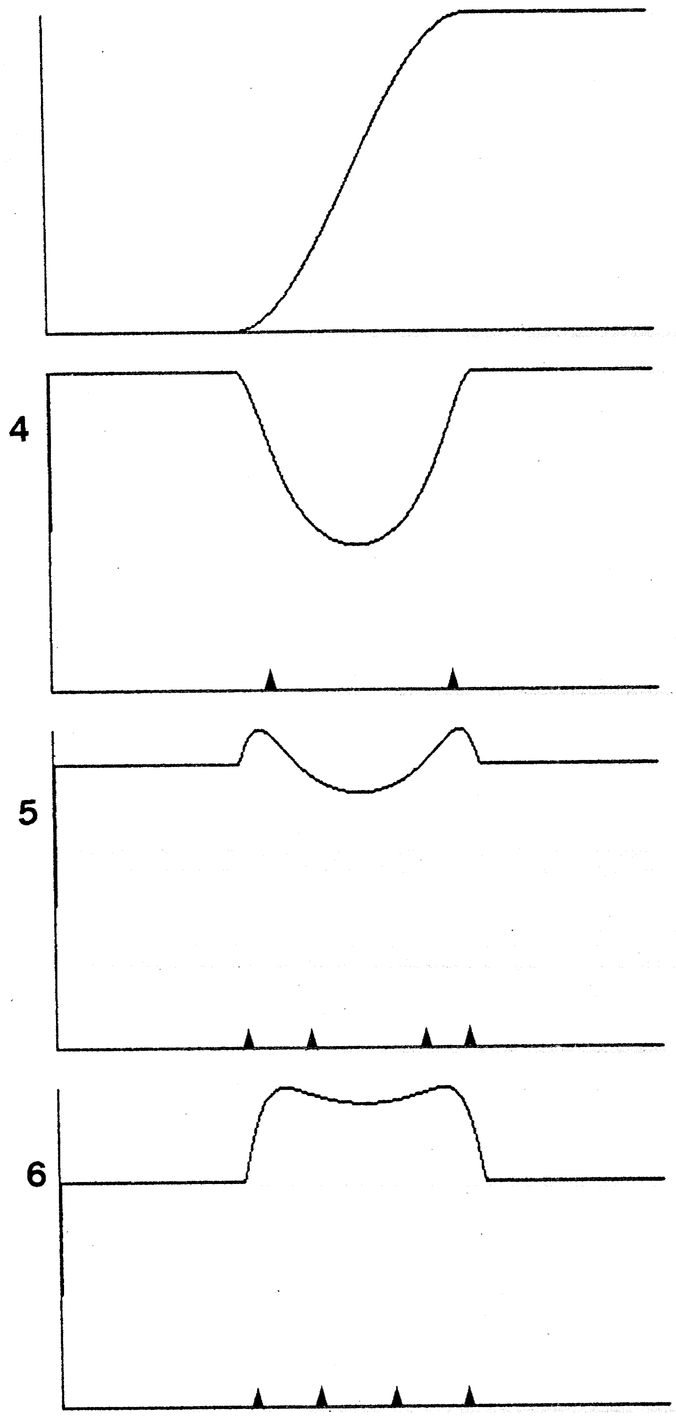

Figure 9. Examples of a Surface with One Inflection. The top figure shows a slice of a surface. The bottom three figures indicate irradiance profiles for different positions of the light source. The positions in the irradiance function which would give rise to a zero-crossing in the convolved image are indicated. (From Grimson [1981b]).

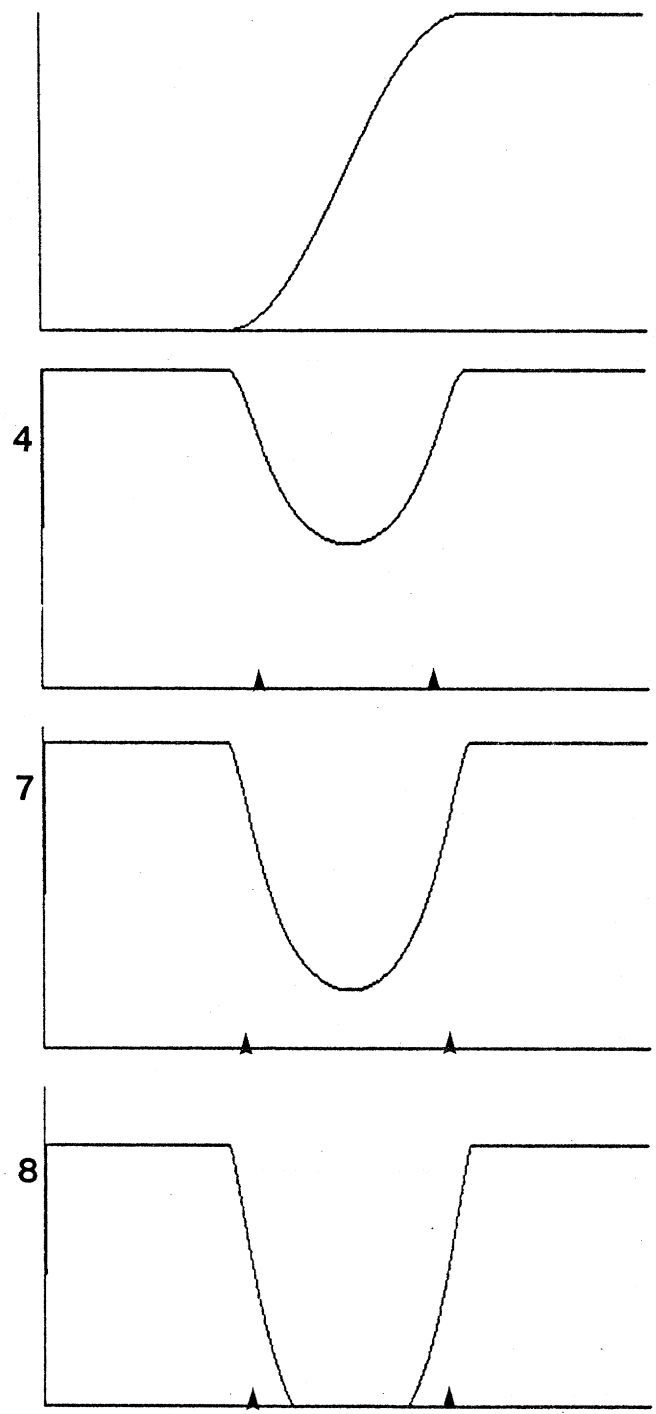

Figure 10. Examples of a Surface with One Inflection. The top figure shows a slice of a surface. The bottom three figures indicate irradiance profiles for different positions of the light source. Note that in case 8, the irradiance profile undergoes a self-shadowing. The positions in the irradiance function which would give rise to a zero-crossing in the convolved image are indicated. (From Grimson [1981b]).

variation in the original surface. The surface consistency constraint can be used to formulate a computational theory of surface interpolation and a specific algorithm for computing the interpolated surface [Grimson, 1982].

7. Appendix

Theorem 1 proved a relationship between any general two-dimensional surface and the probability of a zero-crossing occuring in the corresponding convolved intensities. Because the theorem dealt with any surface, the relationship was somewhat weak. In this appendix, a set of alternate theorems are given which apply in cases where the surface can be considered roughly cylindrical (or developable). The proofs of these theorems are included for completeness.

Theorem 2: Consider a portion of a second differentiable, developable surface oriented along the y axis such that $f_y(x, y) = q = c$, for some constant c. If the following conditions are true:

1. The surface portion contains exactly two inflection points in the x direction, at x_1 and x_2,

2. At the points x_1 and x_2, normalized changes in albedo are dominated by normalized changes in reflectance,

$$\left|\frac{\nabla^2 \rho}{\rho}\right| < \left|\frac{R_p p_{xx}}{R}\right|,$$

3. The reflectance map R does not pass through an extremum in this region of the surface,

4. The reflectance map R is not constant over this region of the surface,

5. The albedo ρ is non-zero,

then there exists a point $x_1 < x' < x_2$ such that $\nabla^2 E(x') = 0$.

Proof: The signum function is defined by:

$$sgn(x) = \begin{cases} 1 & \text{if } x > 0 \\ -1 & \text{if } x < 0 \\ 0 & \text{if } x = 0 \end{cases}$$

From the derivatives of the image equation,

$$\nabla^2 E = (\nabla^2 \rho)R + 2\rho_x R_p p_x + \rho(R_{pp} p_x^2 + R_p p_{xx}).$$

At the inflection points x_1 and x_2, $p_x(x_i) = 0$. Hence, evaluation of the equation yields

$$\nabla^2 E(x_i) = \nabla^2 \rho(x_i) R(p(x_i)) + \rho(x_i) R_p(p(x_i)) p_{xx}(x_i) \quad \text{for } i = 1, 2$$

Condition (2) implies that the albedo changes are negligible in this region, so that the first term may be ignored,

$$sgn(\nabla^2 E(x_i)) = sgn(\rho(x_i) R_p(p(x_i)) p_{xx}(x_i)).$$

Condition (5) implies that $sgn(\rho) = 1$. Observing that

$$sgn(xy) = sgn(x)\, sgn(y),$$

the sign of the convolved intensity function at the surface inflection points is given by

$$sgn(\nabla^2 E(x_i)) = sgn(R_p(p(x_i)) p_{xx}(x_i)).$$

Condition (3) implies that R_p does not change sign in this region of the surface and hence

$$sgn(R_p(p(x_1))) = sgn(R_p(p(x_2))).$$

Note that $p_x = 0$ at x_1, x_2. The fact that there are exactly two inflections on the surface implies that $p_x \not\equiv 0$ over the neighbourhood (x_1, x_2). Thus,

$$sgn(p_{xx}(x_1)) \neq sgn(p_{xx}(x_2)),$$

and thus

$$sgn(\nabla^2 E(x_1)) \neq sgn(\nabla^2 E(x_2)).$$

The second differentiability of the surface implies that there exists a point $x' \in (x_1, x_2)$ such that $\nabla^2 E(x') = 0$. ∎

Corollary 2.1: Suppose one is given a set of known depth points at a set of zero-crossings, along a developable surface. If the albedo ρ is non-zero and the reflectance map R is not constant for this region of the surface, then the surface cannot contain two or more inflection points between any pair of adjacent zero-crossings. ∎

Theorem 3: Consider a portion of a second differentiable developable surface oriented along the y axis such that $f_y(x, y) = c$, for some constant c. If the following conditions are true:

1. *The surface contains exactly one inflection point in the x direction at x_1, and the reflectivity R achieves an extremum at the point x_2, $x_1 \neq x_2$,*
2. *At the point x_1 the normalized changes in albedo are dominated by normalized changes in reflectance of the form,*

$$\left|\frac{\nabla^2 \rho}{\rho}\right| < \left|\frac{R_p p_{xx}}{R}\right|,$$

and at the point x_2 the normalized changes in albedo are dominated by normalized changes in reflectance of the form,

$$\left|\frac{\nabla^2 \rho}{\rho}\right| < \left|\frac{R_{pp} p_x^2}{R}\right|,$$

3. *The reflectance map R is not constant,*
4. *The albedo ρ is non-zero,*

then there exists a point $x_1 < x' < x_2$ such that $\nabla^2 E(x') = 0$.

Proof: As in the proof of the previous theorem, at the point x_1,

$$sgn(\nabla^2 E(x_1)) = sgn(R_p(p(x_1))p_{xx}(x_1)).$$

At the point x_2, $R_p = 0$ so that

$$\nabla^2 E(x_2) = \nabla^2 \rho(x_2) R(p(x_2)) + \rho(x_2) R_{pp}(p(x_2)) p_x^2(x_2).$$

Condition (2) then implies that at this point, the normalized albedo changes are dominated by the normalized reflectance changes,

$$sgn(\nabla^2 E(x_2)) = sgn(\rho(x_2)) sgn(R_{pp}(p(x_2))) sgn(p_x^2(x_2)).$$

Condition (4) implies that $sgn(\rho) = 1$, so that

$$\begin{aligned} sgn(\nabla^2 E(x_2)) &= sgn(R_{pp}(p(x_2))) sgn(p_x^2(x_2)) \\ &= sgn(R_{pp}(p(x_2))) \end{aligned}$$

There are two subcases. In the first subcase, $p_{xx}(x_1) > 0$. Since there is only one inflection point in the surface, this implies that $p(x_1) < p(x_2)$. Then $R_{pp}(p(x_2)) < 0$ implies that

$$R_p(p(x_1)) > R_p(p(x_2)) = 0.$$

Conversely, $R_{pp}(p(x_2)) > 0$ implies that

$$R_p(p(x_1)) < R_p(p(x_2)) = 0.$$

In either case,

$$sgn(R_{pp}(p(x_2))) \neq sgn(R_p(p(x_1))p_{xx}(x_1)).$$

In the second subcase, suppose that $p_{xx}(x_1) < 0$. This implies that $p(x_1) > p(x_2)$. Then $R_{pp}(p(x_2)) < 0$ implies that

$$R_p(p(x_1)) < R_p(p(x_2)) = 0.$$

Conversely, $R_{pp}(p(x_2)) > 0$ implies that

$$R_p(p(x_1)) > R_p(p(x_2)) = 0.$$

In either case,

$$sgn(R_{pp}(p(x_2))) \neq sgn(R_p(p(x_1))p_{xx}(x_1)).$$

Thus, we see that

$$sgn(\nabla^2 E(x_1)) \neq sgn(\nabla^2 E(x_2))$$

and as before the second differentiability of the surface implies that there exists a point $x' \in (x_1, x_2)$ such that $\nabla^2 E(x') = 0$. ▮

Corollary 3.1: Suppose one is given a set of known depth points at a set of zero-crossings, along a developable surface. If the albedo ρ is non-zero and the reflectance map R is not constant for this region of the surface, then if the reflectance map R passes through an extremum, the surface cannot contain any inflection points between any pair of zero-crossings. ▮

Theorem 4: Consider a second differentiable developable surface oriented along the y axis such that $f_y(x, y) = c$, for some constant c. If the following conditions are true:

1. *At the point x_1, the surface becomes self-shadowing, that is $R(x_1) = 0$,*
2. *The reflectivity R achieves an extremum at the point x_2, $x_1 \neq x_2$,*
3. *At the point x_2 normalized changes in albedo are dominated by normalized changes in reflectance,*

$$\left|\frac{\nabla^2 \rho}{\rho}\right| < \left|\frac{R_{pp} p_x^2}{R}\right|,$$

4. *The reflectance map R is not constant over this region,*
5. *The albedo ρ is non-zero,*

then there exists a point $x_1 < x' < x_2$ such that $\nabla^2 E(x') = 0$.

Proof: The fact that the surface becomes self-shadowing implies that there is a region of the surface, beginning at x_1, such that R is constantly zero. The fact that there is an extremum in R for some other point implies that the intensity function must be concave down in the region of the extremum and concave up in the region of self-shadowing. There must be an inflection point in between and hence there must be a point x' such that $\nabla^2 E(x') = 0$. ∎

Theorem 5: Consider a second differentiable developable surface oriented along the y axis such that $f_y(x, y) = c$, for some constant c. If the following conditions are true:

1. *At the point x_1, the reflectivity R achieves an inflection point,*
2. *There exist points $x_0 < x_1 < x_2$ such that reflectance changes dominate albedo changes,*

$$\left|\frac{\nabla^2 \rho}{\rho} + \frac{R_p}{R}\left(2\frac{\rho_x}{\rho}p_x + p_{xx}\right)\right| < \left|\frac{R_{pp}}{R}p_x^2\right|,$$

3. *The first derivative of the surface, p, is monotonic in this region, (i.e. it does not achieve an extremum),*
4. *The reflectance map R is not constant over this region,*
5. *The albedo ρ is non-zero,*

then there exists a point $x_1 < x' < x_2$ such that $\nabla^2 E(x') = 0$.

Proof: The proof is very similar to the previous ones, except that in this case, by condition (2),

$$sgn(\nabla^2 E) = sgn(\rho) sgn(R_{pp}) sgn(p_x^2).$$

at the points x_0, x_2. Then condition (4) implies that

$$sgn(\nabla^2 E) = sgn(R_{pp}).$$

Condition (3) implies that

$$p(x_0) < p(x_1) < p(x_2)$$

or

$$p(x_2) < p(x_1) < p(x_0)$$

In either case, condition (1) then implies that

$$sgn(R_{pp}(p(x_0))) \neq sgn(R_{pp}(p(x_2))).$$

Hence,

$$sgn(\nabla^2 E(x_0)) \neq sgn(\nabla^2 E(x_2)),$$

and as before, the second differentiability of the surface implies that there exists a point $x' \in (x_0, x_2)$ such that $\nabla^2 E(x') = 0$. ∎

8. Acknowledgements

The author wishes to express his gratitude to David Marr, Tommy Poggio, Shimon Ullman, Berthold Horn, Mike Brady, Whitman Richards, Tomas Lozano-Perez, Marilyn Matz and Ellen Hildreth for many useful comments and discussions.

This report describes research done at the Artificial Intelligence Laboratory of the Massachusetts Institute of Technology. Support for the laboratory's artificial intelligence research is provided in part by the Advanced Research Projects Agency of the Department of Defense under Office of Naval Research constract N00014-75-C-0643 and in part by National Science Foundation Grant 79-23110MCS.

9. References

Brady, M. and Grimson, W.E.L. The perception of subjective surfaces, MIT AI Lab, Memo 582, 1981.

Brooks, M.J. Investigating the effects of planar light sources, Dept. of Computer Science, Essex University, Colchester, England, CSM 22, 1978.

Grimson, W.E.L. Computing shape using a theory of human stereo vision, Ph.D. Thesis, Department of Mathematics, Massachusetts Institute of Technology, 1980.

Grimson, W.E.L. "A computer implementation of a theory of human stereo vision," *Phil. Trans. Royal Society of London, B.* 292 (1981a), 217-253.

Grimson, W.E.L. *From Images to Surfaces: A computational study of the human early visual system* , MIT Press, Cambridge, Mass., 1981b.

Grimson, W.E.L. "A computational theory of visual surface interpolation," *Phil. Trans. Royal Soc. London, B* (1982), to appear.

Helmholtz, H. *Physiological Optics, Vol. 3* , Optical Society of America, New York, 1925.

Hildreth, E.C. Implementation of a theory of edge detection, S.M. Thesis, Department of Computer Science and Electrical Engineering, Massachusetts Institute of Technology, 1980.

Horn, B.K.P. Shape from shading: a method for obtaining the shape of a smooth spaque object from one view, MIT Project MAC Technical Report, MAC TR-79, 1970.

Horn, B.K.P. "Obtaining shape from shading information," *The Psychology of Computer Vision, P.H. Winston (ed), McGraw-Hill* (1975), 115-155.

Horn, B.K.P. "Understanding image irradiances," *Artificial Intelligence* 8 (1977), 201-231.

Horn, B.K.P. and Bachman, B.L. Using synthetic images to register real images with surface models, MIT Artificial Intelligence Laboratory, Memo 437, 1977.

Horn, B.K.P. and Sjoberg, R.W. "Calculating the reflectance map," *Applied Optics* **18** (1979), 1770-1779.

Huffman, D.A. "Impossible objects as nonsense sentences," *Machine Intelligence 6, B. Meltzer and D. Michie (eds), Edinburgh University Press* (1971), 295-393.

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

Julesz, B. *Foundations of Cyclopean Vision* , University of Chicago Press, Chicago, 1971.

Longuet-Higgins, H.C. and Prazdny, K. "The interpretation of a moving retinal image," *Proc. Roy. Soc. Lond., B* 208 (1980), 358-397.

Mackworth, A.K. "Interpreting pictures of polyhedral scenes," *Artificial Intelligence* 4, 2 (1973), 121-137.

Mayhew, J.E.W. and Frisby, J.P. "Psychophysical and computational studies towards a theory of human stereopsis," *Artificial Intelligence (Special Issue on Computer Vision)* 17 (1981).

Marr, D. "Early processing of visual information," *Philosophical Transactions of the Royal Society of London* 275, 942 (1976), 483-534.

Marr, D. and Hildreth, E.C. "Theory of edge detection," *Proc. R. Soc. Lond. B* 207 (1980), 187-217.

Marr, D. and Poggio, T. "A theory of human stereo vision," *Proc. R. Soc. Lond. B* 204 (1979), 301-328.

Miles, W.R. "Movement interpretations of the silhouette of a revolving fan," *Am. J. Psych.* 43 (1931), 392-505.

Nicodemus, F.E., Richmond, J.C., Hsia, J.J., Ginsberg, I.W. and Limperis, T. Geometrical considerations and nomenclature for reflectance, National Bureau of Standards, NBS Monograph 160, 1977.

Silver, W. Determining shape and reflectance using multiple images, S.M. Thesis, Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 1980.

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

Wallach, H. and O'Connell, D.N. "The kinetic depth effect," *J. Exp. Psych.* 52, 5 (1953), 571=578.

Wheatstone, C. "Contributions to the physiology of vision. Part I. On some remarkable, and hitherto unobserved, phenomena of binocular vision," *Phil. Trans. R. Soc. Lond.* 128 (1838), 371-394.

Woodham, R.J. Reflectance map techniques for analyzing surface defects in metal castings, MIT AI Lab, TR 457, 1978.