ON SHAPE FROM SHADING METHODS

Some Theoretical Considerations

Alfred M. Bruckstein

EE Report No. 524

Department of Electrical Engineering

Technion, I.I.T.

32000, Haifa, Israel

June, 1985

ABSTRACT

We present two new methods for the recovery, from shading information, of a bivariate function $H(x,y)$ that describes a "nice", almost everywhere differentiable height profile. The given image or shading data is assumed to be a result of Lambertian diffuse reflection of light from the surface. This implies that, if the scene is uniformly illuminated from above, the shading yields information on the cosine of the angle between the vertical and the surface normal at each point. We shall assume that the height along a closed curve is known a-priori. Given this and the shading information in the plane, the shape from shading problem is to determine all height profiles consistent with the data.

One of the new shape-from-shading method that we discuss is based on a recursive way of determining the equal-height or level contours of the surface starting at a given level curve. The other surface recovery algorithm receives shading information over a pixel array and the exact height over a digitized contour and based on this information fits planar patches over each pixel to locally best account for both the data and the surface continuity constraint.

1. Introduction

Shape from shading problems have a long and interesting history. The first researchers to address the problem of determining shape from shading information were apparently those concerned with the photometric analysis of the lunar topography (see [1] and the references therein). With the development of computers and the trend towards endowing them with as much "intelligence" as possible, researchers naturally turned to attempts of mechanically reproducing the visual perception process. Several approaches to the analysis of three-dimensional scenes via their images or shading maps were taken. Many stressed the importance of a huge prior-information data-base in scene analysis and adopted AI-type approaches [2],[3],[4]. Other dealt with the recovery of depth from stereo images [5]. It is clear that shading information plays, along with stereo vision, a crucial role in the depth perception process. The theoretical question of how much depth information can be obtained from a single view of a scene from shading alone, thus arises quite naturally. Algorithms for determining the shape, or depth-profile from a single image produced by realistic shading rules were considered by B.K.P. Horn [1]. His pioneering work on the shape from shading problem was done in the early 70's, at M.I.T.'s Artificial Intelligence Laboratory. He and his group at M.I.T. continued to produce interesting and valuable contributions to this and related fields of research to this date (see e.g. [6], [7]). The more recent work on shape determination from a single image concentrated on the importance of singular points, surface models and occluding boundaries in providing initialization for a basic algorithm devised by Horn, that recovers height on a data-directed curve in the (x,y) plane called a characteristic strip [1], [8]. Also some iterative, relaxation-type techniques were invented, relying on surface smoothness constraints [7], [8].

In this paper we discuss a basic shape from shading problem, stressing the importance of considering the behavior of equal-height-contours. We show that when one such contour is available we can devise a simple algorithm that reconstructs all the equal-height-curves of the surface of interest in a well defined region and also clearly displays the inherent ambiguities of the given problem. We then introduce a recursive surface fitting algorithm that uses only the continuity constraint and works on a digitized picture, i.e. on a pixel-array, again under the assumption that the height is a-priori known on a "digitized" contour.

2. The Problem, Some Geometry And Basic Concepts

Suppose we are given a continuous function of two variables (x,y), $H(x,y)$, describing a surface in three dimensions, as follows

$$z = H(x,y). \tag{2.1}$$

The shaded image of $H(x,y)$ is defined as a light intensity map, $A(x,y)$, so that the value $A(x,y)$ depends on the surface properties, its orientation at (x,y) and the illumination. The shape from shading problem that we address is to recover the function $H(x,y)$ over some region $\mathbb{D}$, from the image $A(x,y)$ over that region and possible some further information, e.g. the values of $H(x,y)$ over some continuous curve in $\mathbb{D}$.

The function $A(x,y)$ is defined via a shading rule. In the case of a surface with Lambertian diffuse reflection properties and uniform illumination $A(x,y)$ is simply the cosine of the angle $\alpha(x,y)$, between the surface normal at (x,y) and the direction from which the light falls on the scene. For simplicity we shall always assume that the illumination is uniform and falls on the surface vertically from above, i.e. from $z = +\infty$. It is customary to define the shading rule via a so-called *reflectivity function*, that characterizes the surface properties and provides an explicit connection between $A(x,y)$ and the surface orientation. A thorough discussion of surface reflectivity properties and various types of reflectivity functions can be found in [6]. Define the directional derivatives of the height profile $H(x,y)$ along the x and y directions as

$$\begin{cases} p(x,y) = \dfrac{\partial}{\partial x}H(x,y) \\ q(x,y) = \dfrac{\partial}{\partial y}H(x,y) \end{cases} \tag{2.2}$$

The surface normal at (x,y) is clearly perpendicular to the plane determined by the vectors $[1,0,p]$ and $[0,1,q]$ therefore it is along the direction of their vector product $[-p,-q,1]$. The normal vector at (x,y) is thus

$$\overline{N}(x,y) = \frac{1}{\sqrt{1+p^2+q^2}}[-p,\ -q,\ 1] \tag{2.3}$$

and the cosine of the angle between $\overline{N}(x,y)$ and the vertical direction $[001]$ is

$$\cos\alpha(x,y) = \frac{1}{\sqrt{1+p^2+q^2}} \tag{2.4}$$

In the Lambertian case with light falling from above we, shall have the shading rule

$$A(x,y) = \frac{1}{\sqrt{1+p^2+q^2}} = R_L(p,q) \tag{2.5}$$

Note that the reflectivity function R is defined on the (p,q) plane - called the "gradient space". A general (not necessarily Lambrctian) shading rule is defined via

$$A(x,y) = R(p(x,y),q(x,y)) \tag{2.6}$$

Consider a plane in three dimensional space defined by a surface normal $\overline{N}$, defined as in (2.3) and the distance from the origin d. The plane equation is

$$z = px + qy + d\sqrt{1+p^2+q^2} = px + qy + z_0 \tag{2.7}$$

In the sequel we shall need another way of parametrizing a planar patch in terms of the intensity $A(x,y)$ it produces in an image. We may, of course parametrize the plane in terms of z_0, $R(p,q)$ and any other function $\tilde{R}(p,q)$ that uniquely determines, together with R, the gradients p and q.

In the Lambertian example we have $R(p,q) = R(p^2+q^2)$, i.e. the shading determines p^2+q^2 at each point (x,y). In order to determine p and q separately we need an angle parameter $\varepsilon\in[0,2\pi]$, since clearly, given p^2+q^2, we shall have

$$\begin{cases} p = (p^2 + q^2)^{1/2}\cos\varepsilon \\ q = (p^2 + q^2)^{1/2}\sin\varepsilon \end{cases} \tag{2.8}$$

Therefore, an equivalent parametization of a planar surface patch is via

$$\{z_0, \quad \cos\alpha = \frac{1}{\sqrt{1+p^2+q^2}}, \quad \varepsilon\in[0,\ 2\pi]\} \tag{2.9}$$

The angle ε is easily seen to provide the direction of the steepest ascent in $H(x,y)$. (See Figure 1). Clearly, if we find a way of determining for each (x,y) the full set of parameters characterizing the tangent plane, we also solved the height reconstruction or shape-from-shading problem.

3. From Shading To Shape

Given the image $A(x,y)$ it is, in general, impossible to unambiguously recover the height profile $H(x,y)$. As an immediate example of ambiguity simply consider the function $-H(x,y)$, which, under a Lambertian shading rule, maps into the same image as $H(x,y)$. Some further information on the function $H(x,y)$ is therefore needed. This is usually given as some smoothness constraint on the surface defined by $z = H(x,y)$ (like, for example C^1 continuity, or C^k continuity) and/or exact or approximate values of $H(x,y)$ at either a discrete set of points $\{(x_i,y_i)\}$, together with the corresponding surface orientations $\{(p_i,q_i)\}$, or on a continuous curve on the (x,y) -plane. Let us first analyse the considerably simpler one-dimensional counterpart of the above defined shape-from-shading problem.

3.1. The 1-D Counterparts

The 1-D counterpart of the shape-from-shading problem is to reconstruct a continuous and almost everywhere differentiable function $H(x)$ from information on the slope of the function at each point. As before the gradient $p(x) = H'(x)$ determines the intensity of the 1-D data image, via a shading rule defined through a reflectivity function $R(p)$. As an example consider the shading rule to be "Lambertian", i.e.

$$A(x) = R(p(x)) = R(p^2(x)) = \frac{1}{1 + p^2(x)} \tag{3.1}$$

This rule provides, via $A(x)$, the absolute value of the derivative at all points where $H(x)$ is differentiable. The further information that we need is sign $\{H'(x)\}$. Clearly we need the complete $H'(x)$ since then the profile $H(x)$ can be reconstructed by integration from any point x_0 at which $H(x_0)$ is known. Suppose however that only $A(x) = |H'(x)|$ is given over an interval $ID = [a,b]$ together with the values $H(a)$ and $H(b)$. If we assume either that $H(x)$ is everywhere differentiable, or that it is unimodal over ID, we can devise methods to recover the height profile completely or to list all profiles that are consistent with the given data. Indeed, if $H(x)$ is continuous and differentiable (or C^1 continuous) we have that the only points at which sign $\{H'(x)\}$ may change value are the x_i's for which $|p(x_i)| = 0$, or $A(x_i) = 1$. These are called the singular points of the image $A(x)$. Suppose the singular points within the interval $[a,b]$ are $\{x_1, x_2, \cdots x_{N-1}\}$. Putting $x_0 = a$ and $x_N = b$ and using the fact that over each of the N intervals $[x_i,x_{i+1}]$, $i = 0,1, \ldots.. N-1$ the sign of $H'(x)$ does not change we have 2^N possible guesses for $H'(x)$ over $[a,b]$. Denote by σ_i the sign assumed for the derivative over the interval $[x_i,x_{i+1}]$. Clearly $\sigma_i \in \{-1, +1\}$ and a sign assignment will be consistent with the data if and only if the integral of the corresponding guess for the derivative, $H'(x)$, will obey (see Figure 2).

$$\int_a^b \widetilde{H}'(x)dx = H(b) - H(a) = \sum_{i=0}^{N-1} \sigma_i \int_{x_i}^{x_{i+1}} |H'(x)|\,dx \qquad (3.2)$$

In this case, the solution of the profile reconstruction problem involves a search over the 2^N possible sign allocations for those that obey (3.2). The problem "given a set of positive numbers $H_i (= \int_{x_i}^{x_{i+1}} |H'(x)|\,dx)$, and $\Delta H = H(b) - H(a)$ find all the sign assignments that satisfy

$$\sum_{i=0}^{N-1} H_i \sigma_i = \Delta H \qquad (3.2a)$$

is quite intractable for very large N. (It is in fact a so-called partition problem, which is known to be NP-complete [10]). However for small N it is not a problem to search over the entire range of sign allocations. If more than one sign assignment exists satisfying (3.2) we have no method to decide which of the corresponding height profiles produced the given data. We did however solve our problem since we obtained a complete characterization of the ambiguities inherent in the data. Suppose now that $H(x)$ is continuous, only *almost* everywhere differentiable but we know it is unimodal, i.e. it has only one local maximum over [a,b]. In this case we can recover $H(x)$ by integrating $|H'(x)|$ from $x = a$ forward and from $x = b$ backwards. At some point $x = x_M$ the integrals

$$\begin{cases} \Lambda_F(x) = H(a) + \int_a^x |H'(\xi)| d\xi \quad \text{and} \\ \Lambda_B(x) = H(b) + \int_x^b |H'(\xi)| d\xi \end{cases} \tag{3.3}$$

have to meet, i.e. we shall have

$$\Lambda_F(x_M) = \Lambda_B \quad (x_M) \tag{3.4}$$

Then we have found the solution of our problem, since clearly (Figure 3), the function

$$H(x) = \begin{cases} \Lambda_F(x) & \text{for } x \le x_M \\ \Lambda_B(x) & \text{for } x > x_M \end{cases} \tag{3.5}$$

is the unique one consistent with both the data and the prior information. This case is in fact simpler to solve than the previous one since we assumed that, over [a,b] the function first monotonically increases then decreases. However, the integration process that solves the problem works for functions which are not necessarily differentiable everywhere. In conclusion, the simple 1-D counterparts of the shape-from-shading problem already show that the problem may be very difficult to handle both computationally and theoretically, unless we have substantial prior information on the nature of the solution.

3.2. The Characteristic-Strip Expansion Method

We return now to the bivariate case. Assume that $z = H(x,y)$ is a smooth surface, i.e. the partials $p(x,y)$ and $q(x,y)$ defined in (2.2) and also second derivatives exist everywhere. Consider now that at some point (x_o,y_o) in the plane we know the height $H(x_o,y_o)$ together with the surface orientation $\{p(x_o,y_o),\ q(x_o,y_o)\}$. Horn observed that in this case one can determine the height profile and the surface orientation along a well-defined curve in the (x,y)-plane called a characteristic. This name is imported from the theory of partial differential equations. Horn's method is in fact a particular case of a general procedure for solving Cauchy-type boundary value problems associated to nonlinear partial differential equations (see e.g. [11]). The characteristic curve is entirely determined by propagating a coupled set of differential equations, driven by the shading data $A(x,y) = R(p(x,y),\ q(x,y))$, from the starting point (x_o,y_o), with initial conditions $\{H(x_o,y_o),\ p(x_o,y_o),\ q(x_o,y_o)\}$. Suppose we take a step in the (x,y)- plane away from (x_o,y_o) so that

$$(x,y) = (x_o,y_o) + (\Delta x,\Delta y) \tag{3.6}$$

Then we have for the change in height

$$\Delta H(x,y) = p(x_o,y_o)\Delta x + q(x_o,y_o)\Delta y \tag{3.7}$$

and also

$$\begin{cases} \Delta p(x,y) = p_x(x_o,y_o)\Delta x + p_y(x_o,y_o)\Delta y \\ \Delta q(x,y) = q_x(x_o,y_o)\Delta x + q_y(x_o,y_o)\Delta y \end{cases} \tag{3.8}$$

Note that $p_y = q_x$ by the smoothness constraint. If, by using the shading information $A(x,y)$, we find a direction $(\Delta x,\Delta y)$ so that, at the new point, both $H(x,y)$ and (p,q) can be determined, then we have a way of computing the height profile on a data-determined curve in the plane. This idea is exactly what underlies the characteristic strip expansion method. Using the chain-rule for differentiation we obtain from (2.6) that

$$\begin{cases} A_x = \left[\frac{\partial}{\partial x}A(x,y)\right] = R_p(p,q) \quad p_x + R_q(p,q) \quad q_x \\ A_y = \left[\frac{\partial}{\partial y}A(x,y)\right] = R_p(p,q)p_y + R_q(p,q)q_y \end{cases} \tag{3.9}$$

Now simply note that if

$$(\Delta x, \Delta y) = (R_p \ \Delta s,\ R_q \ \Delta s) \tag{3.10}$$

for some small Δs , we shall have

$$\begin{cases} \Delta H = (p \cdot R_p + qR_q)\ \Delta s \\ \Delta p = A_x \cdot \Delta s \\ \Delta q = A_y \cdot \Delta s \end{cases} \tag{3.11}$$

This is indeed quite remarkable since (3.11) enables us to propagate for both height and orientation along the curve recursively determined via (3.10). This result is the basis of Horn's shape-from-shading method. He proposed to look at the brightness map $A(x,y)$ and start the height recovery at the singular points - where $A(x,y)$ attains the maximum value of 1, i.e. where $p = q = 0$, (under the Lambertian shading rule). From these points one can propagate outwards in parallel and even use certain neighborhood rules in propagation - such as not allowing crossover of adjacent strips and interpolating new characteristic strips when neighboring strips separate too far. Note that when $p = q = 0$ we have a start-up problem for the algorithm, since (3.10) will not move us out from the singular points. To solve this problem we need to add further assumptions about the behavior of $H(x,y)$ about each such starting-point (i.e. to classify singular points as a local maxima or minima; of course problems arise at saddle points). Implicitly we have to assume the knowledge of the initial slopes (p,q) on a small loop around the singularities [1]. In later work it was stressed that occluding boundaries and other visual clues have a crucial importance in provid-

4. Shape-From-Shading via Equal-Height Contours

It is clear that Horn's characteristic strip expansion method enables one to use various types of prior information - provided it can readily be translated into height and orientation data at a set of points. As a result many ideas for using this method under various circumstances were advanced by Horn and his co-workers and also by Woodham (see e.g. [1],[6]-[8]). Others attempted to exploit different ways of formulating surface continuity and smoothness constraints in order to arrive at practical shape-from-shading algorithms [9]. It seems however that none of the above-quoted works emphasized the potential of using as data and then trying to determine *all* the *equal height contours* of the profile $z = H(x,y)$. An equal height contour or a *level-curve* is a continuous curve in the (x,y)-plane on which the function $H(x,y)$ is constant. If $\{x(\vartheta),\ y(\vartheta)\}\ \vartheta \in \Theta$ is the parametric representation of the curve we have

$$\frac{d}{d\vartheta} H(x(\vartheta),y(\vartheta)) = 0 \tag{4.1}$$

One might argue that such a contour contains a lot of information and is scarcely available to us. This is true, however it can be recognized that many previously proposed algorithms for recovering shape-from-shading require very similar types of prior information (like the surface orientations or the height profile on a closed contour, or occluding boundaries etc.). Also in many practical situations one is quite naturally able to determine equal height contours, or portions of it. As an example, the shores of a lake in a landscape readily provide a closed equal height contour; so is the case when an island raises from the sea - (and in general any shoreline is an equal height curve). Also, in some robot ing just such initial-conditions for the start-up of characteristic strip expansions along several curves in parallel. Ikeuchi and Horn have also analyzed a relaxation-type iterative algorithm using second-derivative surface smoothness constraints together with shading-data and occluding boundary information to recover a height-profile that minimizes a smoothness-measuring cost function. This algorithm uses a special representation of the surface-orientation profile, via the so-called "stereographic projection". It works on a grid of pixels and iteratively assigns the orientations in the stereographic projection plane so as to meet the shading requirements while minimizing orientation changes over neighboring pixels [7]. In the sequel we discuss two new shape-from-shading algorithms. The first one uses the conceptual framework of the characteristic strip expansion method but stresses the importance of recursively determining the equal-height-contours of the profile $H(x,y)$. The other algorithm divides the (x,y)-plane in an array of pixels and recursively determines planar patches over them with parameters adjusted in order to only meet a surface continuity constraint.

vision systems one might be able to provide illumination which actively delineates an equal height contour.

In the sequel we assume that we are given an equal height contour which is almost everywhere differentiable. By definition, along such a curve we have zero height gradient, which yields

$$dH = p dx + q dy = 0 \tag{4.2}$$

Therefore along the given contour $\{x(\vartheta),\ y(\vartheta)\}$ we have determined a relation between the two directional derivatives of the surface p and q. For some ϑ we have, rewriting (4.2),

$$p(x(\vartheta),y(\vartheta)) \cdot \frac{d}{d\vartheta}x(\vartheta) = -q\,(x(\vartheta),y(\vartheta))\,\frac{d}{d\vartheta}y(\vartheta) \tag{4.3}$$

Together with the shading level at that point $A(x(\vartheta),\ y(\vartheta))$ the above relation determines p and q up to an inherent sign ambiguity. Indeed a Lambertian shading rule given by

$$A(x,y) = R(p,q) = \frac{1}{\sqrt{1+p^2+q^2}} \tag{4.4}$$

yields (assuming $y'(\vartheta) \neq 0$)

$$\left\{ \begin{array}{l} p = \pm \left\{ \dfrac{1-A^2}{A^2\left(1+\left(\dfrac{x'}{y'}\right)^2\right)} \right\}^{1/2} \\[2ex] q = -\dfrac{x'}{y'} \cdot p \end{array} \right|_{at\ [x(\vartheta),y(\vartheta)]} \tag{4.5}$$

We get two pairs of solutions corresponding to a certain (p,q) vector and it's negative counterpart. This is an expected result since, at each point on the equal height curve, the same grey level would be produced by the shading rule had the tangent plane have the direction of maximum ascent given either by ε or by $\varepsilon + \pi$. Note also that we could determine the (p,q) pairs up to a similar ambiguity along any continuous path on which the height profile is known a-priori [1]. In case of equal height curves, the direction of the data contour determines the direction of the maximal surface ascent/descent. Suppose we know that the height profile is a mountain raising from the sea. This immediately settles the direction of the steepest ascent as the vector pointing towards the inner region defined by the equal height contour of the shorelines. Using this information we may determine an equal height contour situated a bit above the sea level, and so on we can recursively climb and reconstruct the height profile - provided no "problems" occur. Problems arise if the mountain is not a nice and unimodal profile, and we shall further discuss these issues after a description of the basic profile reconstruction algorithm.

4.1. From a Level Curve to the Ones Nearby

Assume that $[x(\vartheta),\ y(\vartheta)]$ $\vartheta\in[0,1]$ is a closed curve and that, as ϑ goes from 0 to 1 we trace the curve in the counterclockwise direction (see Figure 4). A tangent vector at ϑ is simply given by $[x'(\vartheta),\ y'(\vartheta)]$; the unit normal to it pointing inside the curve will be

$$\vec{n}_\vartheta = \frac{1}{(x'(\vartheta)^2 + y'(\vartheta)^2)^{1/2}}[-y'(\vartheta), x'(\vartheta)] \qquad (4.6)$$

From (4.3) it is clear that $\vec{n}_\vartheta$ we have to go a certain distance, d_ϑ, in order to *climb* a given amount ΔH. If ΔH is small, this distance is quite accurately determined by the shading data alone, since, in the Lambertian case, for example, $A(x(\vartheta),\ y(\vartheta))$ yields the cosine of the angle between the surface normal at $(x(\vartheta),\ y(\vartheta))$ and the vertical direction. As the direction of the maximal ascent is known to be (4.6), we have (see Figure 4b) from geometrical considerations only, that

$$d_\vartheta = \frac{\Delta H \cdot A}{\sqrt{1-A^2}} \tag{4.7}$$

Note that the same result would be obtained by writing (4.2), and substituting (4.5) into it. Then

$$\Delta H = p\ dx + q dy =$$

$$= \frac{\sqrt{1-A^2}}{A} \frac{1}{\sqrt{x'(\vartheta)^2 + y'(\vartheta)^2}} [-y'(\vartheta)dx + x'(\vartheta)dy]$$

and using the requirement that $[dx,dy] = \vec{n}_\vartheta \cdot d_\vartheta$ we obtain

$$\Delta H = \frac{\sqrt{1-A^2}}{A} d_\vartheta$$

which again yields (4.7). (This second derivation is also more general than the geometric argument). In conclusion, given a closed equal-height contour assumed to be at a reference level H_0, a closed contour situated at the level $H_0 + \Delta H$ is determined via

$$[x(\vartheta,\Delta H), y(\vartheta,\Delta H)] = [x(\vartheta), y(\vartheta)] + d_\vartheta \vec{n}_\vartheta = \tag{4.8}$$

$$= [x(\vartheta), y(\vartheta)] + \frac{1}{\sqrt{x'(\vartheta)^2 + y'(\vartheta)^2}} \frac{\Delta H \cdot A_\vartheta}{\sqrt{1-A_\vartheta^2}} [-y'(\vartheta), x'(\vartheta)] .$$

Our derivation in fact leads to a system of first order nonlinear partial differential equations for the functions $x(\vartheta,h)$ and $y(\vartheta,h)$ representing "doubly parametrized" equal height curves in the (x,y) plane. Indeed, if $[x(\vartheta,h), y(\vartheta,h)]$ is defined as a contour corresponding to $H = h$, (4.8) is equivalent to the following set of partial differential equations

$$\frac{\partial}{\partial h}\begin{bmatrix} x(\vartheta,h) \\ y(\vartheta,h) \end{bmatrix} = \tag{4.9}$$

$$\frac{A(x(\vartheta,h),y(\vartheta,h))}{[1-A^2(x(\vartheta,h),y(\vartheta,h))]^{1/2}} \frac{1}{\left[\left(\frac{\partial}{\partial\vartheta}x(\vartheta,h)\right)^2 + \left(\frac{\partial}{\partial\vartheta}y(\vartheta,h)\right)^2\right]^{1/2}} \begin{bmatrix} -\frac{\partial}{\partial\vartheta}y(\vartheta,h) \\ \frac{\partial}{\partial\vartheta}x(\vartheta,h) \end{bmatrix}$$

with initial conditions $[x(\vartheta,0), y(\vartheta,0)] = [x(\vartheta), y(\vartheta)]$. Note that (4.9) is a non-linear initial value problem that has to be integrated to obtain the equal height curves of the profile that yielded the shading $A(x,y)$. We note that it is implicit in our derivations that the surface is smooth enough to provide differentiable equal-height contours at all heights h. It is also assumed that those contours are "well-behaved" as for example in the case of a unimodal $H(x,y)$ over the region of interest (say the interior of the first equal height contour), when they are nicely nested "generalized" rings (see Figure 4). Several remarks should now be made. It is clear that the recursions (4.8) and their differential counterparts (4.9) are valid generally, provided we are given information on which side of the original equal height curve the surface increases. The data $[x(\vartheta), y(\vartheta)]$ can be any curve that is differentiable and if it is not a closed contour we will get, using (4.8), the reconstruction of a well-defined slice of the surface $z = H(x,y)$. If we do start with a closed contour and at some level we obtain a self-intersecting (i.e. not "well behaved") equal height curve this means that we encountered a sadde area which separates peaks or peaks and dips in $H(x,y)$. In this case the contour should be separated into non-intersecting parts and the algorithm may be continued with the separated closed parts as initial equal height curves (see Figure 5 for typical level contour profiles that are encountered).

The practical implementation of the algorithm is, of course, based on (4.8), and the $[x(\vartheta), y(\vartheta)]$ curve is given (perhaps in a suitably chain-coded way) on a finite grid of ϑ values. Then we can use several methods to estimate the derivatives $x'(\vartheta)$ and $y'(\vartheta)$ that appear in the recursion formula. Also we can leave open the choice of the steps in height (ΔH) taken so as to enable the use of various adaptive schemes. (This is useful if an approach to a saddle area is detected). Also, in practice, $A(x,y)$ is known only on a grid of pixels, thus we have to somehow interpolate for the values needed in (4.8) which are points situated on equal height curves. It becomes clear by looking at the system of equations (4.9) that trouble arises when we approach singular point where $A(x,y) = 1$, indicating $p = q = 0$. A singular point can be either a local extremum or a saddle point of some sort. At an isolated singular point we can define an "unsafe" neighborhood and when an equal height curve enters such a neighborhood, we disregard that portion of it but continue to propagate the algorithm from the remaining contour. Some portions of the (x,y)-plane will, of course, remain uncovered using this method. As in the case of the 1-dimensional counterparts discussed in Section 3 the "singular" areas/curves in the plane for which $A(x,y) = 1$ provide boundaries of possible flips in the directions of maximal ascent and a practical shape from shading process should keep track of these and, based on natural constraints on the behavior of equal-height contours, choose the direction assignments which yield consistent final reconstructions. If a-priori we know that the surface is unimodal then, as in the 1-D case discussed, no problems arise. Moreover, we can live with nondifferentiability at a finite set of points along each equal height contour and reconstruct the profile by matching the slices corresponding to differentiable portions. Further issues concerning the practical implementation of the algorithm as well as some results of numerical experiments are discussed in a forthcoming sequel of this report.

4.2. The Relation to the Strip Expansion Method

The differential equations governing the evolution of equal height contours are clearly related to the evolution of characteristic strips. In the characteristic-strip expansion method when $R(p,q) = R(p^2+q^2)$ one determines a path of maximal ascent/descent on the height profile $H(x,y)$. This is so since we have

$$d[x,y] = [R_p, R_q]ds = -\frac{1}{(1+p^2+q^2)^{3/2}}[p,q]ds \tag{4.10}$$

therefore the vector $[dx, dy]$ points to the direction of maximal ascent. We also have that

$$dh = [pR_p + qR_q]ds = -\frac{p^2+q^2}{(1+p^2+q^2)^{3/2}}ds \tag{4.11}$$

In terms of the differential climb dh we can express the direction of the characteristic strip expansion as

$$d[x,y] = \frac{1}{p^2+q^2}[p,q]dh \tag{4.12}$$

This yields

$$\frac{d}{dh}\begin{bmatrix} x \\ y \end{bmatrix} = \frac{\sqrt{p^2+q^2}}{p^2+q^2}\begin{bmatrix} \alpha \\ \beta \end{bmatrix} = \frac{1}{\sqrt{p^2+q^2}}\begin{bmatrix} \alpha \\ \beta \end{bmatrix} \tag{4.13}$$

where $[\alpha,\beta]$ is the unit vector in the direction of the maximal ascent. But we have $A = \frac{1}{\sqrt{1+p^2+q^2}}$, thus $p^2+q^2 = \frac{1-A^2}{A^2}$ and therefore we have

$$\frac{d}{dh}\begin{bmatrix} x \\ y \end{bmatrix} = \frac{\Delta}{\sqrt{1-\Delta^2}}\begin{bmatrix} \alpha \\ \beta \end{bmatrix} \quad . \tag{4.14}$$

where the vector $[\alpha,\beta]$, may be determined, if a ϑ-parameterized equal height curve is available, as $\frac{1}{\sqrt{x'(\vartheta,h)^2+y'(\vartheta,h)^2}}\frac{\partial}{\partial\vartheta}\begin{bmatrix} -y(\vartheta,h) \\ x(\vartheta,h) \end{bmatrix}$. Therefore, if we ask for the evolution of level contours, parameterized as $[x(\vartheta,h), y(\vartheta,h)]$ we obtain again (4.9)

$$\frac{\partial}{\partial h}\begin{bmatrix} x(\vartheta,h) \\ y(\vartheta,h) \end{bmatrix} = \frac{A}{\sqrt{1-A^2}}\frac{1}{\sqrt{x'(\vartheta,h)^2+y'(\vartheta,h)^2}}\frac{\partial}{\partial\vartheta}\begin{bmatrix} -y(\vartheta,h) \\ x(\vartheta,h) \end{bmatrix}.$$

This shows that the curves defined by $[x(\vartheta_o,h), y(\vartheta_o,h)]$ for a given fixed ϑ_o are identical to the characteristic strips. However, note that in the characteristic strip expansion method the parameterization by s is not simply related to the height h. Although Horn mentioned that scale changes are possigble to give various natural interpretations to the s-parameter like arc length etc., the possibility of using h as an equivalent parameter to s was never mentioned. Also the reconstruction of the characteristic strips uses the derivatives of the data field $A(x,y)$ instead of the local direction of the level curves. The use of level contour directions is in our view a very direct exploitation of lateral information if such information is avaiailable. In his 1975 paper Horn proposes another way to use lateral information as obtained via parallel propagation of many characteristic strips emanating from the neighborhood of a singular point, assumed to be a local maximum or minimum. He defines "rings of equal arc length" s about the singular point and uses those rings to interpolate for new characteristic strips when those propagated so far separate too much. The level curve approach is a more natural implementation of this idea. What we are doing is in fact applying a special control on the "speed of expansion" of the characteristic strips to ensure that their "wavefront" is an equal height curve, and use this wavefront to determine the directions of the local maximal ascent/descent on the surface. This replaces the need to differentiate the picture intensity information $A(x,y)$ with natural "laterality" constaints.

We also note that this approach is not limited to level contours. If we would parameterize a general curve on the surface by $\{x(\vartheta,s),y(\vartheta,s),z(\vartheta,s)\}$ where $s=0$ is the data curve we could also propagate the set of equations

$$\frac{d}{ds}\begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} R_p \\ R_q \\ pR_p+qR_q \end{bmatrix} \tag{4.15}$$

together with the nonlinear solutions for (p,q) of the equation

$$\begin{cases} \frac{d}{d\vartheta}z(\vartheta,s) = p\frac{d}{d\vartheta}x + q\frac{d}{d\vartheta}y \\ A(\vartheta,s) = \frac{1}{\sqrt{1+p^2+q^2}} = R(p,q) \end{cases} \tag{4.16}$$

This, of course, is much more tedious and less natural than using level contours.

5. Shape-From-Shading On a Pixel Array

The discussion in Section 4 was based on the assumption that the shading information is given as a continuous function $A(x,y)$. It was also assumed that it is precisely known everywhere on the (x,y)-plane inside a connected domain, say the area enclosed by a level curve. In this section we shall consider the more common, and practical case in which we are given the shading information spatially quantized as an array of grey-levels over picture elements or pixels, $A(i,j)$. In this case we shall assume that we are given the precise height function over a quantized contour in the plane (see Fig. 6). As the underlying height profile is assumed to be a continuous, and relatively smooth function of two variables we shall have that, over a picture element (which is a very small square in the (x,y)-plane) the surface may be well approximated by a planar patch. This plane is determined by a certain normal vector $\overline{N}$ and a constant d, and the shading information over the pixel, $A(i,j)$, provides the cosine of the angle between $\overline{N}$ and the vertical direction. $A(i,j)$ is clearly related to the original surface properties at $(x = i\Delta,\ y = j\Delta)$; it might be an average of $A(x,y)$ on the surface of the pixel or its value at the center. It all depends on the way the approximating plane was chosen. The choice of the plane patch approximation method should, however, become less and less important as the size of the pixels decreases.

Assuming that we are given the $\{A(i,j)\}$ array which approximates the shading information of a continuous profile $H(x,y)$, the shape-from-shading problem may be reformulated as follows. Given the array $\{\cos\alpha(i,j)\}$ of the cosines between $\vec{N}(i,j)$ and the vertical direction find the arrays $z_0(i,j)$ and $\varepsilon(i,j)$ so that the planar patches defined by

$$\{[\cos\alpha(i,j), z_o(i,j), \varepsilon(i,j)]\}_{i,j} \tag{5.1}$$

constitute a good approximation of a continuous profile $H(x,y)$, over a certain region of interest. If prior height information is provided, the reconstructed approximate surface should agree with it, of course. Suppose indeed that the exact heights are given to us over a certain closed "quantized curve" in the plane (Fig. 6). We can devise a simple method for propagating this information into the region of interest R, by using she slope information and only relying on continuity constraints. One can employ the following intuitive fact. Consider a certain pixel (i,j) suppose that we already have some information about the planes approximating the surface $z = H(x,y)$ over some of the neighboring ones. We can use the slope information $\cos\alpha(i,j)$ and search over the other plane defining parameters ($z_o(i,j)$ and $\varepsilon(i,j)$) in order to determine the planar patch over (i,j) that "best" fits in with its neighbors.

To make things more precise consider the pixel (i,j) that spans the area defined by

$$\gamma_{ij} = \left\{(x,y) \,|\, x \in [i\Delta,(i+1)\Delta],\; y \in [j\Delta,(j+1)\Delta]\right\} \tag{5.2}$$

where Δ is the space quantization step. Over γ_{ij}, a planar patch is defined by

$$z = z_o(i,j) + p_{ij}(x - i\Delta) + q_{ij}(y - j\Delta) \tag{5.3}$$

where $z_o(i,j)$ is the height at $(i\Delta,\ j\Delta)$ and p,q are directional derivatives. We are given

$$\cos\alpha(i,j) = \frac{1}{\sqrt{1 + p_{ij}^2 + q_{ij}^2}} = A(i,j) \in [0,1] \tag{5.4}$$

and, if we knew $\varepsilon(i,j)$ we could write the patch equation as follows

$$z = z_o(i,j) + \frac{\sqrt{1 - A(i,j)^2}}{A(i,j)}\Big(\cos\varepsilon(i,j) \cdot (x - i\Delta) + \sin\varepsilon(i,j) \quad (y - j\Delta)\Big) \tag{5.5}$$

In (5.5) we know only $A(i,j)$ and we have to determine the other parameters from neighborhood constraints. If for some of the neighboring pixels these parameters were estimated already, then from each of these pixels we will have a certain "vote" for the height at corresponding corners of γ_{ij}. For example, the pixel $\gamma_{i(j+1)}$ will provide votes for the height at the corners $[i\Delta,\ (j+1)\Delta]$ and $[(i+1)\Delta,\ (j+1)\Delta]$, whereas the pixel $\gamma_{(i+1),(j+1)}$ will only vote for the height at $[(i+1)\Delta,\ (j+1)\Delta]$ in the pixel γ_{ij} (see Figure 7).

When we have votes for at least three different corners of a pixel, we proceed and find a plane that best fits to the vote of the neighbors and is tilted according to $A(i,j)$ as follows.

First, for each corner we list the height that the neighbors voted for. (We can have at most three votes at a corner or none at all). If a planar patch will be fitted in γ_{ij} via $\{z_o, A, \varepsilon\}_{i,j}$, then this patch will provide its own "votes" for the height at the corners of γ_{ij}.

The continuity of $H(x,y)$ requires that we fit the planar patch over γ_{ij} so as to minimize the discrepancy, or the variation between all the resulting height votes at each corner. Thus, we define a cost function, measuring the discrepancy between the votes of the plane fitted over γ_{ij} and the neighbor's votes. Denoting a height vote for the point $(i\Delta,\ j\Delta)$ coming from a pixel $\gamma_{k,l}$ where $(k,l)\in[(i,j),\ (i,j-1),(i+1,j)(i-1,j),(i,j+1)]$ as $V_{ij}^{\gamma_{kl}}$ we can write the votes of the planar patch defined by $\{z_o(i,j),\ A(i,j),\ \varepsilon(i,j)\}$ over the pixel γ_{ij} as $V_{mn}^{\gamma_{ij}}(z_o,\varepsilon)$ (these votes depend parametically on $z_o(i,j)$ and $\varepsilon(i,j)$). The cost function C measuring the smoothness of the fit via $z_o(i,j)$ and $\varepsilon(i,j)$ can be chosen as

$$C(z_o,\varepsilon) = \sum_{\substack{\text{all } (m,n)\text{'s where votes exist}\\ \text{all neighbors voting at } (m,n)}}\sum \left[V_{m,n}^{\gamma_{ij}} - V_{m,n}^{neighbor}\right]^2 \qquad (5.6)$$

In other words we measure by $C(z_o,\varepsilon)$ the squared error between the votes of the pixel and corresponding neighbor votes. Once the cost function is defined we can proceed and determine the unknown parameters $z_o(i,j)$ and $\varepsilon(i,j)$ that minimize this cost function.

Let us consider a typical term of $C(z_o,\varepsilon)$. It has the form (considering (5.5)):

$$\left[\frac{\sqrt{1-A^2}}{A}(\cos\varepsilon\cdot\begin{Bmatrix}0\\or\\1\end{Bmatrix}\cdot\Delta + \sin\varepsilon\cdot\begin{Bmatrix}0\\or\\1\end{Bmatrix}\cdot\Delta) + z_o - V^{neighbor}\right]^2 \qquad (5.7)$$

In order to find the optimal z_o and ε we need to differentiate (5.6) w.r. to ε and z_o. From (5.7) we realize that differentiating w.r. to z_o and setting the (z_o-variation) to zero results in

$$\sum\sum 2\left[\frac{\sqrt{1-A^2}}{A}\cos\varepsilon\begin{Bmatrix}0\\or\\1\end{Bmatrix}\Delta + \sin\varepsilon\begin{Bmatrix}0\\or\\1\end{Bmatrix} + z_o - V^{neighbor}\right] = 0\ .$$

This is a linear equation in z_o and we can solve for it in terms of {$\underset{\sim}{\sin\varepsilon}and\underset{\sim}{\cos\varepsilon}$} and known quantities. Then we can substitute the expression for z_o in the equation resulting from

$$\frac{\partial}{\partial\varepsilon}C(z_o,\varepsilon) = 0$$

and we will obtain a rather complex, nonlinear, equation for ε alone. This equation will contain terms of the form {$\underset{\sim}{\sin\varepsilon}\ \underset{\sim}{\cos\varepsilon}$, $\underset{\sim}{\sin^2\varepsilon}$, $\underset{\sim}{\cos^2\varepsilon}$} and a search process is needed to solve for the optimal ε.

The requirement of fitting a plane only when at least three corners of the pixel have votes from neighbors follows from the following observation: if only two corners of the pixel have votes there remains an inherent ambiguity in the choice of ε. Indeed, if we require a plane to pass through two given heights at two corners and we give its $\cos\alpha$ too, the angle ε is only determined up to an inherent ambiguity. This can be seen from the equation

$$z - z_0 = \sqrt{\frac{1-A^2}{A}}\left(\cos\varepsilon \begin{pmatrix} 0 \\ or \\ 1 \end{pmatrix} \Delta + \sin\varepsilon \begin{pmatrix} 0 \\ or \\ 1 \end{pmatrix} \Delta\right)$$

If, besides z_0 we are given another z at say $x = (i+1)\Delta$, we will have a determined $\cos\varepsilon$, leaving an uncertainty of $\{\varepsilon, 2\pi - \varepsilon\}$.

The shape-from-shading algorithm we propose for pixel arrays is the following.

Given the initial height values along a certain "quantized" curve in the (x,y)-plane, proceed to determine the best fitting planes along the data curve and then propagate the height information using the following process:

> Enter a pixel γ_{ij}. If more than three of its corners have votes, fit a planar patch and distribute the voting to all corners. If not, proceed to a neighboring pixel. When all pixels are fitted with surface patches, reconstruct the height profile as the average value of the four votes at each corner point.

The variation, or squared differences between the votes and their average will provide a measure of the reconstruction quality. This algorithm is nice since it will make a lengthy computation of plane fitting only once for each pixel, thus it is not iterative and furthermore, it is inherently parallel; so a set of processors can work in parallel on several pixels that need plane fitting at a certain stage. Furthermore, a cumulative measure of variation of votes at corners may be used to monitor the surface reconstruction quality as one proceeds. If such a measure grows rapidly indicating that the process is instable one may use some type of backtracking method or quit the shape recovery all-together. A thorough discussion of the practical implementation of this algorithm will be given in the forthcoming sequel of this report.

We also note that an iterative algorithm for numerical shape from shading based on a relaxation scheme was studied by Ikeuchi and Horn [7]. With their method, the given prior information, obtained from occluding boundaries, also slowly diffuses into the pixel array however the algorithm proposed by them is iterative and uses quite strong constraints on slope variation rather than using the surface continuity alone. Also a rather complicated way of representing surface slope, via a stereographic projection was chosen. Our algorithm does not impose such strong smoothness constraints and is also potentially more efficient computationally, being recursive rather than iterative.

6. Conclusion

Two new shape-from-shading algorithm were introduced and discussed. Their implementation and numerical experiments will be the subject of a forthcoming report.

Acknowledgement

I wish to thank Professor Ronald Bracewell of Stanford University for a discussion in which he posed the question whether a surface may be recovered from the cosines of the angles between the normals and the vertical direction. This report may therefore be regarded as my rather partial answer to his question.

References

[1] B.K.P. Horn, "Obtaining Shape From Shading Information", in The Psychology of Computer Vision, Winston, P.H.,(ed), pp.115-155, McGraw-Hill, NY, 1975.

[2] D. Marr, "Vision", W.H.Freeman and Co, SF, 1982.

[3] P.H. Winston, (Editor), "The Psychology of Computer Vision", McGraw-Hill, NY, 1975.

[4] Special Issue of Artificial Intelligence, 17, 1981.

[5] R.J. Woodham, "Photometric Method for Determining Surface Orientation from Multiple Images", Optical Eng.,19, pp.139-144, 1980.

[6] B.K.P. Horn, "Understanding Image Intensities", Artificial Intelligence, 8, pp. 201-231, 1977.

[7] K. Ikeuchi and B.K.P. Horn, "Numerical Shape from Shading and Occluding Boundaries", Artificial Intelligence, 17, pp. 141-184, 1981.

[8] R.J. Woodham, "Analysing Images of Curved Surfaces", Artificial Intelligence, 17, pp. 117-140, 1981.

[9] M.J. Brooks, "Surface Normals from Closed Path", Proc 6th IJCAI, Tokyo, Japan, pp. 98-101, 1979.

[10] M.R. Garey and D.S. Johnson, "Computers and Intractability", W.H.Freeman and Co, SF, 1979.

[11] V.I. Smirnov, "A Course in Higher Mathematics", vol IV, Pergamon Press, Oxford, 1964 (Ch. III).

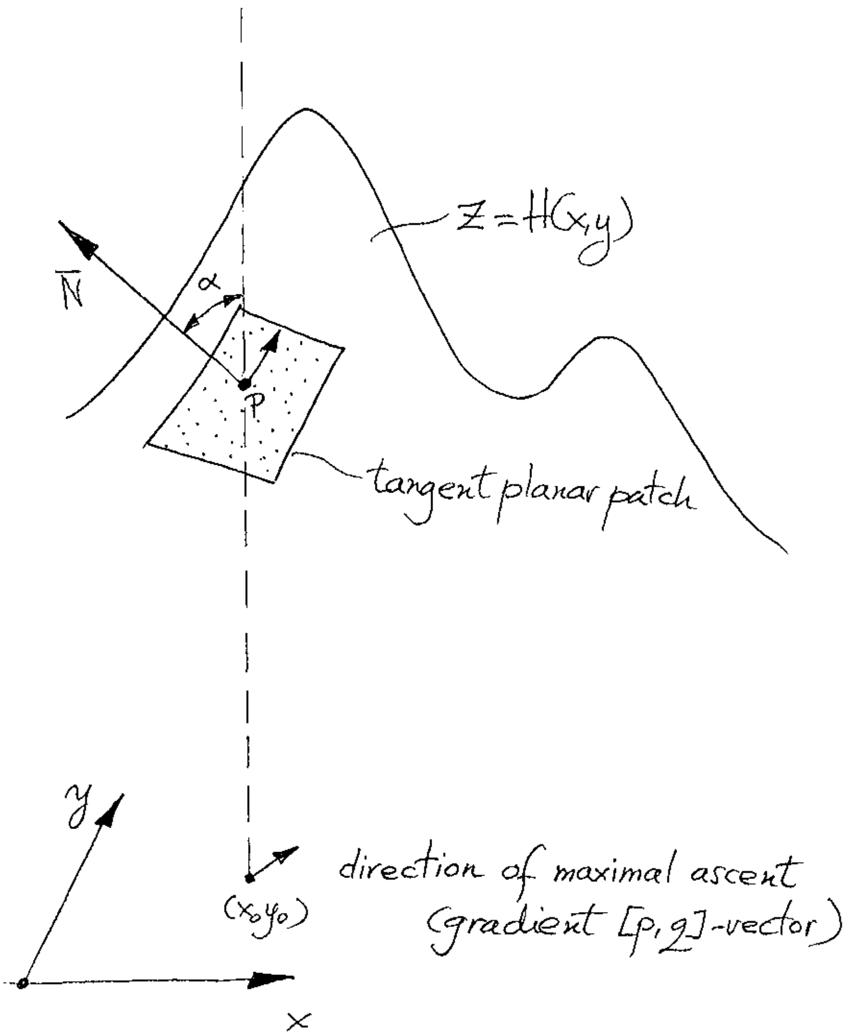

FIGURE 1 : A surface $z = H(x,y)$

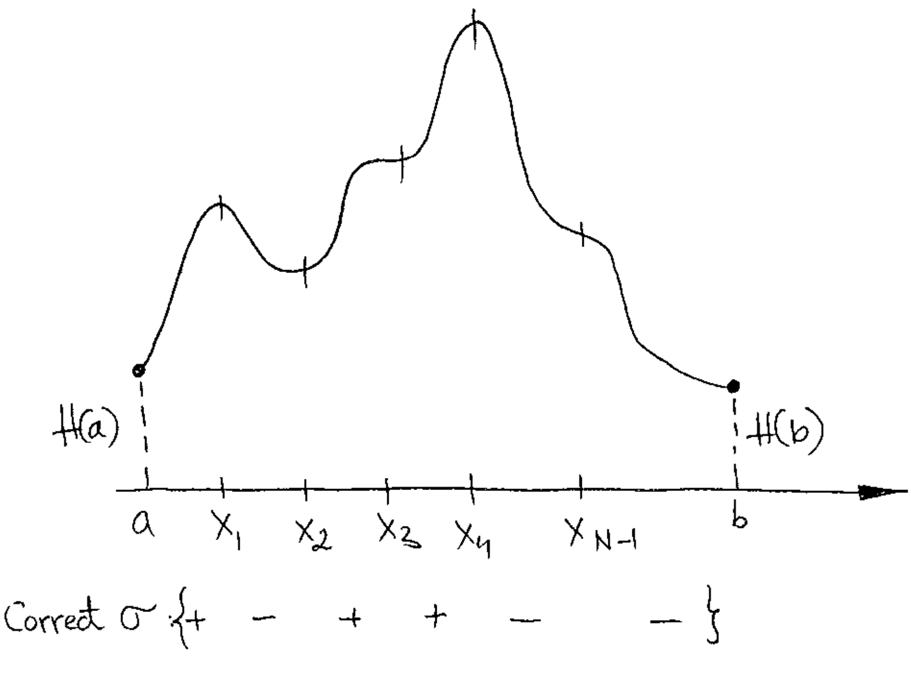

FIGURE 2 : 1-D reconstruction: a search over 2^N sign assignements

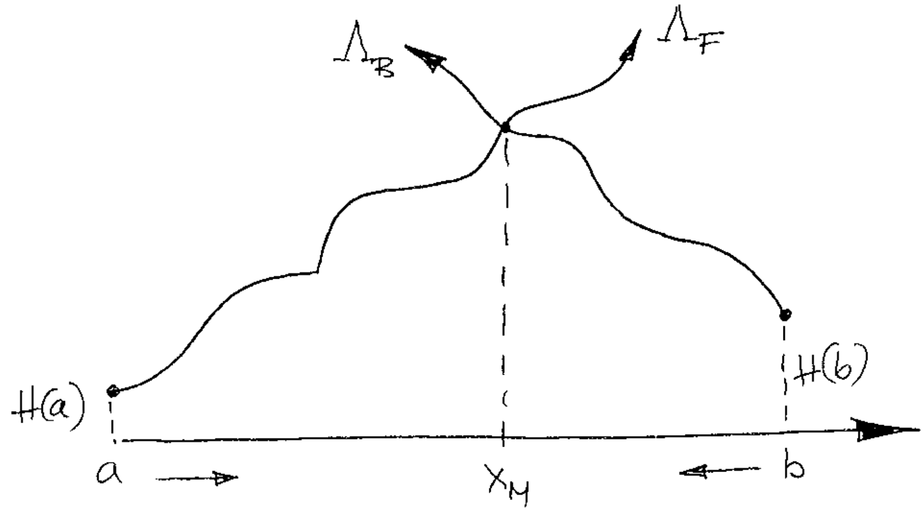

FIGURE 3 1-D reconstruction when H(x) is known to be unimodal.

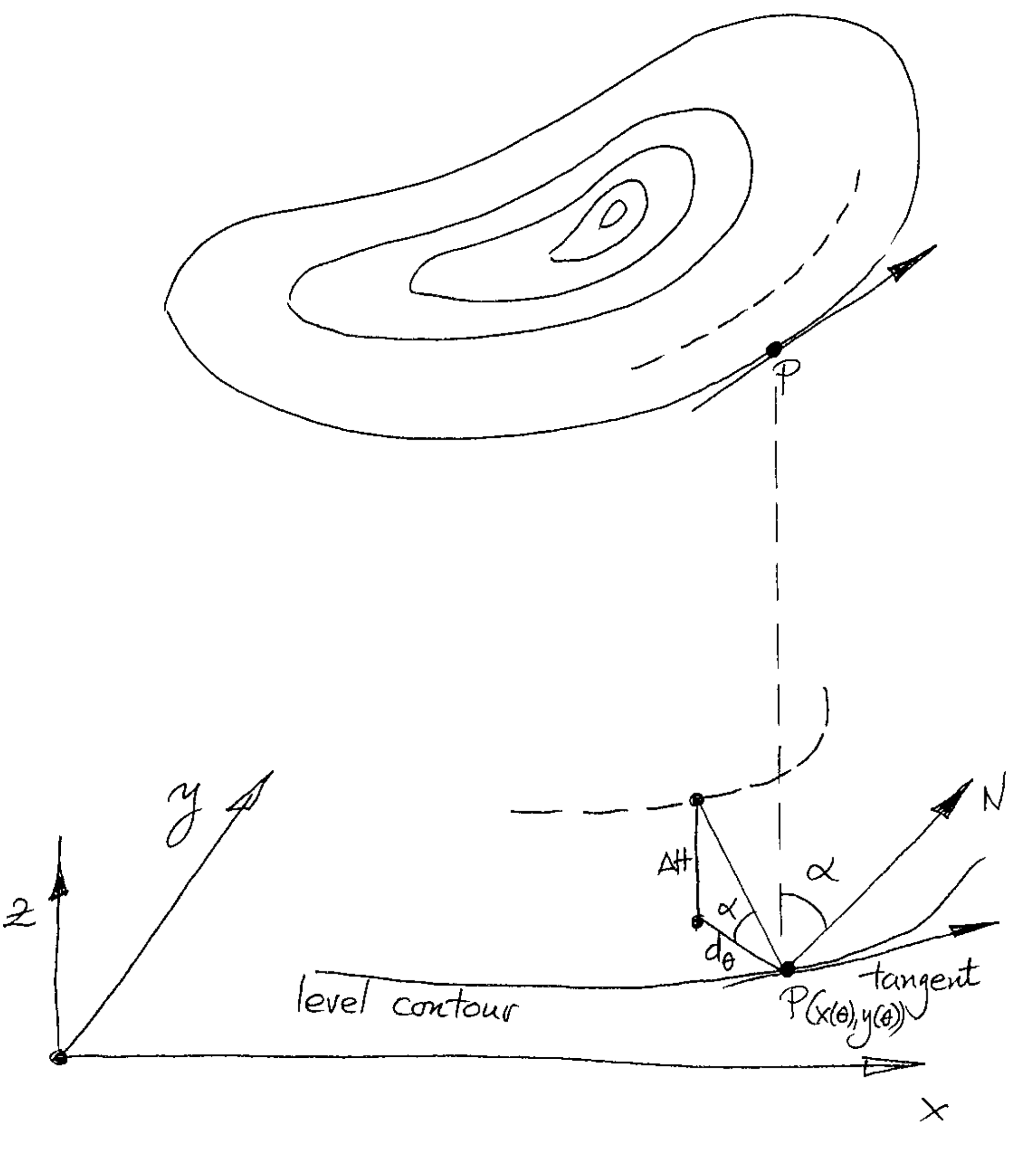

FIGURE 4 Climbing from one level contour to a nearby one.

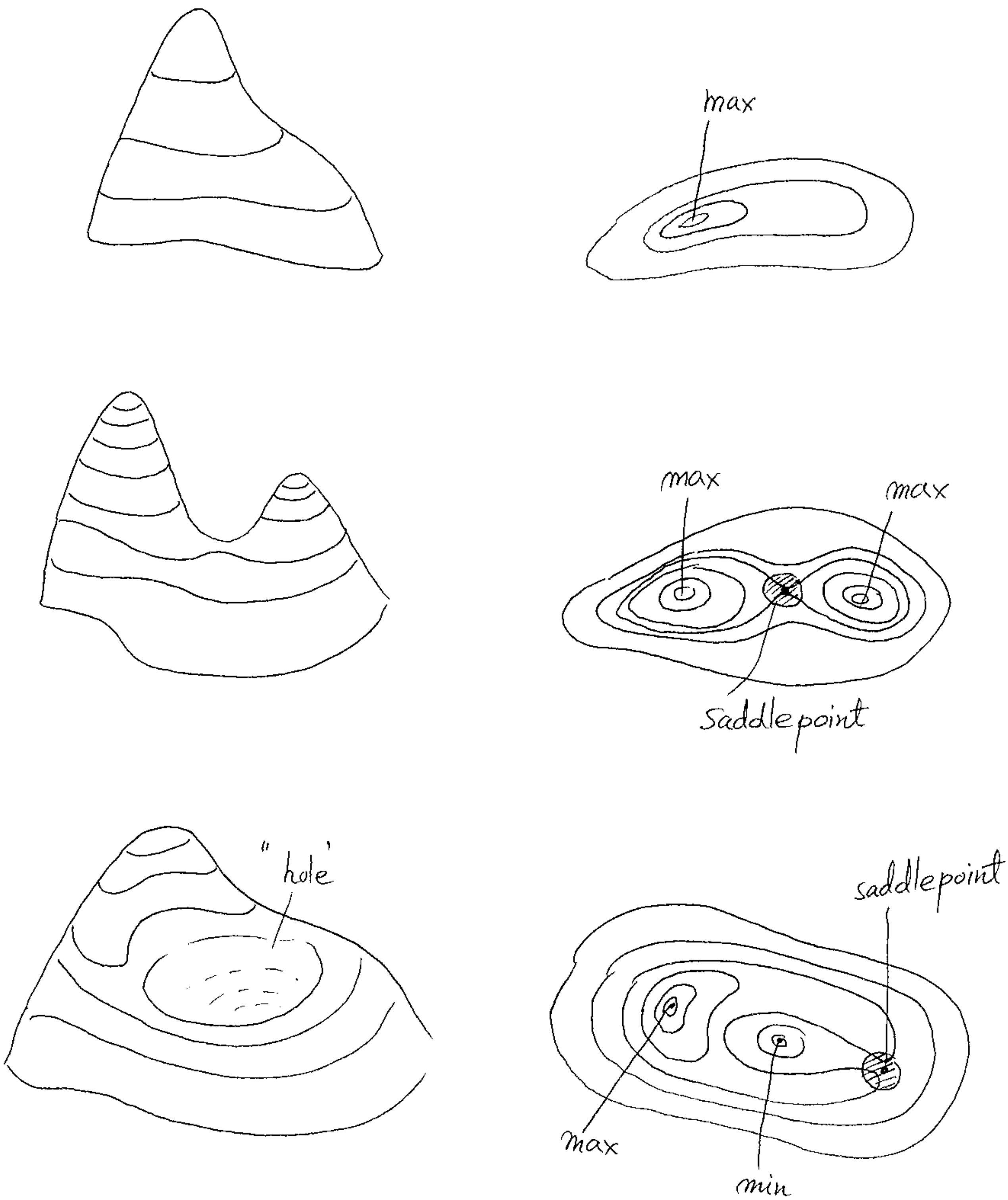

FIGURE 5: Saddle points are dangerous areas for the algorithm.

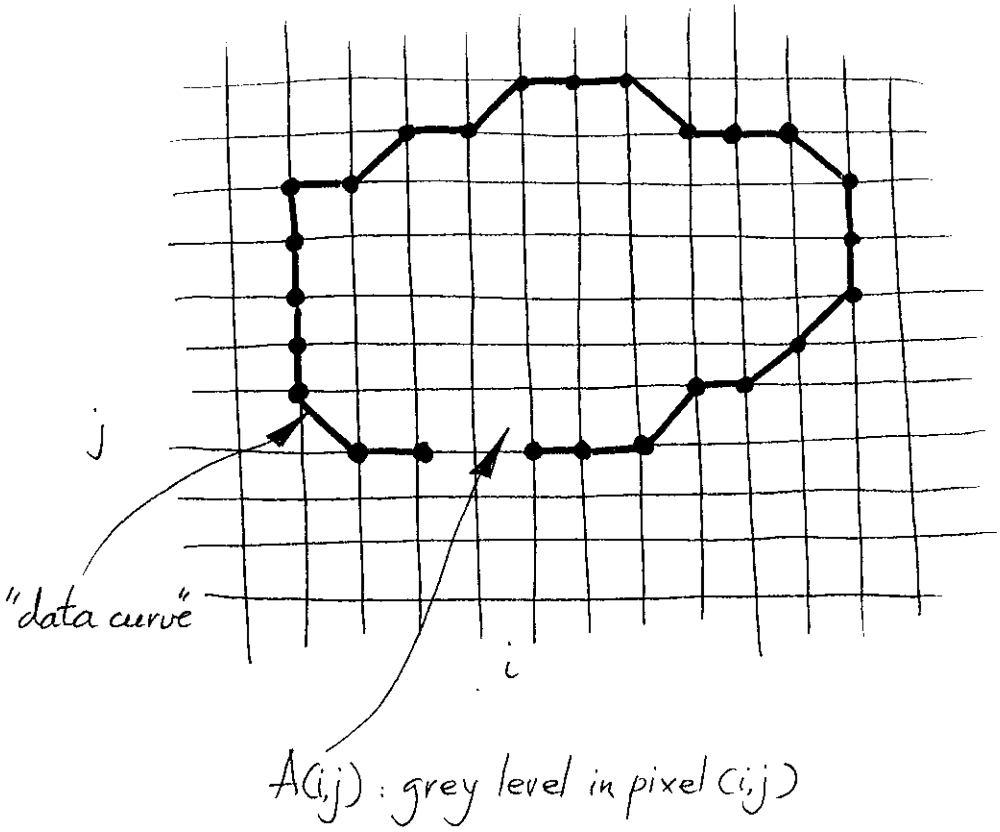

FIGURE 6 : Pixel-array geometry

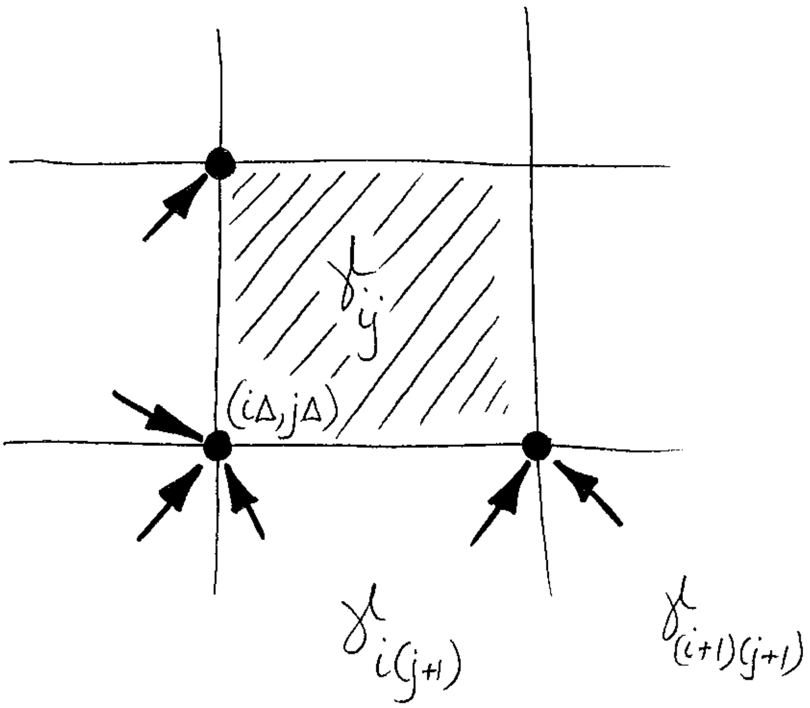

FIGURE 7: The "voting" process for the heights at corners of γ_{ij}.