THE TOPOGRAPHIC PRIMAL SKETCH AND ITS APPLICATION TO PASSIVE NAVIGATION

Thomas J. Laffey

Lockheed Palo Alto Research Labs.
0/52-53 B/201, 3251 Hanover Street, Palo Alto, CA 94304

Robert M. Haralick, Layne T. Watson

Departments of Computer Science and Electrical Engineering,
Virginia Polytechnic Institute and State University,
Blacksburg, VA. 24061

ABSTRACT

A complete mathematical treatment is given for describing the topographic primal sketch of the underlying grey tone intensity surface of a digital image. Each picture element is independently classified and assigned a unique descriptive label, invariant under monotonically increasing gray tone transformations from the set (peak, pit, ridge, ravine, saddle, flat, and hillside), with hillside having subcategories (inflection point, slope, convex hill, concave hill, and saddle hill). The topographic classification is based on the first and second directional derivatives of the estimated image intensity surface. A local, facet model, two-dimensional, cubic polynomial fit is done to estimate the image intensity surface. Zero-crossings of the first directional derivative are identified as locations of interest in the image. Results of the technique applied to digital terrain data and aerial photographs used in the Passive Image Navigation study are presented

1. INTRODUCTION

Representing the fundamental structure of a digital image in a rich and robust way is a primary problem encountered in any general robotics computer-vision system that has to ''understand'' an image. The richness is needed so that shading, highlighting, and shadow information, which are usually present in real manufacturing assembly line situations, are encoded. Richness permits unambiguous object matching to be accomplished. Robustness is needed so that the representation is invariant with respect to monotonically increasing gray tone transformations. Current representations involving edges or the primal sketch as described by Marr (1976; 1980) are impoverished in the sense that they are insufficient for unambiguous matching. They also do not have the required invariance. Basic research is needed to (1) define an appropriate representation, (2) develop a theory that establishes its relationship to properties that three-dimensional objects manifest on the image, and (3) prove its utility in practice. Until this is done, computer-vision research must inevitably be more ad hoc sophistication than science.

The basis of the topographic primal sketch consists of the classification and grouping of the underlying image intensity surface patches according to the categories defined by monotonic, gray tone, invariant functions of directional derivatives. Examples of such categories are peak, pit, ridge, ravine, saddle, flat, and hillside. From this initial classification, we can group categories to obtain a rich, hierarchical, and structurally complete representation of the fundamental image structure. We call this representation the <u>topographic primal sketch</u>.

Why do we believe that this topographic primal sketch can be the basis for computer vision? We believe it because the light-intensity variations on an image are caused by an object's surface orientation, its reflectance, and characteristics of its lighting source. If any of the three-dimensional intrinsic surface characteristics are to be detected, they will be detected owing to the nature of light-intensity variations. Thus, the first step is to discover a robust representation that can encode the nature of these light-intensity variations, a representation that does not change with strength of lighting or with gain settings on the sensing camera. The topographic classification does just that. The basic research issue is to define a set of categories sufficiently complete to form groupings and structures that have strong relationships to the reflectances, surface orientations, and surface positions of the three-dimensional objects viewed in the image.

This research has been supported by National Science Foundation grant MCS-8102872. Experiments were carried out with the support of DARPA contract DAAK70-81-C-0098

1.1. The Invariance Requirement

A digital image can be obtained with a variety of sensing-camera gain settings. It can be visually enhanced by an appropriate adjustment of the camera's dynamic range. The gain setting or the enhancing point operator changes the image by some monotonically increasing function that is not necessarily linear. For example, nonlinear enhancing point operators of this type include histogram normalization and equal probability quantization.

In visual perception, exactly the same visual interpretation and understanding of a pictured scene occurs whether the camera's gain setting is low or high and whether the image is enhanced or unenhanced. The only difference is that the enhanced image has more contrast, is nicer to look at, and is understood more quickly by the human visual system.

This fact is important because it suggests that many of the current, low-level computer-vision techniques, which are based on edges, cannot ever hope to have the robustness associated with human visual perception. They cannot have the robustness, because they are inherently incapable of invariance under monotonic transformations. For example, edges based on zero-crossings of second derivatives will change in position as the monotonic gray tone transformation changes because convexity of a gray tone intensity surface is not preserved under such transformations. However, the topographic categories peak, pit, ridge, valley, saddle, flat, and hillside do have the required invariance.

1.2. Background

Marr (1976) argues that the first level of visual processing is the computation of a rich description of gray level changes present in an image, and that all subsequent computations are done in terms of this description, which he calls the primal sketch. Gray-level changes are usually associated with edges, and Marr's primal sketch has, for each area of gray level change, a description that includes type, position, orientation, and fuzziness of edge. Marr (1980) illustrates that from this information it is sometimes possible to reconstruct the image to a reasonable degree. Unfortunately, as mentioned earlier, edge is not invariant with respect to monotonic image transformations; besides, it is not a rich enough structure. Difficulty, for example, has been experienced in using edges to accomplish unambiguous stereo matching.

The topographic primal sketch we are discussing as a basis for a representation has the required richness and invariance properties and is very much in the spirit of Marr's primal sketch and the thinking behind Ehrich's relational trees (Ehrich and Foith 1978). Instead of concentrating on gray level changes as edges as Marr does, or on one-dimensional extrema as Ehrich and Foith do, we concentrate on all types of two-dimensional gray level variations. We consider each area on an image to be a spatial distribution of gray levels that constitutes a surface or facet of gray tone intensities having a specific surface shape. It is likely that, if we could describe the shape of the gray tone intensity surface for each pixel, then by assembling all the shape fragments we could reconstruct, in a relative way, the entire surface of the image's gray tone intensity values. The shapes that we already know about that have the invariance property are peak, pit, ridge, ravine, saddle, flat, and hillside, with hillside having noninvariant subcategories of slope, inflection, saddle hillside, convex hillside, and concave hillside.

Knowing that a pixel's surface has the shape of a peak does not tell us precisely where in the pixel the peak occurs; nor does it tell us the height of the peak or the magnitude of the slope around the peak. The topographic labeling, however, does satisfy Marr's (1976) primal sketch requirement in that it contains a symbolic description of the gray tone intensity changes. Futhermore, upon computing and binding to each topographic label numerical descriptors such as gradient magnitude and direction, directions of the extrema of the second directional derivative along with their values, a reasonable absolute description of each surface shape can be obtained.

1.3. Facet Model

The facet model states that all processing of digital image data has its final authoritative interpretation relative to what the processing does to the underlying gray tone intensity surface. The digital image's pixel values are noisy sampled observations of the underlying surface. Thus, in order to do any processing, we at least have to estimate at each pixel position what this underlying surface is. This requires a model that describes what the general form of the surface would be in the neighborhood of any pixel if there were no noise. To estimate the surface from the neighborhood around a pixel then amounts to estimating the free parameters of the general form. It is important to note that if a different general form is assumed, then a different estimate of the surface is produced. Thus the assumption of a particular general form is necessary and has consequences.

The general form we use is a bivariate cubic. We assume that the neighborhood around each pixel is suitably fit by a bivariate cubic (Haralick 1981;1982). Having estimated this surface around each pixel, the first and second directional derivatives are easily computed by analytic means. The topographic classification of the surface facet is based totally on the first and second directional derivatives. We classify each surface point as peak, pit, ridge, ravine, saddle, flat, or hillside, with hillside being broken down further into the subcategories inflection point, convex hill, concave hill, saddle hill, and slope. Our set of topographic labels is complete in the sense that every combination of values of the first and second directional derivative is uniquely assigned to one of the classes.

1.4. Previous Work

Detection of topographic structures in a digital image is not a new idea. There has been a wide variety of techniques to detect (a) peaks and pits (spots), (b) ridges and ravines (lines, streaks), (c) hillsides (edges), and other local features. Some of this work includes Fischler (1982), Lee and Fu (1981), Hsu, Mundy, and Beaudet (1978), Toriwaki and Fukurma (1978), Grender (1976), Paton (1975), Johnston and Rosenfeld (1976), Rosenfeld and Kak (1976) and Peuker and Douglas (1975). Detailed discussion of these methods are beyond the scope of this paper. For an excellent discussion of these works the reader is referred to Laffey (1983).

1.5. A Mathematical Approach

From the investigation of previous work, one can see that a wide variety of methods and labels have been proposed to describe the topographic structure in a digital image. Some of the methods require multiple passes through the image, while others may give ambiguous labels to a pixel. Many of the methods are heuristic in nature. The Hsu, Mundy, and Beudet (1978) approach is the most similar to the one discussed here.

Our classification approach is based on the estimation of the first-and second-order directional derivatives. Thus, we regard the digital-picture function as a sampling of the underlying function f, where some kind of random noise is added to the true function values. To estimate the first and second partials, we must assume some kind of parametric form for the underlying function f. The classifier must use the sampled brightness values of the digital-picture function to estimate the parameters and then make decisions regarding the locations of relative extrema of partial derivatives based on the estimated values of the parameters.

In Section 2, we will discuss the mathematical properties of the topographic structures in terms of the directional derivatives in the continuous surface domain. Because a digital image is a sampled surface and each pixel has an area associated with it, characteristic topographic structures may occur anywhere within a pixel's area. Thus, the implementation of the mathematical topographic definitions is not entirely trivial.

In Section 3 we will discuss the implementation of the classification scheme on a digital image. To identify categories that are local one-dimensional extrema, such as peak, pit, ridge, ravine, and saddle, we search inside the pixel's area for a zero-crossing of the first directional derivative. The directions in which we seek the zero-crossing are along the lines of extreme curvature.

In Section 4, we will discuss the local cubic estimation scheme. In Section 5, we will summarize the algorithm for topographic classification using the local facet model. In Section 6, we will show the results of the classifier on digital terrain data and aerial photographs.

2. THE MATHEMATICAL CLASSIFICATION OF TOPOGRAPHIC STRUCTURES

In this section, we formulate our notion of topographic structures on continuous surfaces and show their invariance under monotonically increasing gray tone transformations. In order to understand the mathematical properties used to define our topographic structures, one must understand the idea of the directional derivative discussed in most advanced calculus books. For completeness, we first give the definition of the directional derivative, then the definitions of the topographic labels. Finally, we show the invariance under monotonically increasing gray tone transformations.

2.1. The Directional Derivative

In two dimensions, the rate of change of a function f depends on direction. We denote the directional derivatiye of f at the point (r,c) in the direction β by $f_\beta(r,c)$. It is defined as

$$f'_\beta(r,c) = \lim_{h \to 0} \frac{f(r+h*\sin\beta, c+h*\cos\beta) - f(r,c)}{h}.$$

The direction angle β is the clockwise angle from the column axis. It follows directly from this definition that

$$f'_\beta(r,c) = \frac{\partial f}{\partial r}(r,c) * \sin\beta + \frac{\partial f}{\partial c}(r,c) * \cos\beta.$$

We denote the second derivative of f at the point (r,c) in the direction β by $f''_\beta(r,c)$ and it follows that

$$f''_\beta = \frac{\partial^2 f}{\partial r^2}*\sin^2\beta + 2*\frac{\partial^2 f}{\partial r \partial c}*\sin\beta*\cos\beta + \frac{\partial^2 f}{\partial c^2}*\cos^2\beta.$$

The gradient of f is a vector whose magnitude,

$$\left(\left(\frac{\partial f}{\partial r}\right)^2 + \left(\frac{\partial f}{\partial c}\right)^2\right)^{\frac{1}{2}}$$

at a given point (r,c) is the maximum rate of change of f at that point, and whose direction,

$$\tan^{-1}\left(\frac{\frac{\partial f}{\partial r}}{\frac{\partial f}{\partial c}}\right)$$

is the direction in which the surface has the greatest rate of change.

2.2. The Mathematical Properties

We will use the following notation to describe the mathematical properties of our various topographic categories for continuous surfaces. Let

∇f = gradient vector of a function f;

$||\nabla f||$ = gradient magnitude;

$\omega^{(1)}$ = unit vector in direction in which second directional derivative has greatest magnitude;

$\omega^{(2)}$ = unit vector orthogonal to $\omega^{(1)}$;

λ_1 = value of second directional derivative in the direction of $\omega^{(1)}$;

λ_2 = value of second directional derivative in the direction of $\omega^{(2)}$;

$\nabla f \cdot \omega^{(1)}$ = value of first directional derivative in the direction of $\omega^{(1)}$; and

$\nabla f \cdot \omega^{(2)}$ = value of first directional derivative in the direction of $\omega^{(2)}$.

Without loss of generality, we assume $|\lambda_1| >= |\lambda_2|$.

Each type of topographic structure in our classification scheme is defined in terms of the above quantities. In order to calculate these values, the-first and second-order partials with respect to r and c need to be approximated. These five partials are as follows:

$$\frac{\partial f}{\partial r}, \frac{\partial f}{\partial c}, \frac{\partial^2 f}{\partial r^2}, \frac{\partial^2 f}{\partial c^2}, \frac{\partial^2 f}{\partial r \partial c}.$$

The gradient vector is simply $\frac{\partial f}{\partial r}, \frac{\partial f}{\partial c}$. The second directional derivatives may be calculated by forming the Hessian where the Hessian is a 2*2 matrix defined as

$$H = \begin{vmatrix} \frac{\partial^2 f}{\partial r^2} & \frac{\partial^2 f}{\partial r \partial c} \\ \frac{\partial^2 f}{\partial c \partial r} & \frac{\partial^2 f}{\partial^2 f \partial c^2} \end{vmatrix}.$$

Hessian matrices are used extensively in nonlinear programming. Only three parameters are required to determine the Hessian matrix H, since the order of differentiation of the cross partials may be interchanged. That is

$$\frac{\partial^2 f}{\partial r \partial c} = \frac{\partial^2 f}{\partial c \partial r}$$

The eigenvalues of the Hessian are the values of the extrema of the second directional derivative, and their associated eigenvectors are the directions in which the second directional derivative is extremized. This can easily be seen by rewriting f_β'' as the quadratic form

$$f_\beta'' = (\sin\beta \ \cos\beta) * H * \begin{vmatrix} \sin\beta \\ \cos\beta \end{vmatrix}.$$

Thus,

$$H\omega^{(1)} = \lambda_1 \omega^{(1)} \text{ and } H\omega^{(2)} = \lambda_2 \omega^{(2)}.$$

Furthermore, the two directions represented by the eigenvectors are orthogonal to one another. Since H is a 2*2 symmetric matrix, calculation of the eigenvalues and eigenvectors can be done efficiently and accurately using the method of Rutishauser (1971). We may obtain the values of the first directional derivative in the direction of either extrema of the second directional derivative by simply taking the dot product of the gradient with the appropriate eigenvector:

$$\nabla f \cdot \omega^{(1)}$$
$$\nabla f \cdot \omega^{(2)}$$

There is a direct relationship between the eigenvalues λ_1 and λ_2 and curvature in the directions $\omega^{(1)}$ and $\omega^{(2)}$: When the first directional derivative $\nabla f \cdot \omega^{(i)} = 0$, then $\lambda_i/(1+(\nabla f \cdot \nabla f))^{1/2}$ is the curvature in the direction $\omega^{(i)}$, i = 1 or 2. For further discussion on the relationship of surface curvature to directional derivative, see Laffey (1983).

Having the gradient magnitude and direction and the eigenvalues and eigenvectors of the Hessian, we can describe the topographic classification scheme.

2.2.1. Peak

A peak (knob) occurs where there is a local maxima in all directions. In other words, we are on a peak if, no matter what direction we look in, we see no point that is as high as the one we are on. The curvature is downward in all directions. At a peak the gradient is zero, and the second directional derivative is negative in all directions. To test whether the second directional derivative is negative in all directions, we just have to examine the value of the second directional derivative in the directions that make it smallest and largest. A point is therefore classified as a peak if it satisfies the following conditions:

$$||\nabla f|| = 0, \lambda_1 < 0, \lambda_2 < 0.$$

2.2.2. Pit

A pit (sink, bowl) is identical to a peak except that it is a local minima in all directions rather than a local maxima. At a pit the gradient is zero, and the second directional derivative is positive in all directions. A point is classified as a pit if it satisfies the following conditions:

$$||\nabla f|| = 0, \lambda_1 > 0, \lambda_2 > 0.$$

2.2.3. Ridge

A ridge occurs on a ridge-line, a curve consisting of a series of ridge points. As we walk along the ridge-line, the points to the right and left of us are lower than the ones we are on. Furthermore, the ridge-line may be flat, slope upward, slope downward, curve upward, or curve downward. A ridge occurs where there is a local maximum in one direction. Therefore, it must have negative second-directional derivative in the direction across the ridge and also a zero first-directional derivative in that same direction. The direction in which the local maximum occurs may correspond to either of the directions in which the curvature is ''extremized'', since the ridge itself may be curved. For nonflat ridges, this leads to the first two cases below for ridge characterization. If the ridge is flat, then the ridge-line is horizontal and the gradient is zero along it. This corresponds to the third case. The defining characteristic is that the second directional derivative in the direction of the ridge-line is zero, while the second directional derivative across the ridge-line is negative. A point is therefore classified as a ridge if it satisfies any one of the following three sets of conditions:

$$||\nabla f|| \neq 0, \lambda_1 < 0, \nabla f \cdot \omega^{(1)} = 0$$

or

$$||\nabla f|| \neq 0, \lambda_2 < 0, \nabla f \cdot \omega^{(2)} = 0$$

or

$$||\nabla f|| = 0, \lambda_1 < 0, \lambda_2 = 0.$$

A geometric way of thinking about the definition for ridge is to realize that the condition $\nabla f \cdot \omega^{(i)}=0$ means that the gradient direction (which is defined for nonzero gradients) is orthogonal to the direction $\omega^{(i)}$ of extremized curvature.

2.2.4. Ravine

A ravine (valley) is identical to a ridge except that it is a local minimum rather than maximum in one direction. As we walk along the ravine-line, the points to the right and left of us are higher than the one we are on (see Fig. 2). A point is classified as a ravine if it satisfies any one of the following three sets of conditions:

$$||\nabla f|| \neq 0, \lambda_1 > 0, \nabla f \cdot \omega^{(1)} = 0$$

or

$$||\nabla f|| \neq 0, \lambda_2 > 0, \nabla f \cdot \omega^{(2)} = 0$$

or

$$||\nabla f|| = 0, \lambda_1 > 0, \lambda_2 = 0.$$

2.2.5. Saddle

A saddle occurs where there is a local maximum in one direction and a local minimum in a perpendicular direction A saddle must therefore have positive curvature in one direction and negative curvature in a perpendicular direction. At a saddle, the gradient magnitude must be zero and the extrema of the second directional derivative must have opposite signs. A point is classified as a saddle if it satisifies the following conditions:

$$||\nabla f|| = 0, \lambda_1 * \lambda_2 < 0.$$

2.2.6. Flat

A flat (plain) is a simple, horizontal surface, as illustrated in Fig. 3. It, therefore, must have zero gradient and no curvature. A point is classified as a flat if it satisfies the following conditions:

$$||\nabla f|| = 0, \lambda_1 = 0, \lambda_2 = 0.$$

Given that the above conditions are true, a flat may be further classified as a <u>foot</u> or <u>shoulder</u>. A foot occurs at that point where the flat just begins to turn up into a hill. At this point, the third directional derivative in the direction toward the hill will be nonzero, and the surface increases in this direction. The shoulder is an analogous case and occurs where the flat is ending and turning down into a hill. At this point, the maximum magnitude of the third directional derivative is nonzero, and the surface decreases in the direction toward the hill. If the third directional derivative is zero in all directions, then we are on a flat, not near a hill. Thus a flat may be further qualified as being a foot or shoulder, or not qualified at all.

2.2.7. Hillside

A hillside point is anything not covered by the previous categories. It has a nonzero gradient and no strict extrema in the directions of maximum and minimum second directional derivative. If the hill is simply a tilted flat (i.e., has constant gradient), we call it a <u>slope</u>. If its curvature is positive (upward), we call it a <u>convex hill</u>. If its curvature is negative (downward), we call it a <u>concave hill</u>. If the curvature is up in one direction and down in a perpendicular direction, we call it a <u>saddle hill</u>.

A point on a hillside is an <u>inflection point</u> if it has a zero-crossing of the second directional derivative taken in the direction of the gradient. The inflection-point class is the same as the <u>step edge</u> defined by Haralick (1982), who classifies a pixel as a step edge if there is some point in the pixel's area having a zero-crossing of the second directional derivative taken in the direction of the gradient.

To determine whether a point is a hillside, we just take the complement of the disjunction of the conditions given for all the previous classes. Thus if there is no curvature, then the gradient must be non zero. If there is curvature, then the point must not be a relative extremum. Therefore, a point is classified as a hillside if all three sets of the following conditions are true ('->' represents the operation of logical implication):

$$\lambda_1 = \lambda_2 = 0 \rightarrow ||\nabla f|| \neq 0,$$

and

$$\lambda_1 \neq 0 \rightarrow \nabla f \cdot \omega^{(1)} \neq 0,$$

and

$$\lambda_2 \neq 0 \rightarrow \nabla f \cdot \omega^{(2)} \neq 0.$$

Rewritten as a disjunction of clauses rather than a conjunction of clauses, a point is classified as a hillside if any one of the following four sets of conditions are true:

$$\nabla f \cdot \omega^{(1)} \neq 0,\ \nabla f \cdot \omega^{(2)} \neq 0$$

or

$$\nabla f \cdot \omega^{(1)} \neq 0,\ \lambda_2 = 0$$

or

$$\nabla f \cdot \omega^{(2)} \neq 0,\ \lambda_1 = 0$$

or

$$||\nabla f|| \neq 0,\ \lambda_1 = 0,\ \lambda_2 = 0.$$

We can differentiate between different classes of hillsides by the values of the second directional derivative. The distinction can be made as follows:

SLOPE if $\lambda_1 = \lambda_2 = 0$

CONVEX if $\lambda_1 >= \lambda_2 >= 0,\ \lambda_1 \neq 0$

CONCAVE if $\lambda_1 <= \lambda_2 <= 0,\ \lambda_1 \neq 0$

SADDLE HILL if $\lambda_1 * \lambda_2 < 0$

A slope, convex, concave, or saddle hill is classified as an inflection point if there is a zero-crossing of the second directional derivative in the direction of maximum first directional derivative (i.e., the gradient).

2.2.8. Summary of the Topographic Categories

A summary of the mathematical properties of our topographic structures on continuous surfaces can be found in Table 1. The table exhaustively defines the topographic classes by their gradient magnitude, second directional derivative extrema values, and the first directional derivatives taken in the directions which extremize second directional derivatives. Each entry in the table is either 0, +, –, or *. The 0 means not significantly different from zero; + means significantly different from zero on the positive side; – means significantly different from zero on the negative side, and '*' means it does not matter. The label ''Cannot Occur'' means that it is impossible for the gradient to be nonzero and the first directional derivative to be zero in two orthogonal directions.

From the table, one can see that our classification scheme is complete. All possible combinations of first and second directional derivatives have a corresponding entry in the table. Each topographic category has a set of mathematical properties that uniquely determines it.

(Note: Special attention is required for the degenerate case $\lambda_1 = \lambda_2 \neq 0$, where $\omega^{(1)}$ and $\omega^{(2)}$ can be <u>any</u> two orthogonal directions. In this case, there <u>always</u> exists an extreme direction ω which is orthogonal to ∇f, and thus the first directional derivative $\nabla f \cdot \omega$ is <u>always</u> zero in an extreme direction. To avoid spurious zero directional derivatives, we choose $\omega^{(1)}$ and $\omega^{(2)}$ such that $\nabla f \cdot \omega^{(1)} \neq 0$ and $\nabla f \cdot \omega^{(2)} \neq 0$, unless the gradient is zero.)

Table 1. Mathematical Properties of Topographic Structures

$\|\|\nabla f\|\|$	λ_1	λ_2	$\nabla f \cdot \omega^{(1)}$	$\nabla f \cdot \omega^{(2)}$	Label
0	–	–	0	0	Peak
0	–	0	0	0	Ridge
0	–	+	0	0	Saddle
0	0	0	0	0	Flat
0	+	–	0	0	Saddle
0	+	0	0	0	Ravine
0	+	+	0	0	Pit
+	–	–	–,+	–,+	Hillside
+	–	*	0	*	Ridge
+	*	–	*	0	Ridge
+	–	0	–,+	*	Hillside
+	–	+	–,+	–,+	Hillside
+	0	0	*	*	Hillside
+	+	–	–,+	–,+	Hillside
+	+	0	–,+	*	Hillside
+	+	*	0	*	Ravine
+	*	+	*	0	Ravine
+	+	+	–,+	–,+	Hillside
+	*	*	0	0	Cannot Occur

2.3. The Invariance of the Topographic Categories

For a proof on the invariance of the topographic categories {peak, pit, ridge, ravine, saddle, flat, and hillside}, see Haralick, Watson, and Laffey (1983), or Laffey (1983).

2.4 Ridge and Ravine Continuums

The definitions for ridge and ravine can lead to possibly some unexpected results. For example, all points on a right circular cone, except the vertex, will be labeled ridge. Whether one wishes to call these points ridge points or something else is a matter of taste. These points are classified as ridge points because as one walks up the cone toward the vertex the points to the left and right are lower than the one you are on. The continuum of ridges may or may not be acceptable depending upon your viewpoint. Further work by Haralick (forthcoming) has partially solved this problem.

3.0 THE TOPOGRAPHIC CLASSIFICATION ALGORITHM

The definitions of Section 2 cannot be used directly since there is a problem of where in a pixel's area to apply the classification. If the classification were only applied to the point at the center of each pixel, then a pixel having a peak near one of its corners, for example, would get classified as a concave hill rather than as a peak. The problem is that the topographic classification we are interested in must be a sampling of the actual topographic surface classes. Most likely, the interesting categories of peak, pit, ridge, ravine, and saddle will never occur precisely at a pixel's center, and if they do occur in a pixel's area, then the pixel must carry that label rather than the class label of the pixel's center point. Thus one problem we must solve is to determine the dominant label for a pixel given the topographic class label of every point in the pixel. The next problem we must solve is to determine, in effect, the set of all topographic classes occurring within a pixels's area without having to do the impossible brute-force computation.

For the purpose of solving these problems, we divide the set of topographic labels into two subsets: (1) those that indicate that a strict, local, one-dimensional extremum has occurred (peak, pit, ridge, ravine, and saddle) and (2) those that do not indicate that a strict, local, one-dimensional extremum has occurred (flat and hillside). By one-dimensional, we mean along a line (in a particular direction). A strict, local, one-dimensional extremum can be located by finding those points within a pixel's area where a zero-crossing of the first directional derivative occurs.

So that we do not search the pixel's entire area for the zero-crossing, we only search in the directions of extreme second directional derivative, $\omega^{(1)}$ and $\omega^{(2)}$. Since these directions are well aligned with curvature properties, the chance of overlooking an important topographic structure is minimized, and, more importantly, the computational cost is small.

When $\lambda_1 = \lambda_2 \neq 0$, the directions $\omega^{(1)}$ and $\omega^{(2)}$ are not uniquely defined. We handle this case by searching for a zero-crossing in the direction given by $H^{-1}*\nabla f$. This is the Newton direction, and it points directly toward the extremum of a quadratic surface.

For inflection-point location (first derivative extremum), we search along the gradient direction for a zero-crossing of second directional derivative. For one-dimensional extrema, there are four cases to consider: (1) no zero-crossing, (2) one zero-crossing, (3) two zero-crossings, and (4) more than two zero-crossings. The next four sections discuss these cases.

3.1. Case One: No Zero-Crossing

If no zero-crossing is found along either of the two extreme directions within the pixel's area, then the pixel cannot be a local extremum and therefore must be assigned a label from the set (flat or hillside). If the gradient is zero, we have a flat. If it is nonzero, we have a hillside. If the pixel is a hillside, we classify it further into (inflection point, slope, convex hill, concave hill, or saddle hill). If there is a zero-crossing of the second directional derivative in the direction of the gradient within the pixel's area, the pixel is classified as an inflection point. If no such zero-crossing occurs, the label assigned to the pixel is based on the gradient magnitude and Hessian eigenvalues calculated at the center of the pixel, local coordinates (0,0), as in Table 2.

3.2. Case Two: One Zero-Crossing

If a zero-crossing of the first directional derivative is found within the pixel's area, then the pixel is a strict, local, one-dimensional extremum and must be assigned a label from the set (peak, pit, ridge, ravine, or saddle). At the location of the zero-crossing, the Hessian and gradient are recomputed, and if the gradient magnitude at the zero-crossing is zero, Table 3 is used.

If the gradient magnitude is nonzero, then the choice is either ridge or ravine. If the second directional derivative in the direction of the zero-crossing is negative, we have a ridge. If it is positive, we have a ravine. If it is zero, we compare the function value at the center of the pixel, f(0,0), with the function value at the zero-crossing, f(r,c). If f(r,c) is greater than f(0,0), we call it a ridge, otherwise we call it a ravine.

3.3. Case Three: Two Zero-Crossings

If we have two zero-crossings of the first directional derivative, one in each direction of extreme curvature, then the Hessian and gradient must be recomputed at each zero-crossing. Using the procedure decribed in Section 3.2, we assign a label to each zero-crossing. We call these labels LABEL1 and LABEL2. The final classification given the pixel is based on these two labels and is given in Table 4.

If both labels are identical, the pixel is given that label. In the case of both labels being ridge, the pixel may actually be a peak, but experiments have shown that this case is rare. An analogous argument can be made for both labels being ravine. If one label is ridge and the other ravine, this indicates we are at or very close to a saddle point, and thus the pixel is classified as a saddle. If one label is peak and the other ridge, we choose the category giving us the ''most information,'' which in this case is peak. The peak is a local maximum in all directions, while the ridge is a local maximum in only one direction. Thus, peak conveys more information about the image surface. An analogous argument can be made if the labels are pit and ravine. Similarly, a saddle gives us more information than a ridge or valley. Thus, a pixel is assigned saddle if its zero-crossings have been labeled ridge and saddle or ravine and saddle.

It is apparent from Table 4 that not all possible label combinations are accounted for. Some combinations, such as peak and pit, are omitted because of the assumption that the underlying surface is smooth and sampled frequently enough that a peak and pit will not both occur within the same pixel's area. If such a case occurs, our convention is to choose arbitrarily one of LABEL1 or LABEL2 as the resulting label for the pixel.

Table 2. Pixel Label Calculation for Case One: No Zero-Crossing

$\|\|\nabla f\|\|$	λ_1	λ_2	Label
0	0	0	Flat
+	−	−	Concave Hill
+	−	0	Concave Hill
+	−	+	Saddle Hill
+	0	0	Slope
+	+	−	Saddle Hill
+	+	0	Convex Hill
+	+	+	Convex Hill

Table 3. Pixel Lable Calculation for Case Two: One Zero-Crossing

$\|\|\nabla f\|\|$	λ_1	λ_2	Label
0	−	−	Peak
0	−	0	Ridge
0	−	+	Saddle
0	+	−	Saddle
0	+	0	Ravine
0	+	+	Pit

Table 4. Final Pixel Classification, Case Three: Two Zero-Crossings

LABEL1	LABEL2	Resulting Label
Peak	Peak	Peak
Peak	Ridge	Peak
Pit	Pit	Pit
Pit	Ravine	Pit
Saddle	Saddle	Saddle
Ridge	Ridge	Ridge
Ridge	Ravine	Saddle
Ridge	Saddle	Saddle
Ravine	Ravine	Ravine
Ravine	Saddle	Saddle

3.4. Case Four: More than Two Zero-Crossings

If more than two zero-crossings occur within a pixel's area, then in at least one of the extrema directions there are two zero-crossings. If this happens, we choose the zero-crossing closest to the pixel's center and ignore the other. If we ignore the further zero-crossings, then this case is identical to case 3. This situation has yet to occur in our experiments.

4.0 SURFACE ESTIMATION

In this section we discuss the estimation of the parameters required by the topographic classification scheme of Section 2 using the local cubic facet model (Haralick 1981). It is important to note that the classification scheme of Section 2 and the algorithm of Section 3 are independent of the method used to estimate the first-and second-order partials of the underlying digital image-intensity surface at each sampled point. Results from using basis functions other than the bi-cubic polynomial are presented in (Laffey 1983). In these experiments the cubic model performed best.

4.1. Local Cubic Facet Model

In order to estimate the required partial derivatives, we perform a least-squares fit with a two-dimensional surface, f, to a neighborhood of each pixel. It is required that the function f be continuous and have continuous first-and second-order partial derivatives with respect to r and c in a neighborhood around each pixel in the rc plane.

We choose f to be a cubic polynomial in r and c expressed as a combination of discrete orthogonal polynomials. The function f is the best discrete least-squares polynomial approximation to the image data in each pixel's neighborhood. More details can be found in Haralick's paper (1981), in which each coefficient of the cubic polynomial is evaluated as a linear combination of the pixels in the fitting neighborhood.

To express the procedure precisely and without reference to a particular set of polynomials tied to neighborhood size, we will canonically write the fitted bicubic surface for each fitting neighborhood as

$$\begin{aligned} f(r,c) = {} & k_1 + k_2 r + k_3 c \\ & + k_4 r^2 + k_5 rc + k_6 c^2 \\ & + k_7 r^3 + k_8 r^2 c + k_9 rc^2 + k_{10} c^3, \end{aligned}$$

where the center of the fitting neighborhood is taken as the origin. It quickly follows that the needed partials evaluated at local coordinates (r,c) are

$$\frac{\partial f}{\partial r} = k_2 + 2k_4 r + k_5 c + 3k_7 r^2 + 2k_8 rc + k_9 c^2$$

$$\frac{\partial f}{\partial c} = k_3 + k_5 r + 2k_6 c + k_8 r^2 + 2k_9 rc + 3k_{10} c^2$$

$$\frac{\partial^2 f}{\partial r^2} = 2k_4 + 6k_7 r + 2k_8 c$$

$$\frac{\partial^2 f}{\partial c^2} = 2k_6 + 2k_9 r + 6k_{10} c$$

$$\frac{\partial^2 f}{\partial r \partial c} = k_5 + 2k_8 r + 2k_9 c$$

It is easy to see that if the above quantities are evaluated at the center of the pixel where local coordinates (r,c) = (0,0), only the constant terms will be of significance. If the partials need to be evaluated at an arbitrary point in a pixel's area, then a linear or quadratic polynomial value must be computed.

5. SUMMARY OF THE TOPOGRAPHIC CLASSIFICATION SCHEME

The scheme is a parallel process for topographic classification of every pixel which can be done in one pass through the image. At each pixel of the image, the following four steps need to be performed

1. Calculate the fitting coefficients, k_1 through k_{10}, of a two-dimensional cubic polynomial in an n-by-n neighborhood around the pixel. These coefficients are easily computed by convolving the appropriate masks over the image.

2. Use the coefficients calculated in step 1 to find the gradient, gradient magnitude, and the eigenvalues and eigenvectors of the Hessian at the center of the pixel's neighborhood, (0,0).

3. Search in the direction of the eigenvectors calculated in step 2 for a zerocrossing of the first directional derivative within the pixel's area. (If the eigenvalues of the Hessian are equal and non-zero, then search in the Newton direction.)

4. Recompute the gradient, gradient magnitude, and values of second directional derivative extrema at each zero-crossing. Then apply the labeling scheme as described in Sections 3.1---3.4.

6. RESULTS

In this section, we show the results of the topographic classification on some digital terrain imagery and aerial photographs used in the Passive Image Navigation Study.

6.1 Results on Digtial Terrain Data

In Figure 1 we show the results of the topographic classification algorithm on digitial terrain data which represents a roughly 4x17 mile strip of land and ocean just east of Monterey, California. The actual image resolution is 121x512 pixels. In Figure 1 we show the results of the labeling for several of the categories. The top-most shows the ravines in white, the next shows the riges, and then the peaks are shown. On the bottom the original grey-level picture is shown.

The algorithm shows excellent results on the digital terrain data. The ridges and ravines appear to be robust enough for use in a reference topographic landmark database. Sensed topography would be matched against the reference database for navigation purposes.

6.2 Results on Aerial Photographs

In figure 2 and 3 we show the results of the classifier on a set of aerial photographs. it seems evident that ravines, ridges, and hillsides (slopes) could serve as reference data in an intensity landmark database. Exactly which topographic categraphic are reliable and how they should be linked together and pruned remains a topic of future research.

7. CONCLUSIONS

In this paper, we have given a precise mathematical description of the various topographic structures that which occur in a digital image and have called the classified image the topographic primal sketch. Our set of topographic categories is invariant under gray tone, monotonically increasing transformations and consists of (peak, pit, ridge, ravine, saddle, flat, and hillside), with hillside being broken down further into the subcategories inflection point, slope, convex hill, concave hill, and saddle hill. The hillside subcategories are not invariant under the monotonic transformations.

The topographic label assigned a pixel is based on the pixel's first-and second-order directional derivatives. We use a two-dimensional cubic polynomial fit based on the local facet model to estimate the directional derivatives of the underlying gray tone intensity surface. The calculation of the extrema of the second directional derivative can be done efficiently and stably by forming the Hessian matrix and calculating its eigenvalues and their associated eigenvectors. Strict, local, one-dimensional extrema (such as pit, peak, ridge, ravine, and saddle) are found by searching for a zero-crossing of the first directional derivative in the directions of extreme second directional derivative (the eigenvectors of the Hessian). We have also identified another direction of interest, the Newton direction, which points toward the extremum of a quadratic surface. The classification scheme was found to give satisfactory results on a number of test images.

7.1. Directions for Further Research

Further research on the topographic primal sketch needs to be done to (1) develop better basis functions, (2) make use of fitting error, (3) find a solution for the ridge (ravine) continuum problem, and (4) develop techniques for grouping of the topographic structures. Basis functions worth considering include trigonometric polynomials, polynomials of higher order, and piecewise polynomials of lower order than cubic. The basis functions problem is to find a set of basis functions and an associated inner product for least-squares approximation that can correctly replicate all common image surface features and be simultaneously computationally efficient and numerically stable. Fitting error needs to be used in deciding into which class a pixel falls. Noise causes the fitting error to increase, and increased fitting error increases the uncertainty of the labeling. Also, global knowledge of how the topographic structures fit together could be used to correct the misclassification error caused by noise. The way the neighborhood size affects the surface fitting error and the classification scheme needs to be investigated in detail.

The ridge (ravine) continuum problem needs to be solved. It may be that there is no way to distinguish between a true ridge and a ridge continuum using only the values of partial derivatives at a point. The solution may require complete use of the partial derivatives in a local area about the pixel.

Most important for the use of the primal sketch in a general robotics computer vision system is the development of techniques for grouping and assembling topographically labeled pixels to form the primitive structures involved in higher-level matching and correspondence processes. How well can stereo correspondence or frame-to-frame time-varying image correspondence tasks be accomplished using the primitive structures in the topographic primal sketch? How effectively can the topographic sketch be used in undoing the confounding effects of shading and shadowing? How well will the primitive structures in the topographic sketch perform in the two-dimensional to three-dimensional object matching process?

REFERENCES

[1] Ehrich, R. W., and Foith, J. P. 1978. Topology and Semantics of Intensity Arrays. Computer vision systems, New York: Academic, pp. 111 -- 128.

[2] Fischler, M. A. 1982. Linear delination, Proceedings Image Understanding Workshop, Palo Alto, California, pp 274-282.

[3] Grender, G. C. 1976. TOPO III: A Fortran Program for Terrain Analysis. Comput. Geosci., London:Pergamon Press 2, pp. 195 -- 209.

[4] Haralick, R. M. 1980. edge and region analysis for digital image data. Comput. Graphics Image Processing, New York:Academic, 12(1):60 -- 73.

[5] Haralick, R. M. 1981. The Digital Edge. Proc. 1981 Conf. Pattern Recognition Image Processing. New York: IEEE Computer Society, pp. 285 -- 294.

[6] Haralick, R. M. 1982. Zero-crossing of second directional derivative edge operator. SPIE Proc. Robot Vision, Bellingham:SPIE

[7] Haralick, R. M. Watson, L.T. and Laffey, T. J. 1983. The topographical primal sketch. Int'l Journal of Robotics Research, MIT press, Cambridge, MA (forthcoming)

[8] Haralick, R. M. 1983. Ridges and valley in a digital image. Comp. Graphics and Image Proc., New York:Academic Press forthcoming

[9] Hsu, S., Mundy, J. L. and Beaudet, P. R. 1978. Web representation of image data. 4^{th} Int. Joint Conf. Pattern Recognition. New York:IEEE Computer Society pp. 675 -- 680.

[10] Johnston, E. G. and Rosenfeld, A. 1975. Digital detection of pits, peaks, ridges, and ravines. IEEE Trans. Syst., Man, Cybern. New York:IEEE Trans.Syst., Man, Cybern. 472 -- 480.

[11] Laffey, T. J., Haralick R. M. and Watson, L. T. 1982. Topographic classification of digital image intensity surfaces. Proc. IEEE Workshop Comput. Vision: Theory Contr. New York: IEEE Computer Society. pp. 171 -- 177.

[12] Laffey, T.J. Topographical classification of digital image intensity surfaces. M.S. Thesis, Dept. of Computer Science, Virginia Polytechnic Institute and State University, February, 1983.

[13] Lee, H. C., and Fu, K. S. 1981. The GLGS image representation and its application to preliminary segmentation and pre-attenetive visual search. Proc. 1981 Conf. Pattern Recognition Image Processing. New York:IEEE Computer Society pp. 256 -- 261.

[14] Marr, D. 1976. Early processing of visual information', Philosophical Trans. Royal Soc. London, London:Royal Society, B 275:483 -- 524.

[15] Marr, D. 1980. Visual information processing: The structure and creation of visual representations. Philosophical Tran. Royal Soc. London, London:Royal Society, B 290:199 -- 218.

[16] Paton, K. 1975. Picture description using Legendre polynomials. Comput. Graphics Image Processing, New York:Academic 4(1):40 -- 54.

[17] Peuker, T. K., and Johnston, E. G. 1972 (Nov.). Detection of surface-specific points by local parallel processing of discrete terrain elevation data. Tech. Rept. 206. College Park: University of Maryland Computer Science Center.

[18] Peuker, T. K. and Douglas, D. H. 1975. Detection of surface-specific points by local parallel processing of discrete terrain elevation data Comput. Graphics Image Processing, New York:Academic 4(4):375 - 387.

[19] Rosenfeld, A., and Kak, A.C. 1976. Digital Picture Processing, Academic Press, New York, New York, PP. 306 - 316.

[20] Rutishauser, H. 1971. Jacobi method for real symmetric matrix. Handbook for automatic computation, volume II, linear algebra, ed. J. H. Wilkinson and C. Reinsch. New York:Springer-Verlag.

[21] Strang, G. 1980. Linear algebra and its applications, 2rd ed. New York:Academic, pp.243 - 249.

[22] Toriwaki, J., and Fukumura, T. 1978. Extraction of structural information from grey pictures. Comput. Graphics Image Processing, New York:Academic, 7(1):30 - 51.

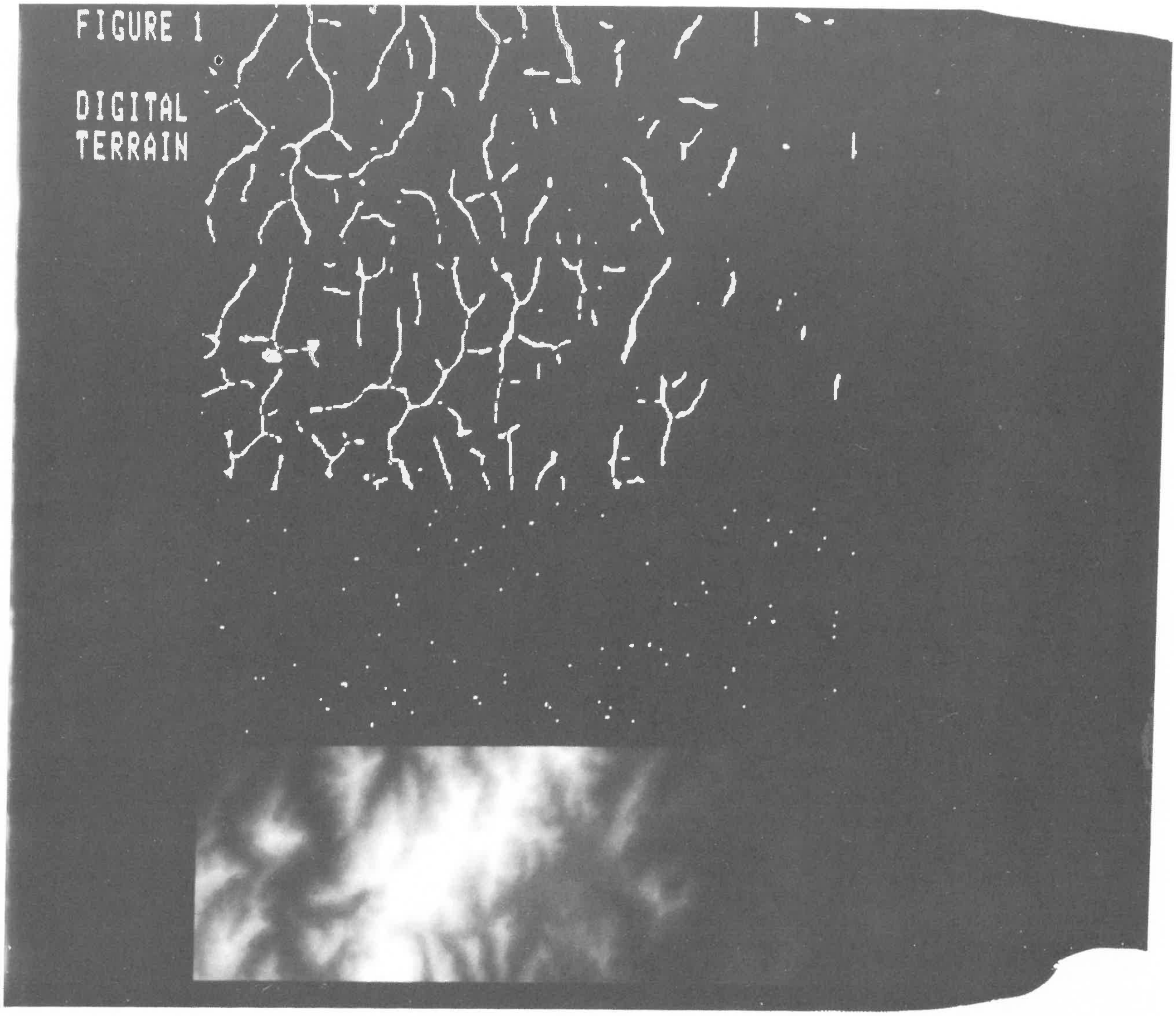
FIGURE 1
DIGITAL
TERRAIN

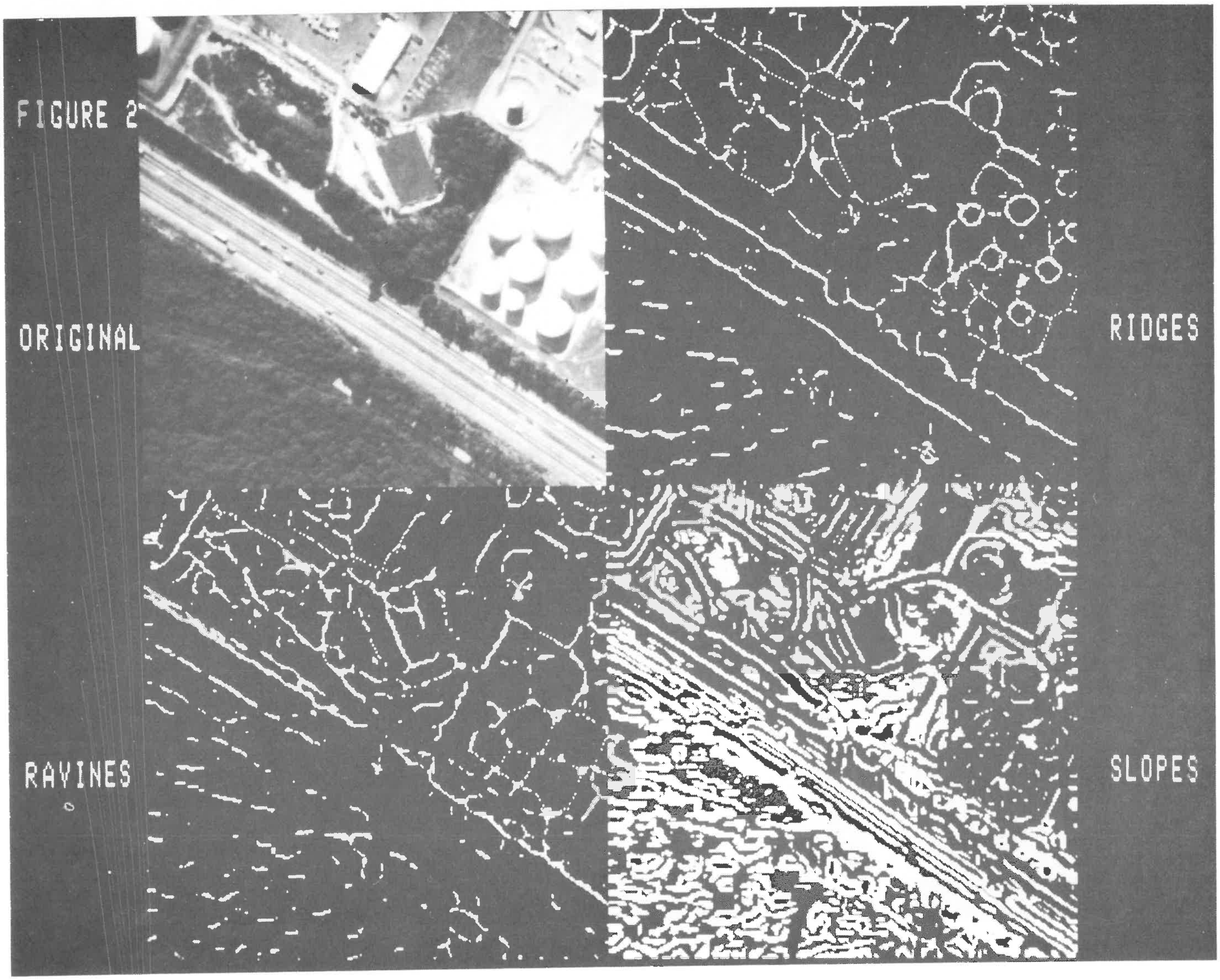
FIGURE 2
ORIGINAL
RIDGES
RAVINES
SLOPES

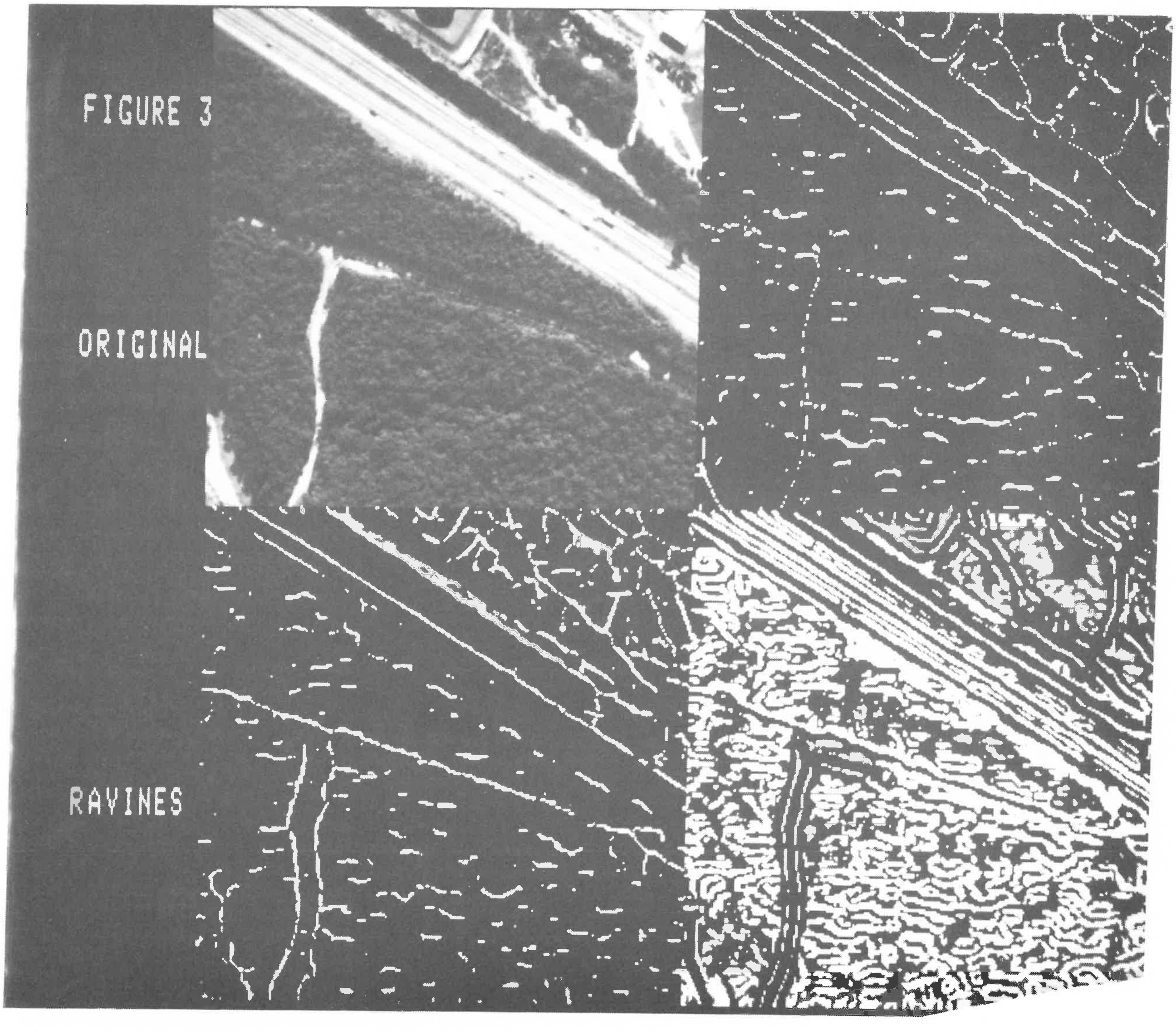
FIGURE 3
ORIGINAL
RAVINES