MACHINE PERCEPTION OF LINEAR STRUCTURE*

Martin A. Fischler and Helen C. Wolf

SRI International
333 Ravenswood Avenue
Menlo Park, California 94025

I INTRODUCTION

In this paper we address a basic problem in machine perception: the tracing of perceptually obvious "line-like" structure appearing in an Image; we present some new ways of looking at this problem and provide techniques that are significantly more general and effective than previously reported methods for this task.

Our approach is a departure from the procedures usually employed in the following important respects: (1) We recognize the typically overlooked distinction between lines and edges; finding lines offers significant simplifications not available when searching for edges. (2) The perceptual primitive we extract from the image to Indicate the local presence of linear structure is a function of total intensity variation over an extended portion of the image rather than Intensity gradient Information which can be extracted with a mask-like "local image operator." While easy to compute, this new primitive appears to be more effective and stable than the almost universally employed gradient based primitives. (3) The overall procedure can be applied to an arbitrary image with essentially no human intervention (parameter tuning or attention focusing) and produces excellent results. Examples of the delineation procedure are shown for aerial, industrial, and radiographic imagery.

II PROBLEM DEFINITION

For the purposes of this paper, we define linear delineation (LD) as the task of generating lists of points for a given 2-D image, such that the points in each list fall sequentially along what any reasonable human observer would describe as a clearly visible "line-like" structure in the Image. Practical examples of this task might be to delineate the roads, rivers, and rail lines in an aerial photograph, or to trace the paths taken by blood vessels in a radiographic angiogram, or to locate the wiring paths on a printed circuit board; however, our goal in this paper is not to look for specific real-world objects or to assign semantic labels to the detected linear structures, but rather to find the most perceptually obvious (to a human observer) occurrences of such structures. We further distinguish between the problems of (1) detecting the edges or contours of extended objects, and (2) delineating those objects whose appearance is adequately represented by a central skeleton — we only address the second problem.

III LINES AND EDGES

While most approaches to LD do not distinguish between lines and edges, and even use edge detection as a necessary first step in the delineation task, a critical concept advanced in this paper is the distinction between line and edge detection.

Edge detection is intuitively based on the concept of finding a discontinuity (in intensity or in some other locally measurable attribute such as color or texture) between two adjacent but distinct regions in an image. However, in a digital representation of an image, we can always fit a smooth surface to the sample values of the integer raster. Thus, edge detection must be based on parameters or thresholds set by assumptions about the nature of the image. (Even if we only mark as edge points those locations at which there are first derivative maxima, or zeros of the second derivative, we must still ultimately make an arbitrary decision in deciding when the corresponding gradient is large enough to be called a discontinuity.)

Intuitively, a "line-like" or linear structure is a (connected) region that is very long relative to its width, and has a ridge or skeleton along which the intensities change slowly and are distinguished from those outside the region; the width need not be constant, but any changes in width should occur In a smooth manner. (To simplify our discussion, we will assume the linear structures are distinguished by their ridge points being brighter than the surrounding background, but any other specified attribute, which is locally detectable, would be an acceptable substitute.) It is important to recognize the fact that a clearly visible line in an image may have no locally detectable edges (and thus no locally measurable width). Finally, it will generally be the case that linear structures have no visible internal detail that is essential to their delineation.

The research reported herein was supported by the Defense Advanced Research Projects Agency (Contract MDA903-83-C-0027), and by the National Science Foundation (Contract ECS-7917028).

IV SMOOTHING, DISTANCE TRANSFORMS, AND THE GRAY SCALE SKELETON

If we can find the edges of a linear structure (LS), we can generate a distance transform and extract a skeleton as the desired delineation (e.g., Rosenfeld [1], Fischler [2]). However, as we noted in the preceding section, LS do not always have locally detectable edges, and, since all of the generally known techniques for deriving a skeleton require a complete contour, some other approach is required.

Surprisingly, something equivalent to a distance transform that works on gray scale images (and on binary images as well) is already available. To achieve our purpose, we need only observe that the intensities in a properly smoothed image can be considered to be the values of an approximate distance transform. What is the best smoothing function for general use? Actually, it doesn't seem to make much difference in many cases. Most digital images have been processed by low-pass optical and electronic systems that have inserted the required minimum level of smoothing. The viewpoint that the smoothed image can be considered to be a distance transform is the essential element. However, If we start with a binary image, or a very noisy image, then additional smoothing is often desirable. Since we are not concerned with blurring edges, and we would like to eliminate or blur any structure or texture internal to the linear regions, we want the smoothing function to have a width of approximately that of the region to be delineated. If the width of the smoothing function is increased further, the thinner linear structures are eventually eliminated. Thus, if we wish to find all possible linear structures without prior knowledge of the content of the image, the processing should be repeated with a set of smoothing filters having a spectrum of widths. Actually, no more than two or three filtering steps should ever be required. For example, to trace all the linear structures (diameters up to 20 pixels) in a noisy radiographic angiogram, a single filter of width 20 was used (see Figure Id). Smoothing introduced by the acquisition process was sufficient to produce excellent results in tracing the linear structures in aerial imagery (see Figures la and lb).

V RIDGES (OR VALLEYS), OPERATORS, AND NEIGHBORHOODS

Having produced an approximate distance transform via smoothing, we now must deal with the problem of locating the ridge points that denote the spines (skeletons) of the LS. When an exact distance transform Is derived from a complete contour, noise is not a problem and the skeleton has assured geometric properties that make it easy to detect; finding the ridge points of an approximate distance transform is considerably more complex.

We traditionally distinguish between locally and globally detectable features: local features are detectable by an intensity pattern which can be observed through a small peep-hole centered on the feature, while global features are ambiguous in a small area. The model or description of the local feature Is generally compiled into an intensity patch (matched filter or operator) which can be convolved with the image to detect the corresponding feature. In the case of an exact distance transform, a 3X3 pixel operator Is sufficient to detect ridge points (a 2X2 operator is sufficient for the Labeled Distance Transform (Fischler [2])); for the approximate distance transform, a small fixed-size operator is ineffective.

The problem of finding the ridge points of an approximate distance transform can be viewed as the problem of finding the ridge points (local maxima) of an exact distance transform to which some amount of noise has been added. We are not concerned about the possibility of making isolated (incoherent) incorrect decisions since we have developed linking and pruning methods capable of eliminating such errors. Our main problem is that we can't count on finding either large local gradients or using known line width to determine some minimum significant gradient threshold, to identify valid ridge points; additionally, noise will Introduce many false local maxima. Thus, we must use total intensity change, rather than rate of change, to detect valid ridge points, and we must have an effective way of determining such total change even in the presence of local variation introduced by noise. (While it is not immediately obvious that total intensity change, rather than rate of change, will recover the perceptually obvious linear features, our experiments indicate that this Is Indeed the case.)

Our approach Is to evaluate two attributes of each of the detected intensity maxima along selected paths In the image (in this case, horizontal and vertical scan lines), which we call the "local" and "global" maxima values. The local maxima value of a point is the total intensity difference from the point to the highest of its immediate left and right intensity miraiраа along the path. The left (right) global maxima value of a point is the total intensity difference from the point to the lowest intensity value found moving to the left (right) prior to encountering a point with an intensity value greater than that of the given point; the global maxima value of the point is the smaller of its left and right global values. In the case of a plateau, only the center point is treated as a maximal point and evaluated as previously described. If a point or (points on) a plateau has an immediate neighbor with a higher intensity value, It Is not a maximal point and it is not assigned either a local or global value (actually, for implementation purposes, it is assigned a zero value); on the other hand, every maximal point will have both a local and global value where the global value equals or exceeds the local value.

What is a large-enough value, of either a local or global maxima, to indicate significance. In our unsmoothed image data, about 1/3 of the points were maximal points, and, in a smoothed image, this percentage is much smaller. Given the linking and pruning techniques we describe in the next section, it might be possible to return all the maximal points in a binary mask and still extract the desired line structure from the background noise contained in such a mask. However, it makes much more sense to first eliminate those maximal points that do not have enough intensity variation to be perceptually distinguishable from a flat background.

Rather than attempting to find some optimal threshold setting (in the sense of maximum noise elimination without any loss of linear structure), which would be difficult or impossible to automatically determine at this level of information organization, we iteratively adjust our threshold settings to satisfy a constraint based on a complexity measure. These program determined thresholds typically allow at least two to ten times the number of ridgepoints (maxima) to be retained above that which would result by manually setting the thresholds to achieve visually acceptable results. The final elimination of "non-significant" maxima is achieved later in the processing at a higher level of organization.

VI CLUSTERING, LINKING, PRUNING, RANKING, AND FINAL DELINEATION

Based on the availability of a binary overlay depicting the locations of the major linear structures contained in the given gray scale image, obtained as described in the preceding section, we have been able to demonstrate that:

(1) The linking step in the delineation process can effectively be based on the single attribute of geometric proximity, and that a clustering or association step, followed by the construction of a Minimum Spanning Tree (MST) through the points of each cluster (Fischler [3]), will correctly link the ridge points along the skeletons of the linear structures.

(2) The desired delineations will be embedded in trees containing additional branches that are either minor linear structures or noise, and that simple pruning techniques can eliminate most of this unwanted detail (tree pruning can produce simplifications that would be difficult, if not impossible, at lower levels of organization of the information).

(3) Having properly linked the ridge points and pruned some of the smaller branches of the resulting trees, we can extract long coherent paths by a decision procedure applied at each node of each tree. This decision procedure, based on the (local to the node) branch attributes of intensity, connectivity, and directionality, assigns path connectivity through a node by splitting off incompatible branches; any remaining ambiguities (more than two branches entering a node) are resolved by choosing those pairings that result in the longest paths.

(4) The paths obtained in the tree partitioning step can be rank ordered with respect to perceptual quality by a metric based on the path attributes of total length, contrast, and continuity.

Details of the procedures mentioned above are described in a journal length paper now in preparation.

VII EXPERIMENTAL RESULTS

Figure 1 presents the results of applying the delineation algorithm to aerial, industrial, and medical images. Our goal was to be able to take imagery from arbitrary domains, and without any human intervention (e.g., parameter adjustment or attention focusing), produce high quality delineation of the obvious linear structures. The results show that we have accomplished much of what we originally intended. The delineations achieved by the uniform parameter settings are quite good in all the images with the exception of the angiogram (which is extremely noisy, and does not really satisfy the criteria of having "clearly visible" linear structure); and even here, by making an appropriate selection within a very small parameter space, we can obtain very good results. By using the values produced by the ranking step of the delineation algorithm, we believe it should be possible to automatically search the parameter space (of approximately 10 to 100 parameter combinations) to optimize the processing for any given image. We are currently investigating this possibility.

VIII CONCLUDING COMMENT

The competence of the ridgepoint based algorithm to abstract the linear structure of an image is apparent by inspection of Figure 1. That ridgepoints should be so effective is, in a sense, a "psychological discovery." It would appear that ridgepoints are important perceptual primitives which may play a significant role in a variety of other tasks (e.g., perception of surface shape). A significant feature of "ridgepoints" is that their locations are independent of any monotonic Intensity transformation of the image, and their geometric configuration is independent of a change of scale. These are precisely the properties we would expect of a perceptual primitive, but are lacking in the commonly employed gradient dependent primitives, such as "edgepoints."

REFERENCES

1. A. Rosenfeld and J.L. Pfalz, "Sequential Operations in Digital Picture Processing," J. ACM, Vol. 13(4), October 1966, pp. 471-494.

2. M.A. Fischler and P. Barrett, "An Iconic Transform for Sketch Completion and Shape Abstraction," Computer Graphics and Image Processing, Vol. 13(4), Aug. 1980, pp. 334-360.

M.A. Fischler, J.M. Tenenbaum, and H.C. Wolf, "Detection of Roads and Linear Structures in Low-Resolution Aerial Imagery Using a Multisource Knowledge Integration Technique," Computer Graphics and Image Processing, Vol. 15(3), March 1981, pp. 201-223.

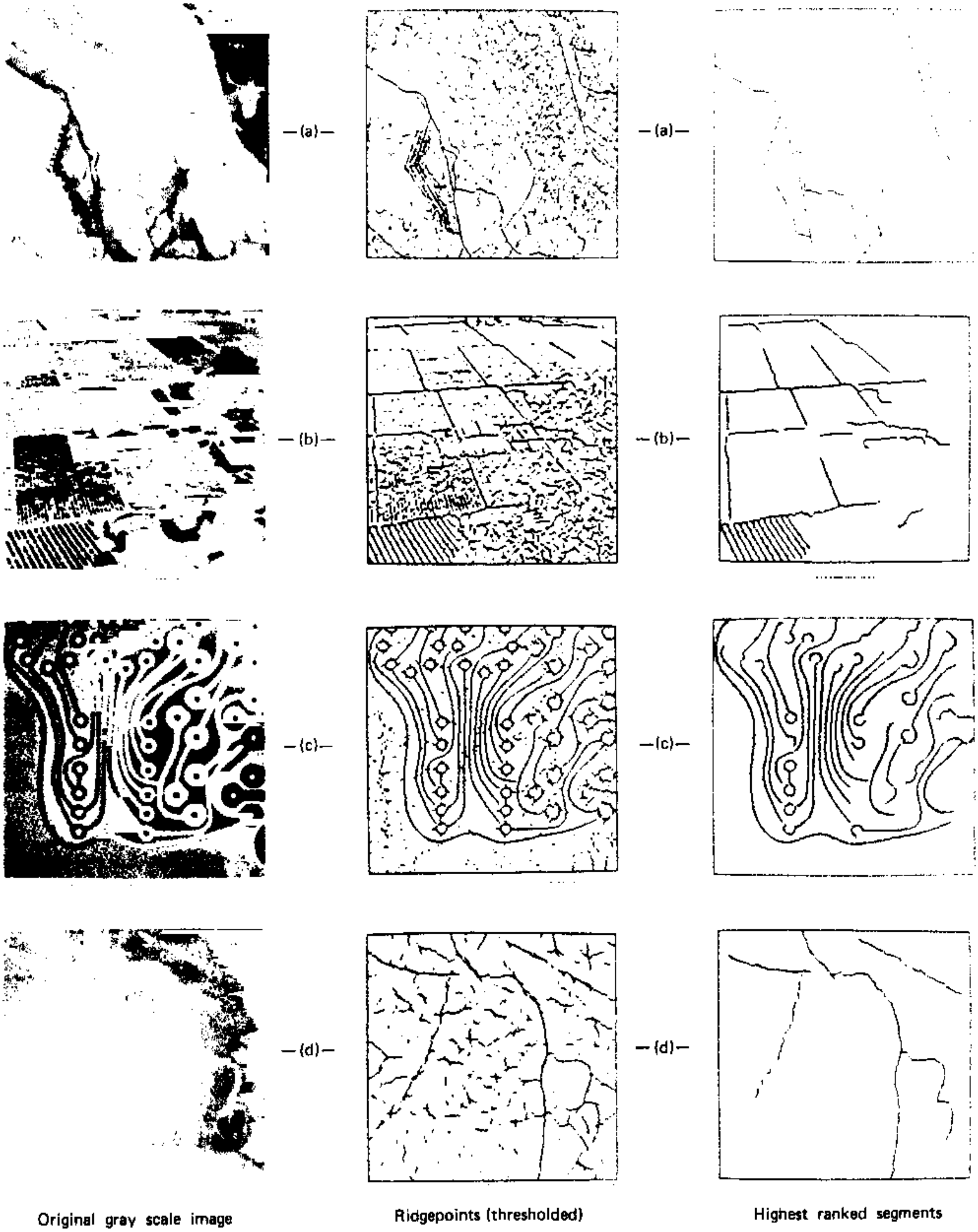

FIGURE 1 EXAMPLES OF THE LINEAR DELINEATION PROCESS