Delineation of Elongated Sub-Patterns
in a Piecewise Constant Foreground

Carlo Arcelli and Giuliana Ramella

Istituto di Cibernetica, CNR
80072 Arco Felice, Napoli, Italy

Abstract. A procedure useful to give evidence to the perceived linear structure of a gray-tone pattern is presented, which allows one to delineate its locally higher intensity regions with a connected set of simple digital lines, qualitatively analogous to the skeleton representation computed in the case of binary images. The pattern is regarded as constituted by a number of regions with constant gray-value, and the skeletonization is based on the detection of suitable pixels on its Distance Transform, computed according to the city-block distance. The set delineating the pattern is found by reducing the set of the skeletal pixels to a one-pixel-thick set, and by pruning part or all of its peripheral branches.

1 Introduction

A stylized version consisting of digital lines has often been used to represent a digital pattern in the case of binary images. However, since the binarization process may introduce shape distorsions that irremediably affect the presence of significant features, there is an increasing interest on the extraction of linear features directly from gray-tone images. We are interested in gray-tone images where the regions that mainly attract the attention of an observer are elongated and appear as having constant gray-value, greater than the gray-value of the neighboring regions. In the global context, the regions with locally higher gray-value could be perceived as lines whose spatial arrangement gives evidence to a still meaningful sketch of the image. By delineation of these elongated regions we mean the extraction of the set of such representative lines, that we call skeleton because of its resemblance with a similarly named representation often used in the case of binary images [1].

A gray-tone pattern has often been interpreted as a Digital Elevation Model (DEM), and the identification on it of the set of digital lines has been related to the detection of topographical features such as ridges, peaks, saddles and so on. For instance, the skeleton can be found by considering the image as a continuous surface, and by using the first and second partial derivatives of this surface to identify the skeletal pixels [2]. Alternatively, one may resort to the repeated application of local operations removing the pixels whose presence is not necessary either to ensure local pattern connectedness, or to mark a ridge configuration. As a result, the background propagates over the pattern and the erosion proceeds from lower gray levels towards higher gray levels until the skeleton is found as a one-pixel-thick subset, possibly in correspondence with locally higher intensity regions [3].

We find the skeleton by looking for specific configurations in the Distance Transform of the pattern. In the case of binary images, the structure of the Distance Transform shows the mode in which the background propagates over the pattern, and gives evidence to the regions where interaction occurs between wave fronts coming from different parts of the background. According to Blum [4], these regions are perceptually meaningful linear sets.

In our case, we regard the pattern as piecewise constant [5], and compute for each region with constant gray-value the Distance Transform, which is of the constrained type [6] whenever there exist adjacent regions with higher gray-value. Computation of the Distance Transform is accomplished according to the city-block distance, by ordered propagation over regions with increasing gray-value.

With respect to the traditional approach to skeletonization based on the Distance Transform, e.g., [7], additional conditions are devised to detect the pixels placed in the saddle configurations of the gray-tone pattern, and to ensure skeleton connectedness.

2 Preliminaries

Let us consider a gray-tone digital image on the square grid, where the pixels are assigned one out of a finite number of integer values, increasing from value 0. The set of pixels with value 0 constitutes the background B, while every connected set of pixels with positive value constitutes a component of the foreground F. Let g_k, k=1,N, denote the gray-value of the pixels of F, and $g(x)$ denote the gray-value of the entity x, where x is either a pixel or a set of pixels. The 8-connectedness and the 4-connectedness are respectively chosen for F and for B. Pixels are denoted by letters, which may also indicate either the corresponding gray-value in the image or the corresponding distance label in the Distance Transform.

The neighbors of a pixel p in F are its 8-adjacent pixels. They constitute the neighborhood $N(p)$ of p and are denoted by $n_1, n_2, ..., n_8$, where the subindexes increase clockwise from the pixel n_1 placed to the left of p.

A path between two pixels p and q in F is a sequence of pixels $p_0, p_1, ..., p_{s-1}, p_s$, such that $p_0 = p$, $p_s = q$, and p_i is a neighbor of p_{i-1}, $i = 1,s$. A 4-path is a path where p_i is a 4-neighbor of p_{i-1}, $i = 1,s$.

A connected set of pixels with gray-value k is said a region at elevation k.

A 4-connected region at elevation k from which any 4-path to pixels not in the region necessarily includes pixels with gray-value greater than k is termed a hollow at elevation k. It is denoted by H_k.

The depth of H_k is the difference in gray-value (h-k), where h is the minimal among the gray-values of a pixel p_r in any 4-path from H_k to pixels not in the region, such that $p_0 \leq p_1 \leq ... < p_s \leq ... \leq p_{r-1} \leq p_r > p_{r+1}$, for p_0 in H_k and p_s not in H_k.

An 8-connected region at elevation k from which any path to pixels not in the region necessarily includes pixels with gray-value smaller than k is termed a plateau at elevation k. It is denoted by PL_k.

The height of PL_k is the difference in gray-value (k-h), where h is the maximal among the gray-values of a pixel p_r in any i-th path from PL_k to pixels not in the region, such that $p_0 \geq p_1 \geq ... > p_s \geq ... \geq p_{r-1} \geq p_r < p_{r+1}$, for p_0 in PL_k and p_s not in PL_k.

Let X and X^c respectively denote a region of F and its complement, moreover let $X^c \supseteq R$, with R a set adjacent to the whole border of X. The Distance Transform of X with respect to the reference set R is the multi-valued set DT(X,R), which differs from X in having each pixel labeled with its distance from R, computed according a chosen distance function. If R is not adjacent to the whole border of X, the transform is called the constrained Distance Transform of X. Examples of Distance Transforms are shown in Figure 1.

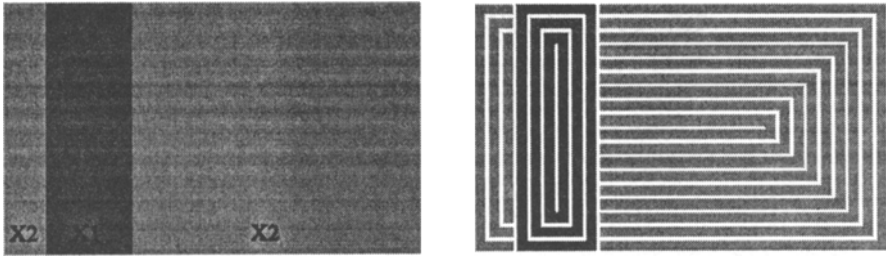Fig. 1. Distance Transforms of the sets X1 and X2, where g(X1)>g(X2), computed with respect to the sets of pixels with gray-value less than g(X1) and less than g(X2), respectively. For a set for which the Distance Transform is constrained, as in the case of X2, the interaction between different wave fronts is as larger as more protruding is the set with respect to the adjacent region(s) with greater gray-value.

We regard the gray-tone pattern as the union of a number of connected single-valued regions, on each of which we compute the Distance Transform with respect to a reference set constituted by the pixels adjacent to the region and belonging to the regions with lower gray-value. We distinguish three types of regions, depending on the gray-value of the regions that are adjacent to the region taken into account:

type 1. all the adjacent regions have smaller gray-value.

type 2. only some of the adjacent regions have smaller gray-value, and with
     respect to the set of the adjacent regions with greater gray-value either

   2a. the region is a protruding extension of that set, or

   2b. the region is not a protruding extension of that set.

type 3. all the adjacent regions have greater gray-value.

It is straightforward to observe that for the regions of *type 1*, the Distance Transform is unconstrained since there are no adjacent regions with higher gray-value. These regions are characterized by locally higher intensity and will certainly include a skeleton branch.

For any region of *type 2*, the Distance Transform is constrained and its computation leads to a pattern of propagating wave fronts where the extent of interaction between the wave fronts is as greater as more protruding is the region itself. See Figure 1. Skeleton subsets will be found where such an interaction occurs.

Let X and Y denote two regions of F, with $g(X)<g(Y)$, and let L_X and L_{XY} be the lengths of the border of X and of the part of the border of X adjacent to Y, respectively. We say that X strongly protrudes from Y if $(L_X-L_{XY}) > 3L_{XY}$. Any region of *type 2* appears as perceptually dominated by the adjacent regions with higher gray-value, and the strength of this dominance may be defined as greater as less protruding is the region itself. We expect to find significant skeleton subsets only in correspondence with the regions which are not strongly dominated. In fact, a protruding region of *type 2a* is naturally perceived as having a locally higher gray-value since in turn, contrary to a region of *type 2b* , it dominates some large enough surroundings.

Finally, for the regions of *type 3* the Distance Transform cannot be computed since the reference set is empty.

Any region of *type 3* is a hollow in the DEM, and we require that its presence be reflected by the skeleton structure only if its depth is significant, i.e., is greater than a threshold value D* depending on the problem domain. When the depth is significant, we regard the hollow as a component of the background. When it is not significant, we change its gray-value to that of the adjacent region with lowest gray-value, and repeat this merging

with the neighboring regions until the hollow is no longer detected. In the first case, we will have the hollow represented by a loop of the skeleton; in the second case, the hollow is filled in and does not influence the topology of the skeleton .

Hollow Detection. Any pixel p such that $p \leq n_i$, i odd, is first identified by a suitable marker. Then, the pixels not belonging to any hollow (i.e., greater than at least a 4-neighbor) are used as seeds from which propagation is accomplished through successive 4-neighbors so as to remove the marker from the pixels of the region with the same gray-level. Every connected set of marked pixels still remaining in F is a hollow.

The hollows with depth not smaller than D* are regarded as components of the background, while the remaining hollows are repeatedly filled in until their depth decreases to zero. To this purpose, the hollows' depth is decreased D*-1 times by increasing D*-1 times the gray-value of the pixels in each hollow. Hollow detection is repeated after each unit increase, to check whether not significant hollows still survive.

Plateau Detection. To ensure perceptual significance to the delineating digital lines, these should be detected in correspondence of sufficiently high plateaux (regions of *type 1*). To this purpose, using a technique similar to the filling in of not significant hollows, a preliminary step of the procedure is devoted to diminish, for an *a priori* given amount H*, the height of the plateaux so that the low ones disappear. The initial gray-value of the pixels of the survived plateaux is then restored by adding H*-1 units to their current gray-value.

3 Method

To avoid finding a skeleton structure too busy, it is convenient that the gray-tone pattern be cleaned of narrow regions of *type 1* and *type 3*. To this purpose a preprocessing, similar to "salt and pepper" noise removal performed when skeletonizing binary images, is accomplished by using shrink and expand operations [8]. Then, after the change of the status of the hollows and of the plateaux, the successive steps are concerned with the distance transformation and the skeletonization. They are briefly described in what follows.

Distance Transformation. The computation is accomplished on the array A, where the image is stored, and is performed according to the city-block distance, by ordered propagation over regions with ascending gray-value. Queues, i. e., first-in-first-out data structures, are employed to this purpose. As a result, the pixels in each region receive a distance label equal to their geodesic distance from a reference set constituted by the adjacent pixels belonging to the regions with lower gray-value. The obtained labeling of the gray-tone pattern originates its Distance Transform, which is the union of the (constrained and unconstrained) Distance Transforms of the regions constituting the pattern itself.

The gray-values g_k (k=1,...,N) of the pattern are detected and stored in increasing order in the 1-dimensional array G. Then, during an initialization pass the pixels interior in the pattern have their value increased of the large positive integer *untrans* ($untrans \gg g_N$). Moreover, the border points of the pattern are traced and stored in a queue Q constituted by a set of queues, a level of priority being attached to each of them (priority queue [9]).

The priority levels correspond to the increasing gray-values in the pattern, and the queue at priority level g_k contains only the pixels with gray value g_k currently under examination.

Gray-values are taken into account in increasing order. Let g_k denote the gray-value under consideration, and $dist_k$ the distance label to be updated of one unit at every propagation step. The initial value of $dist_k$ is $dist_k=0$ for k=1, and is $dist_k = dist_{k-1}$ for k>1. The w_k pixels stored in the queue at the priority level g_k are processed as follows.

For every pixel p, the corresponding value A(p) in array A is set to $dist_k+1$. The 4-neighbors of p are checked, and each neighbor p' such that: A(p')>*untrans* is inserted in the queue according to its gray-value A(p')-*untrans*. When all the initial w_k elements have been

exhausted, $dist_k$ is increased by 1 and, if new pixels have been added to the queue at the current level, the propagation is applied again to these new elements, otherwise the process is applied to the elements of the queue at the next priority level g_{k+1}.

The presence of new elements (which may be added to the queue at the current level) requires a 1-dimensional array of counters. Each elements of this array contains the updated number of pixels for the corresponding gray-value.

Skeletonization. The skeletonization of a single-valued region based on the city-block Distance Transform consists in finding the intrinsic skeletal pixels, defined by certain neighborhood configurations [7], and then growing from some of them monotonically increasing paths terminating on pixels with locally higher distance labels. For the gray-tone pattern, we follow substantially the same procedure, by taking advantage of the monotonic evolution of the distance labels towards regions with higher gray-value.
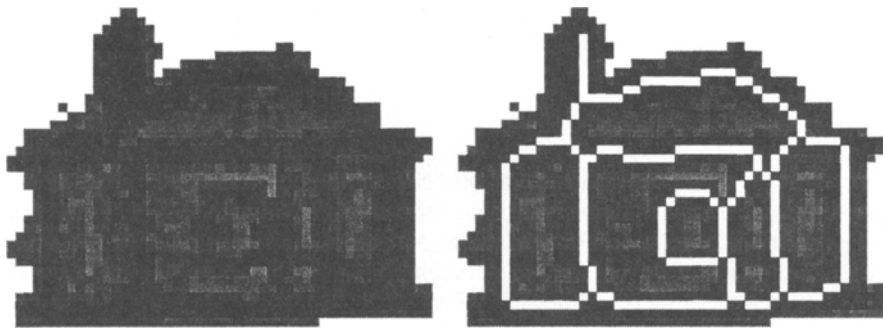

Fig. 2. A gray-tone pattern (left) delineated by its skeleton (right).

However, it is necessary to introduce new conditions in $N(p)$ to ensure the connectedness of the set of the skeletal pixels. In fact, the morphology of the Distance Transform (regarded as a DEM) is no longer as simple and regular as that of the Distance Transform of a single-valued pattern, where the constraints on the distance label distribution in $N(p)$ are well defined. As in the binary case, the set of skeletal pixels is not generally one-pixel-thick, so that removal operations are applied to it in order to obtain a set consisting of simple digital lines.

The conditions we use to detect a skeletal pixel p on the Distance Transform are:

C1. Neither the horizontal nor the vertical pair of 4-neighbors of p are such that one 4-neighbor is greater than p and the other is smaller than p.

C2. There is not exactly one pair of successive 4-neighbors equal to p.

C3. There is more than one 4-component of neighbors with value less than p.

C4. There is at least a triple of successive neighbors, starting from a 4-neighbor, which are all equal to p.

C5. There is more than one 8-component of neighbors with value greater than p.

The procedure to find the set of the skeletal pixels consists of two phases, respectively devoted to the detection of the intrinsic skeletal pixels and of the induced skeletal pixels. For the latter type of pixels, the quality of being skeletal pixels is induced by the presence in their neighborhood of at least one pixel already classified as skeletal.

The intrinsic skeletal pixels are checked parallelwise on the Distance Transform, and a pixel p is detected if the Boolean expression "(*C1* AND *C2*) OR *C3* OR *C4* OR *C5*" is true.

Any induced skeletal pixels is found as the tip of a path originated from an intrinsic skeletal pixel. The paths are grown sequentially, one after the other, and stop as soon as their tip meet some other skeletal pixel, i.e., when in the neighborhood of the tip there is more than one 8-component of skeletal pixels. A pixel p is classified as induced skeletal pixel whenever the Boolean expression "$C1$ OR $C3$" is true. When computing this expression, note that the distance label assigned to the pixels already classified as skeletal in $N(p)$ should be understood as equal to the distance label of p.

Skeleton branches which are not detected in correspondence with elongated sub-patterns with locally higher intensity are clearly superfluous and should be removed. To this purpose, we obtained satisfactory results by employing a simple criterion based on the elongation property. This causes pruning of skeleton branches by removing successive pixels with not decreasing distance labels, until a given number of successive intrinsic skeletal pixel is found. The example in Figure 2 shows a pattern taken from a set of gray-tone images digitized at 16 levels, used to test the performance of the algorithm. Darker pixels are in correspondence with higher intensity regions.

References

1. M.D. Levine: Vision in Man and Machine. New York: McGraw-Hill 1985
2. L. Wang, T. Pavlidis: Detection of curved and straight segments from gray scale topography. CVGIP:Image Understanding 58, 352-365 (1993)
3. K. Abe, F. Mizutani, C. Wang: Thinning of gray-scale images with combined sequential and parallel conditions for pixel removal. In: C. Arcelli, L.P. Cordella, G. Sanniti di Baja (eds.):Visual Form Analysis and Recognition. New York: Plenum 1992, pp 1-10
4. H. Blum: Biological shape and visual science. J. Theor. Biol. 38, 205-287 (1973)
5. A. Rosenfeld: On connectivity properties of grayscale pictures. Pattern Recognition 16, 47-50 (1983)
6. J. Piper, E. Granum: Computing distance transformations in convex and non-convex domains. Pattern Recognition 20, 599-615 (1987)
7. C. Arcelli, G. Sanniti di Baja: A one-pass two-operations process to detect the skeletal pixels on the 4-distance transform. IEEE Trans. PAMI 11, 411-414 (1989)
8. Y. Nakagawa, A. Rosenfeld: A note on the use of local min and max operations in digital picture processing. IEEE Trans. Systems, Man, and Cybernetics 8, 632-635 (1978)
9. D. E. Knuth: The art of computer programming.Vol.3. London: Addison Wesley 1975