DIGITAL TOPOLOGY

AZRIEL ROSENFELD

Digital pictures are rectangular arrays of nonnegative numbers. The analysis of a digital picture usually involves "segmenting" it into parts and measuring various properties of and relationships among the parts. In particular, one often wants to separate out the connected components of a picture subset to determine the adjacency relationships among those components, to track and encode their borders, or to "thin" them down to "skeletons" that have no interiors, without changing their connectedness properties. There are standard algorithms for doing all of these tasks; but to prove that they work, one needs to establish some basic topological properties of digital picture subsets. This paper provides an introduction to the study of such properties, which we call *digital topology*.

1. Introduction. *Digital image processing* or *picture processing* [**1**] is a rapidly growing discipline with broad applications in business (document reading), industry (automated assembly and inspection), medicine (radiology, hematology, etc.), and the environmental sciences (meteorology, geology, land-use management, etc.), among many other fields [**2**]. Most of this work involves picture analysis: given a picture, to construct a description of it in terms of the objects it contains or the regions of which it is composed and their properties and relationships. For example, a printed page is made up of characters on a background; a blood smear on a microscope slide contains blood cells on a background; a chest x-ray shows the heart, lungs, ribs, etc.; a satellite TV image of terrain is composed of terrain types; and so on. The process of decomposing a picture into regions, or into objects and background, is called *segmentation.*

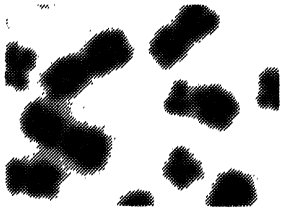
A picture is input to the computer by *sampling* its brightness values at a discrete grid of points, and *digitizing* or *quantizing* these values to a finite number of binary places. The result of this process is called a *digital picture*; it is a rectangular array of discrete values. The elements of this array are called *pixels* (short for "picture elements"), or sometimes simply *points*, and the value of a pixel is called its *gray level*. Segmentation is basically a process of assigning the pixels to classes; one simple way of doing this, called "thresholding," classifies the pixels according to whether or not their gray levels exceed a given threshold value t. Methods of segmenting digital pictures will not be reviewed here; for an introduction to this subject see, e.g., [**1**, Chapter 8].

Once a picture has been segmented into subsets, it can be described in terms of properties of these subsets and relationships among them. Some of these properties depend on the gray levels of the points that belong to a subset, but others are "geometrical" properties which depend only on the positions of these points. Especially basic are *topological* properties of the subsets, involving such concepts as adjacency and connectedness, but not size or shape.

Topological properties of digital picture subsets are useful for a number of reasons. After a subset has been singled out, e.g., by thresholding, one usually wants to further segment it into connected regions, since these often correspond to distinct objects (characters, blood cells, etc.). One may also want to track the borders of these regions, since the sequences of moves around the borders provide a compact encoding of region shape. Alternatively, one may want to "thin" the regions into "skeletons," without changing their connectedness properties, since this too yields a compact representation (e.g., an elongated region is represented by a set of arcs or curves). The adjacency or surroundedness relations among the regions can be compactly

---

Azriel Rosenfeld received his Ph.D. in mathematics from Columbia University in 1957. After ten years in the defense electronics industry, in 1964 he joined the University of Maryland, where he is Research Professor of Computer Science. He edits the journal *Computer Graphics and Image Processing* and is President of the consulting firm ImTech, Inc. He has published ten books and nearly two hundred papers, most of them dealing with the computer analysis of pictorial information.—*Editors*



(a) Photomicrograph of some chromosomes.

```
OLKMPQNKHHIJJJKLJKKJLQIQK2PJFEBDEB9
PLJKJLLJIIJJMHKJKMKLP7NNLB3KGGDCEBB
RMLKIIJKJJKLQOPLLLOU8IBBQN.IGFFCECB
QNMMJKMKKJO1D59ONMSINOOBG1JFFEECDCB
UTQOLKMMKN2LQBJQNN8NNBJ7RLGEEDDCDCC
5D91NLNOPUGQKKL9QO5LBBAQMHFEEFEEDFD
KKBDQQU8CCLOOBBANM1GNDTMKGFFDEEGGHG
NNM8S2JNBNNOOG4PIKL79UPLJHGFFEGHCRO
GLAV15NNBQN92QMKHIIKLQPLKIIIGFHQ55E
5LDUVNKLQI.QNLJIHIHJN61QQQPMIHJ1DNB
M3TRQ5NQK1ONKJHIGIFIRKN3CGE1OKLUNNI
ILQUV6D6VQMLJIFIHIGL9J5GNNQL2NMALNG
JKSC582UQMLLKKHIGIINNNJCJNQBGPLV55C
IM5LONG4TOMLJJJIGJIKT374NBQLBNKIGJK
IM5NBSBN3VTQMKJIGJKJKKMNSCKGRIHFEDE
HK7NBNQIKIDBUOKKJKJHKKKIJLPPLHGECEC
ILR8LBNNQBBQGUMKJIIKJJLJIJJIHGFECDB
KLNP.BDIQLLBNUMKIHJILVUNJHHHHGEECCC
NPRSUV.5INNNERKIHHGJQJNTLHFHHHGGFEC
.B64BA2S.5NN4NJHFHHK4INESIIKLKHIFEC
BONKOOIVQQ1SMKHHFGHNBLKL4LMT3SNGGDC
8NQLNLJVNLKKHHHHHIJKRENJQMUJNI.LFFD
7LEQQQCMLIIIHIJLOOMKIT1NLPCNLBGMGFF
MSQ7A8VLJHGHIJNVCETKHFFIKUNNNQGOIDD
HJHHILMKIHHIJO.JQQ5MHEEGK4NNQN3LGDE
```

(b) Array of numerical values obtained by coarsely sampling (a). Characters blank, 1, 2,...,9, A, B,...,V represent gray levels 0,1,...,31; the same characters overstruck with periods represent levels 32, 33,...,63.

```
                    111
                    1111
                  111111
          111     111111
          1111    11111
111      111111   11111
1111   111111111   111
1111  11111111      1
111  1111111                    111
111  11111                      111
      111            11 111     111
      1             11111111   1111
   111              111111111    11
   1111                 11111
  111111                 111
   111111111
   1111111111
     11111111
        11111        11
 1  11   111         111
1111111              1111
1111111               11     111
 111111                     11111
    11          11          11111
               111          1111
```

(c) Result of thresholding (b) at gray level 40; values <40 are displayed as blanks, values ≥ 40 as 1's.

```
007654565533211012
00676454545552321100111
007656442222
006666434122
071007765443344421
0077700076654433443423 12
07654312
701076654344321
00765444221
06441
```

(d) Move sequences representing the borders of the connected components of 1's in (c); see the end of Section 5 for the code that is used. The starting point for each border is at the leftmost of its uppermost points, and the borders are followed clockwise. Note that two of the components (near the upper left) are diagonally adjacent.

FIG. 1

represented by a graph whose nodes are the regions, and in which two nodes are joined by an arc iff those two regions are adjacent.

Many algorithms exist for segmenting a picture subset into its connected components, border following, thinning, and constructing the adjacency graph of a partition of a picture; see, e.g., [**1**, Chapter 9]. To prove that these algorithms work correctly, or even (in some cases) to state them precisely, it is necessary to establish some of the basic topological properties of digital picture subsets. This paper provides an introduction to the study of such properties, which we call *digital topology*. Of course, this is nothing more than the study of some simple properties of finite sets of lattice points; but it should be of interest because of the widespread practical use of these ideas in digital image processing.

Section 2 introduces the concepts of adjacency and connectedness for digital picture subsets. Section 3 defines digital arcs and curves, and develops some of their basic properties. Section 4 deals with thinning, Section 5 with border following, and Section 6 with adjacency and surroundedness. Only sketches of the proofs are given; a more detailed treatment can be found in [**3**, Chapter 2], which in turn is based on the material in [**4**]–[**7**].

Figure 1 shows a picture of some chromosomes, the corresponding array of coarsely sampled numerical values, a thresholded version of the picture containing ten connected objects, and a list of the move sequences representing the borders of these objects.

2. Connectedness. We begin by formulating the concept of connectedness for subsets of a digital picture Π. For concreteness, we assume that Π is an array of lattice points having positive integer coordinates (x,y), where $1 \leqslant x \leqslant M$, $1 \leqslant y \leqslant N$.

DEFINITION 2.1a. The *4-neighbors* of (x,y) are its four horizontal and vertical neighbors $(x \pm 1, y)$ and $(x, y \pm 1)$.

DEFINITION 2.1b. The *8-neighbors* of (x,y) consist of its 4-neighbors together with its four diagonal neighbors $(x+1, y \pm 1)$ and $(x-1, y \pm 1)$.

Note that if (x,y) is a border point of Π, i.e., if $x=1$ or $M, y=1$ or N, some of these neighbors do not exist. If the points P and Q of Π are neighbors, we call them *(4- or 8-) adjacent*.

DEFINITION 2.2. Let P, Q be points of Π. By a *path* from P to Q we mean a sequence of points $P = P_0, P_1, \ldots, P_n = Q$ such that P_i is a neighbor of P_{i-1}, $1 \leqslant i \leqslant n$.

Note that this is two definitions in one, depending on whether "neighbor" means 4-neighbor or 8-neighbor; we refer to these two types of paths as *4-path* and *8-path*, respectively.

Let S be a nonempty subset of Π. To avoid special cases, we assume that S does not meet the border of Π.

DEFINITION 2.3. We say that P and Q are *connected* in S if there exists a path from P to Q consisting entirely of points of S.

Again, this is two definitions in one; we speak of P and Q being *4-connected* or *8-connected*.

PROPOSITION 2.1. *"Connectedness in S" is an equivalence relation.*

DEFINITION 2.4. The equivalence classes defined by this relation are called the *(connected) components* of S. If S has only one component, it is called *connected*.

Let $\bar{S}$ be the complement of S.

DEFINITION 2.5. The unique component of $\bar{S}$ that contains the border of Π is called the *background* of S; all other components, if any, are called *holes* in S. If S has no holes, it is called *simply connected*.

We shall see shortly that when we study connectedness in digital pictures, both 4- and 8-

definitions must be used—in fact, whichever one we use for S, we must use the other one for $\bar{S}$. If Π were a hexagonal rather than a rectangular array, there would be only one type ("6") of neighbor, connectedness, etc., which we could use for both S and $\bar{S}$; but in practice, only rectangular arrays are used in digital image processing.

Our definition of connectedness is an "arcwise" definition, rather than a definition in terms of open and closed sets. We can, however, define a topology on Π in which the standard notion of connectedness reduces to our definition **[8]**. Let

$$\begin{aligned} U(P) \equiv U(x,y) &= \{P\}, \text{ if } x+y \text{ is odd} \\ &= \{P \text{ and its 4-neighbors}\}, \text{ if } x+y \text{ is even.} \end{aligned}$$

If we take the U's as a basis for the open sets, then a set is connected in the resulting topology iff it is 4-connected as defined above.

3. Arcs and curves. A commonly used method of shape analysis in digital picture processing involves reducing "thick" digital point sets to idealized "thin" forms—e.g., reducing elongated, simply connected objects to arcs, or objects that have a single hole to closed curves. We will discuss "thinning" processes of this sort in Section 4; but first we must introduce digital definitions of arcs and curves.

DEFINITION 3.1. $S \subseteq \Pi$ is called an *arc* if it is connected, and all but two of its points (its "endpoints") have exactly two neighbors in S, while those two have exactly one.

It is easily seen that an arc can be regarded as a path which neither crosses nor "touches" itself —i.e., its points can be numbered $Q_1, \ldots, Q_n$ so that Q_i is a neighbor of Q_j iff $i = j \pm 1$. To rule out degenerate cases, we shall assume that an arc always has at least two points. Readily, S cannot be both a 4-arc and an 8-arc unless it is a horizontal or vertical straight line segment.

PROPOSITION 3.1. *An arc is simply connected.*

REMARK. This proposition is not true if we use 4-connectedness for both the arc and its complement, since the 4-arc

```
PPP
P  P
PP
```

has a 4-hole. The proposition can be proved by induction on the number of points in the arc, using the fact that, if we delete an endpoint from an arc, the result is still an arc (if it has more than one point); the details, which involve an enumeration of cases, will not be given here.

DEFINITION 3.2. $S \subseteq \Pi$ is called a *curve* if it is connected, and each of its points has exactly two neighbors in S.

Readily, we can number the points of a curve $Q_1, \ldots, Q_n$ so that Q_i is a neighbor of Q_j iff $i \equiv j \pm 1$ (modulo n). To rule out degenerate cases, we will assume that a 4-curve always has at least eight points; and an 8-curve, four points. Note that no S can be both a 4-curve and an 8-curve.

PROPOSITION 3.2. *A curve has at most one hole.*

This follows from Proposition 3.1 and the fact that deleting any point from a curve makes it an arc. Note that it, too, is false if we use 4-connectedness for both S and $\bar{S}$; for example the 4-curve

```
PPP
P  PP
PP  P
  PPP
```

has two 4-holes. Indeed, as we shall next see, if we use opposite types of connectedness for S and

$\bar{S}$, then a curve has exactly one hole; but if we use 8-connectedness for both, then the 8-curve

$$\begin{array}{ccc} & P & \\ P & & P \\ & P & \end{array}$$

has no 8-holes.

THEOREM 3.3. *A curve has exactly one hole.*

This is the Jordan Curve Theorem for digital curves. The proof is similar to a standard proof of the theorem for polygons. Let S be a curve, and $P \notin S$; we say that $P \equiv (x,y)$ is "inside" S if the half-line $H_p \equiv \{(z,y) | x \leqslant z \leqslant M\}$ crosses S an odd number of times, and "outside" S otherwise. ("Crosses" must be properly defined, since H_p may meet S in runs of consecutive points; such a run is a crossing if S enters the run from the row above H and exits to the row below H, or vice versa.) It can then be shown that neighboring points of $\bar{S}$ are either both inside or both outside S; hence points in the same component of $\bar{S}$ are either all inside or all outside. The theorem follows from this and the fact that the inside and outside of a curve are both nonempty.

We can also prove

PROPOSITION 3.4. *Every point of a curve S is adjacent (in the sense of $\bar{S}$'s connectedness) to both components of $\bar{S}$.*

This follows readily from the fact that if we delete any point from S, it becomes an arc, which is simply connected.

4. Thinning. The goal of *thinning* is to remove points from a set S without changing the connectedness properties of either S or $\bar{S}$. The class of points which can be safely removed is characterized by the following proposition, in which $N(P)$ denotes the set of 8-neighbors of P.

PROPOSITION 4.1. *The following properties of the point P of S are all equivalent:*

(a) *$S \cap N(P)$ has the same number of components (in the S sense) as $S \cap [N(P) \cup \{P\}]$*

(b) *$\bar{S} \cap N(P)$ has the same number of components (in the $\bar{S}$ sense) as $[\bar{S} \cap N(P)] \cup \{P\}$*

(c) *$S \cap N(P)$ has just one component adjacent to P ("component" and "adjacent" in the S sense)*

(d) *$\bar{S} \cap N(P)$ has just one component adjacent to P ("component" and "adjacent" in the $\bar{S}$ sense)*

(e) *$S - \{P\}$ has the same number of components (in the S sense) as S, and $\bar{S} \cup \{P\}$ has the same number of components (in the $\bar{S}$ sense) as $\bar{S}$.*

DEFINITION 4.1. A point having the properties of Proposition 4.1 is called *simple*.

Evidently, an *isolated* point of S (having no neighbors in S) and an *interior* point of S (having all eight neighbors in S) cannot be simple; while an *end* point of S (having exactly one neighbor in S) is always simple. The proof of the last part of the proposition is not trivial; it requires use of Proposition 3.4.

It follows from Proposition 4.1 that if S is simply connected, and $P \in S$ is not an isolated, interior, or simple point, then $S - \{P\}$ is not connected, but consists of components that are simply connected. Using this observation, we can show, using induction on the number of points in S, that if S is simply connected and has more than two points it must have at least two simple points. In fact, we can show that if S has only two simple points, they must both be ends, and that if S has an interior point it has a simple point that is not an end.

These remarks provide a basis for defining a crude thinning process for simply connected sets S. We repeatedly delete from S simple points that are not ends; each such deletion leaves S simply connected. When no further deletion is possible, S can no longer have interior points and so is (relatively) "thin." Note that the result depends on the order in which the points are deleted. We can also establish the following neat characterization of arcs:

THEOREM 4.2. *S is an arc iff it is simply connected and has exactly two simple points.*

Analogous remarks can be made about connected S's that have only one hole; by Proposition 4.1, if $P \in S$ is not isolated, interior, or simple, then $S - \{P\}$ is either simply connected or not connected. This allows us to prove

THEOREM 4.3. *S is a curve iff it is connected, has exactly one hole, and has no simple points.*

It follows that a connected S having just one hole can be thinned to a curve by repeatedly deleting its simple points. As a further corollary, we obtain a converse to the digital Jordan Curve Theorem:

COROLLARY 4.4. *Let S be connected, let $\bar{S}$ have exactly two components, and let every point of S be adjacent (in the $\bar{S}$ sense) to both of these components; then S is a curve.*

The hypothesis that S is connected is unnecessary, as we shall see in Section 6.

5. Border following. A set $S \subseteq \Pi$ can be represented by specifying its borders; each border can be specified by defining a starting point and a sequence of moves from neighbor to neighbor. This representation, which is often quite compact, is very commonly used in image processing. In this section we define the border representation and give an algorithm for constructing it. We sketch a proof that this algorithm is valid, based on the results of Section 4.

DEFINITION 5.1. The *border* of $S \subseteq \Pi$ is the set of points of S that have 4-neighbors in $\bar{S}$.

One could also define a "thicker" border consisting of points that have 8-neighbors in $\bar{S}$; but the 4-neighbor definition is the one usually used.

The border of S consists, in general, of many parts, since S may have many components, each of which has many holes. In order to define border following, we must single out one of these parts at a time:

DEFINITION 5.2. Let C be a component of S and D a component of $\bar{S}$. The *D-border* of C is the set of points of C that have 4-neighbors in D. We denote this border by C_D.

We now describe an algorithm that successively visits all the points of the D-border of C. We assume that C is 4-connected and D 8-connected; that C has more than one point; and that we are given an initial pair of 4-neighboring points $P_0 \in C$, $Q_0 \in D$, which we assume to be distinctively marked. The algorithm, which we call BF_4, specifies how to find a new point pair (P_{i+1}, Q_{i+1}), given the current pair (P_i, Q_i).

BF_4 operates as follows: Let the 8-neighbors of P_i, in clockwise order starting with Q_i, be $R_{i1} = Q_i, R_{i2}, \ldots, R_{i8}$. Let R_{ij} be the first of the R's that is in C and is a 4-neighbor of P_i (i.e., j is odd); such an R_{ij} must exist, since C is 4-connected and has more than one point. If $R_{i,j-1}$ is in D, take R_{ij} as P_{i+1} and $R_{i,j-1}$ as Q_{i+1}; otherwise, take $R_{i,j-1}$ as P_{i+1} and $R_{i,j-2}$ as Q_{i+1}. If, for some $i > 0$, P_i is P_0 and one of $R_{i1}, \ldots, R_{ij}$ is Q_0, stop.

To illustrate the operation of BF_4, we give a simple example. Let C be the set of P's shown below; the blanks are in $\bar{S}$, while P^* is in S but not in C. Let P_0 be the P on the third row, and let Q_0 be the blank on its left. Then the successive steps of BF_4 are as follows:

Input:

```
    P*
P
P
P   P
```

Step 1.

$$\begin{array}{ccc} & & P^* \\ R_{02} & P & \\ Q_0 & P_0 & \\ & P & P \end{array}$$

Here $R_{03}=P_1$, $R_{02}=Q_1$.

Step 2.

$$\begin{array}{ccc} R_{12} & R_{13} & P^* \\ Q_1 & P_1 & R_{15} \\ & P & R_{16} \\ & P & P \end{array}$$

Here $R_{17}=P_2$, $R_{16}=Q_2$. Note that $R_{14}=P^*$ is in S, but is ignored.

Step 3.

$$\begin{array}{ccc} & & P^* \\ P & & \\ P_2 & & Q_2 \\ P & & P \end{array}$$

Here $R_{22}=P_3$, $R_{21}=Q_3$ $(=Q_2)$. Note that $P_2=P_0$, but the algorithm does not stop since Q_0 is not one of R_{21}, R_{22}, R_{23}.

Step 4.

$$\begin{array}{ccc} & & P^* \\ P & & \\ P & Q_3 & R_{32} \\ P & P_3 & R_{33} \\ R_{36} & R_{35} & R_{34} \end{array}$$

Here $R_{37}=P_4$, $R_{36}=Q_4$.

Step 5.

$$\begin{array}{ccc} & & P^* \\ & P & \\ R_{44} & P & \\ R_{43} & P_4 & P \\ R_{42} & Q_4 & \end{array}$$

Here $R_{45}=P_5$, $R_{44}=Q_5$

Step 6.

$$\begin{array}{ccc} & & P^* \\ & P & \\ Q_5 & P_5 & \\ & P & P \end{array}$$

Here $P_5=P_0$ and $Q_5=Q_0$, so the algorithm stops.

It is easily seen that the successive P_i's chosen by BF_4 are 4-connected to each other in S (though they may not be 4-neighbors); the successive Q_i's are 8-connected to each other in $\bar{S}$; and P_i is 4-adjacent to Q_i. Thus the P_i's are all in C, the Q_i's all in D, and the P_i's are all on the D-border of C. The proof that the P_i's constitute the entire D-border can be outlined as follows: Readily, the operation of BF_4 is unaffected if all points of $\bar{S}$ except those in D (and in the background component) are transferred from $\bar{S}$ to S; hence it suffices to prove the assertion for C's that have at most one hole. For simply connected C's, we can use induction on the number of points in C; by Section 4, C has simple points, and readily if BF_4 works when a simple point

is deleted, it still works when the point is present. For C's with one hole, we can first show that BF_4 works if C is a curve and then use a similar induction argument based on Theorem 4.3.

The algorithm ("BF_8") for the case where C is 8-connected and D 4-connected is very similar. Here we simply let R_{ij} be the first of the R's that is in C, and take $P_{i+1}=R_{ij}$, $Q_{i+1}=R_{i,j-1}$. Thus P_{i+1} is an 8-neighbor of P_i, and Q_{i+1} is 4-connected in $\bar{S}$ to Q. Incidentally, our choice of clockwise order for the R's implies that borders are followed keeping C on the right; thus the outer border of C is followed clockwise, and its hole borders counterclockwise. (On the meaning of "outer border" and "hole border," see Proposition 6.2.)

Since the successive P_i's chosen by BF_4 or BF_8 are 8-neighbors of each other, we can specify the D-border of C by giving the position of the starting point P_0 together with a string of 3-bit numbers $(0,\ldots,7)$ representing the moves from one P_i to the next. For example, we can use the code

$$\begin{matrix} 3 & 2 & 1 \\ 4 & P_1 & 0 \\ 5 & 6 & 7 \end{matrix}$$

to represent these moves (mnemonic: code i corresponds to a move in direction $45i°$). This representation is called a *chain code*.

To reconstruct C from its borders, we need to know the pair of points (P_0, Q_0) and the chain code for each border C_D. It is then straightforward to mark the points of C_D, as well as a band of points in D adjacent to C_D, for each D. When this has been done, it is easy to "color in" the interior of C. Note that if we had not marked the points in D that adjoin C, it would not be easy to decide which side of the D-border of C is interior to C.

6. The adjacency tree. Given $S \subseteq \Pi$, the components of S and $\bar{S}$ partition Π into connected regions. A useful way of (partially) describing a partition of Π is in terms of its *adjacency graph*, which specifies the regions and their adjacencies. When the partition consists of the components of a set and its complement, we can show that its adjacency graph is a tree. It can also be shown that if a component of S and a component of $\bar{S}$ are adjacent, one of them surrounds the other; thus, under the relationship "surrounds," the tree becomes a directed tree. In this section we define these concepts more precisely and sketch the proofs of these assertions (which, incidentally, are true only when we use opposite types of connectedness for S and $\bar{S}$).

DEFINITION 6.1. Let $\mathfrak{S} \equiv \{S_1,\ldots,S_n\}$ be a partition of Π. The *adjacency graph* $\mathcal{G}$ of this partition is the graph whose node set is $\mathfrak{S}$, and in which two nodes S_i, S_j are joined by an arc iff the sets S_i and S_j are adjacent (i.e., some point of S_i is a neighbor of some point of S_j).

When $\mathfrak{S}$ consists of the connected components of S and $\bar{S}$, we shall denote its adjacency graph by $\mathcal{G}_s$. In this case it does not matter whether we use 4-neighbors or 8-neighbors to define the adjacency relationship, since if a component of S and a component of $\bar{S}$ are 8-adjacent, they must also be 4-adjacent.

THEOREM 6.1. *$\mathcal{G}_s$ is a tree.*

Sketch of proof: We must show that $\mathcal{G}_s$ does not contain a cycle. Let T be a component of S and U, V components of $\bar{S}$ that are adjacent to T (or vice versa); then any path from U to V, in the sense of the connectedness of $\bar{S}$, must meet T, since otherwise the regions encountered by the path, together with T, would constitute a cycle. If we knew that U and V had to be in different components of $\bar{T}$, then no path between them could lie entirely in $\bar{T}$. Suppose they were in the same component W of $\bar{T}$; since they are both adjacent to T, they would both have to meet W_T, the T-border of W. But since BF works, we know that W_T is connected (in the $\bar{S}$ sense); and evidently $W_T \subseteq \bar{S}$, since points of $\bar{T}$ that are adjacent to the component T of S cannot be in S. Thus U and V cannot both meet W_T, since they are different components of $\bar{S}$;

hence they cannot be in the same component of $\bar{T}$, so that T separates them, which proves that $\mathcal{G}_s$ has no cycles.

DEFINITION 6.2. Let A, B be any subsets of Π. We say that A *surrounds* B if any 4-path from B to the border of Π meets A.

PROPOSITION 6.2. *Let C, D be adjacent components of S, $\bar{S}$, respectively; then either C surrounds D or D surrounds C. Moreover, exactly one component of $\bar{S}$ surrounds each component of S (and vice versa, for non-background components of $\bar{S}$).*

Sketch of proof: As seen in the proof of Theorem 6.1, two D's cannot be in the same component of $\bar{C}$; hence at most one D can be in the background component, so that all others are in holes and so are surrounded by C. On the other hand, there does exist a D_0 not surrounded by C (e.g., the point just to the right of a rightmost point of C is in such a D_0). On any 4-path from C to the border of Π, let P_i be the last point of C; then P_{i+1} is in some D, but is not surrounded by C, hence is in D_0, so that D_0 surrounds C. These observations also give us

THEOREM 6.3. *Under the relation "surrounds," $\mathcal{G}_s$ can be regarded as a directed tree, with the background component of $\bar{S}$ as root.*

We can also prove the promised stronger version of Corollary 4.4. Let U, V be components of S and W, Z components of $\bar{S}$; then U, V cannot both be adjacent to both W and Z, since this would imply that $\mathcal{G}_s$ contained the cycle U, W, V, Z. Thus if $\bar{S}$ has two components, and every point of S is adjacent to both of them, S cannot have two components, and so is connected. This proves

PROPOSITION 6.4. *If $\bar{S}$ has two components, and every point of S is adjacent to both of them, S is a curve.*

7. Concluding remarks. Many other topics could have been included in this paper; the study of geometrical properties of digital picture subsets ("digital geometry") has a rapidly growing literature. Some additional references on digital topology are [**9**], on homotopy; [**10**], on dimension; [**11**], on genus; [**12**]–[**14**], on shrinking, and [**15**]–[**16**], on thinning; as well as [**17**], which provides an alternative approach to some of the basic results. Other areas of digital geometry deal with natural metrics on digital pictures [**18**]; with perimeter and diameter measurement [**19**]–[**21**] and the isoperimetric inequality [**22**]; and with geodesics [**23**]. There is also considerable literature on convexity (e.g., [**24**]–[**25**]: When can a digital object be the digitization of a convex object?) and straightness (e.g., [**26**]: When can a chain code be the digitization of a straight line?). It is hoped that this paper will help bring this work to the attention of mathematicians.

The support of the National Science Foundation under Grant MCS-76-23763 is gratefully acknowledged, as is the help of Ms. Kathryn Riley in preparing this paper.

References

1. A. Rosenfeld and A. C. Kak, Digital Picture Processing, Academic Press, New York, 1976.

2. A. Rosenfeld, ed., Digital Picture Analysis, Springer, Berlin, 1976.

3. _____, Picture Languages, Academic Press, New York, 1979.

4. _____, Connectivity in digital pictures, J. Assoc. Comput. Mach., 17 (1970) 146–160.

5. _____, Arcs and curves in digital pictures, J. Assoc. Comput. Mach., 20 (1973) 81–87.

6. _____, Adjacency in digital pictures, Information and Control, 26 (1974) 24–33.

7. _____, A converse to the Jordan Curve Theorem for digital curves, Information and Control, 29 (1975) 292–293.

8. F. Wyse et al., Solution to Problem 5712, this MONTHLY, 77 (1970) 1119.

9. J. C. Alexander and A. Thaler, The boundary count of digital pictures, J. Assoc. Comput. Mach., 18 (1971) 105–112.

10. J. P. Mylopoulos and T. Pavlidis, On the topological properties of quantized spaces, J. Assoc. Comput. Mach., 18 (1971) 239–254.

11. S. B. Gray, Local properties of binary images in two dimensions, IEEE Trans. Computers, 20 (1971) 551–561.

12. M. L. Minsky and S. Papert, Perceptrons: An Introduction to Computational Geometry, MIT Press, Cambridge, Mass., 1969.

13. S. Levialdi, On shrinking binary picture patterns, Comm. ACM, 15 (1972) 7–10.

14. C. V. Kameswara Rao, B. Prasada, and K. R. Sarma, A parallel shrinking algorithm for binary patterns, Comp. Graphics Image Proc., 5 (1976) 265–270.

15. R. Stefanelli and A. Rosenfeld, Some parallel thinning algorithms for digital pictures, J. Assoc. Comput. Mach., 18 (1971) 255–264.

16. A. Rosenfeld, A characterization of parallel thinning algorithms, Information and Control, 29 (1975) 286–291.

17. S. Yokoi, J. Toriwaki, and T. Fukumura, An analysis of topological properties of digitized binary pictures using local features, Comp. Graphics Image Proc., 4 (1975) 63–73.

18. A. Rosenfeld and J. L. Pfaltz, Distance functions on digital pictures, Pattern Recognition, 1 (1968) 33–61.

19. A. Rosenfeld, A note on perimeter and diameter in digital pictures, Information and Control, 24 (1974) 384–388.

20. Z. Kulpa, Area and perimeter measurement of blobs in discrete binary pictures, Comp. Graphics Image Proc., 6 (1977) 434–451.

21. P. V. Sankar and E. V. Krishnamurthy, On the compactness of subsets of digital pictures, Comp. Graphics Image Proc., 8 (1978) 136–143.

22. A. Rosenfeld, Compact figures in digital pictures, IEEE Trans. Systems, Man, and Cybernetics, 4 (1974) 221–223.

23. _____, Geodesics in digital pictures, Information and Control, 36 (1978) 74–84.

24. J. Sklansky, Recognition of convex blobs, Pattern Recognition, 2 (1970) 3–10.

25. L. Hodes, Discrete approximation of continous convex blobs, SIAM J. Appl. Math., 19 (1970) 477–485.

26. A. Rosenfeld, Digital straight line segments, IEEE Trans. Computers, 23 (1974) 1264–1269.

COMPUTER SCIENCE CENTER, UNIVERSITY OF MARYLAND, COLLEGE PARK, MD 20742.

FINDING HOW MANY ROOTS A POLYNOMIAL HAS IN (0, 1) OR (0, ∞)

JOHN W. PRATT

1. Introduction and Summary. A real polynomial is given, perhaps of high degree. We are interested in its roots either in (0, 1) or (0, ∞). We suspect there is either none or one. Is there an easier way to prove it than Sturm's theorem if Descartes's rule does not suffice? This paper gives one, involving little more than simple addition.

The question was motivated by "internal rates of return" or "equivalent annual interest rates." Let $c_0, c_1, \ldots, c_n$ be a given sequence of positive and negative cash flows, representing the anticipated after-tax returns at times $0, 1, \ldots, n$ of a project under consideration, or the differences between those of two projects to be compared. At interest rate r, the total accumulated value of the c_j at time n would be $\sum c_j(1+r)^{n-j}$. The corresponding value in present dollars, or "present value," is $\sum c_j x^j$, where $x = 1/(1+r)$ is the one-period "discount factor"; it has advantages of vividness, comparability across horizons n, and for our purposes, of coefficients which are directly at hand. As the appropriate rate r is often debatable, special interest attaches to rates at which the present or equivalently accumulated value is 0, called "internal rates of

The author received his Ph.D. from Stanford (in Statistics) with S. Karlin. After being a Research Associate in Statistics at the University of Chicago, he joined the faculty at Harvard, where he is now a professor in the Graduate School of Business Administration. His research interests are in decision theory, statistics, and probability.—*Editors*