Computer Vision and Image Understanding **86,** 32–51 (2002)
doi:10.1006/cviu.2002.0971

Recovering a Polygon from Noisy Data[1]

Longin Jan Latecki

Department of Computer and Information Sciences, Temple University, Philadelphia, Pennsylvania 19122-6083
E-mail: latecki@temple.edu

and

Azriel Rosenfeld

Center for Automation Research, University of Maryland, College Park, Maryland 20742-3275
E-mail: ar@cfar.umd.edu

Received November 16, 1999; accepted July 30, 2002

Many classes of scenes contain objects that are (approximately) two-dimensional polygons—for example, buildings in an aerial photograph, or flat mechanical parts on a tabletop. This paper deals with the problem of recovering (an approximation to) an unknown polygon from noisy digital data, obtained by digitizing either an image of the (solid) polygon or a sequence of points on its boundary. Note that our goal is to obtain an approximation to the original polygon, not an approximation to the noisy data. We derive constraints on the polygon and on the noisy digitization process under which (approximate) recovery of the polygon is possible. We show that if these constraints are satisfied, the desired approximation can be recovered by selecting a subset of the data points as vertices. We define a vertex elimination process that accomplishes this recovery and give examples of successful recovery of both synthetic and real noisy polygons.

1. INTRODUCTION

This paper deals with the problem of recovering an unknown polygon from noisy digital data, obtained by digitizing either an image of the (solid) polygon or a sequence of points on its boundary.

Many classes of scenes contain objects that are (approximately) two-dimensional polygons—for example, buildings in an aerial photograph, or flat mechanical parts on a

[1] The support of the first author by the German Research Foundation is gratefully acknowledged, as is the help of Janice M. Perrone in preparing this paper.

1077-3142/02 $35.00

tabletop. Given an image of such an object, we can obtain a digital polygon from the image in either of two ways:

(a) Selecting (e.g., by hand) a sequence of points on the boundary of the object, and digitizing the coordinates of these points.

(b) Digitizing the image and thresholding it to segment the object from its background.

In either case, the digital polygon will generally have many more vertices than the original polygonal object. In case (b), the thresholded digital image will contain a simply connected region corresponding to the object, and unless the sides of the object are parallel to the coordinate axes, they will give rise to "staircases" on the border of the region. In case (a), it will usually be desirable to select several points along each side of the object in order to locate the side accurately. Thus in both cases, the digital polygon can be regarded as a "noisy" version of the original polygon. Note that in both cases, the vertices of the digital polygon, like those of the original polygon, are cyclically ordered.

Let P be the original polygon, and let Q be a digital polygon obtained from P by either of the two methods described in the preceding paragraph. Evidently, Q does not determine P, since many different P's could have given rise to Q; thus we cannot expect to be able to unambiguously "recover" P from Q. We can, however, attempt to recover an approximate version of P. Specifically, let us call P' an ϵ-*recovery* of P if there is a one-to-one correspondence between the vertices of P and the vertices of P' such that corresponding vertices are at most ϵ apart. Figure 1 shows an example of an ϵ-recovery of a polygon P from a noisy "hand" digitization of P.

A possible approach to obtaining P' from Q might be to construct P' by fitting lines to the vertices of Q. There is an extensive literature on methods of fitting polygons to given sets of points; for reviews see [1–3]. In most or all of this literature, however, it is not

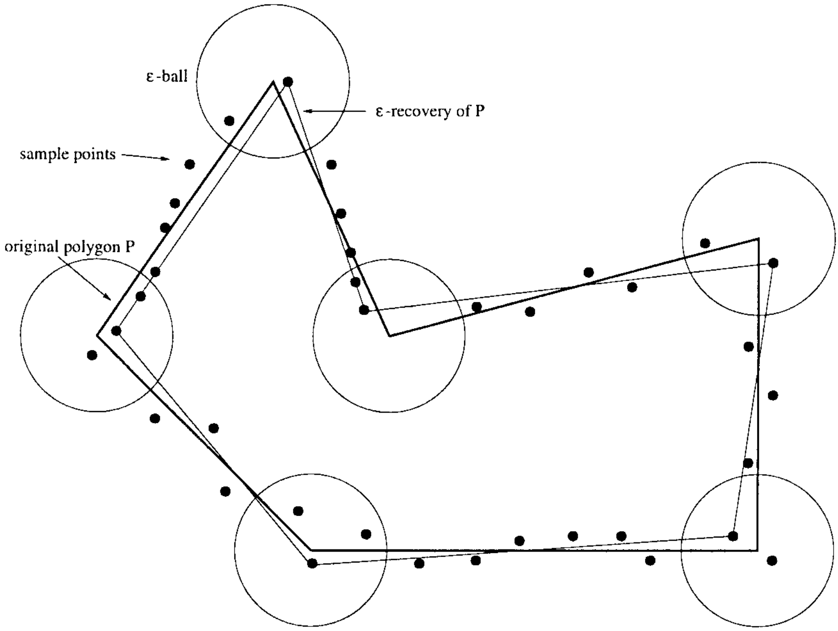

FIG. 1. An ϵ-recovery of a polygon P. Heavy line: The original polygon P. Dots: Points obtained by noisy "hand" digitization of P. Light line: an ϵ-recovery of P; note that it has one vertex within ϵ of each vertex of P.

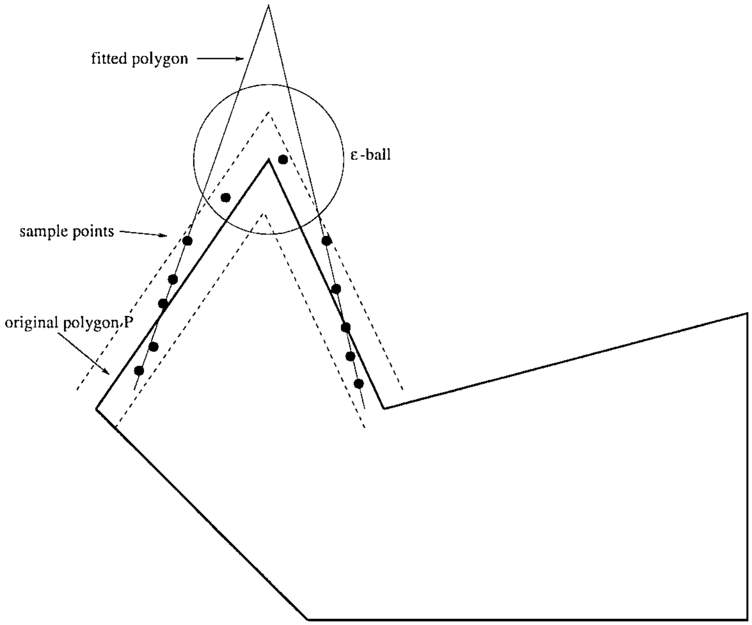

FIG. 2. When lines are fitted to a set of points that lie close to a polygon P, the vertices of the resulting fitted polygon can be arbitrarily far from the vertices of P. Heavy line: The original polygon P. Dots: Points, all of which lie close to two sides of P (within the zone bounded by the dotted lines). The light lines are a good fit to the dots, but their intersection is far from the vertex of P where the two sides meet.

assumed that the given points arise from a (noisy) digitization of an original polygon; the goal is to fit a polygon to the given points, not to recover an (unknown) original polygon. In any case, even if the vertices of Q all lie within ϵ of P, fitting a polygon to these vertices may not yield an ϵ-recovery of P; in fact, the vertices of the fitted polygon can be arbitrarily far away from the vertices of P, as illustrated in Fig. 2. On the other hand, if at least one vertex of Q lies within ϵ of each vertex of P, an ϵ-recovery of P can be obtained if we can somehow select the correct subset of the vertices of Q (as shown in Fig. 1, for example).

In Sections 2.1 and 2.2 (Theorems 1 and 2) we give constraints on the polygon P and the noisy digitization process (for both "hand digitization" and image digitization) that guarantee that the process preserves important geometric features of P; if these constraints are satisfied, ϵ-recovery of P is possible in principle by selecting a subset of the noisy digital data. In Section 3.1 we discuss methods of selecting vertices of a noisy polygon. In Section 3.2 we show (Theorem 3) that if Q is a noisy hand digitization of P, we can always construct an ϵ-recovery P' of P by selecting vertices from Q, provided P and the noise process satisfy appropriate constraints. We have not been able to find a corresponding construction for a noisy image digitization, but our experimental results (Section 3.3) indicate that ϵ-recovery by vertex selection is also possible for such digitizations.

2. CONSTRAINTS ON NOISY POLYGONS THAT PERMIT RECOVERY

A digitization process may fail to preserve important geometric features of a polygon, including its topology. For example, in the case of image digitization, if a side of the polygon P is shorter than the digitization grid length, it will not be detectable in the digitized polygon

Q; if two consecutive sides of P are nearly collinear, they may be indistinguisable from a single side in Q; and if two nonconsecutive sides of P are closer together than the digitization grid length (i.e., P is not "self-avoiding"), they may become adjacent, creating a hole in Q. Similar remarks apply to a noisy "hand digitization," if it displaces the boundary path of P by amounts that are too large relative to the sizes of geometric features of P. In this section we discuss constraints on P and on the digitization process that ensure that digitization preserves the important geometric features of P.

In Section 2.1 we define a noisy "hand-digitization" process that randomly displaces selected points of a closed contour (in particular, a polygonal contour), thus yielding a noisy version of the contour. In Section 2.2 we define a noisy image-digitization process that, when applied to a simply connected region (in particular, a solid polygon), randomly assigns border pixels to the region or to the background, thus yielding a noisy version of the region border. If the given contour or border is a polygon, we show how to ensure that the noisy version preserves both the topology and the geometry of the original polygon, by imposing constraints on the original polygon and the noise process, and "editing" the noisy polygon after it has been generated. As we shall see in Section 3, if these constraints are satisfied, the original polygon P can be recovered from the noisy polygon Q by selecting a subset of the vertices of Q.

2.1. "Hand Digitization"

Let P be a simple closed polygonal contour. To create a noisy version of P, we first choose a set of points $p_1, \ldots, p_n$ on P. We then randomly displace each of the p_i's by a bounded amount. For example, let δ be a vector whose length is unformly distributed in some interval $[0, d]$ and whose direction is uniformly distributed in the interval $[0, 2\pi)$. For each p_i, we randomly choose a vector δ_i from this distribution and replace p_i by $q_i = p_i + \delta_i$. We join the successive q_i's (modulo n); this defines a polygon Q which can be regarded as a noisy version of P. (The p_i's themselves may be randomly chosen points of P. If we allow their positions along P to vary by $\pm d$, we need use only displacements perpendicular to P to define the q_i's. However, complications arise with this definition if p_i is close to a vertex of P. Also, as we shall see below, it is useful to require that the p_i's be a minimum distance apart on P, which may limit our ability to displace them along P.)

Unfortunately, there can be two problems with the Q's that are generated from a given P in this way:

(a) Since Q is constructed by displacing points of P by small amounts (at most d), we would expect that Q should be everywhere close to P. Indeed, the vertices of Q must evidently lie within d of P, but as Fig. 3 shows, the sides of Q can get quite far from P.

(b) Q may not be a simple polygon; its sides may cross each other.

In the following paragraphs we discuss how these problems can be prevented or corrected.

Evidently, if we want to ensure that all of Q lies close to P, we must impose some constraint on how far apart the p_i's can get; as we saw in Fig. 3a, if two consecutive p_i's could be arbitrarily far apart along P, the side $q_i q_{i+1}$ of Q (indeed, the line segment $p_i p_{i+1}$) could get arbitrarily far away from P. On the other hand, if p_i and p_{i+1} lie on the same side s of P, they can be arbitrarily far apart, because q_i and q_{i+1} must be in the d-neighborhood of s, and since this neighborhood is convex, all of $q_i q_{i+1}$ must be in it. Thus our constraint should deal with consecutive p_i's that lie on different sides of P. As we saw in Fig. 3b,

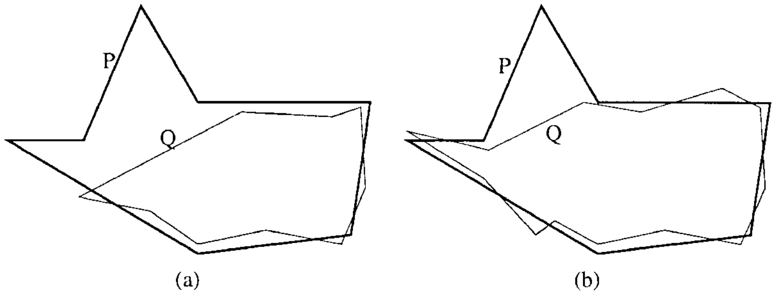

FIG. 3. The sides of Q can be arbitrarily far from P.

if two consecutive p_i's can be arbitrarily far from a vertex v and on different sides of it, $p_i p_{i+1}$ (and $q_i q_{i+1}$) can get arbitrarily far away from v. Thus we shall require that *for every vertex v of P, there must be a sample point within some distance D of it (where distance is measured along P).*

As we shall now see, this implies that every side of Q must lie within distance $D + d$ of some side of P (where distance is measured in the plane). Indeed, let p_i and p_{i+1} be two consecutive sample points. If they are both on the same side s of P, we have already seen that $q_i q_{i+1}$ lies within distance d of s.

Let them be on two consecutive sides r, s of P, and let v be the vertex at which r and s meet. Suppose neither p_i nor p_{i+1} were in the D-neighborhood of v; then their distances from v along P would both be greater than D. But since p_i and p_{i+1} are consecutive, there is no sample point between them along P; hence there is no sample point within distance D of v along P, contradiction. Hence either p_i or p_{i+1}, say the former, is in the D-neighborhood of v; hence both p_i and p_{i+1} are in the D-neighborhood of s, so that both q_i and q_{i+1} are in the $(D + d)$-neighborhood of s, and since this neighborhood is convex it follows that $q_i q_{i+1}$ is in it.

Finally, suppose they are on nonconsecutive sides of P; let t be a side between these two sides, and let u, v be the endpoints of t. Since p_i and p_{i+1} are consecutive sample points, p_i must be the last sample point preceding u, and p_{i+1} must be the first sample point following v. Since there must be sample points within D of u and v along P, and there are no sample points on t, p_i must be in the D-neighborhood of u and p_{i+1} must be in the D-neighborhood of v. Hence p_i and p_{i+1} are both in the D-neighborhood of t, so that q_i and q_{i+1} are in the $(D + d)$-neighborhood of t, and since this neighborhood is convex, it follows that $q_i q_{i+1}$ is in it. This completes our discussion of how to ensure that Q lies close to P.

Ensuring that Q is a simple polygon is more complicated. As we shall see, there are several reasons why sides of Q may cross one another. Some of these crossings can be prevented by suitably constraining P or the noise process; but others cannot be prevented, but rather must be corrected.

If successive sample points p_i are too close together (in fact, closer than $2d$), Q may cross itself, as illustrated in Fig. 4. (If we displace p_i toward p_{i+1} and p_{i+1} toward p_i, both by d, the sides that join q_{i-1} to q_i and q_{i+1} to q_{i+2} may cross.) Note that this type of crossing can occur even if p_{i-1}, p_i, p_{i+1}, p_{i+2} are all on the same side of P. It can be prevented by requiring that *successive p_i's be more than 2d apart* (*where distance is measured along P*). Conversely, if this requirement is satisfied, and p_i, p_{i+1}, p_j, p_{j+1} are all on the same side of P, evidently $q_i q_{i+1}$ and $q_j q_{j+1}$ cannot cross.

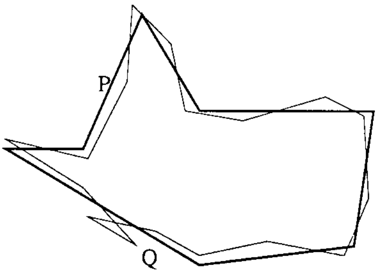

FIG. 4. If successive sample points of P are too close together, Q may cross itself.

To handle crossings that arise from p's that are on different sides of P, we must require that *nonconsecutive sides are never within distance* $2d$ *of each other* (*where distance is measured in the plane*). In general, we call P *e-self-avoiding* if, for any two disjoint sides r, s of P, no point of r is within distance e of any point of s. Note that this implies, in particular, that no vertex v of P is within e of any side s of P unless v is an endpoint of s. (In fact, if P is a simple polygon and not a triangle, this is equivalent to e-self-avoidingness.) This in turn implies that distinct vertices of P cannot be within e of one another, and this in turn implies that every side of P must have length at least e. Evidently, if P is not $2d$-self-avoiding, the sides of Q may cross. (Let r, s be disjoint sides of P that come within $2d$ of one another; then there may exist $p_i p_{i+1}$ on r and $p_j p_{j+1}$ on s such that $q_i q_{i+1}$ crosses $q_j q_{j+1}$; see Fig. 5.).

Self-avoidingness evidently cannot prevent some types of crossings; in particular, it cannot prevent a crossing when p_i, p_{i+1}, p_j, p_{j+1} are on two consecutive sides of P, since such sides are not disjoint, so self-avoidingness does not apply to them. We shall now show that if P is $2(D+d)$-self-avoiding, no other types of crossings are possible. Indeed, suppose $q_i q_{i+1}$ and $q_j q_{j+1}$ lie within distance $D+d$ of two distinct sides s, t of P, and suppose they cross. Since P is $2(D+d)$-self-avoiding, this cannot happen unless s and t share a vertex. Thus we are done unless $q_i q_{i+1}$ and $q_j q_{j+1}$ both lie within distance $D+d$ of a single side r, and neither of them lies within distance $D+d$ of any other side; but we shall now show that in this case, they cannot cross. Note first that p_i, p_{i+1}, p_j, p_{j+1} must all be either on r on the preceding or following side of P, since disjoint sides have disjoint $(D+d)$-neighborhoods. If p_{i+1} were not on r, neither would p_i be, so q_i and q_{i+1} would lie within d of the preceding side, contradiction; hence either p_i lies on the preceding side and p_{i+1} lies on r, or they both lie on r. Similarly, either p_{j+1} lies on the following side and p_j lies on r, or they both lie on r. If p_{i+1} were within distance D of the first endpoint of r, q_{i+1} would be in the $(D+d)$-neighborhood of the preceding side, contradiction; hence p_i must be within distance D of the endpoint, and similarly p_{j+1} must be within distance

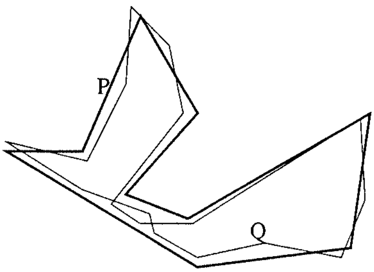

FIG. 5. If P is not $2d$-self-avoiding, Q may cross itself.

D of the second endpoint but p_j cannot be. Let A, C be perpendiculars to r at distance D from its endpoints, and let B be perpendicular to r halfway between p_{i+1} and p_j (which we know are at least $2d$ apart). No matter how sharp an angle the preceding and following sides make with r, p_i must be to the left of A and p_{i+1} to its right, and p_j must be to the left of C and p_{j+1} to its right (and this is certainly true if p_i or p_{j+1} is on r); thus B is between A and C. Hence both p_i and p_{i+1} are at least d to the left of B, and both p_j and p_{j+1} are at least d to its right; hence q_i and q_{i+1} are to the left of B and q_j and q_{j+1} are to its right, so $q_i q_{i+1}$ and $q_j q_{j+1}$ cannot cross.

We have thus shown that *if successive p_i's are more than $2d$ apart on P, P is $2(D+d)$-self-avoiding, and there is a p_i within distance D of every vertex of P*, then two sides $q_i q_{i+1}$ and $q_j q_{j+1}$ of Q can cross only if they lie in the $(D+d)$-neighborhoods of two consecutive sides s, t of P. Such crossings cannot be prevented, but we will now describe how they can be eliminated.

Evidently, such a crossing must lie in the intersection of the $(D+d)$-neighborhoods of s and t. Let this intersection be denoted by $N_{D+d}(s, t)$; we shall require from now on that *for all distinct pairs of consecutive sides of P, these N's are disjoint*. Let Q_{st} be the subset of Q contained in $N_{D+d}(s, t)$. Evidently, the Q_{st}'s are joined by pieces of Q that are simple polygonal arcs. Let Q_i and Q_j be the sides of Q that enter and leave $N_{D+d}(s, t)$. There is a shortest path in $N_{D+d}(s, t)$ that joins Q_i to Q_j, and such a path must be a simple polygonal arc; thus if we eliminate all of Q_{st} except for this path, and do this for each Q_{st}, the resulting simplified Q is a simple polygon Q'. Evidently, each side of Q' lies within distance $D+d$ of some side of P. (The sides inside each $N_{D+d}(s, t)$ that are used to simplify Q all lie within distance $D+d$ of both s and t.)

Finally, we claim that the correspondence of sides of Q' with the sides of P that they lie close to is order-preserving. Note first that by our requirement that consecutive sample points are at least $2d$ apart, when sides of Q' originate from sample points on the same side of P, their order is the same as that of these sample points. Also, by the discussion earlier in this section (showing that there must be a sample point within distance D of every vertex of P), the correspondence between the sides of Q and the sides of P that they lie close to is order-preserving. We obtain Q' from Q by eliminating sides of Q that lie close to a pair of consecutive sides s, t of P, and replacing these sides of Q by line segments (segments of sides of Q) that all lie close to both s and t. Thus any of these new sides of Q' can be associated with either s or t, and this can evidently be done so as to preserve the order of the correspondence.

The results of this section can be summarized in

THEOREM 1. *Let P be a simple polygon which is $2(D+d)$-self-avoiding, and such that for each pair of consecutive sides s, t of P, the intersection of the $(D+d)$-neighborhoods of s and t is disjoint from any other such intersection. Let $p_1, \ldots, p_n$ be points of P such that successive p_i's are at least $2d$ apart and there is at least one p_i within D of each vertex of P. Let q_i be the result of randomly displacing p_i by at most d, let Q be the polygon constructed by joining successive q_i's, and let Q' be obtained by replacing Q_{st}'s by shortest paths as described above. Then Q' is a simple polygon, and every side of Q' lies within distance $D+d$ of some side of P; moreover, this correspondence of the sides of Q' with the sides of P is order-preserving.*

Note that although Q' is close to P, it is not necessarily a D-recovery of P, since it may have too many vertices. In Section 3.2 we will show how to obtain a D-recovery of P by eliminating vertices from Q'.

2.2. Image Digitization

Our second method of creating noisy versions of a simple polygon (or other simply connected shape) is based on the concept of noisy image digitization. Let P be a simple polygon; since P is a simple closed curve, its complement consists of two connected regions, one "inside" (i.e., surrounded by) P and the other "outside" (surrounding) P. Let $\bar{P}$ be the union of P and its "inside" region; thus $\bar{P}$ is a "solid" polygon. Evidently $\bar{P}$ is a closed, simply connected region; we will refer to this region as "black," and to its complement (the "outside" of P) as "white."

To define a noisy digitization of $\bar{P}$, we divide the plane containing $\bar{P}$ into unit-square pixels p_i. A set of pixels $\bar{Q}$ is called a *noisy digitization* of $\bar{P}$ if it has the following properties: If the interior of p_i is all black, p_i belongs to $\bar{Q}$; if its interior is all white, it does not belong to $\bar{Q}$; if its interior is partly black and partly white, it may or may not belong to $\bar{Q}$. In ordinary (nonnoisy) digitization, p_i belongs to $\bar{Q}$ if (e.g.) more than half of its area is black, and does not belong to Q if less than half its area is black; exact ties are decided by some type of rounding. We can introduce nondeterminism into this digitization as follows: p_i belongs to $\bar{Q}$ if more than a certain fraction ρ of its area is black, where $\rho \leq 1/2$; it does not belong to $\bar{Q}$ if more than ρ of its area is white; but if the black fraction of its area is between ρ and $1 - \rho$, we decide (e.g.) randomly whether to make p_i black or white. If $\rho = 1/2$, this is almost the same as nonnoisy digitization, except that exact ties are resolved randomly rather than by rounding. As ρ decreases, the digitization becomes more noisy, because a larger fraction of the decisions are made nondeterministically. If $\rho = 0$, the decisions are all made in this way unless the interior of p_i is entirely black or entirely white. A noisy digitization for which $\rho = 0$ will be called a *random digitization*. [Note that the pixels whose interiors are all black certainly belong to $\bar{Q}$, and those whose interiors are all white certainly do not. We do not allow pixels that lie entirely inside or entirely outside P to be colored randomly, which would result in "salt and pepper noise" in the digital image; the nondeterminism affects only those pixels whose interiors are partly black and partly white, i.e., those pixels that intersect the boundary P of $\bar{P}$.]

Evidently, if $\bar{Q}$ is a random digitization of $\bar{P}$, every border pixel of $\bar{Q}$ either intersects P or is a 4-neighbor of a pixel that intersects P. Thus the "cracks" (unit-length horizontal and vertical line segments) that separate pixels of $\bar{Q}$ from pixels of its complement all lie within distance $\sqrt{2}$ of sides of P. Similarly, the "chain" segments that join the centers of successive border pixels of $\bar{Q}$ all lie within distance $3\sqrt{2}/2$ of sides of P if they are inside $\bar{P}$, and within distance $\sqrt{2}/2$ if they are outside $\bar{P}$. (Note that these chain segments may be either horizontal or vertical and of unit length, or diagonal and of length $\sqrt{2}$.) If we think of the successive cracks or chain segments as defining polygonal arcs, these arcs thus must be close to P.

Unfortunately, random digitization need not preserve the topology of $\bar{P}$. Thus the union of the pixels of $\bar{Q}$ may not be simply connected or even connected, and its border (whether defined by cracks or chain segments) may not be a simple polygon. For example, in the local pattern of pixels shown in Fig. 6a, suppose a side of P crosses d, e, b, and c, so that a is entirely outside P, f is entirely inside it, and b, c, d, e properly intersect it. In a random digitization of $\bar{P}$, a cannot belong to $\bar{Q}$; f must belong to it; and b, c, d, e may or may not belong to it. If b does belong to $\bar{Q}$ but e and c do not, then since the pixel above b cannot belong to $\bar{Q}$, b has no 4-neighbors in $\bar{Q}$—in other words, $\bar{Q}$ is not 4-connected. Thus even the pixels of $\bar{Q}$ that lie along a side of $\bar{P}$ need not be 4-connected. Near a vertex of P, $\bar{Q}$ need not even be 8-connected. For example, if the vertex angle is

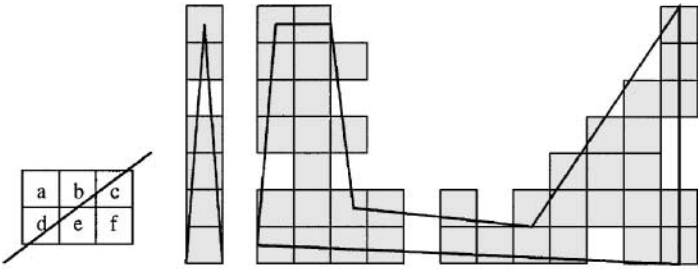

FIG. 6. Random digitization need not preserve connectedness.

very sharp (see Fig. 6b), the part of P near the vertex may be less than one unit wide, so that it intersects a sequence of single pixels. In $\bar{Q}$, these pixels can be either black or white; in particular, a pixel p close to the vertex may be black while a neighbor of p farther from the vertex may be white, so that p is not connected to the rest of $\bar{Q}$. More generally, if two sides of P come within distance $2\sqrt{2}$ of one another, so that they pass through two 4-neighboring pixels p, q (Fig. 6c), and p and q are both white, $\bar{Q}$ may be disconnected. (Note that if the two sides are nonconsecutive, this can happen only if P is not $2\sqrt{2}$-self-avoiding.)

Violations of topology that occur at corners of $\bar{P}$ are hard to prevent by constraining P, as we saw in Section 2.1. However, some types of violations can be prevented, so we next define constraints on P that prevent them. To begin with, we require that *P be $2\sqrt{2}$-self-avoiding*. This implies that the vertices of P are at least $2\sqrt{2}$ apart (so that the sides of P are at least $2\sqrt{2}$ long). In fact, we impose a stronger constraint on the separation of the vertices. Let v be the vertex where the consecutive sides s, t of P meet. Let T_v be the isosceles triangle with vertex v whose equal sides are segments of s and t and whose base has length $2\sqrt{2}$. The sides of T_v must be $\sqrt{2}\sec(\alpha/2)$ long, where α is the vertex angle of T_v; thus if α is very acute, s and t must be very long. We also require from now on that, *for all vertices u, v of P, the triangles T_u and T_v must be at least $2\sqrt{2}$ apart*. (In particular, we assume that there must be a triangle T_v at every vertex; this implies that the sides that meet at every vertex are at least $\sqrt{2}\sec(\alpha/2)$ long, where α is the vertex angle.) Under these assumptions about P, we will show in the following paragraphs that the toplogy violations in $\bar{Q}$ are isolated, and can be eliminated by locally "editing" $\bar{Q}$.

We call a pixel p *special* if p intersects a side of P, and some 2-by-2 block of pixels that contains p intersects two sides of P; otherwise, we call p *regular*. It is easy to see that if p is special, it intersects one of the equal sides of the isosceles triangle T_v associated with some vertex v of P. Thus the border of $\bar{Q}$ consists of runs of regular pixels separated by runs of pixels that intersect T_v's. We shall next show that the runs of regular pixels give rise to portions of the border of $\bar{Q}$ that are simple 4-arcs except possibly at isolated pixels, which can be locally identified.

A run of regular pixels all intersect a single side E of P. Suppose, without loss of generality, that the slope θ of E is in the first octant; i.e., $0 \leq \theta \leq \pi/4$, and $\bar{P}$ lies below E. In each column, E either intersects a single pixel p or intersects two vertically adjacent pixels p and q with p above q (i.e., E crosses the crack between p and q). In the first case, the pixels above p must be white; those below p must be black; and p itself can be either black or white. Thus in this case the column consists of a run of white pixels above a run

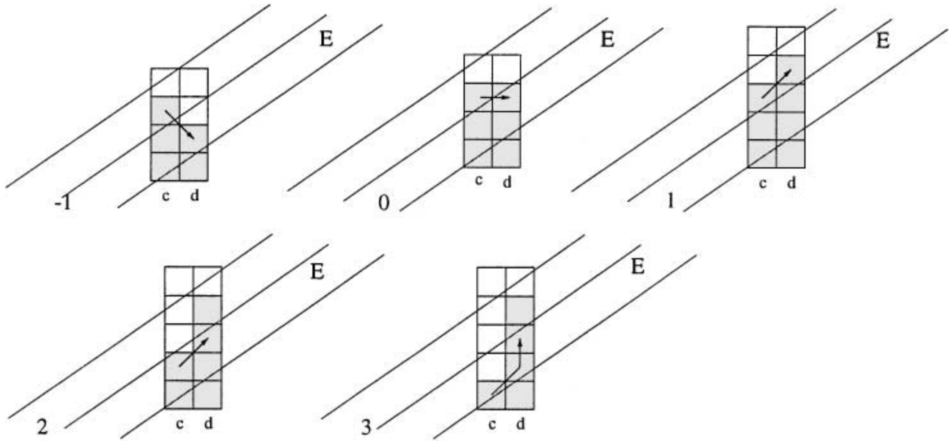

FIG. 7. Height differences between top cracks in successive simple columns.

of black pixels; we call such a column a *simple* column. In the second case, both p and q can be either black or white; the pixels above p must be white and those below q must be black. If p and q are both black or both white, or if p is white and q is black, the column is simple, but if p is black and q is white, the column consists of (from top to bottom) a run of white pixels, a single black pixel (p), a single white pixel (q), and a run of black pixels; we call such a column a *complex* column.

Define the "top crack" of a simple column as the crack that has white above it and black below it, and the "top crack" of a complex column as the upper of the two cracks that have white above them and black below them. It is not hard to see that in two consecutive columns c, d, the difference of the heights of the top cracks is between -1 and $+3$ (see Fig. 7). If c and d are both simple (see Fig. 7), any of these differences can occur, and the resulting pattern of black and white pixels never violates topology (i.e., the black pixels are always simply 4-connected).

We can eliminate all the complex columns by performing the local operation illustrated in Fig. 8. Evidently this operation does not change the 8-connectedness of $\bar{Q}$ or the distance from the border of $\bar{Q}$ to P. (The operation shown in Fig. 8 is designed to handle edges of P in the first octant. To handle the other octants, we use rotations of the operation by multiples of 90°.)

Figures 9, 10, and 11 show all the possible cases of a complex column followed by a simple column; a simple column followed by a complex column; and a complex column followed by a complex column. In Fig. 9, we see that only the height differences -1, 0, and 1 can occur; the other cases (labeled "impossible geometry") cannot occur, because the pixel marked p must be white. After performing the operation of Fig. 8, the complex columns become simple and the height differences remain the same. Note that in the -1 case, the operation also eliminates a topology violation (p is not 4-connected to the rest of $\bar{Q}$).

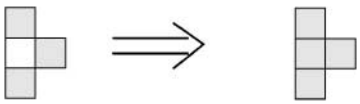

FIG. 8. Elimination of complex columns.

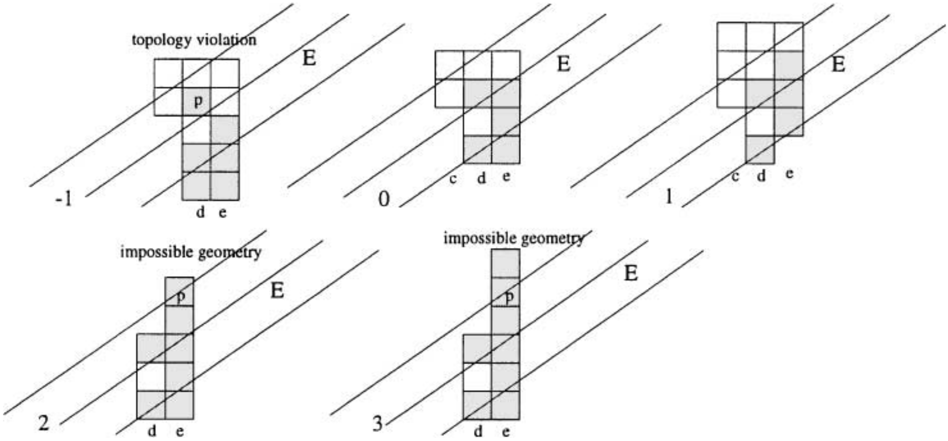

FIG. 9. Cases in which a complex column is followed by a simple column.

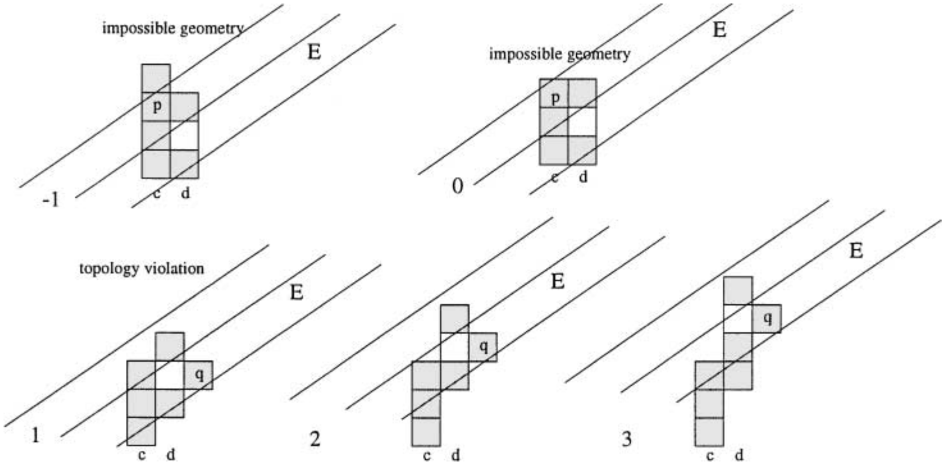

FIG. 10. Cases in which a simple column is followed by a complex column.

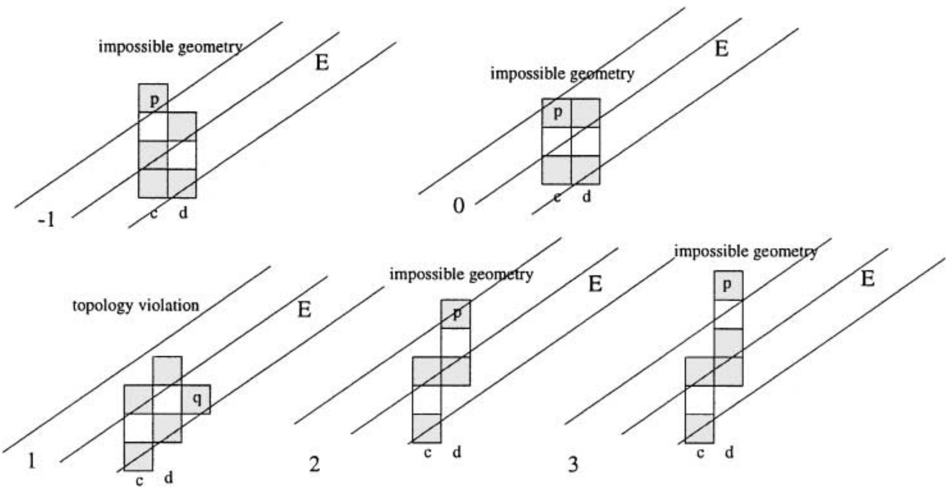

FIG. 11. Cases in which a complex column is followed by a complex column.

In Fig. 10, we see that only the differences 1, 2, and 3 can occur. Note that the pixel marked q, in the next column, must be black. Again, after performing the operation, the complex columns become simple, the height differences remain the same, and the topology violation in the $+1$ case is eliminated. Finally, in Fig. 11, only the difference $+1$ can occur; here again, the operation makes both columns simple, preserves the height difference, and eliminates the topology violation.

Let $\bar{Q}'$ be the result of applying the operation of Fig. 8 to $\bar{Q}$. Based on the legal sequences of simple columns shown in Fig. 7, we see that the chain code $C(\bar{Q}')$ of the border of $\bar{Q}'$ can make only SE, E, NE, and N moves. Hence the chain code has monotonically nondecreasing x coordinates, so that its projection on E is monotonic.

As regards the portion of the border of $\bar{Q}'$ that intersects any T_v, it can be replaced by a simple 4-arc just as we did in Section 2.1. Thus we have shown that we can edit $\bar{Q}$ to make its border a simple polygon Q.

We summarize the results of this section in

THEOREM 2. *Let P be a simple polygon which is $2\sqrt{2}$-self-avoiding. Suppose further that these exist isosceles triangles at the vertices of P that have base lengths $2\sqrt{2}$, and that any two of these triangles are at least $2\sqrt{2}$ apart. Let $\bar{Q}$ be the result of a random image digitization of $\bar{P}$. Then $\bar{Q}$ can be "edited" by performing the operation of Fig.* 8, *yielding $\bar{Q}'$, and then replacing the parts of $C(\bar{Q}')$ in the triangles by shortest paths. The result of this editing is a simple polygon Q. Moreover, all of Q lies close to P (the parts inside $\bar{P}$, within $3\sqrt{2}/2$ of P; the parts outside, within $\sqrt{2}/2$), and it can be projected onto P in an order-preserving manner.*

Here again Q is close to P, but it is not a $3\sqrt{2}/2$-recovery of P because it has too many vertices. In Section 3.3 we will show experimentally that P can be recovered by eliminating vertices from Q.

3. POLYGON RECOVERY BY VERTEX ELIMINATION

In Sections 2.1 and 2.2 we formulated conditions on the polygon P and on the hand or image digitization process that ensure that when P is digitized, the important geometric features of P are preserved. When these conditions are satisfied, Q lies close to P, and it should be possible, in principle, to obtain an ϵ-recovery of P (for some suitable ϵ) from the digitized version Q of P by selecting a subset of the vertices of Q.

In this section we show how an ϵ-recovery of P can be obtained by eliminating "inconspicuous" vertices of Q. In Section 3.1 we discuss possible measures of the conspicuousness of a polygon vertex, and define plausible constraints that such measures should satisfy. In Section 3.2 we prove that if a hand digitization Q of P satisfies the conditions in Theorem 1, a D-recovery of P can be constructed by repeatedly eliminating Q's most inconspicuous vertices. We have not been able to prove an analogous result when Q is a noisy image digitization of P, but in Section 3.3 we show experimentally that in this case too, when Q's most inconspicuous vertices are repeatedly eliminated, an acceptable recovery of P is obtained.

3.1. Vertex Conspicuousness

Let A, B, C be three consective vertices of a polygon. In triangle ABC, let the base AB have length l, and let the sides AC, BC make angles α and β, respectively, with AB, as

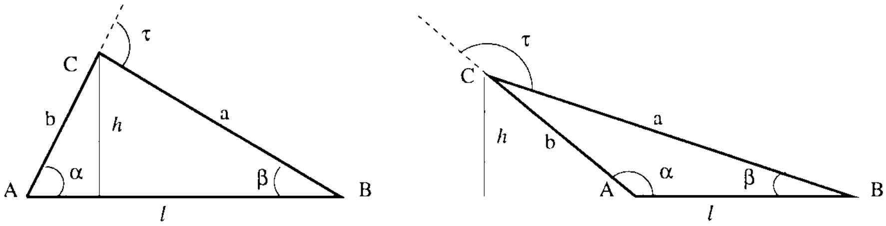

FIG. 12. Two triangles with their parts labeled.

shown in Fig. 12. The exterior angle $\tau = \alpha + \beta$ at vertex C is the *turn angle* at C; evidently we have $\tau < \pi$. The "conspicuousness" of the vertex C should be a quantity that characterizes how different the pair of sides AC and CB is from the side AB. Such a conspicuousness measure can be interpreted as the cost of eliminating vertex C and replacing the pair of sides AC, CB with the side AB. If AC and CB are consecutive sides of a polygon, the conspicuousness of C can also be regarded as a measure of the significance of the contribution of vertex C to the shape of the polygon. We now discuss plausible properties of such a measure.

Let d be the (Euclidean) distance from the vertex C to the base AB. Note that if the base angles α and β are both acute, d is the altitude h; but if one of them is obtuse, d is the length of the side that makes the obtuse angle with the base (see Fig. 12). The conspicuousness K of C may depend on the lengths l, d as well as the angles α, β. We will assume that K is a continuous function of these quantities.

Assume that K is proportional to (linear) size. Since the lengths l and d can vary independently, and both vary linearly with scale, we can study the shape dependence of K, as distinguished from its scale dependence, by considering how K varies with (say) d as l remains constant. It seems intuitively plausible that K is a monotonic function of d and goes to zero as $d \to 0$ (Fig. 13). One could also assume that K becomes arbitrarily large as $d \to \infty$ (Fig. 14), but this assumption seems to be less essential. Other plausible assumptions are that K is monotonic in α and β and that for any given value of $\tau = \alpha + \beta$, K takes on its maximum value when $\alpha = \beta$ (Fig. 15); but these assumptions too seem less essential.

Various simple geometric quantities associated with the triangle ABC have these properties. Perhaps the simplest of these is the distance d itself. To show that, for a given value of $\alpha + \beta$, d takes on its maximum when $\alpha = \beta$ we can use the facts that

$$d = h \frac{l}{(\cot\alpha + \cot\beta)} \quad \text{for}\, \alpha, \beta < \frac{\pi}{2}$$

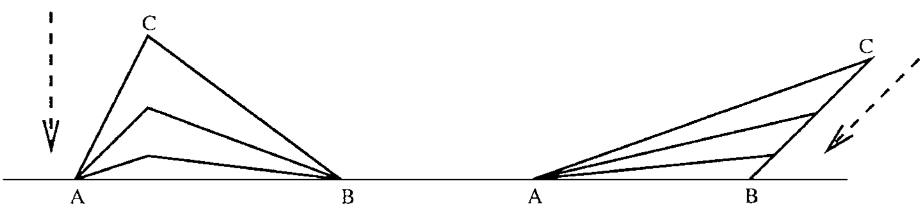

FIG. 13. Conspicuousness goes to zero as d goes to zero.

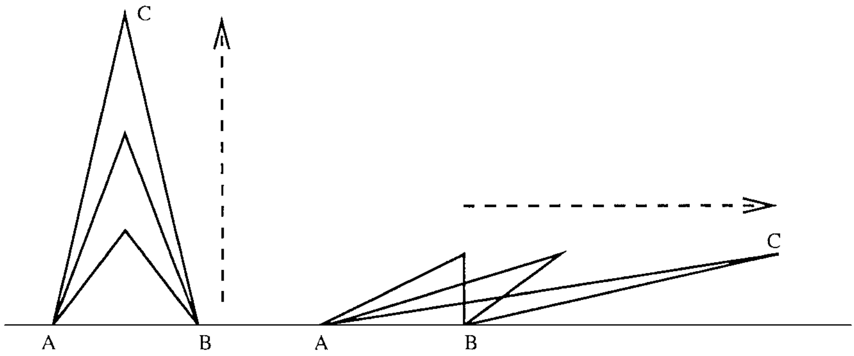

FIG. 14. Conspicuousness grows monotonically with d.

and

$$d = \frac{l \sin\beta}{\sin(\alpha+\beta)} \quad \text{for } \alpha > \frac{\pi}{2}.$$

It should be pointed out that several other quantities also have these properties. One of them is the length of the median, i.e., the line segment joining the vertex C to the midpoint of the base AB. Another such quantity, which we have used in earlier experiments [4, 5], can be defined as follows: Draw a line L through the vertex C, and imagine that each of the sides CA and CB is rotated about C until its other endpoint hits L. Let L_A and L_B be the lengths of the arcs through which CA and CB rotate (i.e., the arcs traced by A and B as they rotate). If we rotate L (around C) toward CA, L_A decreases and L_B increases, and if we rotate it toward CB, the reverse is true; thus there exists an L for which $L_A = L_B$. This common value of L_A and L_B is the K used in [4, 5]; it can be shown that its value is

$$\tau \frac{1}{\frac{1}{a}+\frac{1}{b}} = \frac{\tau ab}{a+b}.$$

Our goal in the next two sections is to show that if Q is a hand or image digitization of P, we can obtain an approximate recovery of P by eliminating the vertices of Q in order of

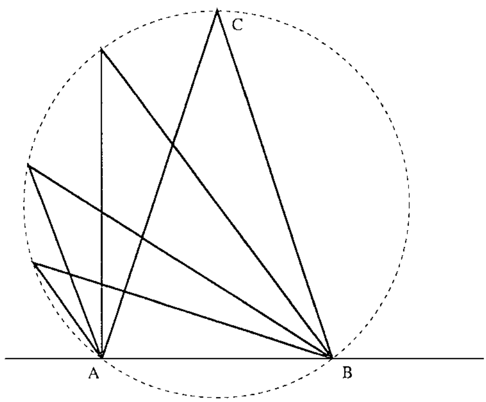

FIG. 15. For a given turn angle (or equivalently, a given vertex angle), conspicuousness may be greatest when the base angles are equal.

their inconspicuousness. To ensure that closeness to P is preserved during the elimination process, we must use a measure of conspicuousness that is a monotonic function of d and goes to zero as $d \to 0$. We have found that it makes little difference which measure we use, as long as it has these properties; in the experiments described below we will use the measure K of [4, 5].

3.2. Polygon Recovery from a Noisy Hand Digitization

In this section Q denotes a noisy hand digitization Q' of P that satisfies the constraints of Theorem 1 (Section 2.1); thus Q lies close to P. We assume that *the side lengths of Q all have length at least E*. We will prove in the following paragraphs that if we eliminate the vertices of Q in order of their inconspicuousness, then even after many of the vertices of Q have been eliminated, the simplified Q remains close to P, as long as it has at least as many vertices as the original P. Thus if P has n vertices, Q has $N > n$ vertices, and we eliminate the $N - n$ least conspicuous vertices from Q, we obtain an n-vertex polygon Q' that lies close to P. This Q' evidently must closely resemble P, so that by constructing Q' we have recovered an approximation to P from its noisy version Q.

Let A', B', C' be three consecutive vertices of Q, obtained by randomly displacing the points A, B, C of P. We call (A', B', C') a *side triangle* if A, B, C are all on the same side of P; otherwise, we call it a *vertex triangle*. We will show in the following paragraphs that

(a) the B''s of the side triangles have conspicuousnesses that are bounded above;
(b) for each vertex v of P (having sides s, t, say) there exists a vertex triangle $A'B'C'$ such that A is on s and C is on t, and the conspicuousness of B' is bounded below, and in fact is greater than the conspicuousnesses of all the side triangles.

It follows from (a) and (b) that if we eliminate the vertices of Q in the order of their inconspicuousness, for every vertex v of P there exists a vertex triangle $A'_v B'_v C'_v$ such that the vertices of all the side triangles will be eliminated before the vertices B'_v of these vertex triangles are eliminated. Thus the resulting simplified Q' will have a vertex close to each vertex of P, so that Q' is a good approximation to P.

LEMMA 1. *If $2d < E$, then the conspicuousness of every B' in a side triangle (A', B', C') is at most $2d$.*

Proof. Let (A', B', C') be a side triangle of Q, obtained by randomly displacing the points A, B, C that lie on the side r of P. Thus A', B', C' lie within distance d of A, B, C, respectively, and $B'A'$, $B'C'$ have lengths at least E. Evidently, since $2d < E$ and since the projection of A', B', C' on r is order preserving (Section 2.1), the base angles of (A', B', C') must be acute and the altitude of (A', B', C') can be at most $2d$; thus the conspicuousness of (A', B', C') is bounded above by $2d$. ■

LEMMA 2. *Let v be a vertex of P with angle V at which sides s and t meet. If*

$$4d \leq E \cos(45^\circ + V/4),$$

then there exists a sample point B within distance D of v such that the vertices of all side triangles whose vertices originate from s and t will be eliminated before B'.

Proof. Since $4d \leq E\cos(45^\circ + V/4) \leq E$, the hypothesis of Lemma 1 is satisfied for sides s and t. Hence Lemma 1 implies that the conspicuousnesses of the side triangles are bounded above by $2d$. We will now show that there exists a sample point B within distance D of v such that the conspicuousness of B' is greater than $2d$.

We recall that there must be at least one sample point on P within distance D of v. Let $B_1, \ldots, B_n$ be all these sample points. Let B be the B_i such that B' is the last B_i' to be eliminated, and let Q^* be a simplified version of Q that contains B' but does not contain any other B_i'.

Let A' and C' be the neighbors of B' in Q^*. Since the other B_i''s have been eliminated, A and C are not within distance D of v. As we shall now see, this allows us to give a lower bound on the conspicuousness of B' in Q^*.

Assume, without loss of generality, that A lies on s and B, C lie on t (see Fig. 16). Since we are looking for a lower bound, we can assume that sides $B'A'$ and $B'C'$ both have length E. If one of the base angles of (A', B', C') is obtuse, the conspicuousness of B' is one of these side lengths, which is E. Since $4d \le E\cos(45° + V/4) \le E$, the conspicuousness of B' is at least $4d$ and so is greater than $2d$.

Now we consider the case in which the base angles of (A', B', C') are acute. In this case the conspicuousness of B' is the altitude of (A', B', C').

We first estimate the altitude of (A, B, C). The closer B is to v, the larger is this altitude. Thus we obtain the smallest altitude if B (on t) is D away from v. The distance of A (on s) from v is greater than D; to obtain the extremal case we assume that A too is D away from v (see Fig. 16).

The angle b at B in triangle (A, B, C) is $90° + V/2$, where the dotted line is the bisector of the angle at vertex v. Consequently, in the extremal case the altitude h of triangle (A, B, C) is E cos $(45° + V/4)$. Thus we have

$$4d \le E\cos(45° + V/4) < h.$$

Since A', B', C' lie within distance d of A, B, C, respectively, we obtain the largest difference between the altitude h' of triangle (A', B', C') and the altitude h of triangle (A, B, C) if A', B', C' are displaced parallel to h, with A', C' displaced toward B and B' displaced toward the base AC. In this case we have $h' = h - 2d$; consequently, we always have $h' > 2d$. Thus the conspicuousness of B' is greater than $2d$.

Since the conspicuousnesses of the side triangles are bounded above by $2d$, we can conclude that the vertices of all the side triangles associated with s and t will be eliminated before B'. ■

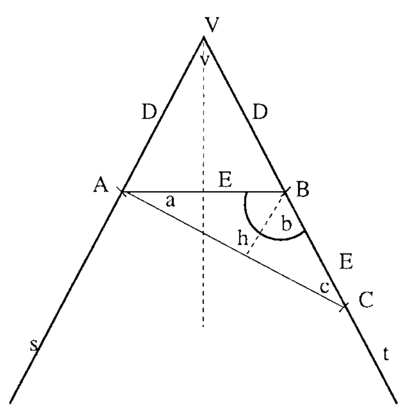

FIG. 16. Lower bound on the conspicuousness of a vertex triangle.

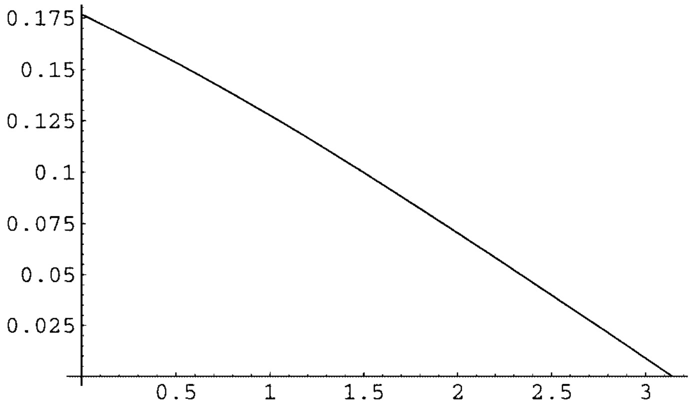

FIG. 17. Relationship between d and the vertex angle V (in radians).

As a simple consequence of Lemma 2 we have

THEOREM 3. *Let $v_1, \dots, v_m$ be the vertices of P, say with vertex angles $V_1, \dots, V_m$. If*

$$4d \leq \min\{E \cos(45^\circ + V_i/4) \mid i = 1, \dots, m\},$$

then there exists a sample point B_i within distance D of every vertex v_i of P such that the vertices of all the side triangles in Q will be eliminated before $B'_1, \dots, B'_m$. Thus the polygon $B_1, \dots, B_m$ is a D-recovery of the original polygon P.

To illustrate the relation between d and E when the hypothesis of Lemma 2 holds, suppose $E = 1$, and regard d as a function of the angle V at a given vertex v of P: $d(V) = \frac{1}{4} \cos \frac{\pi + V}{4}$. The graph of this function for $V \in [0, \pi]$ is shown in Fig. 17. We see, for example, that for $V \approx 2.5 \approx 130^\circ$ we have $d(V) \approx 0.05$, which means that if no angle of P is flatter than 130°, then $d \leq \frac{E}{20}$.

3.3. Polygon Recovery from a Noisy Image Digitization

Theorem 3 assumes that Q is a hand digitization of P, and its proof requires that the side lengths of Q be bounded below (by E). Such a proof cannot be given if Q is a random image digitization of P, because the sides of Q can be short. Nevertheless, as we shall now verify experimentally, in this case too we can recover good approximations to P by eliminating vertices of Q in order of their inconspicuousness.

The noisy digital polygon in our first experiment, shown in Fig. 18 (left), is a polygonal object (with 846 vertices) extracted from a real aerial image; this image was used in [1] as an example of model-based shape recovery. The straight segments in the boundary of this object have been significantly corrupted by noise. The middle image in Fig. 18 shows the result of eliminating all but 50 vertices from this noisy polygon in order of their inconspicuousness. Evidently, the 50-vertex polygon is much less noisy but its shape has changed very little. Figure 18 (right) shows the result of eliminating all but 24 vertices; it can be compared with Fig. 4 of [1]. Evidently, the most significant straight segments of the boundary have all been

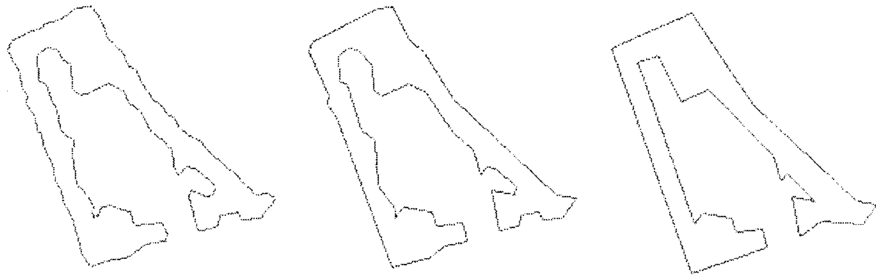

FIG. 18. A noisy polygonal object extracted from a digitized aerial image, and the result of eliminating its vertices in order of their inconspcuousness.

preserved. Note that, unlike the method used in [1], our process did not assume a model for the object.

Figure 19 shows the results of a second experiment in which the original polygon P is known. This polygon, shown in Fig. 19a, has 9 vertices. At some of these vertices the sides

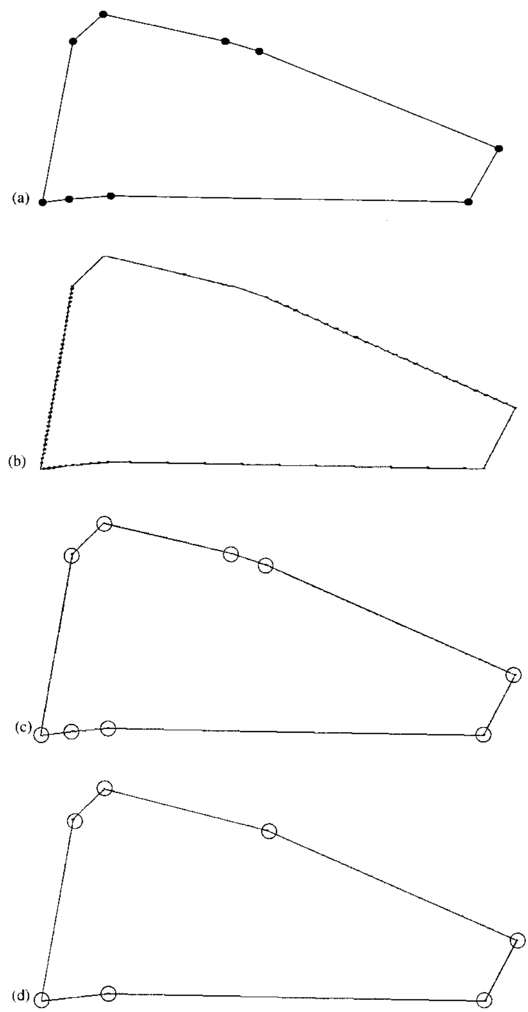

FIG. 19. A polygon with nine vertices (a), and results of eliminating all but 176 (b), 9 (c), and 7 (d) vertices from the digitization of the polygon (which had 1440 vertices).

are short and the angles are extremely flat (i.e., close to π). This polygon was digitized and the boundary of its digital image was extracted in the form of a chain code. The resulting chain-coded polygon Q had 1440 vertices. Figure 19b shows a simplified version of Q that has only 176 vertices, and Fig. 19c shows a version that has only 9 vertices. We see that the polygon in Fig. 19c is a perfect recovery of the original P. Figure 19d shows the result of eliminating two more vertices. Note that although this 7-vertex polygon has fewer vertices, its sides are still very close to those of the original P.

The polygon recovery results obtained in these two experiments, one for a polygon from a real aerial photograph and one for a synthetic polygon, are representative of the results we have obtained in a large number of experiments. Further results and an online Java applet can be found on the Web site [6].

In order to recover P from a noisy hand digitization (in Theorem 3) or image digitization (in Fig. 18), it is convenient to assume that the number of vertices of P is known, so we can use the number of vertices as a stopping criterion in the vertex elimination process. The question arises: if the number of vertices of P is not known, when should we stop eliminating vertices?

One answer to this question is to use a threshold for the maximal distance between the simplified polygon and the noisy polygon Q, as is done in many polygonal approximation algorithms. Another answer is given in [7], where a cognitively motivated shape similarity measure is used to determine a stopping criterion. The idea is that the vertex elimination process is stopped before the simplified polygon becomes significantly dissimilar to the noisy input polygon. A few experimental results based on this approach are shown in Fig. 20.

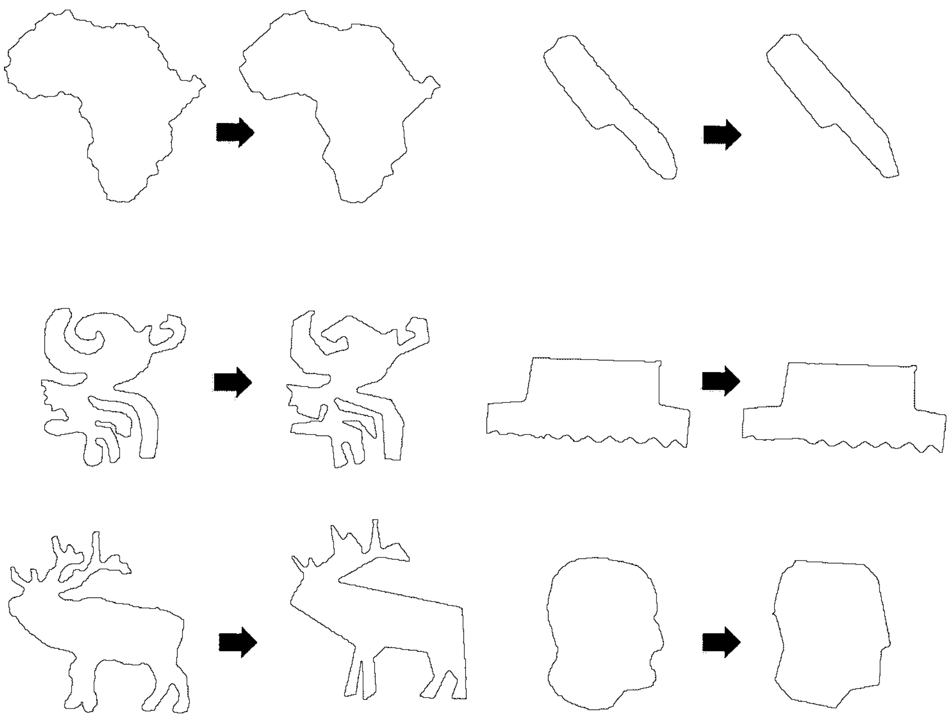

FIG. 20. Polygons obtained when using a shape similarity measure as a stopping criterion in the vertex deletion process.

4. CONCLUDING REMARKS

We have shown in this paper that if we are given a noisy polygon Q obtained from an unknown polygon P, we can construct a simplified polygon Q' that lies within a known distance of P by eliminating vertices from Q in order of their inconspicuousness, as in [4, 5]. More generally, the vertex elimination process can be used not only to restore P, but also to simplify Q, depending on the stopping criterion.

This results of this paper can be generalized in several ways. In two dimensions, we can consider recovery of an approximation to a general rectifiable simple closed curve C, rather than to a polygon P. We could also investigate generalizations of our approach to three dimensions, where P and Q are simple polyhedra whose faces are triangular; such generalizations are a subject for future research.

REFERENCES

1. A. Brunn, U. Weidner, and W. Förstner, Model-based 2D-shape recovery, in *Proc. of 17th DAGM Conference on Pattern Recognition (Mustererkennung), Bielefeld, Germany, 1995*, pp. 260–268, Springer-Verlag, Berlin.
2. R. M. Haralick and L. G. Shapiro, *Computer and Robot Vision*, Sect. 11.5, Addison–Wesley, Reading, MA, 1992.
3. Y. Kurozumi and W. A. Davis, Polygonal approximation by the minimax method, *Comput. Graphics Image Process*. **9**, 1982, 248–264.
4. L. J. Latecki and R. Lakämper, Convexity rule for shape decomposition based on discrete contour evolution, *Comput. Vision Image Understanding* **73**, 1999, 441–454.
5. L. J. Latecki and R. Lakämper, Polygon evolution by vertex deletion, in *Proceedings of International Conference on Scale-Space Theories in Computer Vision, Corfu, Greece, September 1999*.
6. Available at http://www.math.uni-hamburg.de/home/lakaemper/shape.
7. L. J. Latecki and R. Lakämper, Shape similarity measure based on correspondence of visual parts, *IEEE Trans. Pattern Anal. Mach. Intell*. **22**, 2000, 1185–1190.