Journal of Mathematical Imaging and Vision 8, pp. 131–159, 1998.

Preserving Topology by a Digitization Process

Longin Jan Latecki[1] Christopher Conrad[2] Ari Gross[3]

Abstract

The main task of digital image processing is to recognize properties of real objects based on their digital images. These images are obtained by some sampling device, like a CCD camera, and represented as finite sets of points that are assigned some value in a gray-level or color scale. Based on technical properties of sampling devices, these points are usually assumed to form a square grid and are modeled as finite subsets of $\mathbb{Z}^2$. Therefore, a fundamental question in digital image processing is which features in the digital image correspond, under certain conditions, to properties of the underlying objects. In practical applications this question is mostly answered by visually judging the obtained digital images. In this paper we present a comprehensive answer to this question with respect to topological properties. In particular, we derive conditions relating properties of real objects to the grid size of the sampling device which guarantee that a real object and its digital image are topologically equivalent. These conditions also imply that two digital images of a given object are topologically equivalent. This means, for example, that shifting or rotating an object or the camera cannot lead to topologically different images, i.e., topological properties of obtained digital images are invariant under shifting and rotation.

1 Introduction

We begin with an overview of the role topology plays in digital image processing. Currently, there are three areas of research connected to Artificial Intelligence and Cognitive Science which show a growing interest in topology:

(1) vision and image processing,

(2) representation and processing of spatial knowledge,

[1]Department of Computer Science, University of Hamburg, Vogt-Kölln-Str. 30, 22527 Hamburg, Germany
[2]Department of Applied Mathematics, University of Hamburg, Bundesstr. 30, 20146 Hamburg, Germany
[3]Department of Computer Science, Graduate Center and Queens College, CUNY, Flushing, New York 11367, USA

(3) application of topological concepts in cognitive science.

The main question addressed in these areas is: which properties of real objects correspond to properties of their representations. Topological properties play an important role, since they are the most primitive object features and our visual system seems to be well-adapted to cope with topological properties. Psychological evidence has been brought up to show that topological features are primary for human visual perception (see Chen [4], [5], and [6]).

Topological knowledge can lead to the development of practically relevant procedures:

- For example, the digital version of the Jordan curve theorem is relevant in the representation and processing of spatial knowledge. Let us assume that a terrain map is given as a square grid in which the objects are identified, and a robot must bring a box to a location that is surrounded by a fence that the robot cannot cross over. The robot ought to know that his task cannot be achieved. To infer this, it is sufficient for the robot to identify that its starting and goal locations do not belong to the same connected component. This can be done by a connected component labeling algorithm, assuming that the discrete representation of the fence separates the robot's digital imagery into two components. To infer this, it is necessary that the Jordan curve theorem holds in the robot's digital imagery if the pixels representing the fence form a simple closed curve. Note that the robot need not have explicit knowledge about the Jordan curve theorem.

- In many imaging applications topological knowledge can be used to justify correctness of algorithms. For example, Kong and Udupa [17] proved correctness of a surface tracking algorithm, which is used to display a surface of a 3D object on a 2D screen in medicine and engineering. The correct behavior of this algorithm had only been conjectured, but observed nonetheless, in all uses.

- In character recognition, letters in a document can be initially classified according to their topological homotopy types.

- The layout of documents is frequently based on topological predicates, e.g., a specific item must appear within a preassigned box.

- Optical checking of computer chips involves determining if the chip layout has the desired homotopy type.

In image processing and in spatial knowledge representation, continuous objects are represented as finite sets (also called discrete sets), since only finite structures can be handled on computers. Continuous objects and their spatial relations can be characterized using geometric features. Therefore, any useful discrete representation should model the geometry faithfully in order to avoid wrong conclusions. A basic part of geometry is topology.

It is clear that no discrete model can exhibit all the features of the continuous original. Therefore, one has to accept compromises. The compromise chosen depends on the specific application and on the questions which are typical for that application. Digital geometry can be seen as an attempt to evaluate the price one has to pay for discretization. Digital topology is the theoretical basis for understanding topological features of objects in digital images, which must be related to features of the underlying continuous objects. We list some questions which are typically dealt with using digital topology:

- Definition of connectedness and connected components of digital objects.

- Classification of points in a digital set as interior and boundary points, definition of boundaries.

- Definition of Jordan curves (and higher dimensional surfaces) and statement of the digital version of the Jordan curve theorem (and its higher dimensional analogs). Digital versions of the Jordan theorem and its higher dimensional generalizations are proved in Rosenfeld [25], Morgenthaler and Rosenfeld [22], Kong and Roscoe [14, 15], Stout [29], Khalimsky et. al [13], and Herman [12], for example.

- Continuous functions are an important tool of topology. Such functions have no direct counterparts in discrete structures. However, it seems to be useful for many applications to have such a concept as 'discrete continuity'. Two different versions of discrete continuity are defined in Rosenfeld [26] and in Latecki and Prokop [19]. By means of a suitable definition of continuity, one is able to compare topological structures with each other. A powerful concept in topology is homeomorphism. Homeomorphic topological structures are topologically equivalent. So, if a topological problem is investigated, a topological structure is sought in which the problem can be stated and solved as easily as possible.

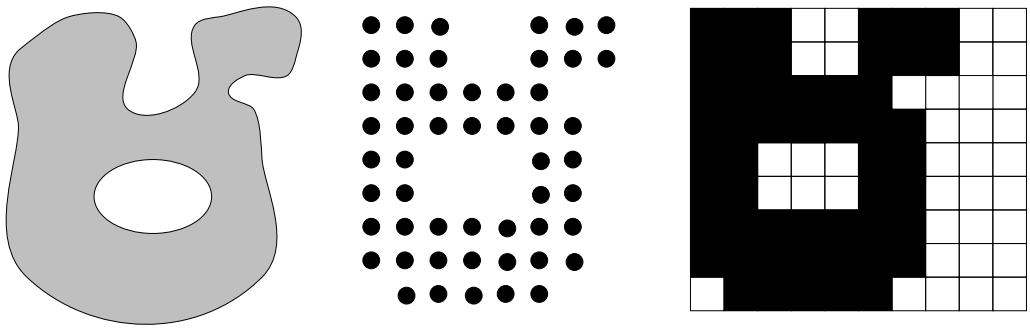
- Digital invariants theory. One important topological invariant is the Euler number. This is, in a certain sense, the only digital invariant which can be decided locally, i.e., in parallel (Minsky and Papert [20]).

- Homotopy theory deals with properties which are invariant under continuous deformations. Translating homotopy from general topology to discrete structures raises a number of questions which have not yet been resolved in a satisfactionary way. However, there are interesting approaches to homotopy in discrete structures, e.g., Boxer [2] defines homotopy directly for a square grid and its higher dimensional versions based on a discrete version of continuity given in Rosenfeld [26].

In digital image processing, properties retrieved from the digital images are assumed to represent properties of the underlying real objects. Practical applications show that this is not always the case. Therefore, the main question which occurs is: under what conditions do certain digital properties represent actual properties of the real object? In this paper, the authors introduce conditions that guarantee that a real object and its digital image are topologically equivalent.

In order to study the topological equivalence of a real continuous object and its digitization, which is a finite set of points, some preliminaries are in order. It is intuitively clear that the real object in Figure 1.a and the digital object in Figure 1.b have the "same topological structure". Based on the technical properties of sampling devices like a CCD camera, digital points representing sensor output are generally assumed to form a square grid and are modeled as points with integer coordinates located in the plane $\mathbb{R}^2$. By a digitization process, these points are assigned some gray-level or color values. By a segmentation process, the digital points are grouped to form digital objects. For example, the digital points are grouped by thresholding gray-level values with some threshold value, i.e., the pixels whose gray-level values are greater than some given threshold value are classified as belonging to a digital object (i.e., assigned the color black). As an output of a digitization and segmentation process, we obtain a binary digital picture, with black points representing the digital object and white points representing the background.

We will identify each black point with a square centered at this point (in such a way that the squares form a uniform cover of the plane). A digital object is then represented as a union of squares which form a subset of the plane. For example, the digital set in Figure 1.b, a finite subset of $\mathbb{Z}^2$, is identified with the union of black squares in Figure 1.c, a subset of $\mathbb{R}^2$. Real objects or their projections are modeled in computer vision as subsets of the plane. Therefore, it makes sense to speak about topological equivalence between a real object (Figure 1.a) and its digital image (Figure 1.c). Thus, the digitization (and segmentation) process is modeled as a mapping from continuous 2D sets representing real objects to discrete sets represented as finite subsets of $\mathbb{Z}^2$, which are identified with finite unions of squares in $\mathbb{R}^2$. Consequently, we can relate topological properties of a continuous 2D object (e.g., a projection of a 3D object) to its digital images interpreted as the union of squares centered at black points.



a) A real object

b) Its digitization represented as a finite set of points

c) Its digitization represented as the union of black squares

Figure 1: An object and its digital image.

Serra [28] considered many kinds of digitizations. He showed that, for a certain class of planar sets, digitizations preserve homotopy. However, he proved this only for subset digitizations in hexagonal grids, where a *subset digitization* of a set A in $\mathbb{R}^2$ is the set of points in $\mathbb{Z}^2$ which are contained in A. To show non-trivial problems connected with digitizations, Serra gave the following title to one of the sections: "To digitize is not as easy as it looks" ([28], p. 211).

Pavlidis [24] was primarily interested in digitizations based on the square sampling grid, since ([24], p.36): "*The most common grid used in picture processing is the square grid consisting of square cells arranged as a chessboard.*" His starting point was Shannon's Sampling Theorem, which is well-known in one-dimensional signal processing. This theorem allows one to determine the size of the sampling interval such that a one-dimensional signal can be exactly reconstructed from its samples. Pavlidis was interested in determining the size of the squares of the sampling grid that guarantees a reconstruction of the "shapes" of image regions ([24], Chapter 7, p. 129):

"*The size of the cells of the sampling grid must be small enough so that the shapes of regions of a given color remain unaltered in reconstructing the image.*"

As the shape preservation criterion, he used topological homeomorphism. Since Pavlidis identifies each point of a digital image with a square centered at this point, a given continuous

set as well as its digital image are both subsets of the plane. First we state his definition of compatibility ([24], Def. 7.4):

A closed planar set A and a square sampling grid whose (square) cells have diameter h are **compatible** *if:*

(a) There exists a number $d > h$ such that for each boundary point x of each connected component R of A, there is a closed ball C with diameter d that is tangent to the boundary of R at x and lies entirely within R.

(b) The same is also true for the closure of the complement of A.

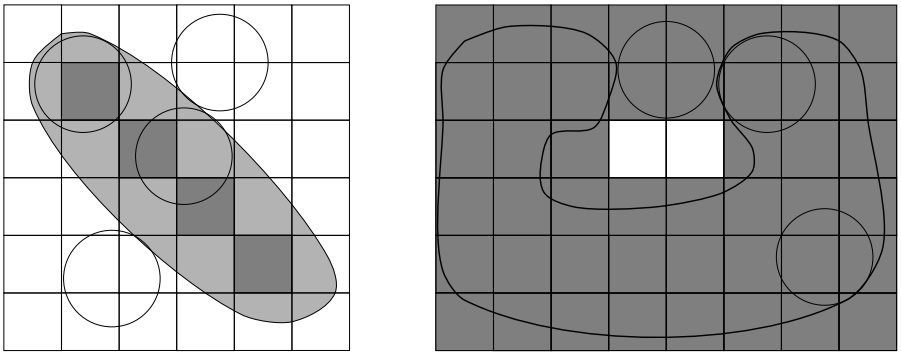
For example, both sets shown in Figure 2 are compatible with the square sampling grid. Using this definition, Pavlidis stated the following theorem ([24], Theorem 7.1):

For a planar set A, the condition of compatibility implies that A and its digitization are topologically equivalent (i.e., homeomorphic).

It is not clearly stated in [24] which digitization process is used in Theorem 7.1. This theorem holds for subset digitization on the square grid, where the grid squares chosen to represent a planar set A are the squares whose centers lie in A.

In this paper, we consider digitizations which are more relevant to practical applications. Our definition of a digitization approximates a real digitization process. Consistent with real sensor output, a digitization is defined with respect to a grid of squares, where each square has diameter r. Associated with each square is a sensor, i.e., the sensor is located at the center of the square. The value of the sensor output depends on the ratio of the area of the object in the square to the total area of the square. We assume that the gray-level values assigned to the sensors are monotonic with respect to the area of the continuous object in the sensor square. This is a standard model of the digitization and segmentation process for CCD cameras if we exclude digitization errors. As an output of a digitization process, we obtain a digital picture (a 2D matrix) with values of the pixels (picture elements) being the values of the corresponding sensors. The pixels in the image are segmented to form digital objects by thresholding gray-level values with some threshold value, i.e., the pixels whose gray-level values are greater than some given threshold value are classified as belonging to a digital object (i.e., assigned a black color). Segmenting objects under this model corresponds to coloring an image point black if the ratio of the area of the object in a square centered at this point to the area of the entire square is greater than some threshold value. This is a standard model of digitization and segmentation process in computer vision, which is also described in Pavlidis [24].

If we use this digitization (and segmentation model), then Pavlidis' theorem is not true, as shown in the following examples. Let A be a strip of width d, where $2h > d > h$, forming a 45^o angle with the square grid, as illustrated in Figure 2(a). If square p is black iff $area(p \cap A)/area(p) > 0.99$ and white otherwise, then the digitization of strip A represented by the gray squares is a digital 8-line, which is not homeomorphic to strip A. Note, however, that A and its digitization are homotopy equivalent. This is not the case for our second example illustrated in Figure 2(b), where a square p is black iff $area(p \cap A)/area(p) > 0$ and white otherwise. Here set A is not even homotopy equivalent to its digitization, represented by gray squares, since A is simply connected, but its digitization is not (there is a white "hole" in it).



(a) (b)

Figure 2: The two sets shown are compatible with the square sampling grid. The set and its digitization are not homeomorphic in (a) and not homotopy equivalent in (b).

In this paper we derive conditions relating properties of continuous objects to the diameter of a square in the grid. If these conditions are satisfied, then the digital object obtained by this digitization (and segmentation) process is guaranteed to be topologically equivalent to the underlying continuous object. Loosely speaking, we show that if we double the diameter of the tangent ball used by Pavlidis, then a continuous object and its digital image will be topologically equivalent. Since we did not see how the proof of Pavlidis' theorem can imply this fact, we developed new proof methods presented in this paper.

2 Parallel Regular Sets

In this section we define a class of subsets of the plane representing "real objects", which we will call parallel regular sets. Let A be a planar set. We denote by A^c the complement of A, by bdA the topological boundary of A, by $intA$ the topological interior of A and by clA the topological closure of A in the usual topology of the plane determined by the Euclidean metric. The connected components of the boundary bdA are called **contours**. We denote by $d(x, y)$ the Euclidean distance of points x, y and by $B(c, r)$ a closed ball of radius r centered at a point c.

The following definition of parallel regular sets is based on the classical concepts in differential geometry of osculating balls and normal vectors, which we define below without using derivatives and limit points.

Definition: We will say that a closed ball $B(c, r)$ is **tangent** to bdA at point $x \in bdA$ if $bdA \cap bd(B(c, r)) = \{x\}$.

We will say that a closed ball $iob(x, r)$ of radius r is an **inside osculating ball** of radius r to bdA at point $x \in bdA$ if $bdA \cap bd(iob(x, r)) = \{x\}$ and $iob(x, r) \subseteq intA \cup \{x\}$ (see Figure 3).

We will say that a closed ball $oob(x, r)$ of radius r is an **outside osculating ball** of radius r to bdA at point $x \in bdA$ if $bdA \cap bd(oob(x, r)) = \{x\}$ and $oob(x, r) \subseteq A^c \cup \{x\}$ (see Figure 3).

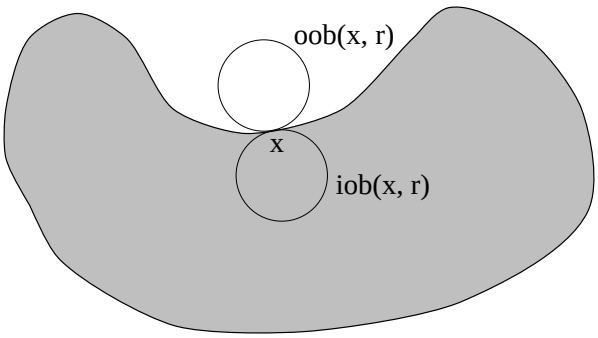


Figure 3: The inside and outside osculating balls of radius r to the boundary of the set A at point x.

Note that x is a boundary point, not the center, of both $iob(x, r)$ and $oob(x, r)$. For example, for every boundary point of a given ball $B(c, s)$ of radius s, there exist inside osculating balls of radii r, where $0 < r < s$. However, $B(c, s)$ itself is not an inside osculating ball for any of its boundary points. Now we define parallel regular subsets of the plane:

Definition: We assume that A is a closed subset of the plane such that its boundary bdA is compact.
A set A will be called **par(r,+)-regular** if there exists an outside osculating ball $oob(x, r)$ of radius r at every point $x \in bdA$.
A set A will be called **par(r,-)-regular** if there exists an inside osculating ball $iob(x, r)$ of radius r at every point $x \in bdA$.
A set A will be called **par(r)-regular** (or **r parallel regular**) if it is par(r,+)-regular and par(r,-)-regular. A set A will be called **parallel regular** if there exists a constant r such that A is par(r)-regular. We will sometimes call parallel regular sets **(spatial) objects**.

In Figure 4 the set A is par(r)-regular while the set B is not par(r)-regular, where r is the radius of the depicted circles. Note that a parallel regular set, as well as its boundary, does not have to be connected.

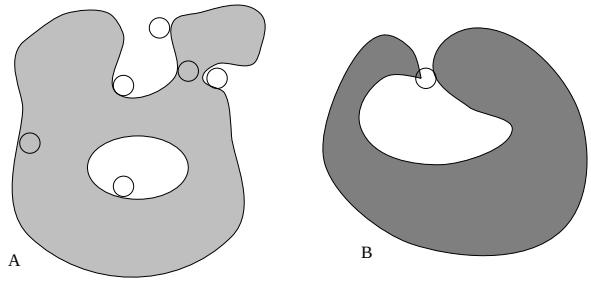


Figure 4: The set A is par(r)-regular while the set B is not par(r)-regular, where r is the radius of the depicted circles.

Definition: Let an outside osculating ball $oob(x, r)$ and an inside osculating ball $iob(x, r)$ exist at some point $x \in bdA$. Let $nl(x)$ be the straight line passing through the centers of

balls $oob(x,r)$ and $iob(x,r)$ (see Figure 5.a). From Proposition 1 (3) below, it follows that $x \in nl(x)$ and $nl(x)$ contains the centers of all balls $oob(x,s)$ and $iob(x,s)$ for every $0 < s \leq r$. Since $nl(x)$ is uniquely determined, we call $nl(x)$ the **normal line** to bdA at point $x \in bdA$. The straight line perpendicular to $nl(x)$ passing through x, we call the **tangent line** to bdA at point $x \in bdA$ and denote it $t(x)$.

In the remaining part of this section, we state some basic properties of parallel regular sets. We first show that outside and inside osculating balls and normal lines are uniquely defined.

Proposition 1 *Let an outside osculating ball $oob(x,r)$ and an inside osculating ball $iob(x,r)$ exist at some point $x \in bdA$.*

(1) If OB is an outside osculating ball at $x \in bdA$ of radius s, where $0 < s \leq r$, we obtain $OB = oob(x,r)$ if $s = r$, and $OB \subseteq oob(x,r)$ and $bd(OB) \cap bd(oob(x,r)) = \{x\}$ if $s < r$.

(2) If IB is an inside osculating ball at $x \in bdA$ of radius s, where $0 < s \leq r$, we obtain $IB = iob(x,r)$ if $s = r$, and $IB \subseteq iob(x,r)$ and $bd(IB) \cap bd(iob(x,r)) = \{x\}$ if $s < r$.

(3) Let $nl(x)$ be the straight line passing through the center points of balls $oob(x,r)$ and $iob(x,r)$ (see Figure 5.a). Then x and the centers of $oob(x,s)$ and $iob(x,s)$ lie on $nl(x)$ for $0 < s \leq r$.

Proof: We prove (1) and (3), the proof of (2) is analogous to the proof of (1). Since $iob(x,r) \subseteq intA \cup \{x\}$, $oob(x,r) \subseteq A^c \cup \{x\}$, and $bdA \cap bd(iob(x,r)) \cap bd(oob(x,r)) = \{x\}$, we obtain $iob(x,r) \cap oob(x,r) = \{x\}$.

We now show that $x \in nl(x)$. Let c_i be the center of $iob(x,r)$ and let c_o be the center of $oob(x,r)$. Then

$$d(c_i, c_o) \leq d(c_i, x) + d(x, c_o) = 2r.$$

Suppose $x \notin nl(x)$. Let y_i be the closest point to x in $bd(iob(x,r)) \cap nl(x)$ and let y_o be the closest point to x in $bd(oob(x,r)) \cap nl(x)$. Then $y_i \neq y_o$, and

$$d(c_i, c_o) = d(c_i, y_i) + d(y_i, y_o) + d(y_o, c_o) > 2r.$$

This is a contradiction. Therefore, we must have $x \in nl(x)$.

Let $t(x)$ be the straight line perpendicular to $nl(x)$ passing through x. Then $oob(x,r) \cap t(x) = \{x\}$ and $iob(x,r) \cap t(x) = \{x\}$.

Let OB be an outside osculating ball at $x \in bdA$ of radius s, where $0 < s \leq r$, i.e., $bdA \cap bdOB = \{x\}$ and $OB \subseteq A^c \cup \{x\}$. We show that if $nl'(x)$ is the straight line passing through the center point of ball OB and point x, then $nl'(x) = nl(x)$. Assume that $nl'(x) \neq nl(x)$, then $int(OB) \cap int(iob(x,r)) \neq \emptyset$ (see Figure 5.b). However, $int(OB) \subseteq A^c$ and $int(iob(x,r)) \subseteq A$. It follows that $OB \subseteq oob(x,r)$ and $bd(OB) \cap bd(oob(x,r)) = \{x\}$ if $s < r$. ■

As a simple consequence of Proposition 1 we obtain:

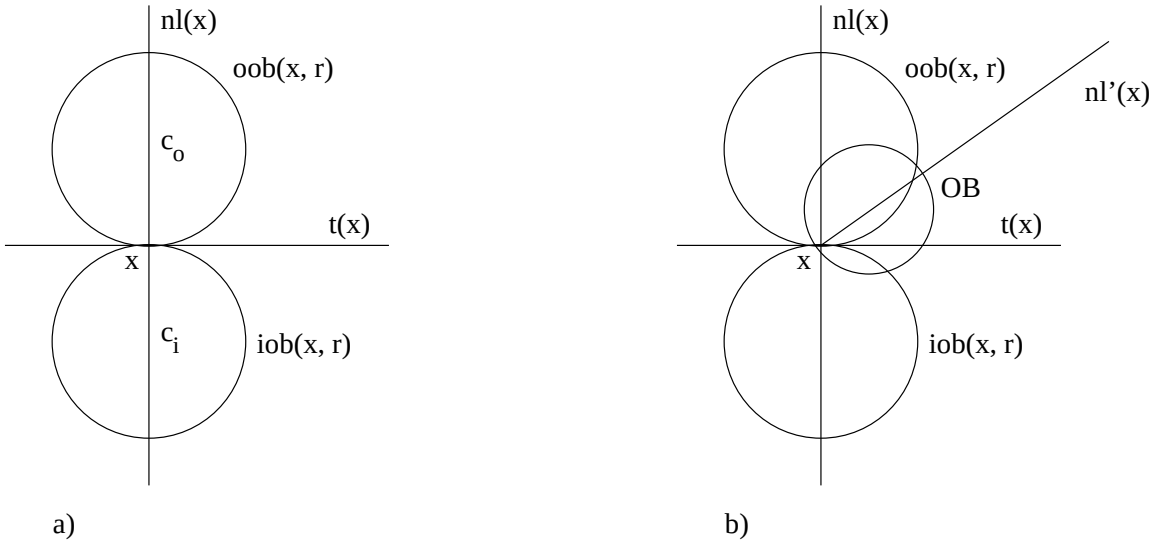


Figure 5:

Corollary 1 *(1) If an outside osculating ball* $oob(x, r)$ *and an inside osculating ball* $iob(x, r)$ *exist at some point* $x \in bdA$*, then there exist exactly one outside osculating ball* $oob(x, s)$ *and exactly one inside osculating ball* $iob(x, s)$ *at* $x \in bdA$ *for every* $0 < s \leq r$*.*

(2) If a set A *is par(r)-regular, then* A *is par(s)-regular for every* $0 < s \leq r$*.* ■

Definition: Let $x \in bdA$. Let an outside osculating ball $oob(x, r)$ and an inside osculating ball $iob(x, r)$ of radius r exist at x for some $r > 0$.

The **outer normal vector** $n(x, s)$ of length s to bdA at $x \in bdA$ is a line segment emanating from x of length s such that $n(x, s) \subseteq nl(x)$ and $n(x, s) \cap int[oob(x, r)] \neq \emptyset$ (see Figure 6).

The **inner normal vector** $-n(x, s)$ of length s to bdA at $x \in bdA$ is a line segment emanating from x of length s such that $-n(x, s) \subseteq nl(x)$ and $-n(x, s) \cap int[iob(x, r)] \neq \emptyset$.

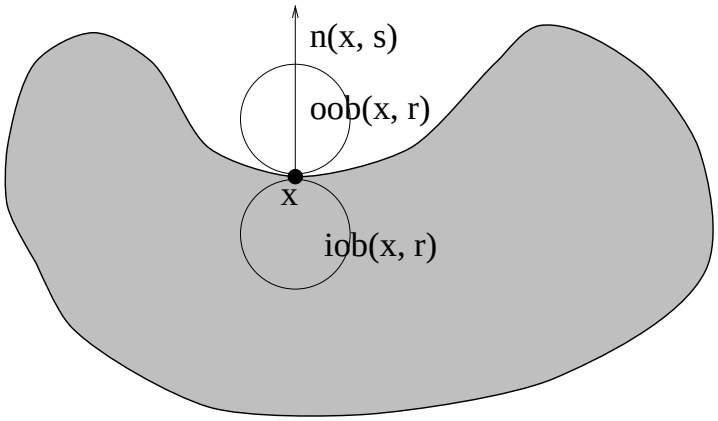


Figure 6: The outer normal vector $n(x, s)$ of length s to bdA at $x \in bdA$.

Observe that if a set A is par(r)-regular, then normal vectors exist at every point $x \in bdA$. On the other hand, it can be that normal vectors exist at every point $x \in bdA$, but the set A is not par(r)-regular for some $r > 0$, since the radius of osculating balls which determine the normal vectors can vary form point to point. For example, it can be that the outer normal vector $n(x, r)$ at some point $x \in bdA$ exists, but there exists no outside normal ball of radius r at x. In this case, since the tangent line $nl(x)$ is well-defined, there must exist an outside normal ball $oob(x, s)$ and an inside normal ball $iob(x, s)$ at $x \in bdA$ for some $s < r$. The relationship between the concepts of osculating balls and normal vectors is exactly described in Theorem 1, below. First we state the following propositions, which are simple consequences of the definition of normal vectors.

Proposition 2 *Let A be a par(r)-regular set and $x \in bdA$. Then, for every $c \in n(x, r)$, $x \in bdA$ is the closest point to c on bdA.* ■

Proposition 3 *Let A be a par(r)-regular set and $c \notin A$. Let $x \in bdA$ be a point with the closest distance from points in bdA to c. Then $c \in n(x, s)$ for every $s \geq d(x, c)$.*

Proof: Let $0 < t < d(c, x)$. Let OB be a closed ball of radius t such that $OB \subseteq B(c, d(c, x))$ and $bd(OB) \cap bd(B(c, d(c, x))) = \{x\}$ (see Figure 7). Since $B(c, d(c, x)) \cap intA = \emptyset$, we obtain that $OB \subseteq A^c \cup \{x\}$ and $bd(OB) \cap bdA = \{x\}$. Therefore, OB is an outer osculating ball at x. Thus, the straight line passing through the center of OB and point x is the normal line $nl(x)$. Since $c \in nl(x)$, we obtain that $c \in n(x, s)$ for every $s \geq d(x, c)$. ■

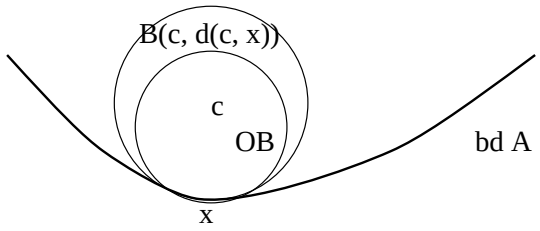


Figure 7: $x \in bdA$ is a point with the closest distance to c on bdA.

We have the following equivalence:

Theorem 1 *A set A is par(r)-regular iff, for every two distinct points $x, y \in bdA$, the outer normal vectors $n(x, r)$ and $n(y, r)$ exist and they do not intersect, and the inner normal vectors $-n(x, r)$ and $-n(y, r)$ exist and they do not intersect.*

For example, in Figure 8, set X is not par(r)-regular while set Y is par(r)-regular, where r is the length of the depicted vectors.

Proof: "$\Rightarrow$" Since A is par(r)-regular, we obtain for every point $x \in bdA$ that the outer normal vector $n(x, r)$ and the inner normal vector $-n(x, r)$ exist. We show that $n(x, r)$ and $n(y, r)$ do not intersect for every two distinct points $x, y \in bdA$. The proof of the same fact

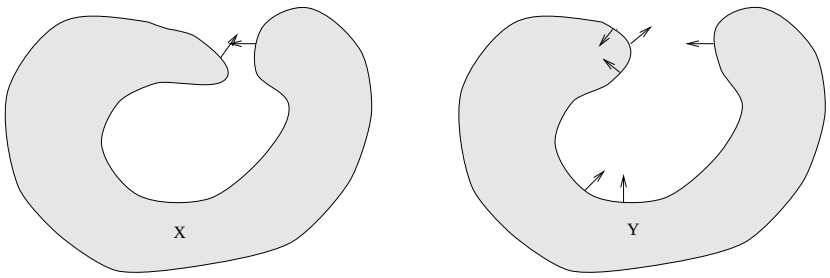


Figure 8: X is not par(r)-regular, but Y is par(r)-regular.

for inner normal vectors is similar. We know that $n(x,r)$ begins at point x and ends at the center of $oob(x,r)$ for every $x \in bdA$.

Assume that there exists two distinct points $x, y \in bdA$ such that $n(x,r)$ and $n(y,r)$ intersect. We show that this contradicts the fact that A is par(r)-regular.

Let $c \in n(x,r) \cap n(y,r)$. If $c = x$, then $x \in n(y,r) \subseteq oob(y,r)$. This contradicts the fact that $oob(y,r) \subseteq A^c \cup \{y\}$. Therefore, $c \neq x$, as well as $c \neq y$.

Let e be the endpoint of $n(x,r)$ and v the endpoint of $n(y,r)$ (see Figure 9). Let $p \in bdA$ be a closest point to c on bdA. Since $p \neq x$ or $p \neq y$, we can assume that $p \neq x$. Since $d(p,c) \leq d(x,c)$, we obtain

$$d(p,e) \leq d(p,c) + d(c,e) \leq d(x,c) + d(c,e) = r.$$

Therefore, $p \in oob(x,r)$, since e is the center of $oob(x,r)$. This contradicts the fact that $oob(x,r) \subseteq A^c \cup \{x\}$.

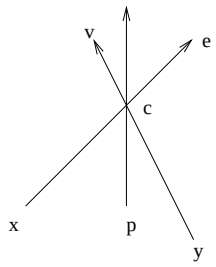


Figure 9:

"$\Leftarrow$" We show that if A is not par(r)-regular and, for every point $p \in bdA$, the outer normal vector $n(p,r)$ and the inner normal vector $-n(p,r)$ exist, then there exists two distinct points $x, y \in bdA$ such that either $n(x,r)$ and $n(y,r)$ intersect or $-n(x,r)$ and $-n(y,r)$ intersect.

Since A is not par(r)-regular, we may assume that there does not exist the outside normal ball of radius r at some point $x \in bdA$. Since there exists the outer normal vector $n(x,r)$ at x, the normal line $nl(x)$ is well-defined, and therefore there must exist the outside osculating ball $oob(x,s)$ and the inside osculating ball $iob(x,s)$ at $x \in bdA$ for some $s < r$. Let e be the endpoint of vector $n(x,r)$ and let OB be a closed ball of radius r centered at e (see

Figure 10). Since OB is not the outer osculating ball at x, $(OB \setminus \{x\}) \cap A \neq \emptyset$. Since $oob(x,s) \subseteq OB$, we obtain $(OB \setminus \{x\}) \cap bdA \neq \emptyset$.

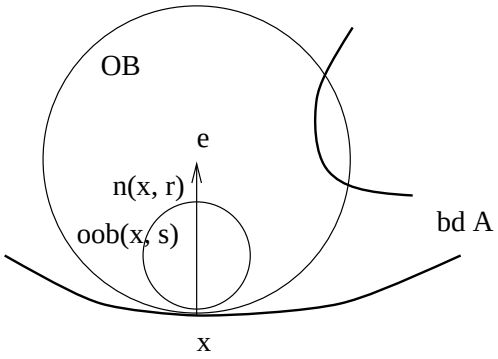


Figure 10:

If $e \in A$, then there exists a point $y \neq x$ such that $y \in bdA \cap n(x,r)$, since $((oob(x,s) \cap n(x,r)) \setminus \{x\}) \subseteq A^c$. Then $n(x,r)$ and $n(y,r)$ intersect.

Therefore, we can assume that $e \notin A$. If $int(OB) \cap A = \emptyset$, then there exists $y \in (bd(OB) \setminus \{x\}) \cap bdA$. In this case, y is a closest point to e on bdA and $d(e,y) = r$.

If $int(OB) \cap A \neq \emptyset$, then let $y \in bdA \cap OB$ be a closest point to e on bdA. Clearly, $d(y,e) < r$, and consequently $y \neq x$.

Thus, in both cases $y \in bdA$ is a closest point to e, $d(y,e) \leq r$, and $y \neq x$. By Proposition 3, $e \in n(y,r)$. Hence $e \in n(x,r) \cap n(y,r)$. ■

Definition: $B(x,r)$ denotes the closed ball of radius r centered at a point x. The **parallel set** of set $A \subseteq \mathbb{R}^2$ with distance r is given by

$$Par(A,r) = A \cup \bigcup \{B(x,r) : x \in bdA\}.$$

This set is also called a **dilation** of A with radius r. We define

$$Par(A,-r) = cl(A \setminus \bigcup \{B(x,r) : x \in bdA\}).$$

For illustration, see Figure 11. The boundaries of $Par(A,r)$ and $Par(A,-r)$ sets are often called offset curves.

It can be shown that $A = Par(Par(A,r),-r) = Par(Par(A,-r),r)$ for a par(r)-regular set A. Thus, a par(r)-regular set A is invariant with respect to morphological operations of opening and closing with a closed ball of radius r as a structuring element (see Serra [28] for definitions). The following proposition motivates the name of parallel regular sets.

Proposition 4 *Let A be a par(r)-regular set. Then (see Figure 11)*

$$Par(A,r) = A \cup \bigcup \{n(a,r) : a \in bdA\} \quad \text{and} \quad Par(A,-r) = cl(A \setminus \bigcup \{-n(a,r) : a \in bdA\}).$$

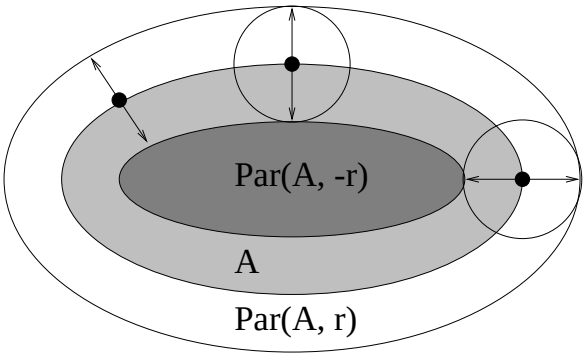


Figure 11: The set A and its parallel sets $Par(A, r)$ and $Par(A, -r)$.

Proof: We show only the first equation; the proof of the second is analogous. It is easy to see that $A \cup \bigcup\{n(a, r) : a \in bdA\} \subseteq Par(A, r)$; simply observe that $n(s, r)$ is contained in the dilation ball $B(s, r)$ for every $s \in bdA$.

It is clear that $A \subseteq Par(A, r)$. So, let $x \in Par(A, r)$ and $x \notin A$. Let $s \in bdA$ be a point having the shortest distance d from x to bdA. Such a point exists, since bdA is compact. Of course, $d \leq r$. By Proposition 3, $x \in n(s, r)$. Thus, $x \in \bigcup\{n(a, r) : a \in bdA\}$. ■

Proposition 5 *Let A be a par(r)-regular set. If x and y belong to two different components of bdA, then $d(x, y) > 2r$.*

Proof: Let $C_1, ..., C_n$ be all connected components of bdA (there is only a finite number of them, since bdA is compact), where $n \geq 2$. For every $i \neq j$, $i, j \in \{1, ..., n\}$, let $d_{ij} : C_i \times C_j \to \mathbb{R}$ be the Euclidean distance d restricted to $C_i \times C_j$. Since d_{ij} is a continuous function on a compact set, there exists $(c_i, c_j) \in C_i \times C_j$ such that $d_{ij}(c_i, c_j) > 0$ is the minimal value of d_{ij}. Let a pair (c_k, c_m), $k \neq m$, be such that $d_{km}(c_k, c_m) \leq d_{ij}(c_i, c_j)$ for all $i, j \in \{1, ..., n\}$ with $i \neq j$. We obtain that $d(x, y) \geq d(c_k, c_m) = d_{km}(c_k, c_m)$ for every x and y belonging to two different components of bdA. We now show that $d(c_k, c_m) \geq 2r$.

Assume that $d(c_k, c_m) \leq 2r$. Consider the closed ball B such that $c_k, c_m \in bdB$ and the line segment $c_k c_m$ is the diagonal of B (see Figure 12). Clearly, the radius of B is not greater than r and $B \cap bdA = \{c_k, c_m\}$.

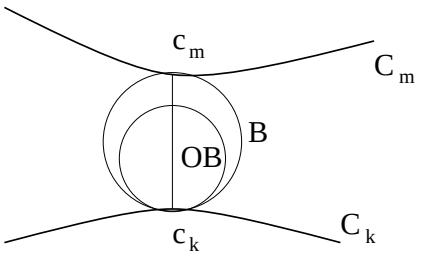


Figure 12:

Therefore, either $B \subseteq A$ or $intB \subseteq A^c$. We assume $intB \subseteq A^c$. The proof in the second case is analogous. Every closed ball OB such that OB is a proper subset of B and

$OB \cap B = \{c_k\}$ is an outside osculating ball of A at c_k. Since the radius of B is not greater than r and the center of B is collinear with all centers of balls OB, we obtain that B is an outside osculating ball of A at c_k. Yet, this contradicts the fact that $B \cap bdA = \{c_k, c_m\}$. Therefore, $d(c_k, c_m) > 2r$, and consequently $d(x, y) > 2r$ for every x and y belonging to two different components of bdA. ■

We can use parallel sets to define Hausdorff distance of planar sets:

Definition: Hausdorff distance d_H of two planar sets A and B is given by

$$d_H(A, B) = \inf\{r \geq 0 : A \subseteq Par(B, r) \text{ and } B \subseteq Par(A, r)\}.$$

3 Digitization and Segmentation Preserving Topology

Let $\mathcal{Q}$ be a cover of the plane with closed squares of diameter r such that if two squares intersect, then their intersection is either their common side or a corner point. A **digital image** can be described a set of points that are located at the centers of the squares of a grid $\mathcal{Q}$ and that are assigned some value in a gray level or color scale. By a **digitization process** we understand a function mapping a planar set X to a digital image. By a **segmentation process** we understand a process grouping digital points to a set representing a digital object. Therefore, the output of a segmentation process can be interpreted as a binary digital image, where each point is either black or white. We assume that digital objects are represented as sets of black points. Thus, the input of a digitization and segmentation process is a planar set X and the output is a binary digital image, which will be called a **digitization** of X with diameter r and denoted $Dig(X, r)$.

In the remainder of this paper, we will interpret a black point $p \in Dig(X, r)$ as a closed (black) square of cover $\mathcal{Q}$ centered at p and the digitization $Dig(X, r)$ as the union of closed squares centered at black points, i.e., $Dig(X, r)$ will denote a closed subset of the plane.

We will treat digitization and segmentation processes satisfying the following conditions relating a planar par(r)-regular set X to its digital image $Dig(X, r)$:

ds1 If a square $q \in \mathcal{Q}$ is contained in X, then $q \in Dig(X, r)$ (i.e., q is black).

ds2 If a square $q \in \mathcal{Q}$ is disjoint from X, then $q \notin Dig(X, r)$ (i.e., q is white).

ds3 If a square q is black and $area(X \cap q) \leq area(X \cap p)$ for some square $p \in \mathcal{Q}$, then square p is black.

These conditions describe a standard model of the digitization and segmentation process for CCD cameras if we exclude digitization errors. In the following, we define some important digitization and segmentation processes satisfying the conditions ds1, ds2, and ds3 above.

Definition: Let X be any set in the plane. A square $p \in \mathcal{Q}$ is black (belongs to a digital object) iff $p \cap X \neq \emptyset$, and white otherwise. We will call such a digital image an **intersection digitization** with diameter r of set X, and denote it with $Dig_\cap(X, r)$, namely $Dig_\cap(X, r) = \bigcup\{p \in \mathcal{Q} : p \cap X \neq \emptyset\}$. See Figure 13 (a), for example, where the union of all depicted squares represents the intersection digitization of an ellipse. With respect to real camera digitization and segmentation, the intersection digitization corresponds to the procedure of

coloring a pixel black iff there is part of the object A in the field "seen" by the corresponding sensor.

Now we consider digitizations corresponding to the procedure of coloring a pixel black iff the object X fills the whole field "seen" by the corresponding sensor. For such digitizations, a square p is black iff $p \subseteq X$ and white otherwise. We will refer to such a digital image of a set X as a **square subset digitization**[4] and denote it by $Dig_{\subset}(X, r)$, where $Dig_{\subset}(X, r) = \bigcup\{p \in \mathcal{Q} : p \subseteq X\}$. In Figure 13 (b), the two squares represent $Dig_{\subset}(X, r)$, where X is an ellipse.

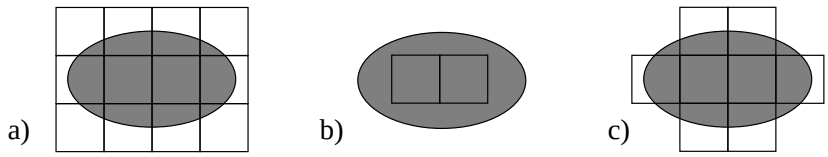


Figure 13: a) The union of all squares represents an intersection digitization of the ellipse. b) The two squares represent a square subset digitization of the ellipse. c) The eight squares represent a digitization of an ellipse with the area ratio equal to 1/5.

Next, let us consider a digitization and segmentation process in which a pixel is colored black iff the ratio of the area of the continuous object in a sensor square to the area of the square is greater than some constant threshold value v. An example is given in Figure 13 (c), where the squares represent a digitization of the ellipse with the ratio equal to 1/5. This process models a segmentation by applying a threshold value to a gray-level digital image for all real devices in which the sensor values can be assumed to be monotonic with respect to the area of the object in the sensor square.

In the following, we briefly review the concept of homotopy equivalence.

Definition: Let X and Y be two topological spaces. Two functions $f, g : X \to Y$ are said to be **homotopic** if there exists a continuous function $H : X \times [0, 1] \to Y$, where $[0, 1]$ is the unit interval, with $H(x, 0) = f(x)$ and $H(x, 1) = g(x)$ for all $x \in X$. The function H is called a **homotopy** from f to g. Sets X and Y are called **homotopy equivalent** or of the same **homotopy type** if there exist two functions $f : X \to Y$ and $g : Y \to X$ such that $g \circ f$ is homotopic with the identity over X (id_X) and $f \circ g$ is homotopic with the identity over Y (id_Y).

Intuitively, a homotopy H from f to g represents a continuous deformation of the map f to g. As a consequence of the properties of homotopy equivalence between planar sets X and Y, there is a complete correspondence between connected components of X and Y and the corresponding components are homotopy equivalent. The Euler characteristic, as well as the fundamental groups of X and Y, are the same (see Naber [23]).

Definition: We say that two topological spaces X and Y are **topologically equivalent** or **homeomorphic** if there exists a bijection $f : X \to Y$ such that f and the inverse function f^{-1} are continuous.

[4]Observe that this digitization differs from subset digitization used by Serra and Pavlidis, where a square p is black iff its center point is contained in X (see page 4).

If two topological spaces X and Y are homeomorphic, then they are homotopy equivalent. We will use topological equivalence as a definition for topology preserving.

Definition: We will say that a digitization $Dig(X, r)$ of some set X is **topology preserving** if X and $Dig(X, r)$ are homeomorphic.

We now consider a special case of homotopy equivalence called a strong deformation retraction. Intuitively, saying that there is a strong deformation retraction from a set X to a set $Y \subseteq X$ means that we can continuously shrink X to Y.

Definition: Let X and $Y \subseteq X$ be two topological spaces. A continuous function $H : X \times [0, 1] \rightarrow X$, where $[0, 1]$ is the unit interval, is called a **strong deformation retraction** of X to Y if $H(x, 1) = x$ and $H(x, 0) \in Y$ for every $x \in X$, and $H(x, t) = x$ for every $x \in Y$ and $t \in [0, 1]$. Y is called a **strong deformation retract** of X.

Note that if Y is a strong deformation retract of X, then Y is homotopy equivalent to X. To see this, take $f : X \rightarrow Y$ to be $f(x) = H(x, 0)$ and $g : Y \rightarrow X$ to be inclusion.

Theorem 2 *Let A be a par(r)-regular set. Then $Par(A, -r)$ is a strong deformation retract of A.*

Proof: If $x \in (A - Par(A, -r))$, then there exists a unique normal vector $-n(a, r)$, for some $a \in bdA$, such that $x \in -n(a, r)$. We define $\pi : (A - Par(A, -r)) \rightarrow bdPar(A, -r)$ by $\pi(x) =$ the end point of the vector $-n(a, r)$, where a is such that $x \in -n(a, r)$. Thus, $\pi(x)$ denotes the single point on $bdPar(A, -r)$ with the closest distance to x, and therefore, π is a metric projection onto $bdPar(A, -r)$.

Let H be a function defined as follows:

$H : A \times [0, 1] \rightarrow A$,

$H(x, t) = x$ for every $x \in Par(A, -r)$ and $t \in [0, 1]$,

$H(x, t) = (1 - t)\pi(x) + tx$ for every $x \in (A - Par(A, -r))$ and $t \in [0, 1]$

Note that $H(x, 1) = x$ for every $x \in A$ and that $H(x, t) = x$ for every $x \in Par(A, -r)$ and $t \in [0, 1]$. Note also that $H(x, 0) = \pi(x)$ for all $x \in (A - Par(A, -r))$. Thus, $H(x, 0) \in Par(A, -r)$ for every $x \in A$.

To prove that H is a strong deformation retraction, it remains to show that H is a continuous function. Clearly, for a fixed x, $H(x, t)$ as a function of t is continuous. If t is fixed, the continuity of $H(x, t)$ as a function of x follows from the upper semi-continuity of the metric projection π, which reduces to continuity of π, since π is a point valued function. This implies that if x and y are close to each other, then $\pi(x)$ and $\pi(y)$ are close to each other, and consequently, the line segments $x\pi(x)$ and $y\pi(y)$ are close to each other. Therefore, H is a strong deformation retraction of A to $Par(A, -r)$. ■

Theorem 3 *Let A be a par(r)-regular set. Then $Par(A, -r)$ is a strong deformation retract of $Dig(A, r)$ for every digital image $Dig(A, r)$ (which satisfies conditions ds1, ds2, and ds3), and $d_H(A, Dig(A, r)) \leq r$, where d_H is Hausdorff distance (see Figure 14 for an illustration).*

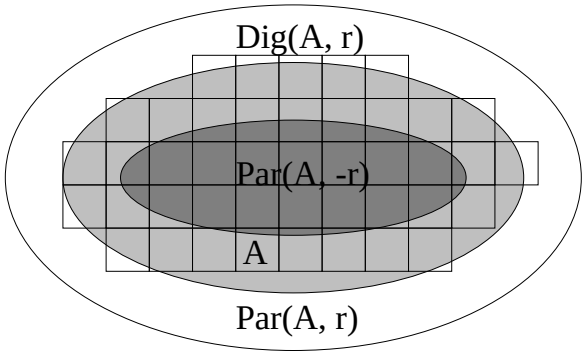


Figure 14: $Par(A, -r)$ is a strong deformation retract of $Dig(A, r)$.

Proof: Let p be a closed square with diameter r such that $p \cap A \neq \emptyset$. Since $p \subseteq B(x, r)$ for every closed ball $B(x, r)$ such that the center $x \in p$, we obtain that $p \subseteq Par(A, r)$. Therefore $Dig_{\cap}(A, r) \subseteq Par(A, r)$. For every closed square p with diameter r, it similarly holds that if $p \cap Par(A, -r) \neq \emptyset$, then $p \subseteq A$. Therefore $Dig_{\cap}(Par(A, -r), r) \subseteq Dig_{\subset}(A, r)$. Since it is clear that $Par(A, -r) \subseteq Dig_{\cap}(Par(A, -r), r)$, we obtain $Par(A, -r) \subseteq Dig_{\subset}(A, r)$. Thus we obtain the following inclusions: $Par(A, -r) \subseteq Dig_{\subset}(A, r) \subseteq Dig_{\cap}(A, r) \subseteq Par(A, r)$. Since by the definition of $Dig(A, r)$ (conditions ds1, ds2), $Dig_{\subset}(A, r) \subseteq Dig(A, r) \subseteq Dig_{\cap}(A, r)$, we obtain that

$$Par(A, -r) \subseteq Dig(A, r) \subseteq Par(A, r).$$

These inclusion relations imply that $d_H(A, Dig(A, r)) \leq r$, since $Par(Par(A, -r), r) = A$.

Let $a \in bdPar(A, -r)$ and let $x \in bdA$ be such that a is the end point of the inner normal vector $-n(x, r)$ (see Figure 15). The outer normal vector at a to $bdPar(A, -r)$ of length $2r$ can be defined as $n(a, 2r) = -n(x, r) \cup n(x, r)$. Clearly, for every two distinct points $a, b \in bdPar(A, -r)$, $n(a, 2r)$ and $n(b, 2r)$ do not intersect.

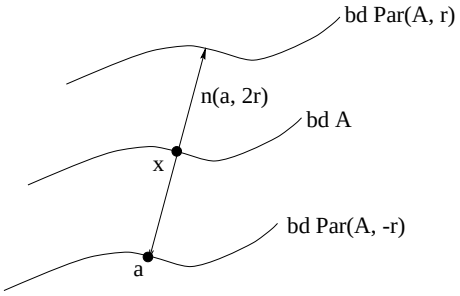


Figure 15: $n(a, 2r) = -n(x, r) \cup n(x, r)$.

We intend to construct a strong deformation retraction

$$D: \ Dig(A, r) \times [0, 1] \rightarrow Dig(A, r)$$

from $Dig(A, r)$ onto $Par(A, -r)$.

In the following, $x \in Dig(A, r) \setminus Par(A, -r)$. Let $p(x)$ be a point on $bdPar(A, -r)$ with the closest distance to x. Let $xp(x)$ be the line segment joining x with $p(x)$. Since $xp(x) \subseteq$

$n(p(x), 2r)$ and the normal vectors $n(a, 2r)$ do not intersects for $a \in bdPar(A, -r)$, $p(x)$ is uniquely determined. The metric projection p is a continuous function from $Dig(A, r) \setminus Par(A, -r)$ to $bdPar(A, -r)$.

If $xp(x) \subseteq Dig(A, r)$, then we could define $D(x, t) = (1 - t)p(x) + tx$ for every $t \in [0, 1]$. However, it can happen that $xp(x) \not\subseteq Dig(A, r)$. Therefore, for every line segment $xp(x)$, we must define a modified path $mp(x, p(x)) \subseteq Dig(A, r)$ from x to $p(x)$. If $xp(x) \subseteq Dig(A, r)$, then $mp(x, p(x)) = xp(x)$.

If $xp(x) \not\subseteq Dig(A, r)$, then there exists two different sides $s_1, s_2 \subseteq bdDig(A, r)$ of some squares in grid $\mathcal{Q}$ such that $xp(x)$ intersects s_1 in point a_1, s_2 in point a_2, and $a_1a_2 \cap Dig(A, r) = \{a_1, a_2\}$, where a_1a_2 is the line segment joining a_1 and a_2. Since $xp(x) \subseteq n(p(x), 2r)$, we can apply Lemma 1, which is stated below. We obtain that the sides s_1, s_2 share a vertex and are perpendicular (see Figure 17). Let c be the common vertex of sides s_1, s_2. We define $mp(x, p(x)) = p(x)a_1 \cup a_1c \cup ca_2 \cup a_2x$ (see Figure 16). If $mp(x, p(x)) \not\subseteq Dig(A, r)$, then we can repeat this construction for $p(x)a_1$ or a_2x. We continue this process until the modified path $mp(x, p(x))$ is contained in $Dig(A, r)$.

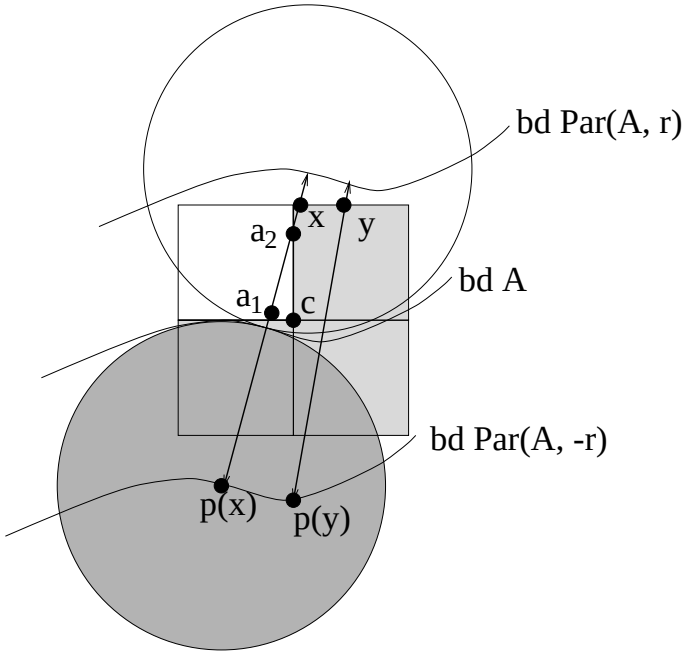


Figure 16: The construction of a modified path $mp(x, p(x)) \subseteq Dig(A, r)$.

Observe that if two points $x, y \in Dig(A, r)$ are close to each other, then the line segments $xp(x)$ and $yp(y)$ are close to each other, since $p(x)$ and $p(y)$ are close to each other by the continuity of the metric projection p. We show that in this case, also the modified paths $mp(xp(x))$ and $mp(yp(y))$ are close to each other. If $n(p(x), 2r)$ and $n(p(y), 2r)$ intersect the same sides s_1, s_2, then $xp(x)$ and $yp(y)$ are clearly close to each other. If $n(p(y), 2r)$ does not intersect the sides s_1, s_2, then $a_1c \cup ca_2$ of $mp(xp(x))$ is contained in the strip region determined by $n(p(x), 2r)$ and $n(p(y), 2r)$ (see Figure 16).

Now we parameterize uniformly $mp(x, p(x))$ with a continuous function

$f_x : [0,1] \to mp(x, p(x))$ such that $f_x(0) = p(x)$ and $f_x(1) = x$.

If $x \in Dig(A,r) \setminus Par(A,-r)$, then we define $D(x,t) = f_x(t)$.

It is easy to observe that for a fixed x, $D(x,t)$ as a function of t is continuous. If t is fixed, the continuity of $D(x,t)$ as a function of x follows from the continuity of the metric projection p, which implies that if x and y are close to each other, then the modified paths $mp(xp(x))$ and $mp(yp(y))$ are close to each other, as shown above.

Finally, we need to establish that D satisfies the other properties of a strong deformation retraction. By definition, $D(x,t) = x$ for every $x \in Par(A,-r)$ and $t \in [0,1]$. Clearly, $D(x,1) = x$ and $D(x,0) \in Par(A,-r)$ for every $x \in Dig(A,r)$. Thus, D is a strong deformation retraction of $Dig(A,r)$ to $Par(A,-r)$. ■

In the proof of Theorem 3 we used the following lemmas:

Lemma 1 *Let A be a par(r)-regular set and let $Dig(A,r)$ be a digital image of A (which satisfies conditions ds1, ds2, and ds3).*
Let s_1, s_2 be two different sides of some squares in grid $\mathcal{Q}$ such that $s_1, s_2 \subseteq bdDig(A,r)$. For every $x \in bdPar(A,-r)$, if $n(x,2r)$ intersects first s_1 in point a_1 and then s_2 in point a_2 such that $a_1a_2 \cap Dig(A,r) = \{a_1, a_2\}$, where a_1a_2 is the line segment joining a_1 and a_2, then the sides s_1, s_2 share a vertex and are perpendicular, and the square determined by s_1, s_2 is white (i.e., does not belong to $Dig(A,r)$). (see Figure 17).

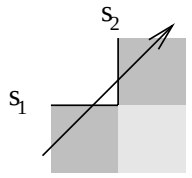


Figure 17: The only possible situation in $Dig(A,r)$.

Proof: Let $x \in bdPar(A,-r)$ and let $v = n(x,2r)$. Since v has length $2r$, v can intersect at most three parallel horizontal grid lines or three parallel vertical grid lines.

We can have 15 cases (modulo reflection and 90^o rotation) which satisfy the assumptions of this lemma. The 7 cases in which sides s_1, s_2 are parallel are shown in Figure 18 and the remaining 8 cases in which sides s_1, s_2 are perpendicular are shown in Figure 19. We show that only the case number 8 is possible.

We show that cases 1-7 and 9-15 are not possible by applying Lemma 2 (which follows). In the following we indicate whether we apply case(a) or case(b) of Lemma 2 and to which sides (see Figures 18 and 19):

1. case(a) applied to sides s_1, s_2.
2. case(a) applied to sides s_3, s_2.

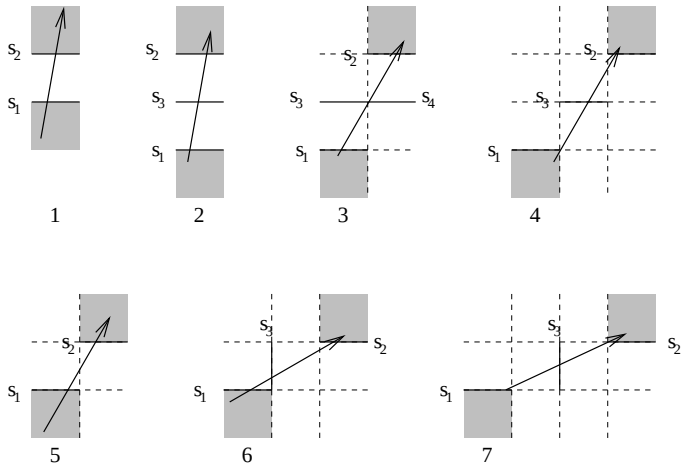


Figure 18: The 7 cases in which sides s_1, s_2 are parallel.

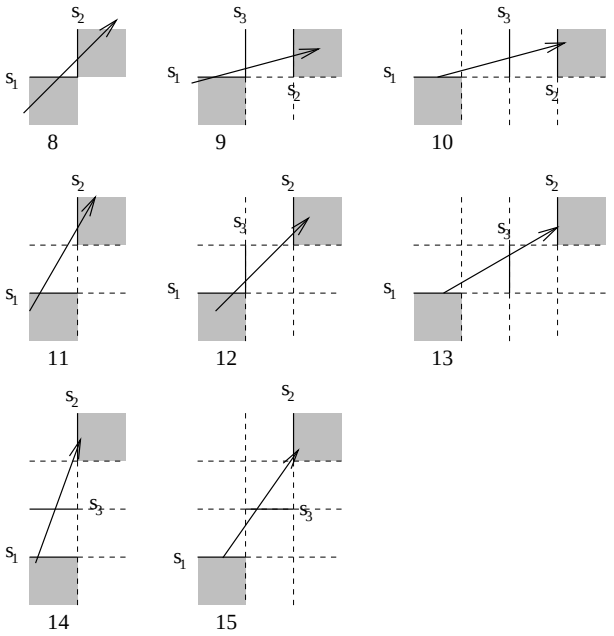


Figure 19: The 8 cases in which sides s_1, s_2 are perpendicular.

3. If v intersects side s_4, then we have case(a) applied to sides s_4, s_2. If v intersects side s_3, then we have case(b) for s_3, s_2.

4. In this case v must intersects s_3. case(b) for s_3, s_2.

5. case(b) for s_1, s_2.

6, 7. case(b) for s_3, s_2.

8. This case is not ruled out by Lemma 1.

9, 10. case(a) for s_3, s_2.

11. case(b) for s_1, s_2.

12, 13, 14, 15. case(b) for s_3, s_2.

The square determined by s_1, s_2 is white (i.e., does not belong to $Dig(A,r)$), since it contains line segment a_1a_2 and $a_1a_2 \cap Dig(A,r) = \{a_1, a_2\}$. ■

Lemma 2 *Let A be a par(r)-regular set and let $Dig(A,r)$ be a digital image of A which satisfies conditions ds1, ds2, and ds3. Let $x \in bdPar(A,-r)$ and let $v = n(x,2r)$ (v is oriented from x to its end point).*

case(a) *Let squares s_1, s_2 of $Dig(A,r)$ share the side $s = s_1 \cap s_2$. Let v intersect first the side parallel and opposite to the side s and then the side s (see Figure 20.case(a)). If square s_2 is black (i.e., $s_2 \in Dig(A,r)$), then s_1 is black.*

case(b) *Let squares s_1, s_2 share a vertex c but not share a side. Let v intersect first a side of s_1 that does not contain c and then a side of s_2 that contains c (see Figure 20.case(b)). If square s_2 is black (i.e., $s_2 \in Dig(A,r)$), then s_1 is black.*

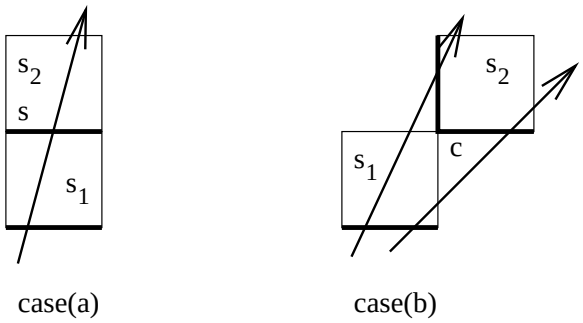


Figure 20: In both cases, if square s_2 is black, then s_1 is black.

Proof: Let m be the midpoint of vector v. Then $m \in bdA$. We show

$$area(A \cap s_1) \geq area(A \cap s_2). \tag{1}$$

Since (see Figure 21)

$$area(A \cap s_1) \geq area(iob(m,r) \cap s_1) \tag{2}$$

and

$$area((oob(m,r))^c \cap s_2) \geq area(A \cap s_2), \tag{3}$$

it is sufficient to show

$$area(iob(m,r) \cap s_1) \geq area((oob(m,r))^c \cap s_2), \tag{4}$$

as then (1) will follow from (4). The inequality (4) is proven in the appendix. ■

Now we are ready to prove our main theorems.

Theorem 4 *Let A be a par(r)-regular set. Then A and $Dig(A,r)$ are homotopy equivalent for every digitization $Dig(A,r)$, and $d_H(A, Dig(A,r)) \leq r$, where d_H is Hausdorff distance (which satisfies conditions ds1, ds2, and ds3).*

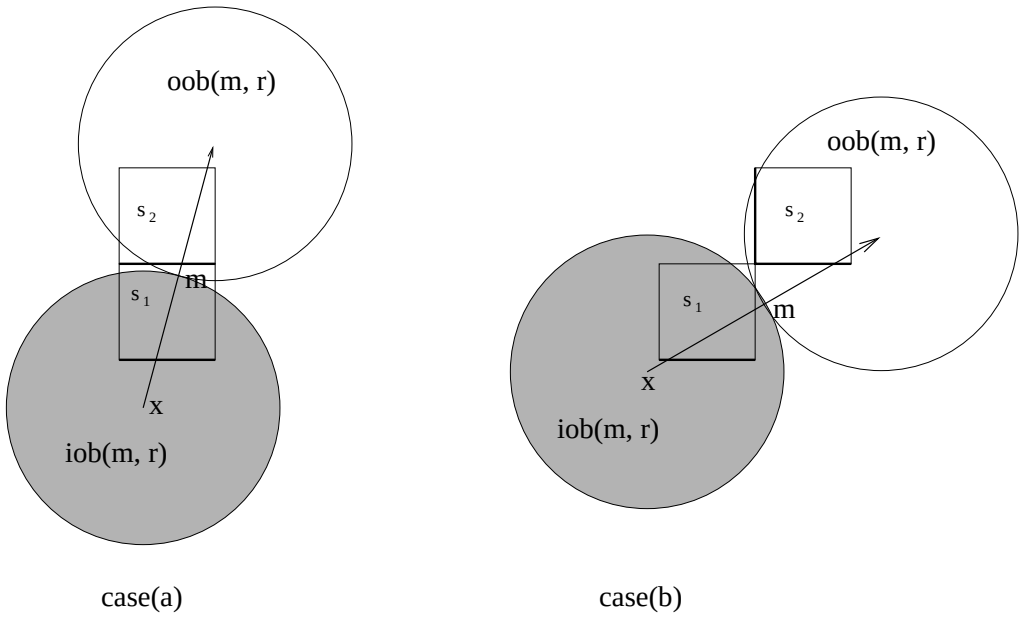


Figure 21: In both cases, $area(iob(m,r) \cap s_1) \geq area((oob(m,r))^c \cap s_2)$.

Proof: By Theorem 2, $Par(A, -r)$ is a strong deformation retract of A, and therefore A and $Par(A, -r)$ are homotopy equivalent. By Theorem 3, $Par(A, -r)$ and $Dig(A, r)$ are homotopy equivalent. Thus, A and $Dig(A, r)$ are homotopy equivalent. ■

For Theorem 5, we need the following concepts:

Definition: We call a closed set A a **bordered 2D manifold** if every point in A has a neighborhood homeomorphic to a relatively open subset of a closed half-plane. A connected component of a 2D bordered manifold is called a **bordered surface**.

We suspect strongly that every par(r)-regular set is a bordered 2D manifold. However, a proof of this assertion would be beyond the scope of this paper. Therefore, in Theorem 5, we explicitly assume that a set A is a bordered 2D manifold.

Theorem 5 *Let A be a par(r)-regular bordered 2D manifold. Then A and $Dig(A, r)$ are homeomorphic for every digital image $Dig(A, r)$ (which satisfies conditions ds1, ds2, and ds3).*

Proof: We will base our proof on the following theorem[5] from Ahlfors and Sario [1], Section 42A, page 98:

"*Two bordered surfaces are topologically equivalent (i.e., homeomorphic) if and only if they agree in character of orientability, number of contours, and Euler characteristic.*"

[5]We would like to thank Prof. Yung Kong (Queens College, CUNY, New York) for pointing out to us both this theorem and its consequences.

Since A and $Dig(A, r)$ are subsets of the plane, they agree in character of orientability. Without loss of generality, we can assume that set A is connected, since if A is not connected, we can apply the following proof to every connected component of A and we have a complete correspondence of components of A and $Dig(A, r)$, since A and $Dig(A, r)$ are homotopy equivalent by Theorem 4. We recall that a connected component of a 2D bordered manifold is a bordered surface. Thus, A is a bordered surface.

By Theorem 4, A and $Dig(A, r)$ agree in Euler characteristic, and since A is connected, $Dig(A, r)$ is also connected. By Theorem 9 (given below), $Dig(A, r)$ is a bordered surface. It remains to show that A and $Dig(A, r)$ agree in number of contours.

For (connected) bordered surfaces in the plane, the Euler characteristic is equal to (2 - the number of contours). This follows, for example, from Theorem 1 of Chapter 13 (p. 91) in Moise [21].[6] Thus, we obtain that A and $Dig(A, r)$ agree in number of contours, since A and $Dig(A, r)$ agree in Euler characteristic. ■

An important consequence of Theorem 5 is the fact that under correct digitization resolution any two digital images of a given spatial object A are topologically equivalent. This means, for example, that shifting or rotating an object or the camera cannot lead to topologically different images, i.e., topological properties of obtained digital images are invariant under shifting and rotation.

Theorem 6 *Let A be a par(r)-regular bordered manifold. Then any two digitizations $Dig^1(A, r)$ and $Dig^2(A, r)$ of A are homeomorphic.*

Proof: By Theorem 5, both $Dig^1(A, r)$ and $Dig^2(A, r)$ are homeomorphic to A. ■

Theorem 7 *Let A be a $\mathcal{C}^2$ subset of the plane (i.e., A is the closure of an open set whose boundary can be described as a disjoint finite union of twice continuously differentiable simple closed curves). Then there always exists a digitization resolution $r > 0$ such that every digitization $Dig(A, r)$ of A is topology preserving.*

Proof: First we show that there always exists $r > 0$ such that, for every $x, y \in bdA$, $n(x, r)$ and $n(y, r)$ do not intersect.

Step 1. Let k_{max} be the maximum of the absolute value of the principal curvatures at every point on bdA (the existence follows from compactness of bdA).

Step 2. Let $t < \frac{1}{k_{max}}$ be fixed. By elementary arguments from differential geometry (see [3], for example), it follows that, for every $x \in bdA$, there exists $e(x) > 0$ such that, for all $y \in bdA$, $d_{bdA}(x, y) < e(x)$ implies that $n(x, t)$ and $n(y, t)$ do not intersect, where d_{bdA} is the intrinsic distance on bdA.

[6] This theorem implies that any k-annulus is homeomorphic to some simple standard k-annulus of our choice, for which we can easily relate the Euler characteristic and the number of contours and show that (Euler characteristic = 2− the number of contours).

Step 3. Let $B_{bdA}(x, e(x))$ be an open ball in the intrinsic distance d_{bdA} on bdA. Since the collection $\{B_{bdA}(x, e(x)) : x \in bdA\}$ is an open covering of a compact set bdA, there exists a finite subcovering $\{B_{bdA}(x_1, e(x_1)), ..., B_{bdA}(x_n, e(x_n))\}$. Therefore, there exists $\epsilon > 0$ such that $\forall x \in bdA\ \exists i \in \{1, ..., n\}\ B_{bdA}(x, \epsilon) \subseteq B_{bdA}(x_i, e(x_i))$. Hence, for every $x, y \in bdA$, $d_{bdA}(x, y) < \epsilon$ implies that $n(x, t)$ and $n(y, t)$ do not intersect.

Step 4. For every $x \in bdA$, let $A(x) = \{y \in bdA\text{: } d_{bdA}(x, y) \geq \epsilon\}$. Since $x \notin A(x)$ and the set $A(x)$ is compact, we obtain that $d(x, A(x)) > 0$, where d is the Euclidean distance in $\mathbb{R}^2$. By compactness of bdA, there exists $\delta > 0$ such that $\delta \leq d(x, A(x))$ for every $x \in bdA$.[7] Hence, for every $x, y \in bdA$, $d_{bdA}(x, y) \geq \epsilon$ implies $d(x, y) \geq \delta$.

Step 5. Let $0 < r < min\{\delta/2, t\}$. For every $x, y \in bdA$, $n(x, r)$ and $n(y, r)$ do not intersect: Let $x \in bdA$. From Step 3, it follows that, for every $y \in bdA$ such that $d_{bdA}(y, x) < \epsilon$, $n(x, r)$ and $n(y, r)$ do not intersect. From Step 4, it follows that, for every $y \in bdA$ such that $d_{bdA}(y, x) \geq \epsilon$, $n(x, r)$ and $n(y, r)$ do not intersect, since $d(x, y) \geq \delta > 2r$.

Applying these steps to clA^c, we obtain that there exists a constant $s > 0$ such that, for every $x, y \in bdA$, $-n(x, s)$ and $-n(y, s)$ do not intersect. Taking $u = min(r, s)$, we obtain that A is u parallel regular, by Theorem 1. The assertion follows from Theorem 5. ■

It is easy to give examples of non par(r)-regular sets which are not topologically equivalent to their digital images. For example, set A in Figure 22(a) is simply connected, but $Dig_{\cap}(A, r)$ represented by gray squares is not simply connected. Of course, one can always find a set X having some special shape which is not par(r)-regular, yet X and $Dig_{\cap}(X, r)$ are homotopy equivalent, like the set presented in Figure 22(b). Although topology was preserved in digitizing the set shown in Figure 22(b), it is clear that important local geometric properties were lost. It can be shown that if a set A is par(r)-regular, then $Dig_{\cap}(A, r)$ will never significantly change its local geometric properties (Gross and Latecki [9]).

Figure 22: (a) A and $Dig_{\cap}(A, r)$ are not homotopy equivalent. (b) X and $Dig_{\cap}(X, r)$ are homotopy equivalent.

[7] We can obtain δ by similar arguments as for ϵ in Step 3 or by the continuity of function $x \longmapsto d(x, A(x))$.

4 Digital Patterns in Digitizations

In this section, we show that if A is a par(r)-regular set, then some digital patterns cannot occur in its digitization $Dig(A, r)$. This is very useful for noise detection, since if these patterns occur, they must be due to noise. So, if in a practical application the resolution r of the digitization is such that the parts of the object which have to be preserved under the digitization are compatible with the square sampling grid, then our results allow for efficient noise detection. By Theorem 8 below, $Dig(A, r)$ is well-composed as defined in Latecki, *et al.* [18]. Well-composed sets have very nice digital topological properties; in particular, a digital version of the Jordan Curve Theorem holds and the Euler characteristic is locally computable. These results imply that many algorithms in digital image processing can be simpler and faster.

Theorem 8 *Let A be par(r)-regular. Then the pattern shown in Figure 23 and its 90^o rotation cannot occur in any $Dig(A, r)$.*

Figure 23: This pattern and its 90^o rotation cannot occur in every $Dig(A, r)$.

Proof: Let c be the common vertex of all four closed squares. We first assume that $c \notin A$ and show that the pattern shown in Figure 23 and its 90^o rotation cannot occur in the configuration of the four squares.

Let S_2 and S_4 be black, and S_1 and S_3 be white as shown in Figure 24.a, where S_1, S_2, S_3, and S_4 are closed squares. We prove that this assumption leads to a contradiction. Since A is closed and $c \notin A$, there is an $e > 0$ such that $B(c, e) \cap A = \emptyset$, where $B(c, e)$ denotes (as always) a closed ball.

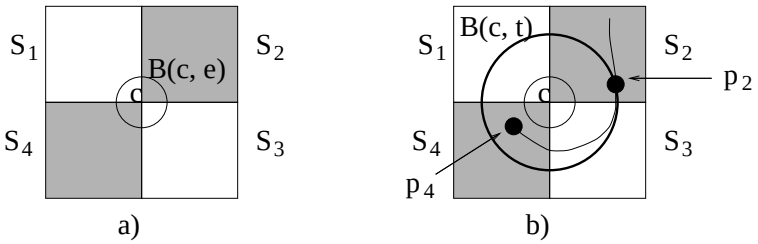


Figure 24: The small circle illustrates ball $B(c, e)$ and the big circle illustrates ball $B(c, t)$.

There must be points of A in both squares S_2 and S_4, since if $S_2 \cap A = \emptyset$, then S_2 would be white by **ds2**. Therefore, $S_2 \cap A \neq \emptyset$, and similarly $S_4 \cap A \neq \emptyset$.

Let p_2 be a point with the shortest distance t to c in $S_2 \cap A$.

Let p_4 be a point with the shortest distance d to c in $S_4 \cap A$.

Clearly, points p_2, p_4 belong to bdA and $t > 0$, $d > 0$, since $c \notin A$, $c \in S_2$, and $c \in S_4$. Without loss of generality, we assume that $t \geq d$. Consider the closed ball $B(c,t)$. We show that p_2 and p_4 belong to two different components of $B(c,t) \cap bdA$. Assume that this is not the case. Then, for some component C of bdA, it follows that $C = arc_1(p_2,p_4) \cup arc_2(p_2,p_4)$, $arc_1(p_2,p_4) \cap arc_2(p_2,p_4) = \{p_2,p_4\}$, and $arc_1(p_2,p_4) \subseteq B(c,t)$ or $arc_2(p_2,p_4) \subseteq B(c,t)$.

Assume $arc_1(p_2,p_4) \subseteq B(c,t)$ and, without loss of generality, assume that $arc_1(p_2,p_4)$ goes through S_3. Then

$arc_1(p_2,p_4) \cap face(S_2,S_3) \cap (B(c,t) - B(c,e)) \neq \emptyset$ and

$arc_1(p_2,p_4) \cap face(S_3,S_4) \cap (B(c,t) - B(c,e)) \neq \emptyset$,

where $face$ denotes the common face of two squares (see Figure 24.b). In this case,

$$arc_1(p_2,p_4) \cap S_3 \subseteq (B(c,t) - B(c,e)) \cap S_3.$$

Since the diameter of square S_3 is r, no component other then C of bdA intersects S_3, by Proposition 5. Therefore, $A \cap S_3$ contains $(S_3 - B(c,t))$ together with part of A between $arc_1(p_2,p_4)$ and $bdB(c,t)$ in S_3. We also have that $A \cap S_2 \subseteq (S_2 - B(c,t))$, since no point in $S_2 \cap A$ is closer to c than distance t (by the definition of constant t). Consequently, $area(A \cap S_3) \geq area(A \cap S_2)$. Thus, square S_3 should be black.

This contradiction implies that $arc_i(p_2,p_4) \not\subseteq B(c,t)$ for $i = 1,2$. Therefore, $bdA \cap B(c,t)$ has at least two components, one containing p_2 and the second containing p_4. In each of these components there is a point with the shortest distance ($\leq t$) to c; call them x_2 and x_4. Then $c \in n(x_2,r) \cap n(x_4,r)$, a contradiction. We have thus shown that

($*$) if $c \notin A$, then the pattern shown in Figure 23 and its 90^o rotation cannot occur in the four squares of $Dig(A,r)$ which have c as their common vertex.

The case in which $c \in A - bdA$ follows directly from the result above applied to the digitization of the complement A^c of A (i.e., the roles of A^c and A are interchanged).

It remains to consider the case in which $c \in bdA$. Let again S_2 and S_4 be black, and S_1 and S_3 be white in $Dig(A,r)$, as shown in Figure 24.a, where S_1, S_2, S_3, and S_4 are closed squares. This implies that

$$\epsilon = min\{area(S_2 \cap A), area(S_4 \cap A)\} - max\{area(S_1 \cap A), area(S_3 \cap A)\} > 0. \quad (5)$$

We denote by $X + v$ the translation of a set X by vector v and by $A \triangleleft B = (A - B) \cup (B - A)$. It is easy to observe that

$$|area(S \cap A) - area((S+v) \cap A)| \leq area(S \triangleleft (S+v)) \quad (6)$$

for every square $S \in \mathcal{Q}$ and every vector v (see Figure 25).

Since $c \in bdA$, there are points of the complement A^c in every neighborhood of c. Therefore, there exists a vector v such that $c + v \notin A$ and $area(S \triangleleft S') < \epsilon/2$ for every square $S \in \mathcal{Q}$, where $S' = S + v$. As a consequence of this fact and inequalities (5) and (6), we obtain

$$min\{area(S'_2 \cap A), area(S'_4 \cap A)\} - max\{area(S'_1 \cap A), area(S'_3 \cap A)\} > 0. \quad (7)$$

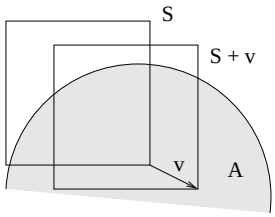


Figure 25: $|area(S \cap A) - area((S + v) \cap A)| \leq area(S \triangleleft (S + v))$.

Therefore, S'_2 and S'_4 are black, and S'_1 and S'_3 are white, in the digitization $Dig'(A, r)$ of A with respect to the square cover $\mathcal{Q}$ translated by v. This contradicts $(*)$, since $c + v \notin A$ and $c + v$ is the common vertex of the four squares. The obtained contradiction proves that if $c \in bdA$, then the pattern shown in Figure 23 and its 90^o rotation cannot occur in the four squares of $Dig(A, r)$ which have c as their common vertex. ■

Theorem 9 *Let A be par(r)-regular. Then $Dig(A, r)$ is a bordered 2D manifold.*

Proof: Since the configuration shown in Figure 23 (and its 90^o rotation) cannot occur in $Dig(A, r)$ by Theorem 8, there exist only three 2×2 configurations of boundary squares in $Dig(A, r)$ shown in Figure 26 (modulo reflection and 90^o rotation).

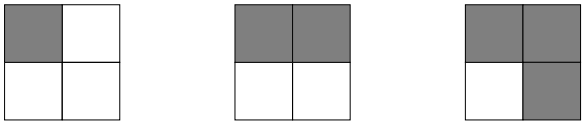


Figure 26: The only possible 2×2 configurations of boundary squares in $Dig(A, r)$ of a par(r)-regular set A (modulo reflection and 90^o rotation).

Therefore, if we view $Dig(A, r)$ as a subset of $\mathbb{R}^2$, every point in $Dig(A, r)$ has a neighborhood homeomorphic to a relatively open subset of a closed half-plane. Hence $Dig(A, r)$ is a bordered 2D manifold. ■

5 Conclusions

In this paper, we gave conditions on the correct digitization resolution which guarantee that topology is preserved under a digitization and segmentation process. For a par(r)-regular set A, A and each of its digital images $Dig(A, r)$ are topologically equivalent. We also proved that Hausdorff distance of sets A and $Dig(A, r)$ is less or equal to r. This results have many consequences. For example, under correct digitization resolution any two digital images of a given spatial object A are topologically equivalent. Furthermore, $Dig(A, r)$ is well-composed, i.e., the checker board digital patterns cannot occur in $Dig(A, r)$. In Latecki *et al.* [18] it is shown that well-composed sets have very nice digital properties, which imply that many

algorithms for digital image processing can be simpler and faster. Well-composedness can be useful for noise detection, since if the neighborhood of a boundary point contains a checker board digital pattern, it must be due to noise. For a large class of 2D objects, which includes projections of some real 3D objects, a constant r can be computed such that they are par(r)-regular.

Our definition of a digitization models a real digitization process. However, we did not consider blurring effects to the underlying objects. Thus, we modeled perfectly focused object boundaries. We are currently extending our results to the digitizations of objects with blurred boundaries (see Gross and Latecki [10]).

A Appendix

The goal of this appendix is to prove the inequality (4) in the proof of Lemma 2 (Figure 21):

$$area(s_2 - oob(m, r)) \leq area(iob(m, r) \cap s_1).$$

We will prove this inequality in **case (a)**; the proof in **case (b)** is based on analogous arguments. We first restate the assumptions of **case (a)** in Lemma 2.

Let squares s_1 and s_2 have a side with endpoints C and O in common, i.e., $CO = s_1 \cap s_2$. Let the other side of s_1 that is parallel to CO have endpoints D and E, and the other side of s_2 that is parallel to CO have endpoints B and A (see Figure 27). We assume that vector $v = n(x, 2r)$ first intersects side DE and then CO. Without loss of generality, we can assume that the length of the sides of a square (in the square grid) is 1, and therefore, $r = \sqrt{2}$ and the length of vector v is $2\sqrt{2}$. Let M be the midpoint of v. The beginning point of v is the center of ball $iob(M, \sqrt{2})$ of radius $\sqrt{2}$, the endpoint of v is the center of ball $oob(M, \sqrt{2})$ of radius $\sqrt{2}$, and $iob(M, \sqrt{2}) \cap oob(M, \sqrt{2}) = \{M\}$. We will denote $iob(M, \sqrt{2})$ by bb (black ball) and the its boundary by bc (black circle). Similarly, we will denote $oob(M, \sqrt{2})$ by wb (white ball) and the its boundary by wc (white circle). We will denote the center of bb by Z_b and the center of wb by Z_w.

We introduce the Cartesian coordinate system such that the origin is at point O, i.e., $O = (0, 0)$, and $A = (0, 1)$ and $C = (-1, 0)$. The goal of this appendix is to prove

Theorem 10

$$area(s_2 - wb) \leq area(s_1 \cap bb). \tag{8}$$

Adding $area(s_1 - bb)$ to both sides of (8) yields

$$area(s_2 - wb) + area(s_1 - bb) \leq 1, \tag{9}$$

since $area(s_1) = area(s_2) = 1$. Inequality (9) is trivially true when $s_2 \subseteq wb$ or $s_1 \subseteq bb$. Therefore, we assume that $s_2 \not\subseteq wb$ and $s_1 \not\subseteq bb$. We will prove inequality (9) for the following four cases:

(I.1) ((M is below or on line CO) and ($O \in wb$ or $C \in wb$)) or

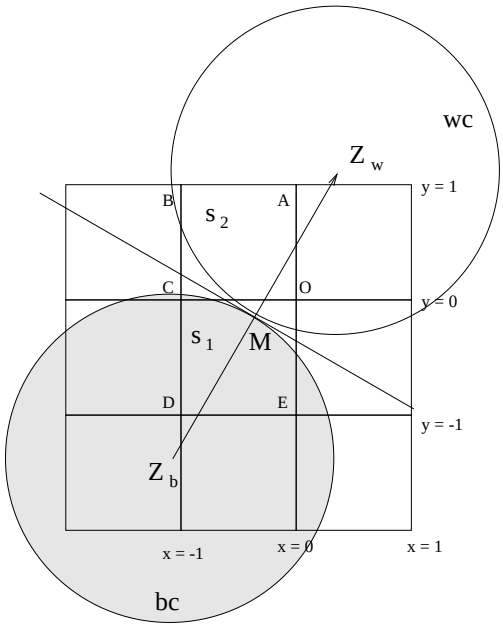


Figure 27:

(I.2) ((*M* is below or on line *CO*) and (*O* $\notin$ *wb* and *C* $\notin$ *wb*)) or

(II.1) ((*M* is above *CO*) and (*O* $\in$ *bb* or *C* $\in$ *bb*)) or

(II.2) ((*M* is above *CO*) and (*O* $\notin$ *bb* and *C* $\notin$ *bb*)).

The inequality (8) for cases (I.2) and (II.2) is proved in Lemma 5 below. By Lemma 6 below, case (I.1) reduces to the case in which ($O \in v$ or $C \in v$). By Lemma 7 below, case (II.1) reduces to the case in which ($D \in v$ or $E \in v$). Lemma 8 below demonstrates how to prove (9) for these cases. We begin with the following simple geometric fact.

Lemma 3 *Let point X lie outside a closed ball B with center Z. Let* $H(XZ)$ *denote one of the closed half planes determined by the straight line XZ. Let* B' *be a closed ball B rotated around X such that the center* Z' *of* B' *is not contained in* $H(XZ)$ *and* $H(XZ) \cap B \cap B' \neq \emptyset$ *(see Figure 28(a)). Then* $H(XZ) \cap B' \subset H(XZ) \cap B$.

Proof: The circles *bdB* and *bdB*$'$ intersect at exactly two points. The line L that passes through the two intersection points of *bdB* and *bdB*$'$ goes through the midpoint of line segment ZZ' and is perpendicular to ZZ' (see Figure 28(a)). Let $H(L)$ be the half plane determined by L that contains point Z. Since two distinct circles can intersect at no more than two points and $Z' \notin H(L)$, we obtain $H(L) \cap B' \subset H(L) \cap B$. Since $H(XZ) \cap B \subset H(L) \cap B$, we have $H(XZ) \cap B' \subset H(XZ) \cap B$. ■

Using arguments similar to those of Lemma 3, we can also prove

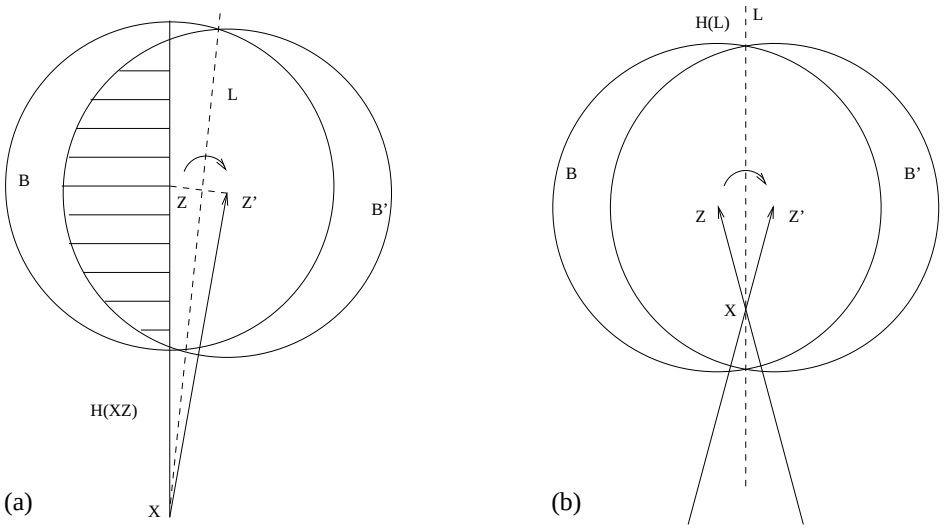


Figure 28:

Lemma 4 *Let point X lie inside a closed ball B with center Z, where $X \neq Z$. Let B' be a closed ball B rotated around X such that $B' \neq B$ (see Figure 28(b)). Let L be the straight line through the two intersection points of bdB and bdB' and let $H(L)$ be the half plane determined by L that contains point Z. Then L goes through X and $H(L) \cap B' \subset H(L) \cap B$.*
■

Lemma 5 *If the midpoint M of vector v is on or below line CO and $O \notin wb$ and $C \notin wb$, then $area(s_2 - wb) \leq area(s_1 \cap bb)$. If M is on or above line CO and $O \notin bb$ and $C \notin bb$, then $area(s_2 - wb) \leq area(s_1 \cap bb)$.*

Proof: We prove only the first part. The proof of the second part is analogous with wb and bb interchanged. Let M be below or on line CO and $O \notin wb$ and $C \notin wb$. First we will bound the $area(s_2 - wb)$ from above.

Since wc only intersects line CO (i.e., line $y = 0$) between points C and O, the point S with the smallest y-coordinate in wb is contained in $wb \cap s_1$ (see Figure 29(a)). Let wb', v', and S' denote wb, v, and S translated parallel to y-axis (by vector $\ll 0, y \gg$ with $0 \leq y$) such that $S' \in CO$. Since $wb' \cap s_2 \subseteq wb \cap s_2$, we have $area(s_2 - wb) \leq area(s_2 - wb')$. Now we will bound $area(s_2 - wb')$ from above.

We have the situation in which $S' \in CO$ is the point with the smallest y-coordinate in wb'. If S' is different from C and O, then $s_2 - wb'$ is a union of two connected sets, say L and R, such that $L \cap R = \{S'\}$ and L is to the left and R to the right of S' (see Figure 29(b)). If $area(L) \leq area(R)$, then $area(s_2 - wb') = area(L) + area(R)$ can only increase if we translate wb' parallel to x-axis in the negative direction, i.e., by vector $\ll x, 0 \gg$, where $x < 0$. Let wb'_x and S'_x denote wb' and S' translated by vector $\ll x, 0 \gg$ such that $S'_x = C$ (see Figure 30(a)). Then $area(s_2 - wb') < area(s_2 - wb'_x)$, and consequently

$$area(s_2 - wb) \leq area(s_2 - wb'_x). \tag{10}$$

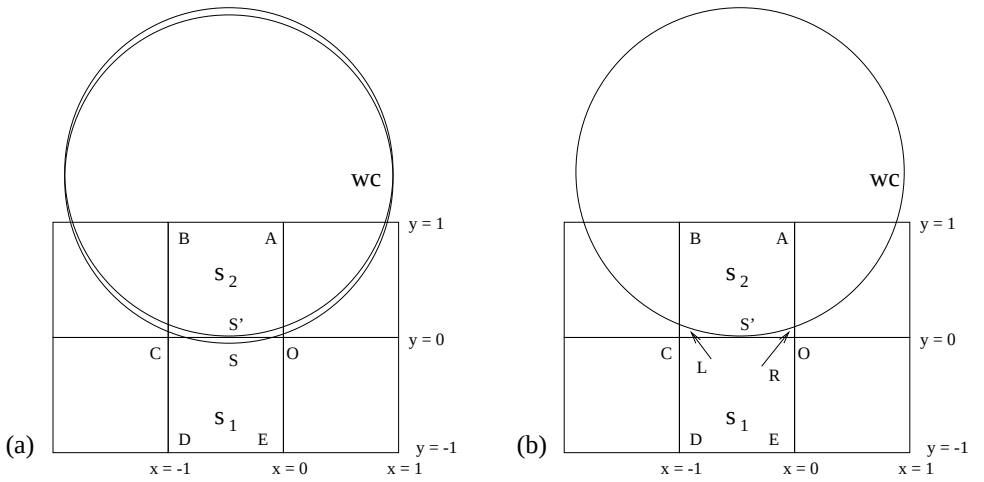


Figure 29:

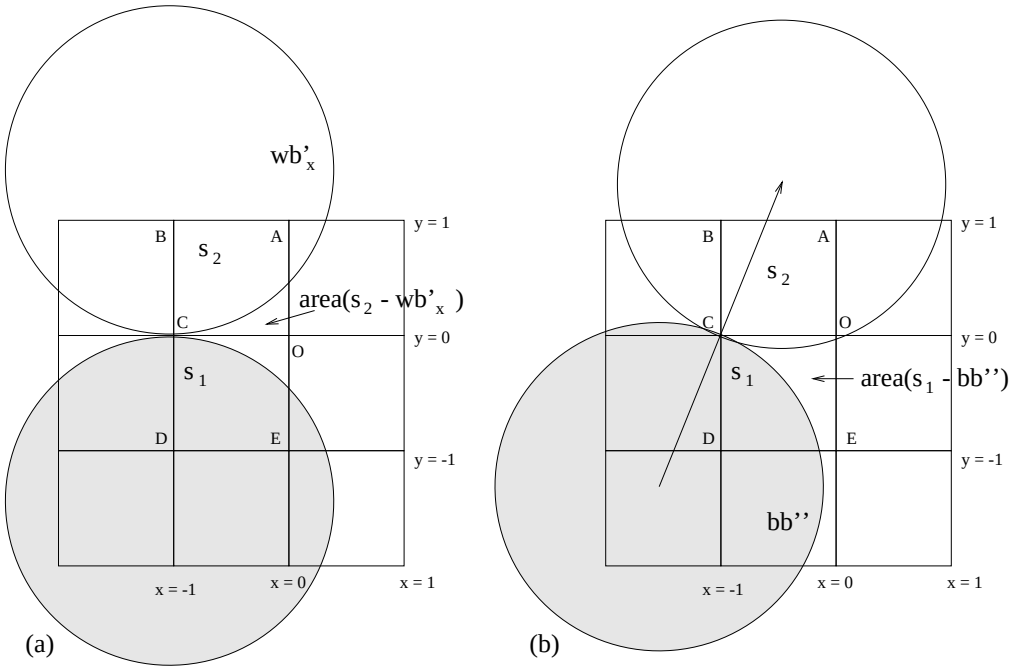


Figure 30:

Now we will show that $area(s_2 - wb'_x) < area(s_1 \cap bb)$. Since M is on or below line CO and $O \notin wb$ and $C \notin wb$, we have $wc \cap BC \neq \emptyset$ and $wc \cap AO \neq \emptyset$. Let $wc \cap BC = \{S\}$ and $wc \cap AO = \{T\}$ (see Figure 31(a)).

Let wb_α, v_α, and M_α denote wb, v, and M rotated around the center Z_b of bb by angle α. Let $wc_\alpha \cap BC = \{S_\alpha\}$ and $wc_\alpha \cap AO = \{T_\alpha\}$. Since points S and T are in two different half planes determined by vector v, we have the following implications by Lemma 3:

If $\alpha > 0$, then $|CS_\alpha| < |CS|$ and $|OT_\alpha| > |OT|$. If $\alpha < 0$, then $|CS_\alpha| > |CS|$ and $|OT_\alpha| < |OT|$.

Therefore, there exists angle α such that $|CS_\alpha| = |OT_\alpha|$. (Since bb does not change its

location during this rotation, $area(s_1 \cap bb)$ remains constant.) Let wb', v', M', and bb' denote wb_α, v_α, M_α, and bb translated parallel to y-axis by vector $\ll 0, -|CS_\alpha| \gg$ (see Figure 31(b)). Then $wc' \cap BC = \{C\}$ and $wc' \cap AO = \{O\}$, and $area(s_1 - bb) \leq area(s_1 - bb')$. We also have that $M' \in wc' \cap s_1$, since M and M_α both lie in between lines $x = -1$ and $x = 0$.

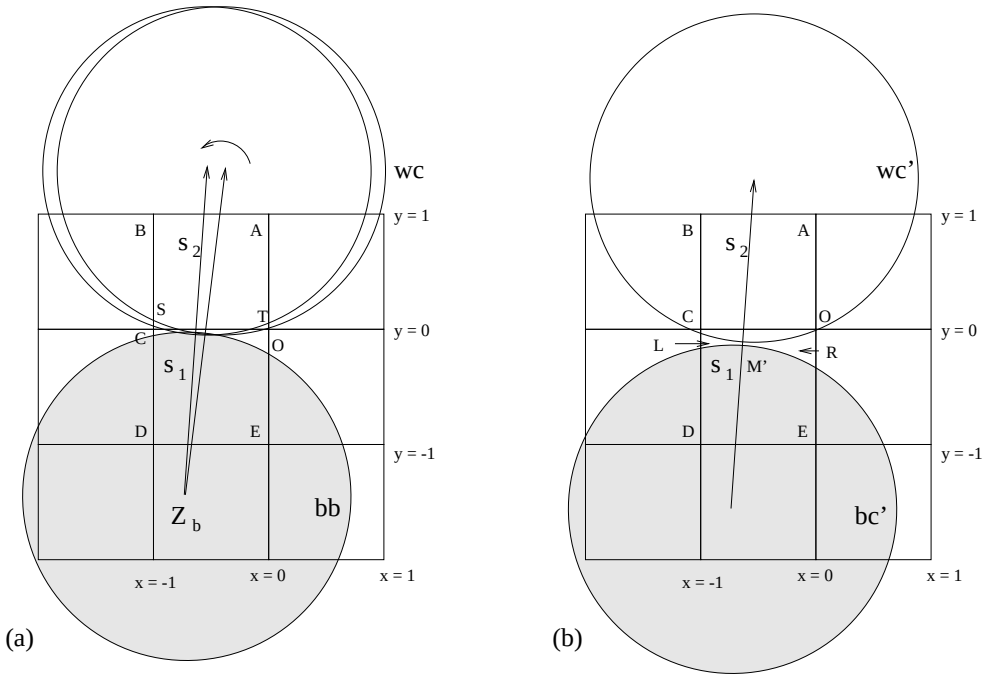


Figure 31:

We assume that M' is distinct from C and O. Observe that $(s_1 - bb') - (wb' \cap s_1)$ is a union of two disjoint connected sets, say L and R, such that L is to the left and R to the right of M'.

If $area(L) \leq area(R)$, then $area((s_1 - bb') - (wb' \cap s_1)) = area(L) + area(R)$ can only increase if BB is rotated clockwise around Z_w (i.e., to the left). Let wb'', v'', M'', and bb'' denote wb'_r, v'_r, M'_r, and bb' rotated clockwise around Z_w by the smallest angle such that $M'' = C$ (see Figure 30(b)). The configuration wb'', v'', M'', and bb'' shown in Figure 30(b) is uniquely determined for every starting configuration wb'_r, v'_r, M'_r, and bb', where $area(L) \leq area(R)$. Thus, we obtain

$$area((s_1 - bb') - (wb' \cap s_1)) < area((s_1 - bb'') - (wb' \cap s_1)),$$

and consequently $area(s_1 - bb') \leq area(s_1 - bb'')$. Therefore, $area(s_1 - bb) \leq area(s_1 - bb'')$, which implies

$$area(s_1 \cap bb'') \leq area(s_1 \cap bb). \tag{11}$$

By calculating $area(s_2 - wb'_x)$ and $area(s_1 \cap bb'')$, we obtain $area(s_2 - wb'_x) \leq area(s_1 \cap bb'')$ (this inequality is visualized in Figure 32). By (10) and (11), we obtain $area(s_2 - wb) \leq area(s_1 \cap bb)$. ■

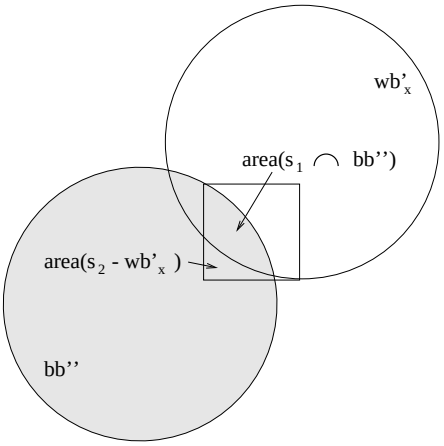


Figure 32:

Lemma 6 *Let (M be on or below line CO) and (O ∈ wb or C ∈ wb). Then*

$$area(s_2 - wb) + area(s_1 - bb) \leq area(s_2 - wb') + area(s_1 - bb),$$

where wb′ and v′ are wb and v rotated around Z_b such that either O ∈ v′ or C ∈ v′ (see Figure 33(a)).

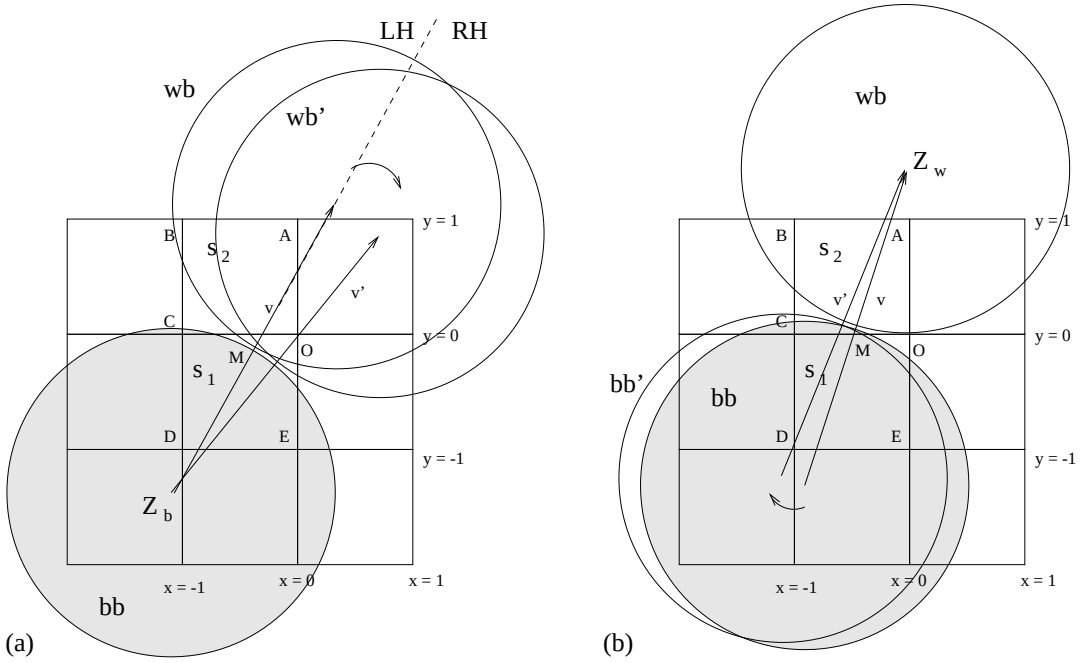


Figure 33:

Proof: Observe first that ($O \in wb$ or $C \in wb$) iff

1. ($O \in wb$ and $C \notin wb$) or

2. ($O \notin wb$ and $C \in wb$) or

3. ($O \in wb$ and $C \in wb$).

Since the distance between M and the endpoint Z_w of vector v is not greater than $\sqrt{2}$ and the angle of v (with x-axis) is in the interval $[45^o, 135^o]$, the endpoint Z_w of v must be contained in squares $s_2 \cup \ ... \ \cup s_7$, where $s_2, \ ... \ , s_7$ are closed squares named as shown in Figure 34. If $Z_w \in s_2 \cup s_5 \cup s_6 \cup s_7$, then one of these squares is contained in wb, and consequently their common corner A is contained in wb. Therefore, at least one of the points A and B is always contained in wb.

It can be calculated that if $C \notin wb$, then $Z_w \notin s_3 \cup s_4$, and therefore $A \in wb$. Similarly, if $O \notin wb$, then $B \in wb$. This implies that ($O \in wb$ or $C \in wb$) iff

1. ($O \in wb$ and $C \notin wb$ and $A \in wb$) or

2. ($O \notin wb$ and $C \in wb$ and $B \in wb$) or

3. ($O \in wb$ and $C \in wb$ and ($A \in wb$ or $B \in wb$)).

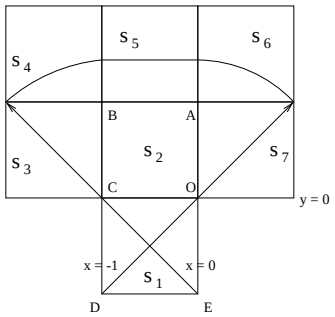


Figure 34:

These three cases imply the following two cases

(A) ($O \in wb$ and $A \in wb$) or

(B) ($C \in wb$ and $B \in wb$).

Below, we will show this lemma for case (A). The proof for case (B) is analogous. We thus assume case (A), which implies that line segment AO is contained in wb. Since the lemma is trivially true when $O \in v$, we will assume that $O \notin v$. Let wb' and v' be wb and v rotated around Z_b such that $O \in v'$ (see Figure 33(a)).

We denote by LH the closed half plane determined by vector v which does not contain O, and by RH the complement of LH. Since $O \in v'$ and the center Z'_w of wb' is the endpoint of v', we have $Z'_w \notin LH$. By Lemma 3, $LH \cap wb' \subset LH \cap wb$. Therefore,

$$LH \cap wb' \cap s_2 \subset LH \cap wb \cap s_2. \tag{12}$$

Since M is on or below line CO, the line segment $v \cap s_2$ is contained in wb. Since line segment OA is contained in wb, the convex hull of $v \cap s_2$ and OA is contained in wb. Since $RH \cap s_2$ is contained in this convex hull, $RH \cap s_2 \subset RH \cap wb$. Hence $RH \cap s_2 \subseteq RH \cap wb \cap s_2$. Since $RH \cap wb' \cap s_2 \subseteq RH \cap s_2$, we obtain

$$RH \cap wb' \cap s_2 \subseteq RH \cap wb \cap s_2. \tag{13}$$

By (12) and (13),

$$wb' \cap s_2 \subset wb \cap s_2,$$

which implies $s_2 - wb \subset s_2 - wb'$. Therefore,

$$area(s_2 - wb) < area(s_2 - wb'),$$

which implies

$$area(s_2 - wb) + area(s_1 - bb) \leq area(s_2 - wb') + area(s_1 - bb).$$

■

Using the same arguments as in Lemma 6, we can also prove

Lemma 7 *Let (M be on or above line CO) and ($O \in bb$ or $C \in bb$). Then*

$$area(s_2 - wb) + area(s_1 - bb) \leq area(s_2 - wb) + area(s_1 - bb'),$$

where bb' and v' are bb and v rotated around Z_w such that either $E \in v'$ or $D \in v'$ (see Figure 33(b)). ■

By Lemmas 6 and 7, it is sufficient to prove inequality (8) for the following cases:

(I.1′) ((M is below or on line CO) and ($O \in v$ or $C \in v$)) or

(II.1′) ((M is above CO) and ($D \in v$ or $E \in v$)).

Lemma 8 proves (8) when M lies below or on line CO and $O \in v$. Since the proofs for the remaining three cases are analogous, they are omitted.

Lemma 8 *Let the midpoint M of v lie on or below line CO and $O \in v$ (see Figure 35(a)). Then*

$$area(s_2 - wc) \leq area(s_1 \cap bc).$$

Proof: We assume that point Z_w (the endpoint of v and the center of the white circle wc) has coordinates (p, q), where $0 \leq p, q$. Under this assumption, the white circle wc satisfies the equation $(x - p)^2 + (y - q)^2 = 2$. The point Z_b, which is the starting point of v and the center of the black circle bc, has coordinates $(-a, -b)$, where $0 \leq a, b$. Thus, the black circle bc satisfies the equation $(x + a)^2 + (y + b)^2 = 2$.

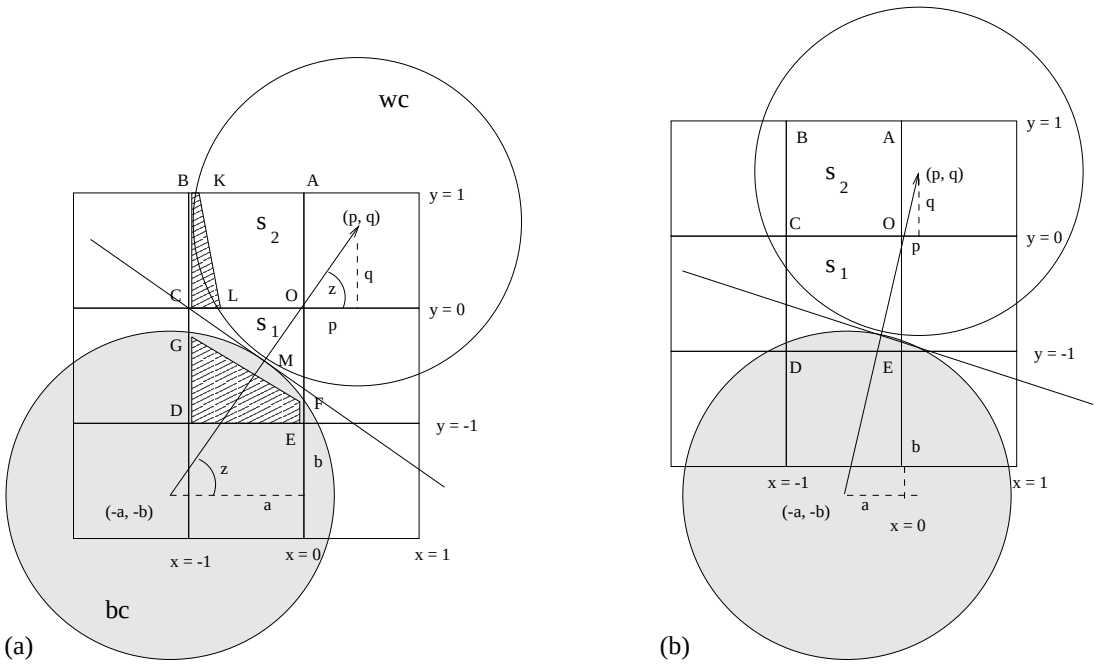


Figure 35:

Observe that $a \leq b$ and $p \leq q$, since the angle of vector v (with x-axis) is in the interval $[45^o, 90^o]$. Since the length of v is $2\sqrt{2}$, we have $\sqrt{a^2+b^2} + \sqrt{p^2+q^2} = 2\sqrt{2}$. Since M lies below the line segment CO, M belongs to line segment $Z_bO \subseteq v$. Therefore, we have $\sqrt{p^2+q^2} \leq \sqrt{2}$, which implies $p, q \leq \sqrt{2}$, and $\sqrt{2} \leq \sqrt{a^2+b^2}$. Note also that $\sqrt{a^2+b^2} \leq 2\sqrt{2}$ and $a, b \leq 2\sqrt{2}$.

Since the distance between M and the endpoint Z_w of vector v is not greater than $\sqrt{2}$ and the angle of v (with x-axis) is in the interval $[45^o, 90^o]$, the endpoint Z_w of v must be contained in squares $s_2 \cup s_5 \cup s_6 \cup s_7$, where s_2, s_5, s_6, s_7 are closed squares as shown in Figure 34. Therefore, one of these squares is contained in wb, and consequently its corner A. Since $O \in wb$, side AO is contained in wb. Therefore, there are the following configurations in which wc can intersect the sides of square s_2:

$s_2(1)$: $wc \cap BA \neq \emptyset$ and $wc \cap CO \neq \emptyset$ (see Figure 35(a)),

$s_2(2)$: $wc \cap BA = \emptyset$ and $wc \cap CO \neq \emptyset$; in this case $wc \cap CB \neq \emptyset$,

$s_2(3)$: $wc \cap BA \neq \emptyset$ and $wc \cap CO = \emptyset$; in this case $wc \cap CB \neq \emptyset$,

$s_2(4)$: $wc \cap BA = \emptyset$ and $wc \cap CO = \emptyset$; in this case $s_2 \subset wb$, since all corner points of s_2 are in wb.

Since midpoint M of v lies below line segment CO, there are the following configurations in which bc can intersect the sides of square s_1:

$s_1(1)$: $bc \cap CD \neq \emptyset$ and $bc \cap OE \neq \emptyset$ (see Figure 35(a)),

$s_1(2):\ bc \cap CD = \emptyset$ and $bc \cap OE \neq \emptyset$,

$s_1(3):\ bc \cap CD \neq \emptyset$ and $bc \cap OE = \emptyset$,

$s_1(4):\ bc \cap CD = \emptyset$ and $bc \cap OE = \emptyset$.

It is necessary to consider the conjunction of the cases for s_2 with the cases for s_1. We will give a detailed proof of this lemma only in case $s_2(1)s_1(1)$. The remaining cases will be omitted, since their proofs follow similar arguments or are comparatively simpler. We thus assume

$s_2(1)s_1(1)$: $wc \cap BA \neq \emptyset$, $wc \cap CO \neq \emptyset$, $bc \cap CD \neq \emptyset$, and $bc \cap OE \neq \emptyset$.

Let wc intersect side CO at a point L and side BA at a point K, and let bc intersect sides CD and OF at points G and F, respectively (see Figure 35(a)). Since $area(s_2 - wc) \leq |BC|\frac{|BK|+|CL|}{2}$, $|DE|\frac{|EF|+|DG|}{2} \leq area(s_1 \cap bc)$, and $|BC| = |DE| = 1$, it is sufficient to show

$$|BK| + |CL| \leq |EF| + |DG|. \tag{14}$$

First, we express $|BK| + |CL|$ in algebraic form. To find the intersection points of the white circle with sides BA and CO of square s_2, we solve the equation $(x-p)^2+(y-q)^2 = 2$ for $y = 0$ and $y = 1$. If $y = 0$, then $x^2 - 2px + p^2 + q^2 - 2 = 0$ and $x_{1,2} = p \pm \sqrt{2-q^2}$. Since we need the smaller x, we obtain $x = p - \sqrt{2-q^2}$ by $y = 0$. Thus $L = (p - \sqrt{2-q^2}, 0)$ is an intersection point of wc with CO.

If $y = 1$, then $x^2-2px+p^2+q^2-2q-1 = 0$ and $x_{1,2} = p \pm \sqrt{-q^2+2q+1}$. Since we need the smaller x, we obtain $x = p - \sqrt{-q^2+2q+1}$ by $y = 1$. Thus $K = (p - \sqrt{-q^2+2q+1}, 1)$ is an intersection point of wc with side BA. Therefore,

$$|BK|+|CL| = p - \sqrt{-q^2+2q+1} + 1 + p - \sqrt{2-q^2} + 1 = 2 + 2p - \sqrt{-q^2+2q+1} - \sqrt{2-q^2} \tag{15}$$

To calculate $|EF| + |DG|$, we find the intersection points of the black circle with sides CD and OE of square s_1, i.e., we solve the equation $(x + a)^2 + (y + b)^2 = 2$ for $x = 0$ and $x = -1$.

If $x = 0$, then $y^2 + 2by + a^2 + b^2 - 2 = 0$ and $y_{1,2} = -b \pm \sqrt{2-a^2}$. Since we need the greater value of y, we obtain $y = -b + \sqrt{2-a^2}$ by $x = 0$. Thus $F = (0, -b + \sqrt{2-a^2})$ is an intersection point of bc with side OE. We also obtain that $a \leq \sqrt{2}$.

If $x = -1$, then $y^2 + 2by + a^2 + b^2 - 2a - 1 = 0$ and $y_{1,2} = -b \pm \sqrt{-a^2+2a+1}$. Since we need the greater value of y, we obtain $y = -b + \sqrt{-a^2+2a+1}$ by $x = -1$. Thus $G = (-1, -b + \sqrt{-a^2+2a+1})$ is an intersection point of bc with side CD. Therefore,

$$|EF|+|DG| = -b+\sqrt{2-a^2}+1-b+\sqrt{-a^2+2a+1}+1 = 2-2b+\sqrt{2-a^2}+\sqrt{-a^2+2a+1}. \tag{16}$$

In order to show that $area(s_2 - wc) \leq area(s_1 \cap bc)$, using equations (15) and (16), it is sufficient to show that

$$2p - \sqrt{-q^2+2q+1} - \sqrt{2-q^2} \leq -2b + \sqrt{2-a^2} + \sqrt{-a^2+2a+1} \tag{17}$$

Inequality (17) is equivalent to the following

$$0 \leq -2p + \sqrt{-q^2 + 2q + 1} + \sqrt{2 - q^2} - 2b + \sqrt{2 - a^2} + \sqrt{-a^2 + 2a + 1} \qquad (18)$$

Before we prove (17), we will show that if $area(s_2 - wc) \neq 0$, then $\sqrt{a^2 + b^2} \leq \sqrt{5}$ and $\sqrt{p^2 + q^2} \geq 2\sqrt{2} - \sqrt{5}$.

We first assume that $a \geq 1$. If we had $b < 1$, then $DE \cap v = \emptyset$. Therefore, $b \geq 1$. Since $bc \cap OE \neq \emptyset$ and $O \notin bb$, we have $E \in bb$, which implies (see Figure 35(a)):

$$a^2 + (b-1)^2 \leq 2 \Rightarrow 0 \leq b \leq 2 \text{ and } a^2 + (b-1)^2 \leq 2 \Rightarrow a^2 + b^2 \leq 2b + 1 \leq 5.$$

Therefore $\sqrt{a^2 + b^2} \leq \sqrt{5}$.

We show that the assumptions $0 \leq a < 1$ and $\sqrt{a^2 + b^2} > \sqrt{5}$ imply $s_2 \subset wb$. Since $\sqrt{a^2 + b^2} + \sqrt{p^2 + q^2} = 2\sqrt{2}$, we obtain $\sqrt{p^2 + q^2} \leq 2\sqrt{2} - \sqrt{5} (\approx 0.592)$. Therefore, the endpoint of vector v (i.e. the center of wb) is contained in the square to the right of s_2 that shares a side with s_2 (see Figure 35(b)). Since this square is then contained in wb, its corners A and O are contained in wb.

Since the assumptions $0 \leq a < 1$ and $\sqrt{a^2 + b^2} \geq \sqrt{5}$ imply $b \geq 2$, we obtain $\frac{q}{p} = \frac{b}{a} > 2$ if $p > 0$, and therefore $2p < q$ for $p \geq 0$. Since $2p < q \leq \sqrt{p^2 + q^2} \leq 2\sqrt{2} - \sqrt{5}$, we have $2p < 2\sqrt{2} - \sqrt{5}$. The corner point C of s_2 is contained in wb iff $(1+p)^2 + q^2 \leq 2$ (see Figure 35(b)), which is true:

$$(1+p)^2 + q^2 = 1 + 2p + p^2 + q^2 < 1 + (2\sqrt{2} - \sqrt{5}) + (2\sqrt{2} - \sqrt{5})^2 (\approx 1.943) \leq 2.$$

The corner point B of s_2 is contained in wb iff $(1+p)^2 + (1-q)^2 \leq 2$ (see Figure 35(b)). Since $2p < q$ and $p^2 < \frac{1}{4}q^2$, we obtain

$$(1+p)^2 + (1-q)^2 = 2 + 2p - 2q + p^2 + q^2 \leq 2 - q + \frac{5}{4}q^2.$$

Since $2 - q + \frac{5}{4}q^2 \leq 2$ iff $q(\frac{5}{4}q - 1) \leq 0$, which is true for $0 \leq q \leq \frac{4}{5}$, we obtain that $(1+p)^2 + (1-q)^2 \leq 2$. Since all corner points of s_2 are contained in wb, $s_2 \subset wc$. This inconsistency implies that if $0 \leq a < 1$, then $\sqrt{a^2 + b^2} < \sqrt{5}$. Since $\sqrt{a^2 + b^2} + \sqrt{p^2 + q^2} = 2\sqrt{2}$, we obtain $\sqrt{p^2 + q^2} \geq 2\sqrt{2} - \sqrt{5}$. We have thus shown that $\sqrt{a^2 + b^2} \leq \sqrt{5}$ and $\sqrt{p^2 + q^2} \geq 2\sqrt{2} - \sqrt{5}$.

Now we will prove inequality (18) when the angle z of v (with x-axis) is in the interval $[\frac{\pi}{3}, \frac{\pi}{2}]$ (see Figure 35(a)). Let $r = \sqrt{p^2 + q^2}$. Then $2\sqrt{2} - \sqrt{5} \leq r \leq \sqrt{2}$. In this case $q = r\sin(z)$, $p = r\cos(z)$, $b = (2\sqrt{2} - r)\sin(z)$, and $a = (2\sqrt{2} - r)\cos(z)$. In order to show that (18) holds it is sufficient to show that $f(r, z) \geq 0$, where

$$\begin{aligned} f(r,z) &= -2\,r\,\cos(z) + \sqrt{2 - \left(2\sqrt{2} - r\right)^2 \cos(z)^2} \\ &\quad + \sqrt{1 + 2\left(2\sqrt{2} - r\right)\cos(z) - \left(2\sqrt{2} - r\right)^2 \cos(z)^2} - \\ &\quad 2\left(2\sqrt{2} - r\right)\sin(z) + \sqrt{2 - r^2\sin(z)^2} + \sqrt{1 + 2\,r\,\sin(z) - r^2\sin(z)^2}. \end{aligned}$$

Since the second derivative of f with respect to r is negative

$$\frac{\partial^2 f(r,z)}{\partial r} = -\frac{\left(2\sqrt{2}-r\right)^2\cos(z)^4}{\left(2-\left(2\sqrt{2}-r\right)^2\cos(z)^2\right)^{\frac{3}{2}}} - \frac{\cos(z)^2}{\sqrt{2-\left(2\sqrt{2}-r\right)^2\cos(z)^2}} - \frac{\left(-2\cos(z)+2\left(2\sqrt{2}-r\right)\cos(z)^2\right)^2}{4\left(1+2\left(2\sqrt{2}-r\right)\cos(z)-\left(2\sqrt{2}-r\right)^2\cos(z)^2\right)^{\frac{3}{2}}} - \frac{\cos(z)^2}{\sqrt{1+2\left(2\sqrt{2}-r\right)\cos(z)-\left(2\sqrt{2}-r\right)^2\cos(z)^2}} - \frac{r^2\sin(z)^4}{\left(2-r^2\sin(z)^2\right)^{\frac{3}{2}}} - \frac{\sin(z)^2}{\sqrt{2-r^2\sin(z)^2}} - \frac{\left(2\sin(z)-2r\sin(z)^2\right)^2}{4\left(1+2r\sin(z)-r^2\sin(z)^2\right)^{\frac{3}{2}}} - \frac{\sin(z)^2}{\sqrt{1+2r\sin(z)-r^2\sin(z)^2}},$$

f is a concave function, i.e., for every $r \in [2\sqrt{2}-\sqrt{5},\sqrt{2}]$ and every $z \in [\frac{\pi}{3},\frac{\pi}{2}]$ we have

$$\forall t \in [0,1]\ f(r,z) \geq (1-t)f(2\sqrt{2}-\sqrt{5},z) + tf(\sqrt{2},z),$$

and consequently

$$f(r,z) \geq min\{f(2\sqrt{2}-\sqrt{5},z), f(\sqrt{2},z)\}.$$

Since it can be shown that the two one-variable functions

$$f(2\sqrt{2}-\sqrt{5},z) = 2\left(-2\sqrt{2}+\sqrt{5}\right)\cos(z) + \sqrt{2-5\cos(z)^2} + \sqrt{1+2\sqrt{5}\cos(z)-5\cos(z)^2} - 2\sqrt{5}\sin(z) + \sqrt{2-\left(2\sqrt{2}-\sqrt{5}\right)^2\sin(z)^2} + \sqrt{1+2\left(2\sqrt{2}-\sqrt{5}\right)\sin(z)-\left(2\sqrt{2}-\sqrt{5}\right)^2\sin(z)^2}$$

and

$$f(\sqrt{2},z) = -2\sqrt{2}\cos(z) + \sqrt{2}\sqrt{\cos(z)^2} + \sqrt{2\sqrt{2}\cos(z)-\cos(2z)} - 2\sqrt{2}\sin(z) + \sqrt{2}\sqrt{\sin(z)^2} + \sqrt{\cos(2z)+2\sqrt{2}\sin(z)}$$

are positive for $z \in [\frac{\pi}{3},\frac{\pi}{2}]$, we obtain that $f(r,z) \geq 0$ for $r \in [2\sqrt{2}-\sqrt{5},\sqrt{2}]$ and $z \in [\frac{\pi}{3},\frac{\pi}{2}]$.

Finally we will prove inequality (17) when the angle z of v (with x-axis) is in the interval $[\frac{\pi}{4},\frac{\pi}{3}]$. Let $c = \sqrt{a^2+b^2}$, i.e., c is the distance from the center Z_b of bb to point O. As we have shown, $\sqrt{2} \leq c \leq \sqrt{5}$.

By Lemma 4 applied for $X = O$ to wb, the left side of inequality (17) is maximal for angle $z = \frac{\pi}{4}$. In this case, we have $p = q \Rightarrow \sqrt{p^2+q^2} = \sqrt{2}q = 2\sqrt{2}-c \Rightarrow q = 2-\frac{c}{\sqrt{2}}$. Therefore, the left side of inequality (17) is less than or equal to

$$4-\sqrt{2}c-\sqrt{1+\sqrt{2}c-\frac{c^2}{2}}-\sqrt{-2+2\sqrt{2}c-\frac{c^2}{2}} \tag{19}$$

In the following, we will bound from below the right side of inequality (17). By substituting $b = \sqrt{c^2 - a^2}$ in the right side of (17), we obtain

$$r(a, c) = -2\sqrt{c^2 - a^2} + \sqrt{2 - a^2} + \sqrt{-a^2 + 2a + 1}. \tag{20}$$

Since $a \leq \sqrt{2}$ and $a = c\cos z$ implies $a \geq \sqrt{2}\cos(\frac{\pi}{3}) = \frac{\sqrt{2}}{2}$, we obtain $a \in [\frac{\sqrt{2}}{2}, \sqrt{2}]$. We will show that the function $r(a, c)$ obtains its minimum at $a = \frac{\sqrt{2}}{2}$ for every $c \in [\sqrt{2}, \sqrt{5}]$.

First, we further restrict the domain of a. Let $c \in [\sqrt{2}, \sqrt{5}]$ be fixed. Let bb_z^c denote a bb such that the distance from the center Z_b to O is c and the angle of vector v_z (which contains Z_bO) with the x-axis is z. By Lemma 3 applied for $X = O$ to bb_z^c, if $E \notin bb_{\frac{\pi}{3}}^c$, then $E \notin bb_z^c$ for every $z \in [\frac{\pi}{4}, \frac{\pi}{3}]$.

Since we consider case $s_2(1)s_1(1)$, we only need to consider those c for which there exists at least one z such that $E \in bb_z^c$ (see Figure 35(a)). Therefore, we must have $E \in bb_{\frac{\pi}{3}}$, which means that

$$a^2 + (b-1)^2 = a^2 + (\sqrt{3}a - 1)^2 = 4a^2 - 2\sqrt{3}a + 1 \leq 2,$$

since in this case $\frac{b}{a} = \tan(\frac{\pi}{3}) = \sqrt{3}$. By solving this inequality, we obtain $\frac{2\sqrt{3}-2\sqrt{7}}{8} \leq a \leq \frac{2\sqrt{3}+2\sqrt{7}}{8}$, which implies that $a \in [\frac{\sqrt{2}}{2}, \frac{2\sqrt{3}+2\sqrt{7}}{8}]$ (since $\frac{2\sqrt{3}+2\sqrt{7}}{8} \approx 1.094 \leq \sqrt{2}$).

To find the minimum of (20), we split this expression into two parts. We observe first that the minimum of $\sqrt{-a^2 + 2a + 1}$ for $a \in [\frac{\sqrt{2}}{2}, \frac{2\sqrt{3}+2\sqrt{7}}{8}]$ is obtained at $a = \frac{\sqrt{2}}{2}$ and is equal to $\sqrt{\frac{1}{2} + \sqrt{2}}$.

To find the minimum of the remaining part of (20), we will treat $g_c(a) = -2\sqrt{c^2 - a^2} + \sqrt{2 - a^2}$ as a function of one variable a with parameter c. We calculate

$$g_c'(a) \;=\; -\frac{a}{\sqrt{2 - a^2}} + \frac{2\,a}{\sqrt{-a^2 + c^2}}$$

and

$$g_c''(a) \;=\; -\frac{a^2}{(2 - a^2)^{\frac{3}{2}}} - \frac{1}{\sqrt{2 - a^2}} + \frac{2\,a^2}{(-a^2 + c^2)^{\frac{3}{2}}} + \frac{2}{\sqrt{-a^2 + c^2}}.$$

We obtain $g_c'(a) = 0 \Leftrightarrow a = 0$ or $a = \pm\sqrt{\frac{8-c^2}{3}}$. Since $g_c''(0) = \frac{2\sqrt{2}-c}{\sqrt{2}c} > 0$ for every $c \in [\sqrt{2}, \sqrt{5}]$, g_c has a minimum at $a = 0$ and maxima at $a = \pm\sqrt{\frac{8-c^2}{3}}$.

Since the image of the function $a(c) = \sqrt{\frac{8-c^2}{3}}$ for $c \in [\sqrt{2}, \sqrt{5}]$ is in the interval $[1, \sqrt{2}]$, the function $g_c(a)$ obtains a maximum for $a \in [1, \sqrt{2}]$. If $g_c(a)$ obtains a maximum for $a \in [\frac{2\sqrt{3}+2\sqrt{7}}{8}, \sqrt{2}]$, then g defined over $[\frac{\sqrt{2}}{2}, \frac{2\sqrt{3}+2\sqrt{7}}{8}]$ obtains the minimum at $a = \frac{\sqrt{2}}{2}$. If $g_c(a)$ obtains a maximum for $a \in [1, \frac{2\sqrt{3}+2\sqrt{7}}{8})$, then g obtains the minimum at one of the endpoints of interval $[\frac{\sqrt{2}}{2}, \frac{2\sqrt{3}+2\sqrt{7}}{8}]$. Since $g(\frac{2\sqrt{3}+2\sqrt{7}}{8}) \geq g(\frac{\sqrt{2}}{2})$ for every $c \in [\sqrt{2}, \sqrt{5}]$, function g defined over $[\frac{\sqrt{2}}{2}, \frac{2\sqrt{3}+2\sqrt{7}}{8}]$ obtains the minimum at $a = \frac{\sqrt{2}}{2}$. Since $g(\frac{\sqrt{2}}{2}) = \sqrt{\frac{3}{2}} - 2\sqrt{c^2 - \frac{1}{2}}$, we obtain that $r(a, c)$ is greater than or equal to

$$\sqrt{\frac{3}{2}} - 2\sqrt{c^2 - \frac{1}{2}} + \sqrt{\frac{1}{2} + \sqrt{2}} \tag{21}$$

for every $c \in [\sqrt{2}, \sqrt{5}]$.

In order to prove inequality (17) it is sufficient to show that (19) $\leq$ (21), which is true, since it can be shown that the one-variable function $i(c) = (21) - (19)$ is greater than or equal to 0 for every $c \in [\sqrt{2}, \sqrt{5}]$. ■

B Acknowledgments

The quality of this paper has been substantially improved due to the reviewers' comments.

References

[1] L. V. Ahlfors and L. Sario, *Riemann Surfaces*. Princeton University Press, Princeton, New Jersey, 1960.

[2] L. Boxer, Digitally Continuous Functions, *Pattern Recognition Letters*, 15:833-839, 1994.

[3] M. P. do Carmo, *Differential Geometry of Curves and Surfaces*, Prentice-Hall, Englewood Cliffs, New Jersey, 1976.

[4] L. Chen, Topological structure in Visual Perception. *Science*, 218:669–700, 1982.

[5] L. Chen, Topological structure in the perception of apparent motion. *Perception*, 14:197–208, 1985

[6] L. Chen, Holes and wholes: A reply to Rubin and Kanwisher. *Perception and Psychophysics*, 47:47–53, 1990.

[7] U. Eckhardt and L. Latecki, Digital Topology. In *Current Topics in Pattern Recognition Research, Research Trends*, Council of Scientific Information, Vilayil Gardens, Trivandrum, India, 1994.

[8] A. Gross, Analyzing Generalized Tubular Surfaces, *Proceedings of SPIE's Conference on Intelligent Robots and Computer Vision*, Boston, 1994.

[9] A. Gross and L. Latecki, Digitizations Preserving Topological and Differential Geometric Properties. *Computer Vision and Image Understanding* 62, 370–381, Nov. 1995.

[10] A. Gross and L. Latecki. A Realistic Digitization Model of Straight Lines. *Computer Vision and Image Understanding*, to appear 1997.

[11] Ch. Habel, Prepositional and depictorial representations of spatial knowledge: The case of path-concepts. In R. Studer (ed.): *Logics and Natural Language.*, Springer-Verlag, Berlin, 1990.

[12] G. Herman, Discrete Multidimensional Jordan Surfaces. *CVGIP: Graphical Models and Image Processing*, 54:507–515, 1992.

[13] E. Khalimsky, R. Kopperman, and P. R. Meyer, Computer Graphics and Connected Topologies on Finite Ordered Sets. *Topology and its Applications*, 36:1-17, 1990.

[14] T. Y. Kong and A. W. Roscoe. Continuous Analogs of Axiomatized Digital Surfaces. *Computer Vision, Graphics, and Image Processing*, 29:60–86, 1985.

[15] T. Y. Kong and A. W. Roscoe. A theory of binary digital pictures. *Computer Vision, Graphics, and Image Processing*, 32:221–243, 1985.

[16] T. Y. Kong and A. Rosenfeld, Digital topology: Introduction and survey. *Computer Vision, Graphics, and Image Processing*, 48:357–393, 1989.

[17] T. Y. Kong and J. K. Udupa, A Justification of a Fast Surface Tracking Algorithm. *Computer Vision, Graphics, and Image Processing: Graphical Models and Image Processing*, 54:162–170, 1992.

[18] L. Latecki, U. Eckhardt, and A. Rosenfeld, Well-Composed Sets. *Computer Vision and Image Understanding*, 61:70–83, 1995.

[19] L. Latecki and F. Prokop, Semi-proximity Continuous Functions in Digital Images. *Pattern Recognition Letters*, 16:1175–1187, 1995.

[20] M. Minsky and S. Papert, *Perceptrons. An introduction to Computational Geometry*, The MIT Press: Cambridge, Mass., 1969.

[21] E. E. Moise, *Geometric Topology in Dimensions 2 and 3*, Springer-Verlag, 1977.

[22] D. G. Morgenthaler and A. Rosenfeld. Three-dimensional digital topology. *Information and Control*, 51:227–247, 1981.

[23] G. L. Naber, *Topological methods in Euclidean spaces*, Cambridge University Press, Cambridge, 1980.

[24] T. Pavlidis, *Algorithms for Graphics and Image Processing*, Springer–Verlag, Berlin, 1982.

[25] A. Rosenfeld, Digital topology. *American Mathematical Monthly*, 86:621–630, 1979.

[26] A. Rosenfeld, 'Continuous' functions on digital pictures, *Pattern Recognition Letters*, 4:177-184, 1986.

[27] A. Rosenfeld and A.C. Kak, *Digital Picture Processing*, Academic Press, New York, 1982.

[28] J. Serra, *Image Analysis and Mathematical Morphology*, Academic Press, New York, 1982.

[29] L. N. Stout, Two Discrete Forms of the Jordan Curve Theorem. *American Mathematical Monthly*, 95:332–336, 1988.