A New Convexity Measure for Polygons

Joviša Žunić*, Paul L. Rosin

Computer Science Department, Cardiff University
Queen's Buildings, Newport Road, P.O. Box 916
Cardiff CF24 3XF, Wales, U.K.

{J.Zunic, Paul.Rosin}@cs.cf.ac.uk

Abstract

Convexity estimators are commonly used in the analysis of shape. In this paper we define and evaluate a new convexity measure for planar regions bounded by polygons. The new convexity measure can be understood as a "boundary based" measure and in accordance with this it is more sensitive to measured boundary defects than the so called "area based" convexity measures. When compared with the convexity measure defined as the ratio between the Euclidean perimeter of convex hull of the measured shape and the Euclidean perimeter of the measured shape then the new convexity measure also shows some advantages – particularly for shapes with holes.

The new convexity measure has the following desirable properties:

- the estimated convexity is always a number from $(0, 1]$;
- the estimated convexity is 1 if and only if the measured shape is convex;
- there are shapes whose estimated convexity is arbitrary close to 0;
- the new convexity measure is invariant under similarity transformations;
- there is a simple and fast procedure for computing the new convexity measure.

Keywords: Shape, polygons, convexity, measurement.

1 Introduction

Shape is a crucial component in many areas of scientific analysis [4, 6], with examples including geomorphology [15], powder particle characterisation [10], and biology [2]. This paper is

*Joviša Žunić is also with Mathematical Institute of Serbian Academy of Sciences and Arts, Belgrade.

concerned with the measurement of the *convexity* of polygons, which can be considered as one of the basic descriptors of shape [20] (alongside others such as compactness, circularity, and rectangularity) and has received some attention over the years [3, 21]. A convexity measure can be used for a variety of applications, for instance shape decomposition [11, 18] which in turn can be used to compute shape similarity and has been applied to object indexing [12].

In general, a planar shape S is said to be convex if it has the following property: If points A and B belong to S then all points from the line segment $[AB]$ belong to S as well. The smallest convex set which includes a shape S is called the *convex hull of* S and it is denoted as $\mathbf{CH}(S)$ (see Fig. 1). The previous two definitions suggest the following two possibilities

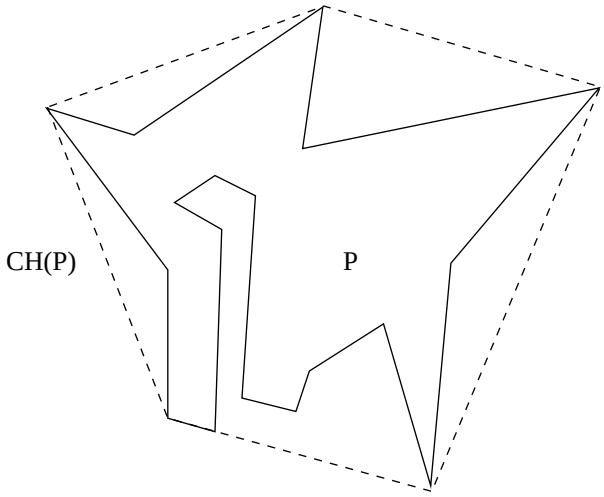


Figure 1: Non convex polygon P and its convex hull $\mathbf{CH}(P)$ (dashed line).

for convexity measurements of planar shapes.

Definition 1 *For a given planar shape S its convexity measure $\mathcal{C}_1(S)$ is defined to be the probability that for randomly chosen points A and B from S all points from the line segment $[AB]$ also belong to S, under the assumption that A and B are chosen uniformly.*

Definition 2 *For a given planar shape S, its convexity measure $\mathcal{C}_2(S)$ is defined to be*

$$\mathcal{C}_2(S) \;=\; \frac{Area(S)}{Area(\mathbf{CH}(S))}.$$

Both convexity measures defined above have the following desirable properties:

- **i)** the convexity measure is a number from $(0, 1]$;
- **ii)** the convexity measure of a given shape equals 1 if and only if this shape is convex;
- **iii)** there are shapes whose convexity measure is arbitrary close to 0;
  (i.e., there is no a gap between 0 and the minimal possible convexity measured);
- **iv)** the convexity measure of a shape is invariant under similarity transformations of this shape.

But there is also some "bad" properties of the above definitions. The main objection to the Definition 1 is that $\mathcal{C}_1(S)$ is difficult to compute, even if S is a polygon. In practice, the convexity measure given by Definition 2 is the one that is mostly used, and appears in textbooks [20]. $\mathcal{C}_2(S)$ is easy and efficient to compute and is very robust with respect to noise. Its "discrete version" where the real objects are presented as finite sets of points and the area of object is simply estimated by using the number of points which fall into it is also widely used in practical applications.

On the other hand, the convexity measurement given by Definition 2 does not detect huge defects on boundaries of shapes which have a relatively small impact on the shape areas. Two simple examples are given in Fig. 2. Setting $t = 1 - h$, then for small enough values of h the polygons $P(h)$ and $T(1-h, h)$ have the same perimeter and almost the same area. Nevertheless, this leads to the following inconsistent relation between convexity estimates

$$\lim_{h\to 0} \mathcal{C}_2(P(h)) = 1 \; > \; \lim_{h\to 0} \mathcal{C}_2(T(1-h,h)) = \frac{2}{3}.$$

The anomaly $\lim_{h\to 0} \mathcal{C}_2(P(h)) = 1$ can be avoided by a modification of Definition 2 as follows.

Definition 3 *Let a planar set S be given. Let $\mathbf{MCS}(S)$ denote a convex subset of S with the maximum possible area. Then, the convexity measure $\mathcal{C}'_2$ of S is defined as*

$$\mathcal{C}'_2(S) = \frac{Area(\mathbf{MCS}(S))}{Area(S)}.$$

The previous measure also satisfies **i)**, **ii)**, **iii)**, and **iv)** and gives the reasonable estimate $\lim_{h\to 0} \mathcal{C}'_2(P(h)) = \frac{1}{2}$. On the other hand $\lim_{h\to 0} \mathcal{C}'_2(T(1-h,h)) = 1$ is a bad estimate, and what is more important the computation of $Area(\mathbf{MCS}(S))$ is very difficult.

The previously mentioned possibilities for convexity measurement can be understood as "area based" measures, and consequently, they are expected to be robust with respect to boundary defects (caused by noise for example). On the other hand, if a convexity measure is based on the shape boundary, then it is likely to be more sensitive to the boundary properties than the measures $\mathcal{C}_1$ and $\mathcal{C}_2$. Such a sensitivity can be a very useful property of the measurement – see again Fig. 2 for an example. The following boundary based convexity measure computed as the ratio of the Euclidean perimeter of a given shape S and the Euclidean perimeter of the convex hull of S seems to be a natural solution.

Definition 4 *For a given planar shape S, its convexity measure $\mathcal{C}_3(S)$ is defined to be*

$$\mathcal{C}_3(S) = \frac{\mathcal{P}er_2(\mathbf{CH}(S))}{\mathcal{P}er_2(S)}$$

where $\mathcal{P}er_2(\mathbf{CH}(S))$ and $\mathcal{P}er_2(S)$ are the Euclidean perimeters of the boundaries of $\mathbf{CH}(S)$ and S, respectively.

By the last definition we have the more acceptable situation $\lim_{h\to 0} \mathcal{C}_3(P(h)) = \frac{2}{3}$ and $\lim_{h\to 0} \mathcal{C}_3(T(1-h,h)) = \frac{3+\sqrt{5}}{6} \approx 0.87$.

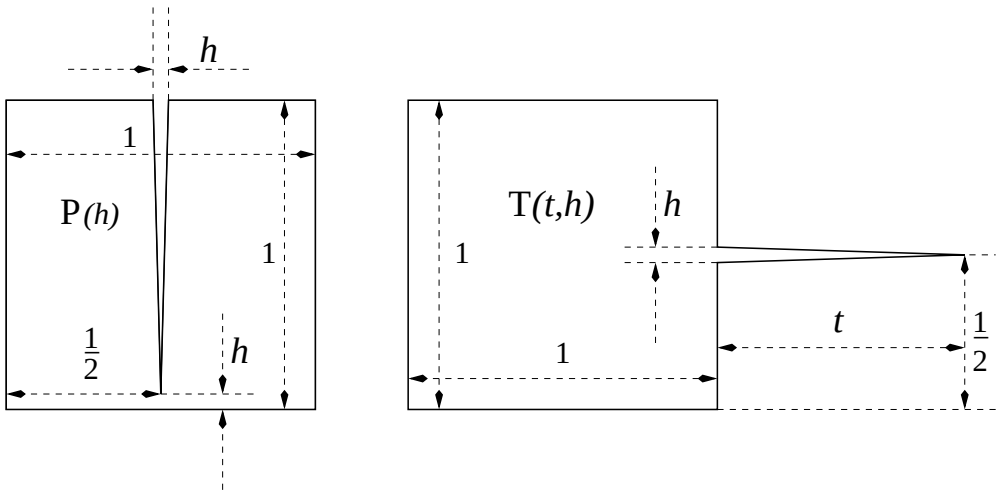


Figure 2: $\mathcal{C}_1(P(h)) \to \frac{1}{2}$ as $h \to 0$ is reasonable, but $\mathcal{C}_2(P(h)) \to 1$ as $h \to 0$ is not. $\mathcal{C}_1(T(t,h)) \to 1$ as $h \cdot t \to 0$ is poor for a big t, while it could be acceptable that $\mathcal{C}_2(T(t,h)) \to \frac{2}{2+t}$ with any fixed t and $h \to 0$.

While convexity measures $\mathcal{C}_1$, $\mathcal{C}_2$, $\mathcal{C}_2'$, and $\mathcal{C}_3$ follow directly from the definition of convex shapes in literature there already exist measures which are not such direct consequences of the definition.

For instance, Stern [21] proposed an area based convexity measure. In an attempt to incorporate the entire topology of the polygon into the measure he considered the set of all mutually visible pairs of points contained in the polygon. To define the measure for a given polygon P a probability density function is defined as

$$f(x,y) = \frac{Area(V((x,y),P))}{\iint\limits_{(x,y)\in P} Area(V((x,y),P))\; dy\; dx}$$

where $V((x,y),P)$ is the set of points $(\tilde{x},\tilde{y})$ such that the line segment $[(x,y),(\tilde{x},\tilde{y})]$ is completely included in P. The polygonal entropy of P is defined as

$$E(P) = -\iint\limits_{(x,y)\in P} f(x,y) \ln f(x,y)\; dx\; dy.$$

Finally, the convexity measure is defined as the ratio

$$\frac{E(P)}{\ln Area(P)}$$

or as an arbitrary normalization of it.

Although the above convexity measure is theoretically appealing, in practice its main disadvantage is the computational burden required in its determination. An approximate (but faster) estimation of the proposed convexity measure was also described.

In [3] Boxer described two ways for measuring the deviation of a given n-gon from convexity, and gave $\mathcal{O}(n)$ algorithms for their computation. Both measures are boundary based.

Roughly speaking, they estimate polygonal convexity by using the distances of the vertices of the measured polygon from the convex hull of this polygon. The first measure is called a *simple index of non-convexity*, and is formally defined in the following way:
Let P be an n-gon whose vertices in circular order are $A_0, A_1, \ldots A_n = A_0$, and let $A_{i_1}, A_{i_2}, \ldots, A_{i_k}$ be vertices of $\mathbf{CH}(P)$, where $0 \leq i_0 < i_i < \ldots < i_k \leq n$. The simple index of non-convexity of P is then defined to be

$$v(P) = \max_{0 \leq j < k} \left\{ \max_{i_j \leq i < i_{j+1}} \{ d((x_1, x_2), [A_{i_j} A_{i_{j+1}}]) \mid (x_1, x_2) \in [A_i A_{i+1}] \} \right\}$$

where $d((x_1, x_2), [A_{i_j} A_{i_{j+1}}]) = min\{\sqrt{(x_1 - x)^2 + (y_1 - y)^2} \mid (x, y) \in [A_{i_j} A_{i_{j+1}}]\}$, as usual. It is obvious that $v(P) = 0$ if and only if P is convex. The second measure is called the *total index of non-convexity* and it is a slight modification of the previous one, replacing the outer maximum operation by a summation.

In this paper we define a new easily computable convexity measure for polygons which is also computed from the boundaries of measured shapes. It satisfies the requirements **i)**, **ii)**, **iii)**, and **iv)**, but has some advantages with respect to the previous convexity measure – particularly in measuring shapes with holes. In Sections 5 and 6 the new measure $\mathcal{C}(S)$ will be compared against the above measures $\mathcal{C}_2(S)$ (in order to illustrate the difference between a boundary based measure and an area based measure) and $\mathcal{C}_3(S)$ (in order to illustrate possible differences between two boundary based measures).

2 Definitions and Denotations

Throughout the paper it will be assumed that all considered shapes are planar bounded compact sets with a non empty interior. A polygon means a compact planar area bounded by a simple polygonal line (which also belongs to the polygon because compactness is assumed). In the subsection 5.4 polygons with holes are considered. A polygon with holes is obtained as a set difference of a given polygon and polygons (one or more) which are subsets of it.

We will use the following denotations. For a given n-gon P having vertices denoted by $A_0, A_1, \ldots, A_n = A_0$, its edges will be denoted $e_i = [A_{i-1}, A_i]$ for $i = 1, 2, \ldots, n$.

The Euclidean length of the straight line segment $e = [(x_1, y_1), (x_2, y_2)]$ is $l_2(e) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$, while the length of e according to the l_1 metric is $l_1(e) = |x_1 - x_2| + |y_1 - y_2|$, i.e., $l_1(e)$ equals the sum of the projections of e onto the coordinate axes.

$\mathcal{P}er_2(P)$ will denote the Euclidean perimeter of P, while $\mathcal{P}er_1(P)$ will denote the perimeter of P in the sense of the l_1 metric. So,

$$\mathcal{P}er_2(P) = \sum_{e_i \text{ is an edge of } P} l_2(e_i) \qquad \text{and} \qquad \mathcal{P}er_1(P) = \sum_{e_i \text{ is an edge of } P} l_1(e_i).$$

Since isometric polygons do not necessarily have the same perimeter under the l_1 metric, we shall use $\mathcal{P}er_1(P, \alpha)$ for the l_1 perimeter of the polygon which is obtained by rotating P by the angle α with the origin as the center of rotation. If the same rotation is applied to the edge e, the l_1 perimeter of the obtained edge will be denoted as $l_1(e, \alpha)$.

If the oriented angle between the positively oriented x-axis and the vector $\overrightarrow{A_{i-1}A_i}$ is denoted by ϕ_i $(i = 1, 2, \ldots n)$, then obviously $l_1(e_i) = l_2(e_i) \cdot (|\cos\phi_i| + |\sin\phi_i|)$, while $l_1(e_i, \alpha) = l_2(e_i) \cdot (|\cos(\phi_i + \alpha)| + |\sin(\phi_i + \alpha)|)$.

The line determined by points A and B will be denoted as $l(A, B)$.

The minimal rectangle with edges parallel to the coordinate axes which includes a polygon P will be denoted by $\mathbf{R}(P)$.

If a polygon P is rotated by an angle α around the origin then $\mathbf{R}(P, \alpha)$ denotes the minimal rectangle with edges parallel to coordinate axes which includes the rotated polygon. Even though $\mathcal{P}er_1(\mathbf{R}(P, \alpha)) = \mathcal{P}er_2(\mathbf{R}(P, \alpha))$ we will use $\mathcal{P}er_2(\mathbf{R}(P, \alpha))$ rather than $\mathcal{P}er_1(\mathbf{R}(P, \alpha))$.

3 A New Convexity Measure

In this section we define a new convexity measure for polygons. First, we give a simple lemma and after that we proceed with a theorem which gives a useful characterisation of convex polygons (for another implications of it see [1]).

Lemma 1 *The inequality*

$$\mathcal{P}er_2(\mathbf{R}(P)) \;\leq\; \mathcal{P}er_1(P)$$

holds for any polygon P.

Proof. Let A, B, C, and D be four vertices belonging to pairwise different edges of $\mathbf{R}(P)$ (some of them may coincide). Let Q be the 4-gon with vertices A, B, C, and D. Then $\mathcal{P}er_2(\mathbf{R}(P)) = \mathcal{P}er_1(\mathbf{R}(P)) = \mathcal{P}er_1(Q) \leq \mathcal{P}er_1(P)$ finishes the proof. ▯

Theorem 1 *A given polygon P is convex if and only for any choice of the coordinate system the l_1 perimeter of P equals the perimeter of the minimal rectangle whose edges are parallel to the coordinate axes and which includes P, i.e.,*

$$\mathcal{P}er_2(\mathbf{R}(P, \alpha)) = \mathcal{P}er_1(P, \alpha) \qquad \textit{for any} \quad \alpha \in [0, 2\pi].$$

Proof. If a given polygon P is convex then the projections of the edges of P onto the x and y axes exactly cover the boundary of the minimal rectangle whose edges are parallel to the coordinate axes (see Fig. 3). Since the sum of such projections equals both the l_1 perimeter of the polygon P and the Euclidean perimeter of $\mathbf{R}(P)$ the convexity of P implies $\mathcal{P}er_1(P) = \mathcal{P}er_2(\mathbf{R}(P))$ independently of the choice of the coordinate system, or equivalently, $\mathcal{P}er_1(P, \alpha) = \mathcal{P}er_2(\mathbf{R}(P, \alpha))$ for any $\alpha \in [0, 2\pi]$.

On the other hand, if P is not convex, then there exist points A and B from the interior of P such that the line segment $[AB]$ is not completely contained in P. If the line determined by A and B is chosen to be one of coordinate axes, say u, then the projections of edges of P onto coordinate axis v (perpendicular to u) must overlap (see Fig. 4). In other words, the strict inequality $\mathcal{P}er_1(P) > \mathcal{P}er_2(\mathbf{R}(P))$ holds. This completes the proof. ▯

Theorem 1 gives a useful characterisation for convex polygons and gives the basic idea for the polygon convexity measurement described in this paper. In the first stage, the inequality

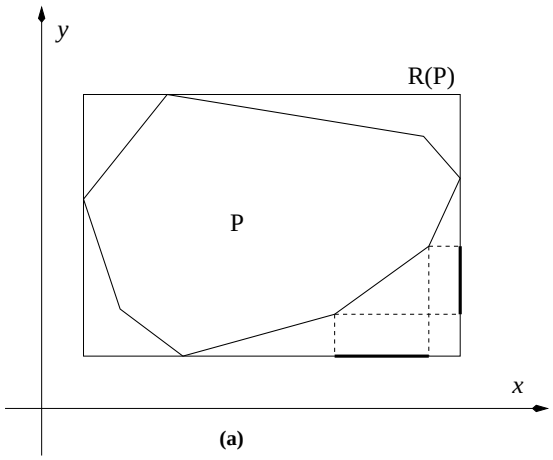


Figure 3: If a given polygon P is convex then $\mathcal{P}er_1(P) = \mathcal{P}er_2(\mathbf{R}(P))$.

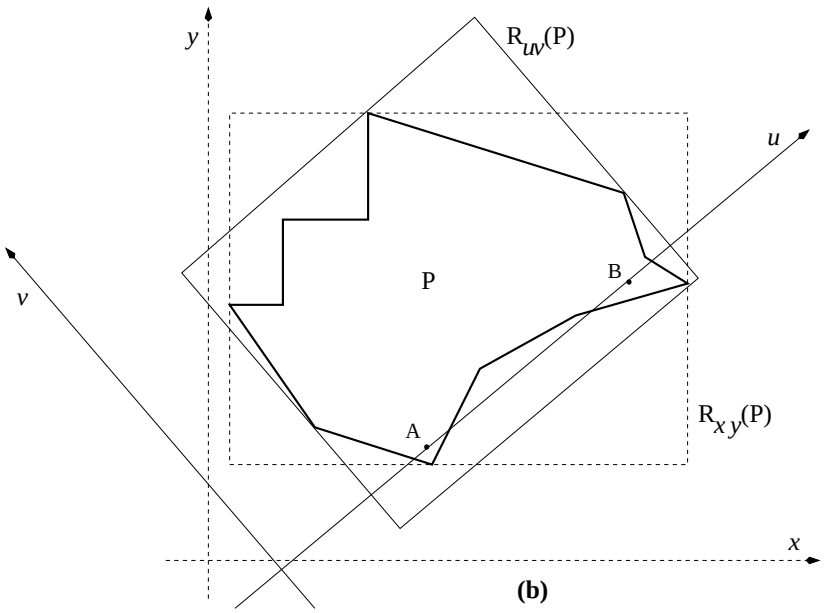


Figure 4: If x and y are chosen to be the coordinate axes then $\dfrac{\mathcal{P}er_2(\mathbf{R}_{xy}(P))}{\mathcal{P}er_1(P)} = 1$. Since P is not convex there is another choice of the coordinate axes, say u and v, such that the strict inequality $\dfrac{\mathcal{P}er_2(\mathbf{R}_{uv}(P))}{\mathcal{P}er_1(P)} < 1$ holds.

$\mathcal{P}er_1(P) \geq \mathcal{P}er_2(\mathbf{R}(P))$ (see Lemma 1) suggests that the fraction $\dfrac{\mathcal{P}er_2(\mathbf{R}(P))}{\mathcal{P}er_1(P)}$ can be used as a convexity measure for polygons since it is a number from $(0, 1]$, it is defined for any polygon P, it can be calculated easily, and for any convex polygon it equals 1. But, on the other hand, this ratio can depend strongly of the choice of the coordinate system (see Fig. 9 (c) and Fig. 10 (c)) – which is not a desirable property. Also, this ratio can be equal to 1 for non-convex polygons (see Fig. 4 for an example) which is not acceptable for a convexity

measure. These problems can be avoided by considering $\min\limits_{\alpha\in[0,2\pi]} \dfrac{\mathcal{P}er_2(\mathbf{R}(P,\alpha))}{\mathcal{P}er_1(P,\alpha)}$ instead of $\dfrac{\mathcal{P}er_2(\mathbf{R}(P))}{\mathcal{P}er_1(P)}$.

Now we can define a new convexity measure for polygons which satisfies **i)**, **ii)**, **iii)**, and **iv)**.

Definition 5 *For a given polygon P its convexity $\mathcal{C}(P)$ is defined as:*

$$\mathcal{C}(P) \;=\; \min_{\alpha\in[0,2\pi]} \frac{\mathcal{P}er_2(\mathbf{R}(P,\alpha))}{\mathcal{P}er_1(P,\alpha)}.$$

The following theorem summarises the desirable properties of the polygon convexity measure proposed here.

Theorem 2 *For any polygon P we have:*

i) $\mathcal{C}(P)$ *is well defined and* $\mathcal{C}(P) \in (0,1]$*;*

ii) $\mathcal{C}(P) = 1$ *if and only if* P *is convex;*

iii) $\inf\limits_{P\in\Pi}(\mathcal{C}(P)) = 0$*, where* Π *denotes the set of all polygons;*

iv) $\mathcal{C}(P)$ *is invariant under similarity transformations.*

Proof. Since $\dfrac{\mathcal{P}er_2(\mathbf{R}(P,\alpha))}{\mathcal{P}er_1(P,\alpha)}$ is a continuous function on α (for more details see Section 4) it must reach its minimum on a closed interval $[0,2\pi]$. So, $\mathcal{C}(P)$ is well defined. Since $\mathcal{C}(P) > 0$ is trivial and because $\mathcal{C}(P) \le 1$ follows from Lemma 1, item *i)* is proved.
Item *ii)* is a direct consequence of Theorem 1.
To prove item *iii)* consider the polygon P_n from Fig. 5. Trivially,

$$\mathcal{P}er_2(\mathbf{R}(P_n,\alpha)) \le 4\cdot diameter\_of(P_n) = 4\cdot\sqrt{2}\cdot n, \qquad \text{for any} \quad \alpha\in[0,2\pi].$$

On the other hand,

$$\mathcal{P}er_1(P_n,\alpha) \ge \mathcal{P}er_1(P_n,0) = 2\cdot n + 2\cdot n\cdot(n-1) + 2 = 2\cdot n^2 + 2,$$

and finally,

$$0 \le \lim_{n\to\infty} \mathcal{C}(P_n) \le \lim_{n\to\infty} \frac{4\cdot\sqrt{2}\cdot n}{2\cdot n^2+2} = 0$$

proves $\inf\limits_{P\in\Pi}(\mathcal{C}(P)) = 0$.
In order to prove *iv)* note that $\mathcal{C}(P)$ is invariant under all isometric transformations – which follows from the definition. Also, $\dfrac{\mathcal{P}er_1(\mathbf{R}(P,\alpha))}{\mathcal{P}er_1(P,\alpha)}$ and consequently $\mathcal{C}(P)$ are invariants under any transformation of the form $(x,y)\to(\lambda\cdot x,\lambda\cdot y)$ for any choice of $\lambda\neq 0$, P, and α. That completes the proof. ▥

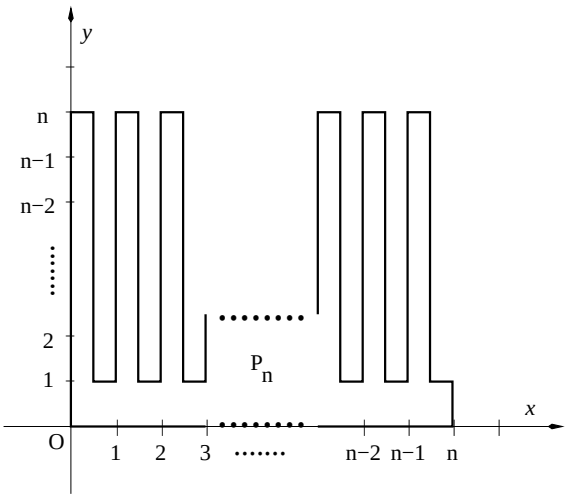


Figure 5: For a big enough n the convexity measure $\mathcal{C}(P_n)$ of the polygon P_n is arbitrary close to 0.

4 Computation of $\mathcal{C}(P)$

It is obvious that for a given polygon P its new convexity measure

$$\mathcal{C}(P) \;=\; \min_{\alpha\in[0,2\pi]} \frac{\mathcal{P}er_2(\mathbf{R}(P,\alpha))}{\mathcal{P}er_1(P,\alpha)}$$

can be computed numerically to an arbitrary precision (of course, a higher precision requires a higher time complexity). On the other hand, although computation of $\mathcal{C}(P)$ is a nonlinear optimization problem, it turns out that the exact value of $\mathcal{C}(P)$ can be computed in $\mathcal{O}(n^2)$ time, if P is an n-gon.

In the rest of this section we describe how to compute (effectively and efficiently) the exact value of $\mathcal{C}(P)$. We need some additional investigation of the functions $\mathcal{P}er_1(P,\alpha)$ and $\mathcal{P}er_2(\mathbf{R}(P,\alpha))$ which depend only on α if P is fixed. They will be analysed separately in the next two subsections. A simple procedure for $\mathcal{C}(P)$ computation comes as a direct consequence of such an analysis.

4.1 Analysis of $\mathcal{P}er_1(P,\alpha)$

Let an edge e_i $(1 \le i \le n)$ of a given n-gon P be given. Trivially,

$$l_1(e_i,\alpha) = l_2(e_i)\cdot(|\cos(\phi_i+\alpha)|+|\sin(\phi_i+\alpha)|) = a(\alpha)\cdot l_2(e_i)\cdot\cos(\phi_i+\alpha)+b(\alpha)\cdot l_2(e_i)\cdot\sin(\phi_i+\alpha)$$

where $a(\alpha)$ and $b(\alpha)$ take $+1$ or -1 depending on α. Consequently, there is an integer $k \le 4\cdot n$ and a sequence $0 \le \alpha_1 < \alpha_2 < \ldots < \alpha_k \le 2\pi$ such that

$$\mathcal{P}er_1(P,\alpha) = \sum_{i=1}^{n} a_{j,i}\cdot l_2(e_i)\cdot\cos(\phi_i+\alpha) + b_{j,i}\cdot l_2(e_i)\cdot\sin(\phi_i+\alpha) \qquad \text{if } \alpha\in[\alpha_j,\alpha_{j+1}] \quad (1)$$

where $\alpha_{k+1} = \alpha_1 + 2\pi$ and $\{a_{j,i}, b_{j,i} \mid 1 \leq j \leq k,\ 1 \leq i \leq n\} \subset \{+1, -1\}$. Since $l_2(e_i)$ and ϕ_i are constants we can conclude (from (1)) that there are some numbers c_j and d_j, $(j = 1, 2, \ldots, k)$ such that

$$\mathcal{P}er_1(P, \alpha) = c_j \cdot \cos\alpha + d_j \cdot \sin\alpha \qquad \text{for} \quad \alpha \in [\alpha_j, \alpha_{j+1}], \quad j = 1, 2, \ldots, k \tag{2}$$

$(\alpha_{k+1} = 2 \cdot \pi + \alpha_1)$. Obviously,

$$c_j = \sum_{i=1}^{n} l_2(i) \cdot (a_{j,i} \cdot \cos\phi_i + b_{j,i} \cdot \sin\phi_j) \quad \text{and} \quad d_j = \sum_{i=1}^{n} l_2(e_i) \cdot (-a_{j,i} \cdot \sin\phi_i + b_{j,i} \cdot \cos\phi_j),$$

for $j = 1, 2, \ldots, k$.

Note 1 *For any angle $\alpha_p \in \{\alpha_1, \alpha_2, \ldots, \alpha_k\} \subset [0,\ 2\pi]$ there is an edge e_q, $1 \leq q \leq n$ such that after the rotation for the angle α_p the edge e_q becomes parallel to one of coordinate axes. Some of the angles $\{\alpha_1, \alpha_2, \ldots, \alpha_k\}$ can be coincident.*

Two examples displayed on Fig. 9 (b) and Fig. 10 (b) illustrate that $\mathcal{P}er_2(P, \alpha)$ behaves as described by (2).

4.2 Analysis of $\mathcal{P}er_2(\mathbf{R}(P, \alpha))$

The construction of optimal rectangles which include a polygon P is already well studied in the literature – and various approaches exist [9, 13]. A related problem is the determination of the diameter of a given polygon P, and a very simple algorithm was presented in [16]. The diameter of P is defined to be the greatest distance between parallel lines of support of P. A line L is a line of support of P if the interior of P lies completely to one side of L. A pair of vertices is an antipodal pair if it admits parallel lines of support. Preparata and Shamos' algorithm generates all antipodal pairs by a procedure which resembles rotating a pair of dynamically adjustable parallel support lines once around the polygon P. This idea is generalised in [22] where two orthogonal pairs of line supports (called *orthogonal calipers*) are formed around the polygon solving several geometric problems. What is important for us is that the same procedure can be used here in order to obtain the intervals $[\beta_i, \beta_{i+1}]$, $i = 1, 2, \ldots, m$ for which four vertices of P (more precisely, four vertices of $\mathbf{CH}(P)$ since $\mathcal{P}er_2(\mathbf{R}(P, \alpha)) = \mathcal{P}er_2(\mathbf{R}(\mathbf{CH}(P), \alpha))$) forming two pairs of antipodal points belonging to the boundary of $\mathbf{R}(P, \alpha)$ remain the same for $\alpha \in [\beta_i, \beta_{i+1}]$. That further implies that $\mathcal{P}er_2(\mathbf{R}(P, \alpha))$ is of the form

$$\mathcal{P}er_2(\mathbf{R}(P, \alpha)) = g_i \cdot \cos\alpha + f_i \cdot \sin\alpha \qquad \text{for} \quad \alpha \in [\beta_i, \beta_{i+1}], \quad i = 1, 2, \ldots, m, \tag{3}$$

and $\beta_{m+1} = 2 \cdot \pi + \beta_1$.

We refer to Fig. 6 for an illustration. Let δ_1 be the angle between the positively oriented x-axis and the edge $[AB]$. Also, let $\delta_2 = \min\{\angle(JBC),\ \angle(KDE),\ \angle(GFL),\ \angle(HGI)\}$ (in a situation as in Fig. 6, $\delta_2 = \angle(KDE)$).

If the line $l(I, J)$ is chosen to be a support line then B, F and D, G are the antipodal pairs which determine two orthogonal pairs of line supports: $l(I, J), l(L, K)$ and $l(J, K), l(I, L)$.

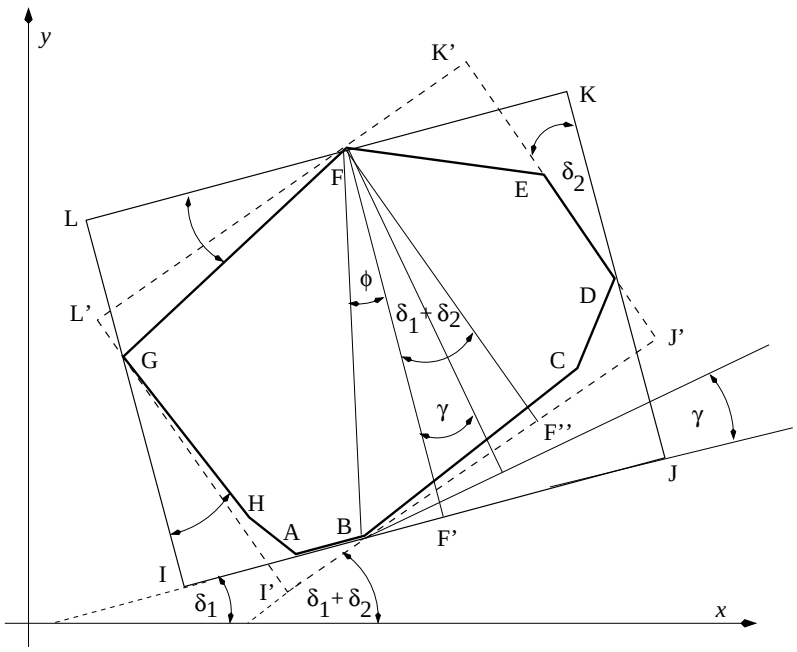


Figure 6: An illustration that $\mathcal{P}er_2(\mathbf{R}(P,\alpha))$ is of the form described by (3).

If the support line $l(I,J)$ is rotated into a new position around the vertex B the pairs B,F and D,G remain antipodal until the rotation γ angle varies from 0 to δ_2 (see Fig. 6).

For $\gamma \in [0,\delta_2]$ the width of the minimal rectangle which includes P varies from $l_2([FF'])$ to $l_2([FF''])$, more precisely this width is

$$l_2([FF'])\cdot\cos(\phi+\gamma) = (l_2([FF'])\cdot\cos\phi)\cdot\cos\gamma + ((-1)\cdot l_2([FF'])\cdot\sin\phi)\cdot\sin\gamma \quad \text{for } \gamma \in [0,\delta_2].$$

Since $l_2([FF'])$ and ϕ are constants which do not depend on γ, and by noticing that an analogous expression can be derived for the height of such a minimal rectangle we have proved that $\mathcal{P}er_2(\mathbf{R}(P,\alpha))$ can be expressed as in (3).

Two examples displayed in Fig. 9(a) and Fig. 10(a) illustrate that $\mathcal{P}er_1(\mathbf{R}(P,\alpha))$ behaves as described by (3).

Note 2 *For any angle $\beta_i \in \{\beta_1,\beta_2,\ldots,\beta_m\} \subset [0,\ 2\pi]$ there is an edge e_i, $1 \le i \le m$ of $\mathbf{CH}(P)$ such that after the rotation for the angle β_i the edge e_i becomes parallel to one of coordinate axes.*

4.3 A procedure for $\mathcal{C}(P)$ Computation

By using results from the previous two sub-sections, we can prove a useful theorem which shows that $\mathcal{C}(P)$ can be computed by comparing the values of $\dfrac{\mathcal{P}er_2(\mathbf{R}(P,\gamma))}{\mathcal{P}er_1(P,\gamma)}$ when γ belongs to a finite number point set consisting of no more than $5\cdot n$ points, where n is the number of vertices of P.

Theorem 3 *The convexity measure $\mathcal{C}(P)$ of a given given n-gon P can be computed as*

$$\mathcal{C}(P) = \min\left\{ \frac{\mathcal{P}er_2(\mathbf{R}(P,\gamma_i))}{\mathcal{P}er_1(P,\gamma_i)} \mid i = 1,2,\ldots,l \right\},$$

where $\{\gamma_1, \gamma_2, \ldots, \gamma_l\}$ is the set union of the sets $\{\alpha_j \mid i = 1,2,\ldots,k\}$ and $\{\beta_i \mid i = 1,2,\ldots,m\}$ which are defined in (2) and (3).

Proof. Let $\gamma_1 < \gamma_2 < \ldots < \gamma_l$ be the ordered sequence of angles from the set union $\{\alpha_j \mid i = 1,2,\ldots,k\} \cup \{\beta_i \mid i = 1,2,\ldots,m\}$ (obviously $l \leq k+m \leq 5 \cdot n$), where α_j $(1 \leq j \leq k)$ and β_i $(1 \leq i \leq m)$ are determined as in (2) and (3), respectively. Further, from (2) and (3)

$$\frac{\mathcal{P}er_2(\mathbf{R}(P,\gamma))}{\mathcal{P}er_1(P,\gamma)} = \frac{\tilde{g}_i \cdot \cos\gamma + \tilde{f}_i \cdot \sin\gamma}{\tilde{c}_i \cdot \cos\gamma + \tilde{d}_i \cdot \sin\gamma}$$

for some constants $\tilde{c}_i, \tilde{d}_i, \tilde{g}_i, \tilde{f}_i$ for $\gamma \in [\gamma_i, \gamma_{i+1}]$ and for $i = 1,2,\ldots,l$.

The first derivative

$$\left(\frac{\tilde{g}_i \cdot \cos\gamma + \tilde{f}_i \cdot \sin\gamma}{\tilde{c}_i \cdot \cos\gamma + \tilde{d}_i \cdot \sin\gamma}\right)' = \frac{\tilde{f}_i \cdot \tilde{c}_i - \tilde{g}_i \cdot \tilde{d}_i}{(\tilde{c}_i \cdot \cos\gamma + \tilde{d}_k \cdot \sin\gamma)^2}$$

together with $\mathcal{P}er_1(P,\gamma) = \tilde{c}_i \cdot \cos\gamma + \tilde{d}_i \cdot \sin\gamma > 0$ shows that $\dfrac{\mathcal{P}er_2(\mathbf{R}(P,\gamma))}{\mathcal{P}er_1(P,\gamma)}$ does not have any local extrema inside all intervals (γ_i, γ_{i+1}) where $i = 1,2,\ldots l$, and consequently it reaches its minimum at one of the points from $\{\gamma_i \mid i = 1,2,\ldots,l\}$. ▥

The previous theorem shows that there is a trivial procedure for computing the new convexity measure of a given n-gon in an $\mathcal{O}(n^2)$ time. Namely, it is enough to take the minimum among a finite number of values

$$\frac{\mathcal{P}er_2(\mathbf{R}(P,\gamma_1))}{\mathcal{P}er_1(P,\gamma_1)}, \quad \frac{\mathcal{P}er_2(\mathbf{R}(P,\gamma_2))}{\mathcal{P}er_1(P,\gamma_2)}, \quad \ldots, \quad \frac{\mathcal{P}er_2(\mathbf{R}(P,\gamma_l))}{\mathcal{P}er_1(P,\gamma_l)}.$$

Also, we note that this minimum can be obtained if the ratio $\dfrac{\mathcal{P}er_2(\mathbf{R}(P))}{\mathcal{P}er_1(P)}$ is computed for all choices of coordinate system such that one of coordinate axes is parallel to one of edges of the polygon P or one of the edges of $\mathbf{CH}(P)$. The examples from Fig. 9(c) and Fig. 10(c) verify the proven statement that $\dfrac{\mathcal{P}er_2(\mathbf{R}(P,\alpha))}{\mathcal{P}er_2(P,\alpha)}$ has no local extrema on the defined intervals (γ_i, γ_{i+1}), for $1 \leq i \leq l$.

5 Further Analysis, Comparison and Examples

This section contains several examples with synthetic data in order to illustrate applicability of the new convexity measure. It is divided into four subsections. The first two subsections are related to the initial examples: $P(h)$ and $T(t,h)$ from Fig. 2. The third subsection shows an example when the new measure gives no acceptable results. The fourth subsection compares the boundary based convexity measures $\mathcal{C}$ and $\mathcal{C}_3$. Situations where the new measure $\mathcal{C}$ is advantageous over $\mathcal{C}_3$ are pointed out.

5.1 The Graph of $\mathcal{C}(P(h))$

The polygon $P(h)$ is defined in Fig. 2, while its measured convexity $\mathcal{C}(P(h))$ as a function of h is given on Fig. 7. It is easy to check that the minimal measured convexity is obtained in the limit case $h = 0$ when $\lim_{h\to 0} \mathcal{C}(P(h)) = \frac{2}{3}$ holds. Starting from $h = 0$ the function $\mathcal{C}(P(h))$ monotonically increases from $\frac{2}{3}$ and reaches the maximum $\mathcal{C}(P(h)) = 1$ for $h = 1$ when $P(1)$ coincides with a square.

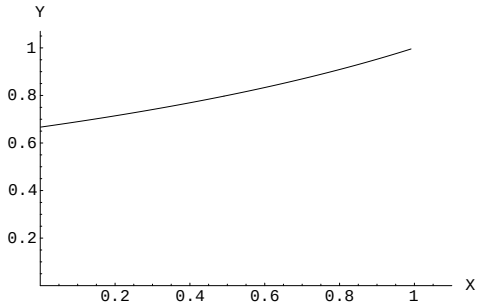


Figure 7: The measured convexity $\mathcal{C}(P(h))$ monotonically increases from $\frac{2}{3}$ to 1, as h varies from 0 to 1.

5.2 Convexity Measure of $T(t, h)$

Taking now $T(t, h)$ from Fig. 2, the measured convexity $\mathcal{C}(T(t, h))$ depends on two variables. Let us consider the case when h is arbitrary small. Then, we can define the function $\mathcal{C}(T(t, 0))$ as

$$\mathcal{C}(T(t, 0)) = \lim_{h\to 0} \mathcal{C}(T(t, h)$$

whose graph is given in Fig. 8. Obviously, $\mathcal{C}(T(t, 0))$ is not a monotonic function. More specifically, if t varies from 0 to 0.5 then $\mathcal{C}(T(t, 0))$ monotonically decreases from the maximum value 1 (when $T(0, 0)$ coincides with a square) to the minimum value 0.8 reached for $t = 0.5$, i.e., $\mathcal{C}(T(0.5, 0)) = \min_{t\in[0,\infty)} \mathcal{C}(T(t, 0)) = 0.8$. The measured convexity $\mathcal{C}(T(t, 0))$ monotonically increases on the interval $t \in (\frac{1}{2}, \infty)$. On this interval $\mathcal{C}(T(t, 0))$ does not reaches the maximum, but for a large enough t the function $\mathcal{C}(T(t, 0))$ is arbitrarily close to 1, or more precisely

$$\lim_{t\to\infty} \mathcal{C}(T(t, 0)) = 1.$$

The last equality can be seen as desirable. Fig. 9 and Fig. 10 illustrate the results of the subsections 4.1, 4.2, and 4.3. For two chosen values $t = 0.35$ and $t = 0.5$ they display the graphs of the three functions whose form was studied there.

5.3 An Example of Unsatisfactory Measurement

We showed in Section 1 that area based methods such as $\mathcal{C}_2(P)$ are not sensitive to huge defects on the shape's boundary if they have little impact on the area. While the proposed

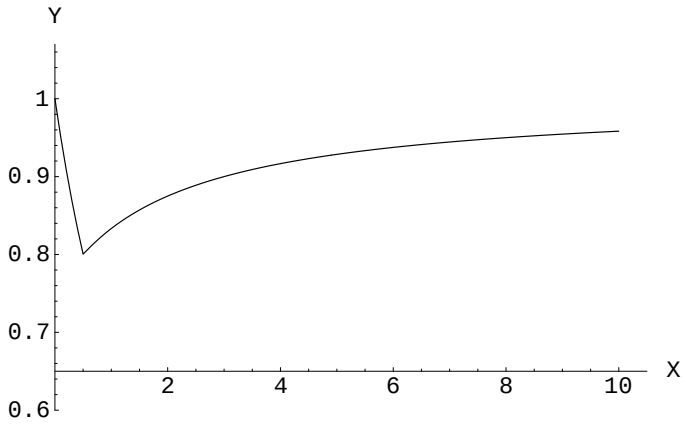
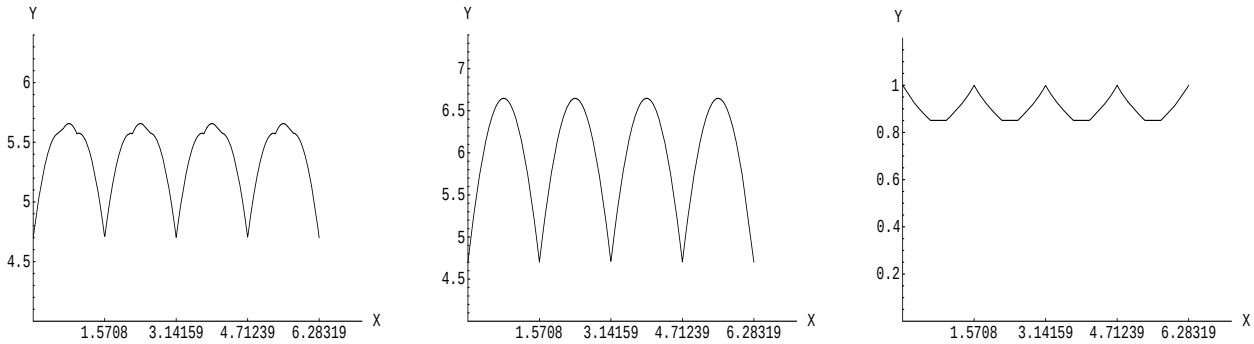


Figure 8: The measured convexity $\mathcal{C}(T(t,0))$.



(a) $\mathcal{P}er_2(\mathbf{R}(T(0.35,0),\alpha)$ (b) $\mathcal{P}er_1(T(0.35,0),\alpha)$ (c) $\frac{\mathcal{P}er_2(\mathbf{R}(T(0.35,0),\alpha)}{\mathcal{P}er_2(T(0.35,0),\alpha)}$

Figure 9: Graphs showing the behaviour of $\mathcal{P}er_2(\mathbf{R}(T(t,h),\alpha))$, $\mathcal{P}er_1(T(t,h),\alpha)$, and $\frac{\mathcal{P}er_2(\mathbf{R}(T(t,h),\alpha))}{\mathcal{P}er_1(T(t,h),\alpha)}$, when α varies from 0 to 2π with fixed $t = 0.35$ and $h \to 0$.

method $\mathcal{C}(P)$ rectifies this anomaly there are of course situations when it produces poor results, as shown in Fig. 11.

Since

$$\min_{\alpha\in[0,2\pi]} \mathcal{P}er_2(\mathbf{R}(P_{n,m},\alpha)) = 2\cdot(n+m) \qquad \text{and} \qquad \max_{\alpha\in[0,2\pi]} \mathcal{P}er_1(P_{n,m},\alpha) = 4\cdot(n+m)-4$$

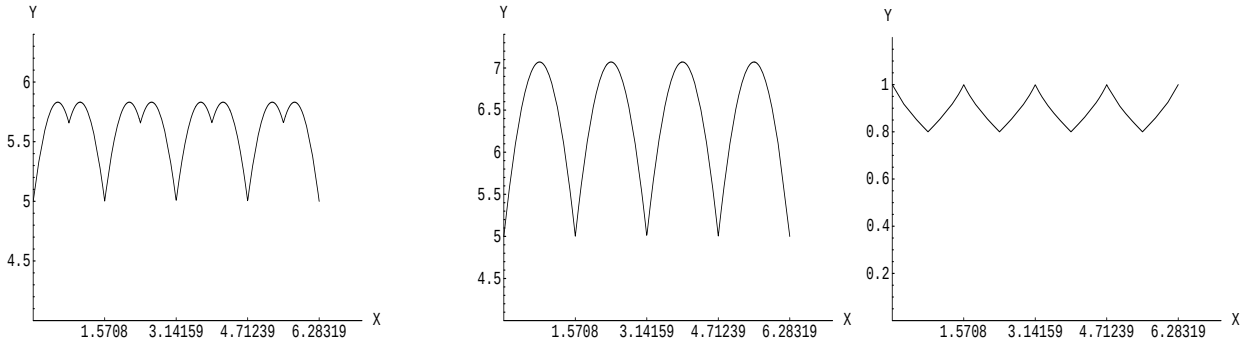
and because a lower bound (derived as follows)

$$\mathcal{C}(P_{n,m}) = \min_{\alpha\in[0,2\pi]} \frac{\mathcal{P}er_2(\mathbf{R}(P,\alpha))}{\mathcal{P}er_1(P_{n,m},\alpha)} \leq \frac{n+m}{2\cdot(n+m)-2}$$

is reached for $\alpha = 0$, it can be concluded that

$$\lim_{(m+n)\to\infty} \mathcal{C}(P_{n,m}) = \frac{1}{2}.$$

The previous equality could be acceptable for $m = 2$ or $m = 3$ and $n \to \infty$, but it is reasonable to expect that the convexity of such a shape for large m and n should be close to 1 – as it is estimated by $\mathcal{C}_1(P_{n,m})$, $\mathcal{C}_2(P_{n,m})$, and $\mathcal{C}_2'(P_{n,m})$. Note that $\mathcal{C}_3(P_{n,m}) = \frac{1}{\sqrt{2}}$.



(a) $\mathcal{P}er_2(\mathbf{R}(T(0.5,0),\alpha)$ (b) $\mathcal{P}er_1(T(0.5,0),\alpha)$ (c) $\frac{\mathcal{P}er_2(\mathbf{R}(T(0.5,0),\alpha)}{\mathcal{P}er_2(T(0.5,0),\alpha)}$

Figure 10: Graphs showing the behaviour of $\mathcal{P}er_2(\mathbf{R}(T(t,h),\alpha))$, $\mathcal{P}er_1(T(t,h),\alpha)$, and $\frac{\mathcal{P}er_2(\mathbf{R}(T(t,h),\alpha))}{\mathcal{P}er_1(T(t,h),\alpha)}$, when α varies from 0 to 2π with fixed $t = 0.5$ and $h \to 0$.

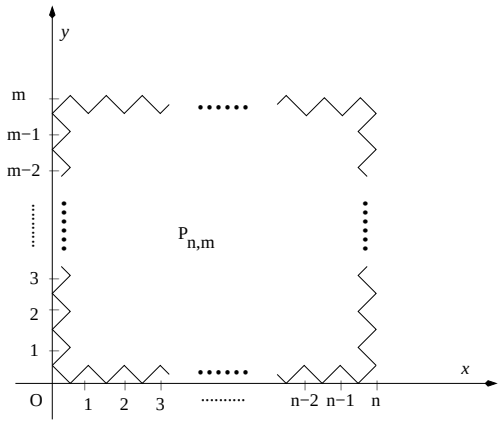


Figure 11: $P_{n,m}$ has the measured convexity $\mathcal{C}(P_{n,m}) = \frac{m+n}{2\cdot n+2\cdot m-2}$.

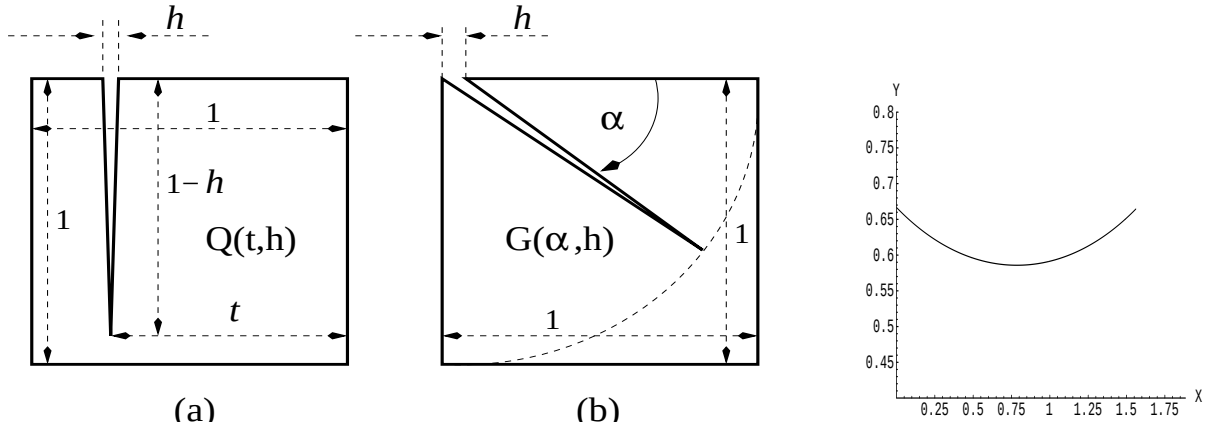
5.4 Comparison Between Boundary Based Measurements

In this subsection we compare $\mathcal{C}(P)$ and $\mathcal{C}_3(P)$ which are both boundary based convexity measures. As mentioned they are expected to be more sensitive to large boundary defects than the area based measures.

With two examples we will illustrate that the new measure $\mathcal{C}(P)$ is more sensitive than $\mathcal{C}_3(P)$. By the way, such an estimate follows from the definitions as well. Namely, Definition 4 says that all shapes with the same convex hull and the same Euclidean perimeter have the same measured convexity. It implies, as an example, that the measured convexity of polygons $Q(t, h \to 0)$ and $G(\alpha, h \to 0)$ does not depend on t and α respectively. That is,

$$\mathcal{C}_3(Q(t,h\to 0)) = \frac{2}{3} \text{ for any } t \in (0,1) \quad \text{and} \quad \mathcal{C}_3(G(\alpha,h\to 0)) = \frac{2}{3} \text{ for any } \alpha \in (0,\frac{\pi}{2}).$$

However, the $\mathcal{C}$ measure can differentiate between shapes with the same convex hull and Euclidean perimeter. Referring again to Fig. 12, it can be easily derived that $\mathcal{C}(Q(t, h \to 0)) = \frac{2}{3}$ independently of t, but if $\mathcal{C}$ is applied to the polygon $G(\alpha, h \to 0)$ then $\mathcal{C}(G(\alpha, h \to$



(c)

Figure 12: (a) $\mathcal{C}(Q(t,h)) = \mathcal{C}_3(Q(t,h)) = 2/3$ for any $t \in (0,1)$ if $h \to 0$. (b) $\mathcal{C}(G(\alpha,t))$ strongly depends on α, while $\mathcal{C}_3(Q(t,h)) = 2/3$ for all $\in (0,1)$ and $h \to 0$; (c) The graph of $\mathcal{C}(G(\alpha,h))$ when $h \to 0$.

0)) strongly depends on α. The graph of $\mathcal{C}(G(\alpha, h \to 0))$ when $\alpha \in (0, \pi/2)$ is given in Fig. 12(c).

Note that the new convexity measure can be applied to shapes whose boundaries consist of several polygonal lines (see Fig. 13), i.e., to the shapes which are unions or set differences of polygonal areas. From the viewpoint of practical applications it means that it is possible to measure the convexity of shapes with holes. The perimeter of such shapes is defined to be the sum of the length of all boundary lines.

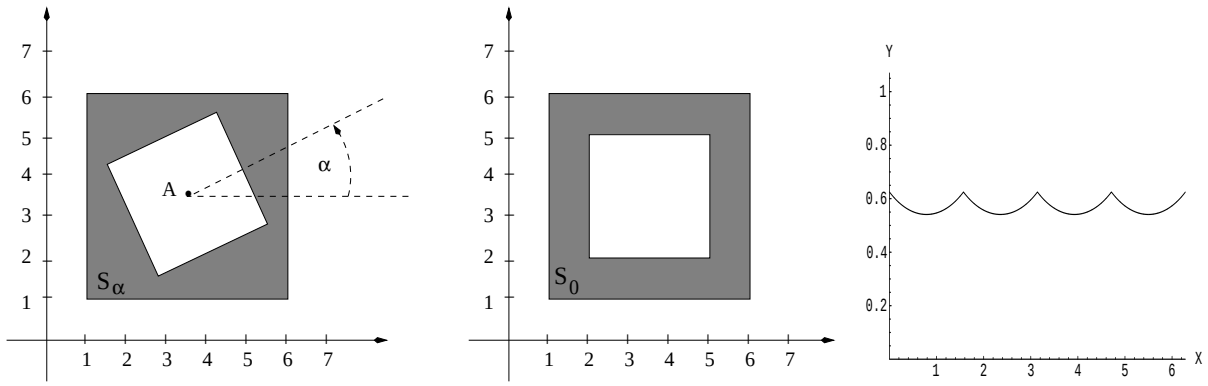


Figure 13: The estimate $\mathcal{C}(S_\alpha) < \mathcal{C}(S_0)$ (if $\alpha \neq 0$) is more appropriate than the equality $\mathcal{C}_3(S_\alpha) = \mathcal{C}_3(S_0)$ which holds for any α. The right figure displays the graph of $\mathcal{C}(S_\alpha)$ for $\alpha \in [0, 2\pi]$.

Let T_1 and T_2 be two isometric polygonal subsets of a given polygon P. Then the set differences $P \setminus T_1$ and $P \setminus T_2$ have the same convexity measured by $\mathcal{C}_3$, i.e., $\mathcal{C}_3(P \setminus T_1) = \mathcal{C}_3(P \setminus T_2)$, for all isometric (moreover isoperimetric) $T_1 \subset P$ and $T_2 \subset P$. But $\mathcal{C}(P \setminus T_1) = \mathcal{C}(P \setminus T_2)$ is not guaranteed. Consider the example from Fig. 13. The boundary of the shape S_α consists of two squares. The first one has the vertices $(1,1)$, $(6,1)$, $(6,6)$, and $(1,6)$ while the second one is obtained by the rotating the square with vertices $(2,2)$, $(5,2)$, $(5,5)$, and

$(2, 5)$ for an angle α around the point $A = (3.5,\ 3.5)$. The equality $\mathcal{C}_3(S_\alpha) = 5/8$ holds for any choice of α but the new convexity measure gives different values for non-isometric S_α and S_β. Moreover, $\mathcal{C}(S_\alpha)$ reaches its maximum for $\alpha = k \cdot \frac{\pi}{2}$, with $k = 0, 1, 2, 3$, its minimum is at $\alpha = \frac{\pi}{4} + k \cdot \frac{\pi}{2}$, with $k = 0, 1, 2, 3$, which seems to be natural and which is in accordance with the measure $\mathcal{C}'_2$.

6 Experimental Results

We start with a qualitative comparison between the two measures $\mathcal{C}$ and $\mathcal{C}_2$. Figs. 14 and 15 show 35 shapes ordered into decreasing convexity by the measures. To reduce the sensitivity of $\mathcal{C}$ to noise, minor fluctuations have been removed by first simplifying the boundary using Ramer's [17] polygonal approximation algorithm. A threshold of maximum deviation equal to three is used in all cases and the widths of the shapes are between 100–300 pixels.

First off, by manual inspection the results demonstrate that the new convexity measure produces sensible results. Although a comparison of the rankings from $\mathcal{C}$ and $\mathcal{C}_2$ reveals many similarities (as expected) some differences are also evident. It can be seen that $\mathcal{C}$ is stricter than $\mathcal{C}_2$ regarding shapes that are roughly convex apart for relatively narrow indentations. For example, the third, fifth, and seventh shapes in the first row in Figs. 15 score highly according to $\mathcal{C}_2$ and lower according to $\mathcal{C}$. Also evident, the more the shape is fragmented by the intrusions the lower its score. In contrast, protrusions are penalised rather less by $\mathcal{C}$ than $\mathcal{C}_2$, as shown by examples such as the tennis racket (third shape in the first row) and the "L" shape (eighth shape in the first row) in Fig. 14. Shapes containing substantial indentations *and* protrusions such as the music example (third last shape in the last row in Fig. 14) are rated low by both measures.

For a second, quantitative test we perform classification of *diatoms*, which are unicellular algae found in water, and have applications in forensics, geology, ecological monitoring etc. [5]. The mixed genera set from the ADIAC project was used, consisting of 808 contours covering 38 taxa. The contours vary in size, containing between 134–1277 pixels. All are processed as before by finding their polygonal approximations with a maximum deviation of three pixels. Figure 16 shows an evenly sampled selection after ranking by decreasing convexity according to $\mathcal{C}$. Unlike $\mathcal{C}_2$ the measure $\mathcal{C}$ can also be applied to take interior detail into account. In figure 17 the original images are shown in the first row, and below are the boundaries combined with internal structure obtained by edge detection. Again, the shapes are ranked by $\mathcal{C}$, and a very different ordering is obtained compared to figure 16. Notionally the internal structures are like infinitely thin cuts into the shape's interior, and so the diatoms with denser internal detail are assigned lower convexity values.

Next, classification of the diatoms into the 38 taxa was performed using Murthy *et al.*'s oblique decision trees (OC1) [14] and 100-fold cross-validation. The first row of values in table 1 shows the accuracies using $\mathcal{C}$, $\mathcal{C}_2$ and $\mathcal{C}_3$ applied separately to the outer contour. Classification was also carried out using $\mathcal{C}$ applied to the combined boundaries and internal contours (as shown in figure 17) and also using the pair of $\mathcal{C}$ values obtained from just the boundary data and also the combined data. Although for this classification task $\mathcal{C}$ has less

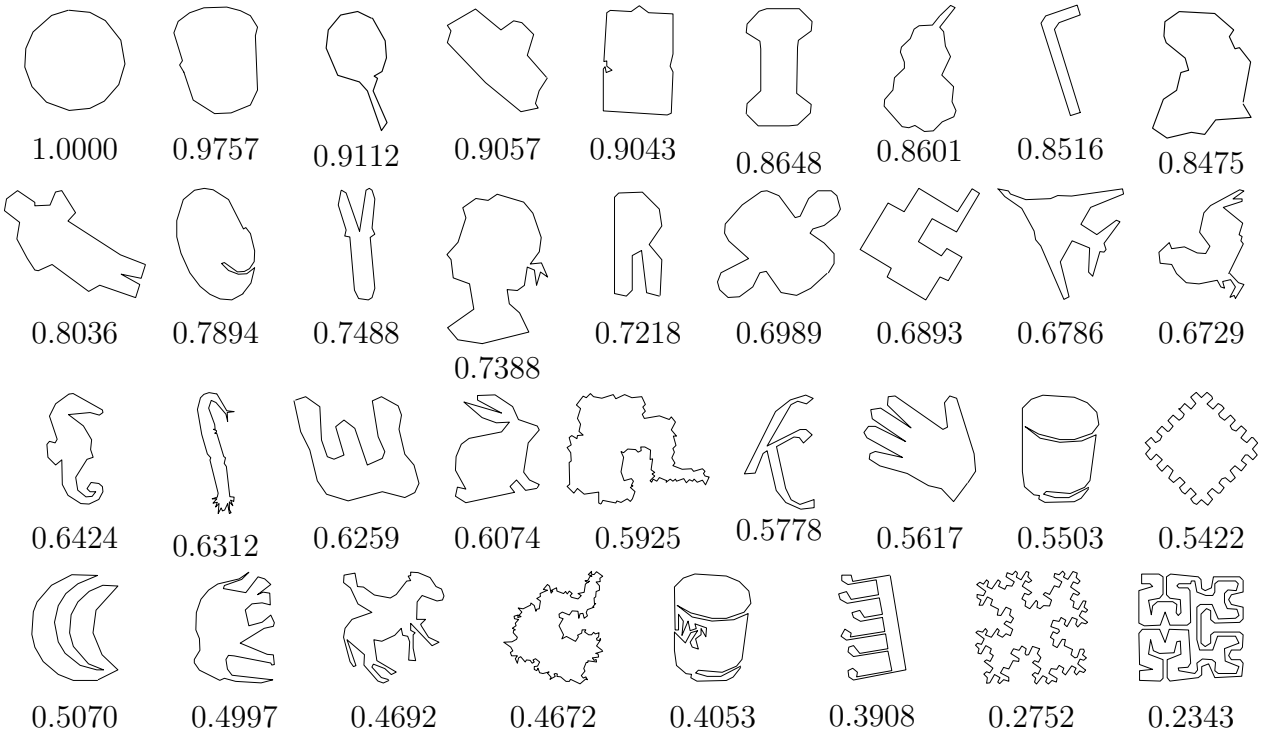


Figure 14: Shapes ranked by the $\mathcal{C}$ measure.

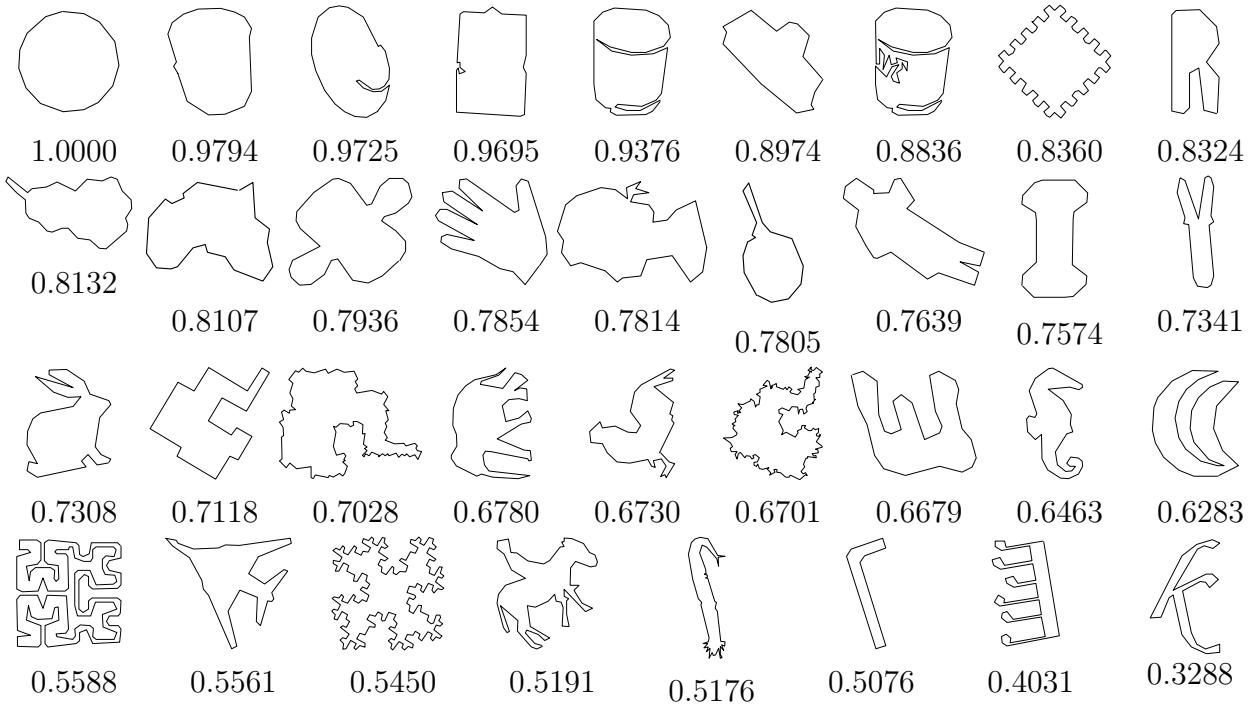


Figure 15: Shapes ranked by the measure $\mathcal{C}_2$.

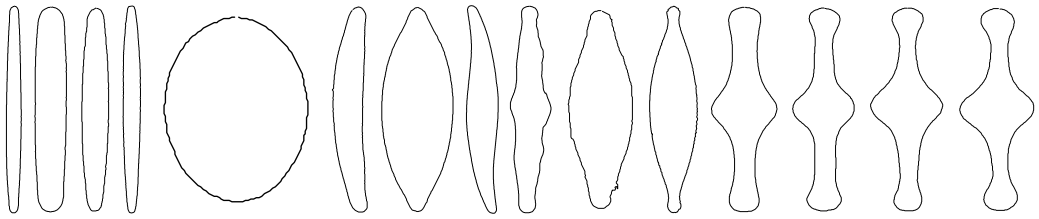


Figure 16: Diatoms ranked in decreasing convexity by $\mathcal{C}$ applied to their boundaries.

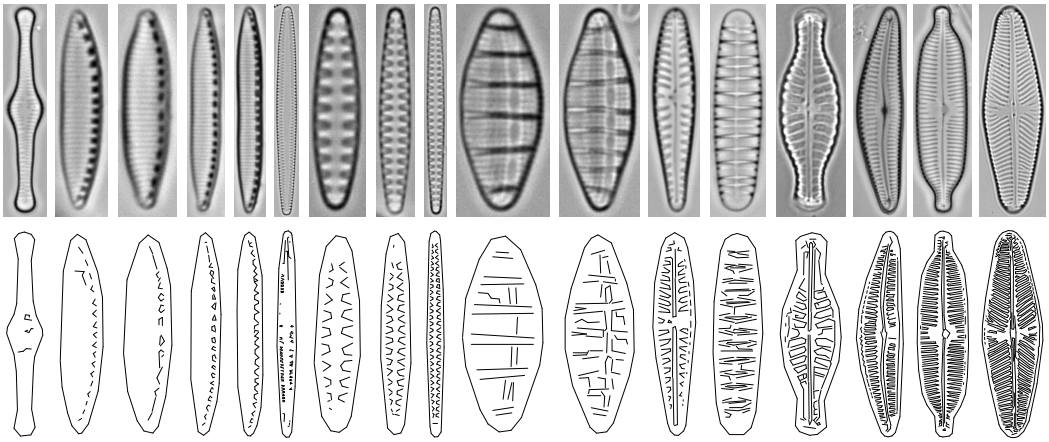


Figure 17: Diatoms ranked in decreasing convexity by $\mathcal{C}$ applied to both their boundaries and their internal contours. The top shows the original images, the bottom row shows the input to $\mathcal{C}$.

discriminating power than $\mathcal{C}_2$ when applied to the boundaries, its application to the interior provides an independent set of shape measurements that enables the combination of the two values to substantially improve the classification accuracy. The second row lists the results obtained when an additional set of standard and recent shape measures[1] was used by the classifier along with convexity. Again an increase in accuracy is achieved using $\mathcal{C}$ applied both internally and externally.[2]

As seen in figure 16 the diatoms have fairly smooth boundaries. To demonstrate the sensitivity and effectiveness of $\mathcal{C}$ to indentations it is applied in the last example to *desmids*, another type of algae. They are not as flat as diatoms, making fully in-focus images problematic, and so drawings are still extensively used by biologists for classification. As a small scale classification task we have taken the drawings of the *Micrasterias* taxon from West

[1]The additional shape measures used were: circularity, ellipticity, rectangularity, triangularity [19] aspect ratio, compactness, convexity, eccentricity, the first four rotation, translation, and scale moment invariants, four rotation, translation, and scale moment invariants [20], the first three affine moment invariants [8].

[2]Fischer and Bunke [7] reported better accuracies than ours (in excess of 90%). However, they used diatom specific boundary features (e.g. 10 descriptors for valve endings) and internal textural details (to capture the diatom's "ornamentation"). Moreover, by applying bagging they further increased the performance of the decision tree classifiers.

	$\mathcal{C}_2(S)$ bdry	$\mathcal{C}_3(S)$ bdry	$\mathcal{C}(S)$ bdry	$\mathcal{C}(S)$ intr	$\mathcal{C}(S)$ intr & $\mathcal{C}(S)$ bdry
Alone	17.82	12.50	13.99	21.78	43.69
Other Features	77.72	78.59	76.11	79.58	81.81

Table 1: Classification accuracies for 808 diatoms using the OC1 decision tree and 100-fold cross validation. Measures were applied to the shape boundary (bdry) and interior (intr).

and West's [23] comprehensive flora, and selected all nine species containing at least four drawings. The data set is limited to 43 sample outlines, each containing 1000–10000 pixels, with 4–7 drawings for each species. Although high resolution images were used, the narrow indentations were still sometimes truncated during boundary extraction, complicating the recognition task; an example of each species is shown in figure 18.

Classification into the nine species was carried out using OC1 as above, except that leave-one-out testing was performed due to the small data set. As can be seen in table 2 the single convexity measures are not sufficient for discrimination. Combining pairs of convexity measures brings a significant improvement, and emphasises that the area-based and perimeter-based measures provide complementary information. Using $\mathcal{C}_2(S)$ and $\mathcal{C}(S)$ together provided the best result with 55.81% accuracy. Incorporating $\mathcal{C}_3(S)$ as well provided a relatively small gain: combining the three measures achieved 58.40% accuracy. Of course, in order to gain higher accuracies a larger training set and a larger set of shape measures would be necessary.

	$\mathcal{C}_2(S)$	$\mathcal{C}_3(S)$	$\mathcal{C}(S)$
$\mathcal{C}_2(S)$	37.21	48.84	55.81
$\mathcal{C}_3(S)$	–	25.58	16.28
$\mathcal{C}(S)$	–	–	18.60

Table 2: Classification accuracies for 43 desmids using the OC1 decision tree and leave-one-out cross validation.

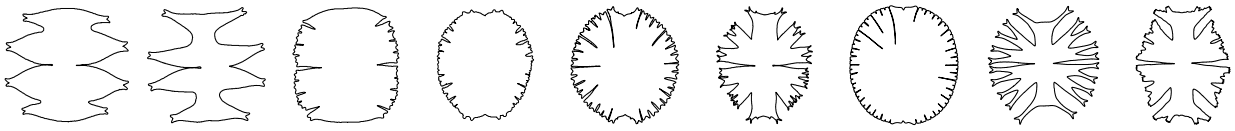


Figure 18: Examples of *Micrasterias* desmid boundaries.

7 Concluding Remarks

In this paper a new convexity measure has been proposed for describing shapes. In contrast to the most common approach (the ratio of the area of the shape to the area of its convex

hull) it is based on the shape's boundary perimeter rather than its area. Theoretical and experimental analysis shows that it performs well. Compared to area based approaches it is more sensitive to deep indentations into shapes, especially if they are thin (i.e. of negligible area).

Moreover, compared to the boundary based convexity measure $\mathcal{C}_3$ it gives better results if applied to shapes with holes or additional internal edges. Namely, while the relative position of the holes or edges inside the shape has no effect on the convexity measured by $\mathcal{C}_3$ they have an impact on the convexity measure proposed here. That provides a significant advantage in shape representation.

Acknowledgements

The diatom data was kindly provided by the ADIAC project; CEC contract MAS3-CT97-0122. The desmid data was kindly provided by the DIADIST project; BBSRC grant 754/BIO14262.

References

[1] D.M. Acketa and J. Žunić. On the maximal number of edges of digital convex polygons included into an (m, m)-grid. *Journal of Combinatorial Theory Series A*, 69(2):358–368, 1995.

[2] F. Bookstein. *Morphometric Tools for Landmark Data: Geometry and Biology*. Cambridge University Press, 1991.

[3] L. Boxer. Computing deviations from convexity in polygons. *Pattern Recognition Letters*, 14:163–167, 1993.

[4] L. da F. Costa and R.M. Cesar Jr. *Shape Analysis and Classification*. CRC Press, 2001.

[5] J.M.H. du Buf and M.M. Bayer, editors. *Automatic Diatom Identification*. World Scientific, 2002.

[6] S. Hyde *et al. The Language of Shape*. Elsevier, 1997.

[7] S. Fischer and H. Bunke. Identification using classical and new features in combination with decision tree ensembles. In J.M.H. du Buf and M.M. Bayer, editors, *Automatic Diatom Identification*, pages 109–140. World Scientific, 2002.

[8] J. Flusser and T. Suk. Pattern recognition by affine moment invariants. *Pattern Recognition*, 26:167–174, 1993.

[9] H. Freeman and R. Shapira. Detemining the minimum-area encasing rectangle for an arbitrary closed curve. *Comm. ACM*, 18(7):409–413, 1975.

[10] A.E. Hawkins. *The Shape of Powder-Particle Outlines*. J. Wiley and Sons, 1993.

[11] L.J. Latecki and R. Lakämper. Convexity rule for shape decomposition based on discrete contour evolution. *Computer Vision and Image Understanding*, 73(3):441–454, 1999.

[12] L.J. Latecki and R. Lakämper. Shape similarity measure based on correspondence of visual parts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1185–1190, 2000.

[13] R.R. Martin and P.C. Stephenson. Putting objects into boxes. *Computer Aided Design*, 20:506–514, 1988.

[14] S.K. Murthy, S. Kasif, and S. Salzberg. System for induction of oblique decision trees. *Journal of Artificial Intelligence Research*, 2:1–33, 1994.

[15] A.F. Pitty. *Geomorphology.* Blackwell, 1984.

[16] F.P. Preparata and M.I. Shamos. *Computational Geometry.* Springer-Verlag, 1985.

[17] U. Ramer. An iterative procedure for the polygonal approximation of plane curves. *Computer Graphics and Image Processing*, 1:244–256, 1972.

[18] P.L. Rosin. Shape partitioning by convexity. *IEEE Transactions on Systems, Man and Cybernetics*, part A, 30(2):202–210, 2000.

[19] P.L. Rosin. Measuring shape: Ellipticity, rectangularity, and triangularity. *Machine Vision and Applications*, 14:172–184, 2003.

[20] M. Sonka, V. Hlavac, and R. Boyle. *Image Processing, Analysis, and Machine Vision.* Chapman and Hall, 1993.

[21] H.I. Stern. Polygonal entropy: a convexity measure. *Pattern Recognition Letters*, 10:229–235, 1989.

[22] G.T. Toussaint. Solving geometric problems with the rotating calipers. In *Proc. IEEE MELECON '83*, pages A10.02/1–4, 1983.

[23] W. West and G.S. West. *A Monograph of the British Desmidiaceae.* The Ray Society, London, 1904-1923.