A Probabilistic Convexity Measure

Paul L. Rosin[a] Joviša Žunić,[b, *]

July 21, 2015

[a] School of Computer Science, Cardiff University, Cardiff CF24 3AA, Wales, U.K.
tel/fax: +44 (0)29 2087 5585/4598
e-mail: Paul.Rosin@cs.cf.ac.uk

[b] Department of Computer Science, University of Exeter, Exeter EX4 4QF, U.K.
tel/fax +44 (0) 1392 26 4044/4067
e-mail: J.Zunic@ex.ac.uk

Abstract

In order to improve the effectiveness of shape based classification there is an ongoing interest in creating new shape descriptors or creating new measures for descriptors that are already defined and used in shape classification tasks. Convexity is one of the most widely used shape descriptors and also one of the most studied in the literature. There are already several defined convexity measures. The most standard one comes from the comparison between a given shape and its convex hull. There are also some nontrivial approaches.

In this paper we define a new measure for shape convexity. It incorporates both area based and boundary based information, and in accordance with this it is more sensitive to boundary defects than exclusively area based convexity measures. The new measure has several desirable properties and it is invariant under similarity transformations. When compared with convexity measures that trivially follow from the comparison between a measured shape and its convex hull then the new convexity measure also shows some advantages – particularly for shapes with holes.

Keywords: Shape, convexity, measurement, shape classification.

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

1 Introduction

Shape is a recurrent theme in computer vision, and is still an active area of research even after thirty years. This paper only considers two dimensional shapes (and by shape we mean a compact planar set with a nonempty interior), but the techniques discussed could be directly applied to higher dimensions. A recent popular application of shape analysis to 3D data is the retrieval of objects from large collections of 3D surface models, such as the Princeton Shape Database [11].

There are many ways to characterise shape. One possibility is to decompose the object into parts and describe the parts and their interrelationships. Alternatively, more holistic or global approaches treat the object as a single entity. Either the shape is then described by a set of values (e.g. Osada *et al.*'s approach was to represent the shape signature of a 3D object as a distribution of sample distances [11]) or else by a single scalar as a compact description of some salient aspect of shape such as circularity, complexity, regularity, symmetry, etc.

This paper is an example of the latter approach, and defines a new *convexity* measure. Over the years several convexity measures have been developed (e.g. [2, 5, 6, 9, 14, 15, 17, 21, 23, 24]) and have been used for applications such as shape decomposition [8, 20], figure/ground segmentation [13], and pornography blocking [1]. Basically, these alternative convexity measures are derived from some suitable definition of convex shapes or by comparing measured shapes with their convex hulls.

The most standard definition of convex shapes is the following.

Definition 1 *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.*

Based on the previous definition it is reasonable to define a convexity measure as follows.

Measure 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.*

The convexity measure defined above has the following desirable properties:

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

- the convexity measure of a shape is invariant under similarity transformations of this shape.

On the other hand, the main objection to Definition 1 is that the exact computation $\mathcal{C}_1(S)$ is not easy even if S is bounded by a polygon.

As mentioned above, shape convexity can be measured by comparing the measured shape with its *convex hull*, where the convex hull of a given shape S is the smallest convex set which includes S. As usual, the convex hull of S is denoted as $\mathbf{CH}(S)$ (see figure 1).

Considering a shape and its convex hull, the following two convexity measures follow naturally. The first one can be understood as an area based one, while the second is boundary based.

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

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

In practice, the $\mathcal{C}_2$ convexity measure is the one that is mostly used, and appears in textbooks [22]. $\mathcal{C}_2(S)$ 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 measure $\mathcal{C}_2$ is not able to detect huge defects on boundaries of shapes which have a relatively small impact on the shape areas. A simple example is given in figure 2 (a) (see [24]). For small enough values of h the polygons $P(t,h)$ have the measured convexity $\mathcal{C}_2$ very close to 1, i.e.,

$$\lim_{h\to 0} \mathcal{C}_2(P(t,h)) = 1 \qquad \text{for all} \quad t \in (0,1).$$

For many applications the above estimate is not acceptable. Such an "anomaly" as $\lim_{h\to 0} \mathcal{C}_2(P(t,h)) = 1$ can be avoided by using another convexity measure that is derived by comparing the perimeter of the considered shape and the perimeter of its convex hull. We give the following definition.

Measure 3 *For a given planar shape S, its convexity measure $\mathcal{C}_3(S)$ is defined to be*

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

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

By the last definition we have a more acceptable situation for the shape presented in figure 2 (a): $\lim_{h\to 0} \mathcal{C}_3(P(t,h)) = \frac{2}{3}$. However, for applications to classification it is a disadvantage that the last estimate does not depend on t. Even worse, for a fixed $h \in (0,1)$ the shape $P(t,h)$ changes significantly as t varies through $(h/2, 1-h/2)$ but the measured convexity

$$\mathcal{C}_3(P(t,h)) = \frac{4}{4-h+\sqrt{5h^2-8h+4}}$$

is independent of t. The convexity measure defined here will overcome such a disadvantage.

While convexity measures $\mathcal{C}_1$, $\mathcal{C}_2$, and $\mathcal{C}_3$ follow directly from the definition of convex shapes and their convex hulls, in the literature there exist measures which are not derived in such a straightforward way – see [2], [23], and [24].

In this paper we define a new convexity measure which incorporates both area based and boundary based aspects. It satisfies the basic requirements listed after Measure 1 but has some advantages with respect to the previous convexity measure – particularly in measuring shapes with holes and shapes with deep intrusions.

The paper is organised as follows. The new measure $\mathcal{C}$ is defined in Section 2. The computation of $\mathcal{C}$ will be discussed in Section 3. Some possible modifications of the proposed measure are discussed in Section 4. In Section 5 several experimental results are provided. Section 6 contains concluding remarks.

2 A New Convexity Measure

In this section we define a new convexity measure. It follows naturally from the following definition of shape convexity.

Definition 2 *A planar shape S is said to be convex if it has the following property: If points A and B belong to the boundary of S then all points from the line segment $[AB]$ belong to the shape S.*

It is easy to see that Definition 1 and Definition 2 are equivalent. Based on Definition 2 it is reasonable to define a convexity measure as follows.

Measure 4 *For a given planar shape S its convexity measure $\mathcal{C}(S)$ is defined to be the probability that for randomly chosen points A and B from the boundary of S all points from the line segment $[AB]$ also belong to S, under the assumption that A and B are chosen uniformly along the boundary of S.*

Note. If other probability distributions (rather than the uniform distribution) is chosen, then we can generate other convexity measures. In the rest of the manuscript the uniform distribution will be assumed if not mentioned.

The convexity measure is obviously boundary based since it is defined in terms of points on the boundary. However, it also has an area based aspect since the randomly generated line segments can be considered to be probes of the area of the interior of the shape.

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

Theorem 1 *For any shape S we have:*

i) $\mathcal{C}(S) \in (0, 1]$*;*

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

iii) for any $\alpha \in (0, 1]$ *there is a shape* S *such that* $\mathcal{C}(S) = \alpha$*;*

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

Proof. The items $i,$) ii), and iv) follow easily from the definition.

In order to prove iii) just consider the ring $R(r)$ from figure 3(a). Applying elementary mathematics, the measured convexity $\mathcal{C}(R(r))$ can be expressed as

$$\mathcal{C}(R(r)) = \frac{2}{\pi(r+1)} \arccos r$$

while r varies from 0 to 1. Indeed, by the notations from figure 3(b), the angle β is equal to $\arccos r$, since $|OA| = r$ and $|OB| = 1$ i.e. the length of the arc BC is $\arccos r$. Thus, the probability that the line segment $[AX]$ belongs to $R(r)$ is $\dfrac{2 \arccos r}{2\pi r + 2\pi}$, where due to symmetry A can be considered fixed as shown in figure 3(b), and X lies on the boundary of $R(r)$ which consists of the two circles. Similarly, the probability that the edge EY (where Y belongs to the boundary of $R(r)$) is $\dfrac{4 \arccos r + 2r \arccos r}{2\pi r + 2\pi}$. By using those facts and the total probability theorem (see the footnote 1) the last equality holds. The proven equality implies iii). □

2.1 Measure Behaviour Illustrated by Synthetic Examples

In order to illustrate the behaviour of the new convexity measure, we consider a few synthetic examples given in figure 2 and figure 3 that enable an easy analytical computation.

The shape $P(t,h)$ from figure 2 (a) depends on the parameters $t \in (0,1)$ and $h \in (0,1)$. For a very small h the situation corresponds to a real shape with a very deep intrusion. It is easy to show that

$$\lim_{h\to 0} \mathcal{C}(P(t,h)) = \frac{2t^2 - 2t + 5}{9}, \qquad t \in (0,1).$$

The minimum measured convexity is for $t = 0.5$ and it is 0.5. For t very close to 0 (or very close to 1) the estimated convexity is close to 1, i.e. the shape is estimated as almost convex. Those estimates could be understood as reasonable ones.

Note. It is important to mention that the measured convexity depends on t, which is an advantage with respect to the convexity measure $\mathcal{C}_3$. Notice that even though $\mathcal{C}_3$ is boundary based, it is insensitive to the changes of boundary of $P(t,h)$ caused by the varying of t. Indeed, $\lim_{h\to 0} \mathcal{C}_3(P(t,h)) = 2/3$ for any $t \in (0,1)$. As mentioned $\lim_{h\to 0} \mathcal{C}_2(P(t,h)) = 1$ is also independent on t.

If we consider the shape $T(t,h)$ from the same figure, then

$$\lim_{h\to 0} \mathcal{C}(T(t,h)) = \frac{4 + t^2}{(2+t)^2}, \qquad \text{for each fixed} \quad t > 0.$$

This can be derived easily. Such a behaviour could be understood as reasonable. More precisely, for very small t (let us say, if $t \to 0$) it gives a measured convexity very close to 1 (i.e. the shape is estimated as "almost convex"). Also, when t is very large (let us say $t \to \infty$), the estimated convexity is again close to 1 which is acceptable because then the main part of boundary of $T(t \to \infty, h \to 0)$ degenerates into two coincident straight line segments. The minimum convexity is measured for $t = 2$ and this minimum is 0.5. Note that the convexity measure $\mathcal{C}_2(T(t, h \to 0))$ also depends on t and it is $(3 + \sqrt{4t^2+1})/(4+2t)$.

A circular ring (grey area) is presented in figure 3 (a). As it has been mentioned in the proof of Theorem 1, if the bigger circle has a radius equal to 1 and the radius of the smaller circle is r then the measured convexity of such a ring is $\frac{2}{\pi} \cdot \arccos r$ for $r \in (0,1)$. That is in accordance with our expectation that a very low measured convexity corresponds to r close to 1 while a measured convexity close to 1 corresponds to a very small r.

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

In this section we consider the computation aspects of the new measure. Our measure is, in part, boundary based and consequently the shape boundary plays a key role in the

computation of the measure. It is clear that when working with an image of an arbitrary real shape S the boundary $\mathbf{B}(S)$ is unknown – i.e. it is not given in an analytical way. Theoretically speaking, if we have such analytical expressions the measured convexity $\mathcal{C}$ can be computed. A standard approach could be a suitable decomposition of the boundary of the given shape and application of the *total probability theorem* [1] In practice, such a scenario could be applied only to shapes with simple boundaries. In fact, this is what was done in the case of the shapes in figure 2 and figure 3.

The situation is even worse when the measured shape S is presented in the corresponding binary image [7] – there is always an inherent loss of information. Even if we could properly extract the boundary points in order to apply the previous scenario we should know the boundary length between the selected points that made a proper boundary decomposition. But shape boundary length estimation (from a binary image) is also a very difficult problem (for a comparative study see [3]).

Thus, it seems that the most acceptable way to estimate $\mathcal{C}(S)$ is to approximate the boundary of S by a polygon and then estimate the measured convexity (of such a polygonal approximation) by applying a standard numerical approach (that uses a number of randomly generated sample points). That is because the exact computation of the convexity of the obtained polygonal approximation could be very complex if such a polygonal approximations has many vertices. How to approximate a curve by a polygonal line is well-studied problem (for an overview see [19]) with several very efficient solutions. Now, we recommend the following, very simple algorithm for an estimation of $\mathcal{C}$.

ESTIMATION OF $\mathcal{C}(S)$

Input: A binary image of a shape S and a number N that depends on the required precision that should be reached.

Step 1 Find a polygonal approximation $\mathbf{PA}(S)$ of the boundary of S;

Step 2 Find the perimeter $\mathcal{P}er(\mathbf{PA}(S))$ of $\mathbf{PA}(S)$ and fix an arbitrary point X on $\mathbf{PA}(S)$;

Step 3 Generate N pairs

$(a_1, b_1), (a_2, b_2), \ldots, (a_N, b_N)$

of random numbers from the interval $[0, \mathcal{P}er(\mathbf{PA}(S))]$;

[1] Given n mutually exclusive events $B_1, \ldots, B_n$ whose probabilities sum to unity, and an arbitrary event A. The formula of total probability says that $P(A) = \sum_{i=1}^{n} P(A|B_i)P(B_i)$ where $P(A|B_i)$ is the conditional probability of A assuming B_i.

(a) For any pair (a_i, b_i) $\quad i = 1, 2, \ldots, N$ $\quad$ determine the points Y_i and Z_i lying on $\mathbf{PA}(S)$ such that
– the part of $\mathbf{PA}(S)$ between points X and Y_i (taken in the counterclockwise order, for example) has the length a_i;
and
– the part of $\mathbf{PA}(S)$ between points X and Z_i (taken in the counterclockwise order) has the length b_i;

(b) Check if the straight line segment $[Y_i, Z_i]$ belongs to the area bounded by $\mathbf{PA}(S)$;

Output: The convexity $\mathcal{C}(S)$ is estimated as the number of edges that belong to the area bounded by $\mathbf{PA}(S)$ divided by N.

Note. Let us notice that the previous algorithm works if shape S consists of several disjoint parts (i.e. S is not connected) or/and if S has holes. In such situations $\mathbf{PA}(S)$ consists of all polygonal approximations of all parts (i.e. including all holes) $\mathbf{PA}(A_i)$ $(i = 1, 2, \ldots, k)$, listed in an arbitrary order. On each $\mathbf{PA}(A_i)$ an arbitrary point X_i should be fixed. The interval $[0, \mathcal{P}er(\mathbf{PA}(S))]$ can be expressed as the union of k intervals I_1, I_2, ..., I_k such that

$$[0, \mathcal{P}er(\mathbf{PA}(S))] = I_i \cup I_2 \cup \ldots \cup I_k =$$

$$[0, \mathcal{P}er(\mathbf{PA}(A_1))] \cup [\mathcal{P}er(\mathbf{PA}(A_1)), \mathcal{P}er(\mathbf{PA}(A_1)) + \mathcal{P}er(\mathbf{PA}(A_2))] \cup$$

$$\ldots \cup \left[\sum_{1 \leq i \leq k-1} \mathcal{P}er(\mathbf{PA}(A_i)), \mathcal{P}er(\mathbf{PA}(S)) \right].$$

To determine the points Y_i and Z_i from **Step 3** of the algorithm, the interval I_{v_0} to which a_i belongs should first be determined. Then the point Y_i is on the distance $a_i - \sum_{1 \leq j < v_0 - 1} length\_of\_I_j$ from the point X_{v_0}. Z_i is determined in the analogous way.

4 A Modified Convexity Measure

Some modifications of the new convexity measure are possible. One idea comes from [15]. The authors used an idea of generating pairs of points from a measured shape S and then checking if certain points on the corresponding line segment belong to the shape S. They use a parameter α to determine the exact location of the point on the line segment. Formally, it can be formulated as follows.

Measure 5 *Fix $\alpha \in (0,1)$. For a given shape S its convexity measure is defined as the probability that for randomly chosen points $A \in S$ and $B \in S$, the point $\alpha A + (1-\alpha)B$ also belongs to the shape S.*

It is easy to see that a shape is convex if and only if its measured convexity is equal to 1. Of course, infinitely many different measures can be obtained by varying the parameter α. A very similar approach is to define a convexity measure which does not take into account the complete line segment or a single point of it but does take a sub-segment of it. Also, wishing to keep the benefit of the easier generation of random points from the boundary, rather than interior points from the shape interior, we give the following definition.

Measure 6 *Fix $\alpha \in [0, 0.5]$. For a given shape S its convexity measure $\mathcal{C}_\alpha$ is defined as the probability that for randomly chosen points A and B from the boundary of S, the straight line segment $[\alpha A + (1-\alpha)B, (1-\alpha)A + \alpha B]$ completely belongs to S.*

Computation of $\mathcal{C}_\alpha$ follows the algorithm given in Section 3, except that in **Step 3b** the line segment $[Y_i, Z_i]$ is replaced by the sub-segment $[\alpha Y_i + (1-\alpha)Z_i, (1-\alpha)Y_i + \alpha Z_i]$.

Again, it is easy to see that a shape is convex if and only if its measured convexity is equal to 1 independently of which $\alpha \in [0, 1/2]$ has been chosen.

Obviously, if $\alpha = 0$ then $\mathcal{C}_0$ is equal to the convexity measure $\mathcal{C}$ introduced in Section 2. In the case $\alpha = 0.5$, i.e, when the line segment $[AB]$ is scaled by the maximum amount $\alpha = 0.5$ it collapses to its midpoint. This provides a similar type of convexity measure to that proposed by Rahtu *et al.* [15] (see Measure 5), which is based on the probability of a midpoint lying inside the polygon. The difference between the two measures is that they use different distributions of tested line segments.

The implication $\alpha \leq \beta \Rightarrow \mathcal{C}_\alpha \leq \mathcal{C}_\beta$ is an easy consequence of the definition of $\mathcal{C}_\alpha$. A practical interpretation of this implication is that a higher value of α would penalise less heavily the intersections between the generated line segments and the shape exterior if the intersections occur closer to the shape boundary. That can be useful when working with shapes having poorly extracted boundaries – or with shapes with irregular or textured boundaries. This observation will be verified on fractal examples in the next section.

5 Experimental Results

While the maximum convexity value of $\mathcal{C}$ of one is reached for any convex polygon, what shapes produce values approaching zero? A trivial example is given in figure 3. The presented

circular ring $R(r)$ is expected to have the measured convexity tending to zero when its area tends to zero (i.e. when $r \to 1$). Another class of such shapes are fractals; as demonstrated in figure 4 which shows the computed convexity for a Koch snowflake, recursively generated at a finite number of levels of detail. At the highest level of detail that was generated (i.e. 10) there were over 780,000 edges in the polygon. As the fractal's detail is increased the shape becomes more convoluted, thereby reducing the probability that the tested line segments do not intersect the boundary. However, if only part of the line segments are tested then the measure become less sensitive to minor deviations from convexity, as can be seen in figure 4.

Even for complex shapes the convexity measure converges reasonably quickly. This is demonstrated in figure 5 which analyses the Koch snowflake with 2, 3, 4, and 5 levels of detail, the polygons containing between 12 and 768 edges. It can be seen that a relatively accurate estimate is obtained after testing only 1000 line segments. To ensure high accuracy the experiments in this paper have used the conservative number of 100000 line segment tests.

The following show examples of the application of the proposed convexity measure along with results from several other convexity measures described in the literature by Žunic and Rosin [24] ($\mathcal{C}_J$), Rahtu *et al.* [15] ($\mathcal{C}_R$), Rosin and Mumford [21] ($\mathcal{C}_Q$), as well as the convex hull based methods $\mathcal{C}_2$ and $\mathcal{C}_3$.

Figure 6 illustrates with a set of simple synthetic polygons how the different convexity measures are affected by changes to the data:

- The area based methods ($\mathcal{C}_2$, $\mathcal{C}_Q$, and $\mathcal{C}_R$) are invariant under affine transformations.
- The convex hull methods ($\mathcal{C}_2$ and $\mathcal{C}_3$) are insensitive to rotations and translations of holes (or concavities so long as the transformations do not affect the convex hull).
- Measure $\mathcal{C}_J$ is invariant to translations of holes (or concavities as above), but is sensitive to rotations.
- Measure $\mathcal{C}_Q$ uses a fitted convex polygon that is something like the convex hull or convex skull depending on whether the input polygon mostly contains intrusions or protrusions. It is insensitive to rotations and translations of both intrusions and protrusions.
- The new measure $\mathcal{C}$ is sensitive to rotations and translations of both intrusions and protrusions.

In figure 7 the differences in the various convexity measures are highlighted by applying them to a wide range of shapes. We note first that all measures correctly assign the only

perfectly convex shape a maximum score, but differ in the ranking of the remaining shapes. It can be observed that the perimeter based methods ($\mathcal{C}_3$, $\mathcal{C}_J$) overestimated the convexity of the "L" shape (second last shape in the second row). On the other hand, those convexity measures give shapes that contain many narrow but deep intrusions (such as the first shape in the second row) much lower convexity scores than other measures. Note that the rankings by $\mathcal{C}_2$ and $\mathcal{C}_Q$ (which are both area measures based on a convex polygon) are similar. Also, $\mathcal{C}_R$ and $\mathcal{C}_{0.5}$ are similar – which is expected since they are based on similar principles.

measures	accuracy
$\mathcal{C}_2$, $\mathcal{C}_3$, $\mathcal{C}_J$	58.40%
$\mathcal{C}_{0.0}$	55.81%
$\mathcal{C}_{0.0}$, $\mathcal{C}_{prot}$	69.77%
$\mathcal{C}_{0.0}$, $\mathcal{C}_{prot}$, $\mathcal{C}_Q$	74.42%,
$\mathcal{C}_{0.0}$, $\mathcal{C}_{prot}$, $\mathcal{C}_Q$, $\mathcal{C}_3$	83.72%

Table 1: Classification accuracies for 43 desmids using the nearest neighbour classifier with Mahalanobis distances and leave-one-out cross validation.

A final demonstration of the convexity measures is given for the small scale classification task carried out in our previous paper on convexity [24]. Nine species of a type of algae called desmids were analysed. The data set contained 43 outlines, each containing 1000–10000 pixels, with 4–7 drawings for each species, and were taken from West and West's [25] flora. Some examples are shown in figure 8. As before, the outlines were preprocessed by performing polygonal approximation using Ramer's algorithm [16], and a threshold of 3.

The polygonal approximation is mainly intended to substantially reduce the number of polygon vertices, thereby speeding up the line intersection tests. A second consequence is that noise and fine detail is eliminated. If the polygonal approximation algorithm's threshold parameter is set higher to remove significant detail then the measured convexity value could be increased significantly. Note that increasing α for the $\mathcal{C}_\alpha$ measure has a similar effect, as shown in figure 9; also when the threshold is increased to 40 the ordering of the desmids becomes closer to $\mathcal{C}_{0.5}$ applied to a fine polygonal approximation (figure 8). The advantage of using a coarse polygonal approximation rather than applying $\mathcal{C}_\alpha$ with large α to a fine polygonal approximation is computational efficiency. However, the drawback is that the convexity measure becomes increasingly sensitive to the performance of the polygonal

approximation algorithm as alternative coarse polygonal approximations with similar fitting errors could vary substantially. For that reason we prefer to retain an accurate polygonal approximation, and modify α as required. Consequently, the $\mathcal{C}_\alpha$ measure will be insensitive to variations in the polygonal approximation. That is, alternative polygonal approximation algorithms, or modifications to the data (e.g. due to noise, rotation, quantisation) will all produce an output polygon with a small (e.g. Hausdorff) distance from the input, even if the selection of vertices in the outputs vary. This is demonstrated in figure 10 in which the different polygonal approximation algorithms have produced slight variations in output which are only apparent on close examination. A single parameter value was set for each algorithm, and so the numbers of vertices for each outline varies across the algorithm outputs. Nevertheless, the $\mathcal{C}$ values for each shape compared across the algorithms are fairly similar, as are the orderings.

Previously we showed that combining the three convexity measures $\mathcal{C}_2$, $\mathcal{C}_3$, and $\mathcal{C}_J$ achieved 58.40% classification accuracy. In the current paper classification was performed using a nearest neighbour classifier[2] with Mahalanobis distances, and leave-one-out accuracy values computed. To find a good set of convexity features sequential forward search was employed, in which all unused features are considered, and the feature, which combined with the current feature set maximises the classification rate, is added to the current feature set [12]. Table 1 shows that the best single convexity measure for the classification of desmids was $\mathcal{C}_{0.0}$. Furthermore, combining certain other convexity features, namely $\mathcal{C}_{prot}$,[3] $\mathcal{C}_Q$ and $\mathcal{C}_3$ was sufficient to substantially increase classification accuracy (combining further convexity measure resulted in a drop in accuracy). This demonstrates that although the different convexity measures are all designed to measure the same basic characteristic of shapes, they nonetheless actually capture different aspects (which was also implied by the differing invariances displayed in figure 6). Of course, classification accuracy could be increased still further by incorporating additional non-convexity shape measures.

[2]We found that the nearest neighbour classifier was preferable to our previous use of the OC1 decision tree [10] since the results of the latter depended on the ordering of the data. Given the data matrix of samples versus features, permuting either the rows or columns could affect the classification rate, although generally only by a small amount.

[3]$\mathcal{C}_{prot}$ is a measure of protrusion based on the same convex polygon from which the $\mathcal{C}_Q$ convexity measure is calculated [21]. Note, also that only the first stage of the optimisation process for determining the convex polygon was carried out (no vertex perturbation was carried out), as the subsequent refinement stage was shown to be not cost-effective [21].

6 Concluding Remarks

We have described a new shape convexity measure $\mathcal{C}_\alpha$ which varies from most other such measures in several ways:

- it incorporates both area based and boundary based information;
- it is sensitive to rotations and translations of both intrusions and protrusions to the shape;
- it incorporates a parameter α that determines its sensitive to boundary fluctuations.

Moreover, it is simple to implement. Although its computation is based on the Monte Carlo method, it is reasonably efficient to compute. Running time for a C implementation applied to a relatively complex polygon containing 200 edges after polygonal simplification takes about 5 seconds on a 2.0 GHz Pentium 4 using 100000 line segment tests. However, the number of tests (and consequently the running time) can be reduced by one or two orders of magnitude with only small changes to the measured convexity values. The polygonal approximation algorithm stage takes a negligible computation time: about 40 milliseconds for an input outline containing 4000 vertices.

References

[1] A. Bosson, G.C. Cawley, Y. Chan, and R. Harvey. Non-retrieval: Blocking pornographic images. In *Conf. Image and Video Retrieval*, volume 2383 of *Lecture Notes in Computer Science*, pages 50–60. Springer, 2002.

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

[3] D. Coeurjolly and R. Klette. A comparative evaluation of length estimators of digital curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(2):252–257, 2004.

[4] T.J. Deveau. Reducing the number of points in a plane curve representation. In *Proc. AutoCarto VII*, pages 152–160, 1985.

[5] A. Held and K. Abe. On approximate convexity. *Pattern Recognition Letters*, 15(6):611–618, 1994.

[6] R. Kakarala. Testing for convexity with Fourier descriptors. *Electronics Letters*, 34(14):1392–1393, 1998.

[7] R. Klette and A. Rosenfeld. *Digital Geometry*. Morgan Kaufmann, San Francisco, 2004.

[8] 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.

[9] R.R. Martin and P.L. Rosin. Turning shape decision problems into measures. *Int. J. Shape Modelling*, 10(1):83–113, 2004.

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

[11] R. Osada, T.A. Funkhouser, B. Chazelle, and D.P. Dobkin. Shape distributions. *ACM Trans. Graph.*, 21(4):807–832, 2002.

[12] Devijver P.A. and J.V. Kittler. *Pattern Recognition: A statistical approach*. Prentice-Hall, 1982.

[13] H. Pao, D. Geiger, and N. Rubin. Measuring convexity for figure/ground separation. In *Int. Conf. Computer Vision*, pages 948–955, 1999.

[14] H.K. Pao and D. Geiger. A continuous shape descriptor by orientation diffusion. In *Proc. Workshop on Energy Minimization Methods in Computer Vision and Pattern Recognition*, volume 2134, pages 544–559. LNCS, 2001.

[15] E. Rahtu, M. Salo, and J. Heikkilä. A new convexity measure based on a probabilistic interpretation of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(9):1501–1512, 2006.

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

[17] C. Ronse. A bibliography on digital and computational convexity (1961-1988). *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(2):181–190, 1989.

[18] A. Rosenfeld and E. Johnston. Angle detection on digital curves. *IEEE Transactions on Computers*, 22:875–878, 1973.

[19] P.L. Rosin. Techniques for assessing polygonal approximations of curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(6):659–666, 1997.

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

[21] P.L. Rosin and C.L. Mumford. A symmetric convexity measure. *Computer Vision and Image Understanding*, 103(2):101–111, 2006.

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

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

[24] J. Žunić and P.L. Rosin. A new convexity measurement for polygons. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(7):923–934, 2004.

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

[26] C.M. Williams. An efficient algorithm for the piecewise linear approximation of planar curves. *Computer Graphics and Image Processing*, 8:286–293, 1978.

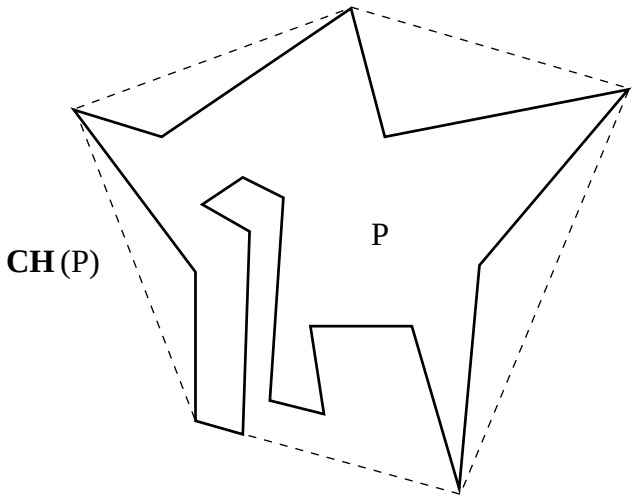

Figure 1: Non-convex polygon P and its convex hull $\mathbf{CH}(P)$ (dashed line).

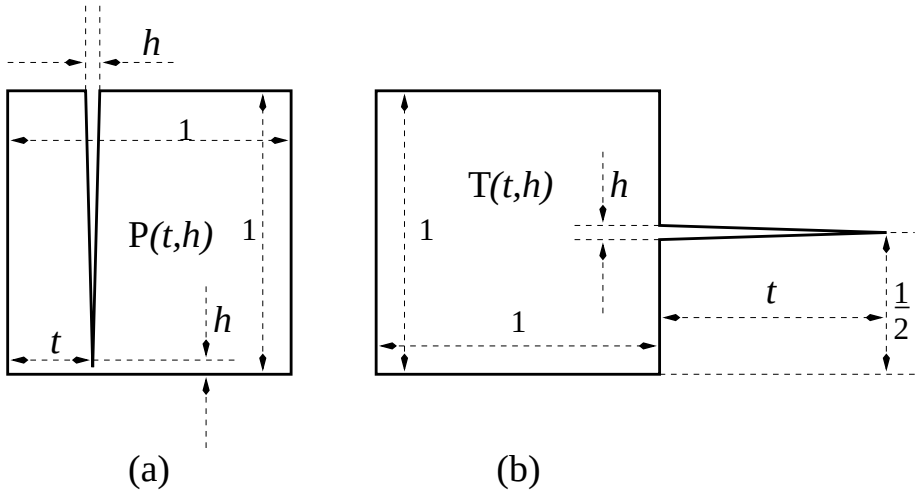

Figure 2: The measure $\mathcal{C}_1$ gives an acceptable $\lim_{h\to 0}\mathcal{C}_1(P(t,h)) = 1-2t+2t^2$ and unacceptable $\lim_{h\to 0}\mathcal{C}_1(T(t,h)) = 0$ (for any fixed t).The measure $\mathcal{C}_2$ gives an unacceptable $\lim_{h\to 0}\mathcal{C}_2(P(t,h)) = 1$ and acceptable $\lim_{h\to 0}\mathcal{C}_2(T(t,h)) = \frac{2}{2+t}$ (for any t).The measure $\mathcal{C}_3$ gives $\lim_{h\to 0}\mathcal{C}_3(P(t,h)) = 2/3$ and $\lim_{h\to 0}\mathcal{C}_1(T(t,h)) = (3+\sqrt{4t^2+1})/(4+2t)$ (both acceptable).

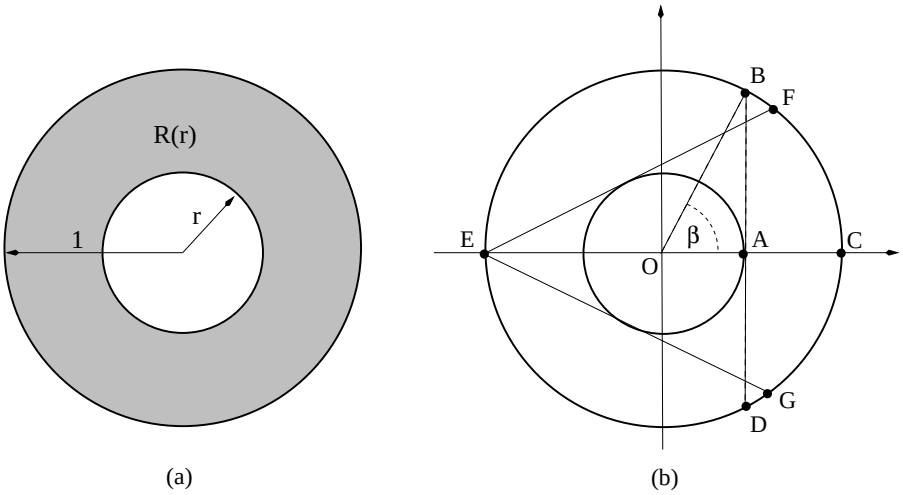

Figure 3: (a) When r varies through $(0,1)$ the measured convexity $\mathcal{C}(R(r))$ varies through the same interval, as well. (b) Notations are given in order to make an easier derivation of $\mathcal{C}(R(r))$.

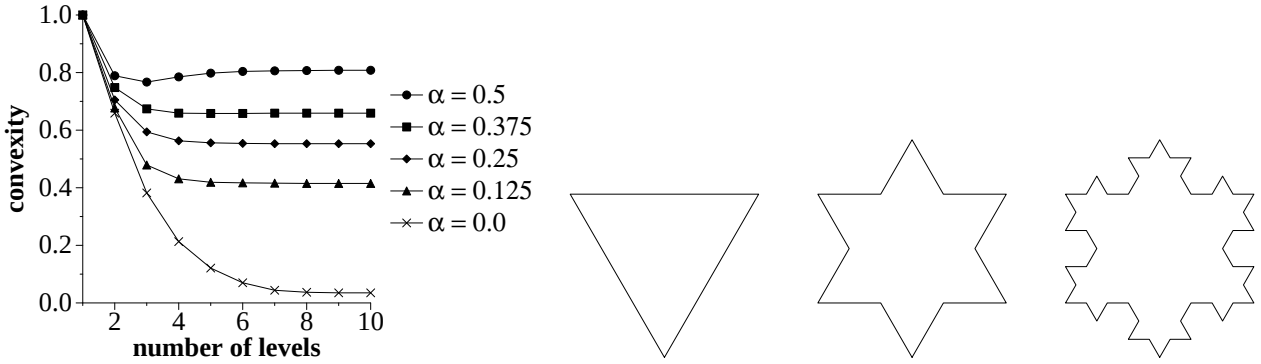

Figure 4: Convexity values $\mathcal{C}$ for a Koch snowflake. As the number of levels of detail is increased $\mathcal{C}_{0.0}$ approaches a value close to zero. When the tested line segments are shrunk then the measure becomes less sensitive (although not totally insensitive) to minor deviations from convexity. Thus, in the extreme case where the lines are shrunk to their midpoints, the $\mathcal{C}_{0.5}$ value drops initially after the first level (i.e. the convex triangle), but thereafter its value does not vary much. The Koch snowflake is shown generated at each of the first three levels of detail.

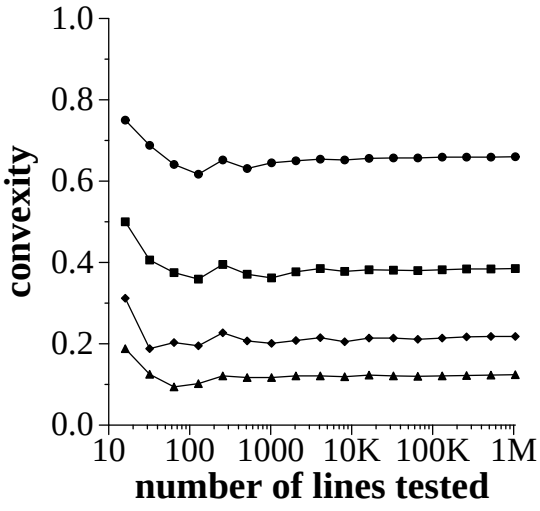

Figure 5: Graph showing convergence of the convexity measure $\mathcal{C}_{0.0}$ for the Koch snowflake with 2, 3, 4, and 5 levels of detail (given by the plotted lines in descending order).

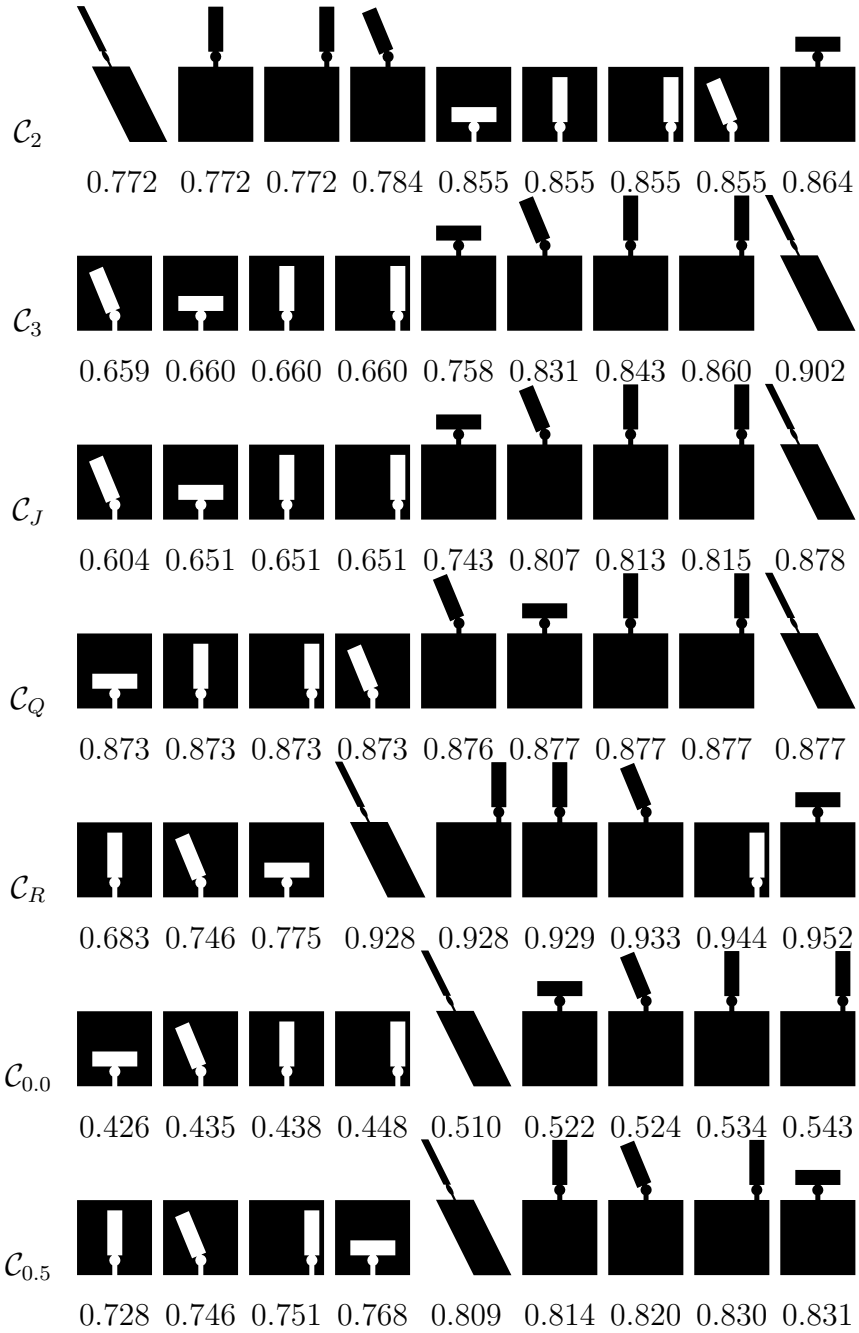

Figure 6: Convexity values for simple synthetic shapes, showing the effects of rotation and translation of intrusions/protrusions, and global skew.

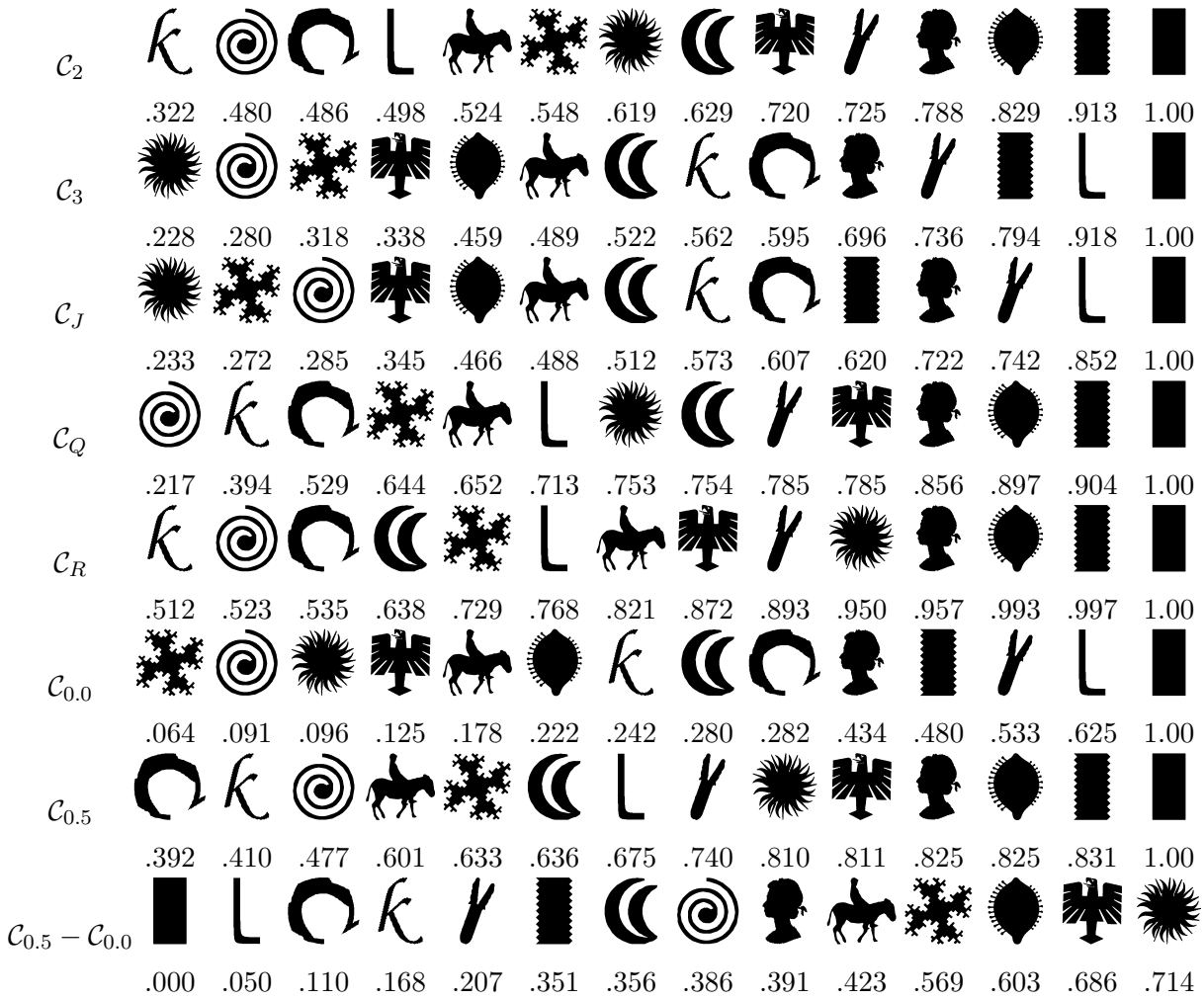

Figure 7: A variety of shapes ranked into ascending order by several convexity measures.

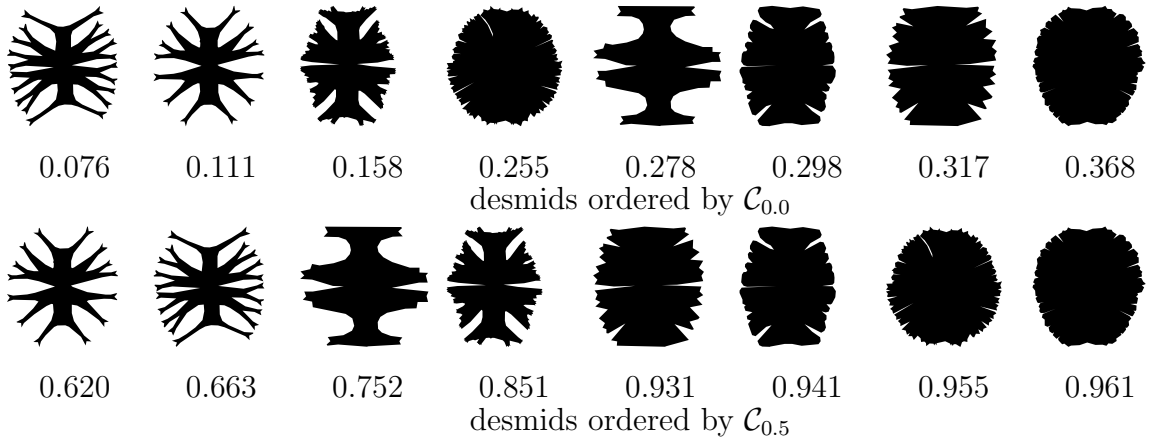

Figure 8: A sample of the 43 desmids (taxon *Micrasterias*) ordered by the new convexity measures. Although $\mathcal{C}_{0.0}$ is very sensitive to minor boundary fluctuations, $\mathcal{C}_{0.5}$ is relatively insensitive.

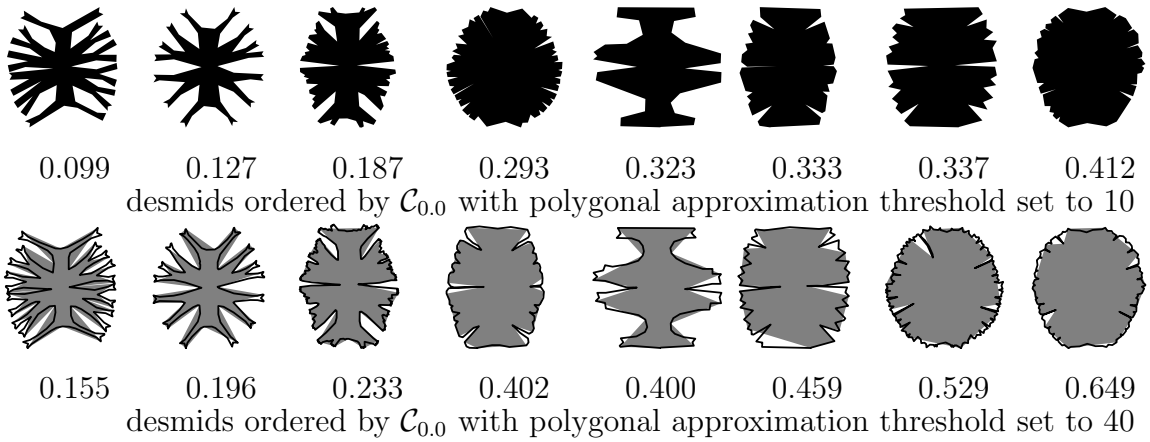

Figure 9: The effect of making coarser polygonal approximations is to eliminate many of the shapes' concavities, increasing the measured $\mathcal{C}_{0.0}$ values. At the high threshold used in the second row the polygonal approximation becomes very crude, as shown by the overlay of the true outline.

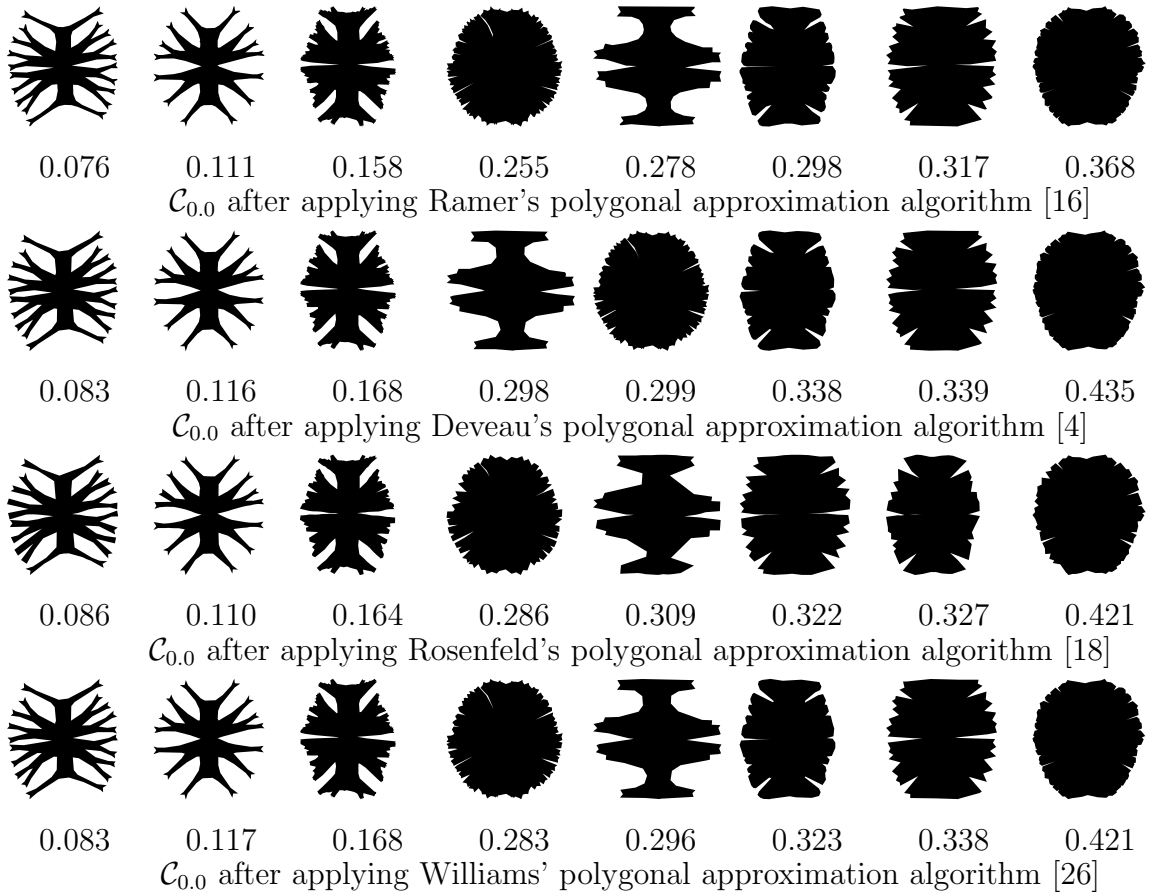

Figure 10: The desmid outlines are preprocessed by different polygonal approximation algorithms before ordering by $\mathcal{C}_{0.0}$.