A New Convexity Measure Based on a Probabilistic Interpretation of Images

Esa Rahtu, Mikko Salo, and Janne Heikkilä

Abstract

In this article we present a novel convexity measure for object shape analysis. The proposed method is based on the idea of generating pairs of points from a set, and measuring the probability that a point dividing the corresponding line segments belongs to the same set. The measure is directly applicable to image functions representing shapes, and also to gray-scale images which approximate image binarizations. The approach introduced gives rise to a variety of convexity measures, which makes it possible to obtain more information about the object shape. The proposed measure turns out to be easy to implement using the Fast Fourier Transform and we will consider this in detail. Finally, we illustrate the behavior of our measure in different situations and compare it to other similar ones.

Index Terms

Shape analysis, object classification, affine invariance

A New Convexity Measure Based on a Probabilistic Interpretation of Images

I. Introduction

Shape analysis is one of the major fields of computer vision research. Techniques of shape analysis are widely applied, for example in classification, calibration, image registration, segmentation, etc. There is also plenty of demand in other disciplines which use pattern recognition techniques, like biology, mathematical morphology, medical image analysis [1], and integral geometry, [2], [3]. Common features which are used to describe shape include compactness, circularity, rectangularity and convexity. Here, we concentrate on measuring object convexity. We recall that an object is said to be convex if the line segment between any two points in the object belongs to the same object. For more information about convexity see [4].

Convexity in image processing has been studied for quite some time and convexity measures and their applications are discussed in several publications, like [5], [6], [7], [8], [9], [10]. In this paper, we present a novel framework for constructing convexity measures, based on probabilistic ideas. This technique was originally introduced in [11], but here we define more general measures and consider their properties carefully, comparing them with other similar methods. However, when comparing these techniques, one should keep in mind that desirable properties of a convexity measure may vary according to the specific application. For example, in some cases, strong reactivity to already slight changes in the shape is required, but in other applications, we may need strong robustness against such alterations. It is also reasonable to require some basic properties of the convexity measures to ensure that they behave well and produce sensible results. We begin by introducing these properties in a similar manner to that in [9].

Definition 1: A convexity measure should have the following basic properties:

1) *The convexity measure is a number from* $(0, 1]$.
2) *The convexity measure of a given shape is equal to* 1 *if and only if the shape is convex.*
3) *There are shapes whose convexity measure is arbitrarily close to* 0*, implying that there is*

no gap between 0 *and the minimal possible value of the measure.*

4) *The convexity measure is invariant under appropriate geometric transformations of the shape.*

We continue by reviewing some previously proposed convexity measures, which are divided into two primary groups following [9]. This partition is made in order to illustrate the main characteristics of the methods, and to give a guideline for choosing the proper measure (or measures) for a particular application.

Definition 2 (Area based convexity measures): For a given planar shape K*, where* $K \subseteq \mathbf{R}^2$ *is a compact set, we define*

$$C_{ch}(K) = \frac{|K|}{|\mathrm{ch}(K)|},$$
$$C_{mcs}(K) = \frac{|\mathrm{mcs}(K)|}{|K|},$$

where $\mathrm{ch}(K)$ *is the convex hull of* K*,* $|K|$ *is the area of a set* K*, and* $|\mathrm{mcs}(K)|$ *is the supremum of the areas of convex subsets of* K*.*

Further, let X *and* Y *be independent random variables drawn uniformly from the set* K*, and let* $[x, y]$ *be the line segment between* x *and* y*. We define a third convexity measure as*

$$C_{ls}(K) = P([X, Y] \subseteq K),$$

where $P(A)$ *denotes the probability of the event* A*.*

In general, the measures in Definition 2 can be thought to be area-based, since their values depend on the areas of different parts of the object. Consequently such measures are usually tolerant with respect to small defects in the objects, which can be caused for example by noise or insufficient segmentation. From these measures C_{ch} and C_{ls} are particularly interesting. The measure C_{ch} is practical and robust, appearing frequently in the literature and applications. Also, C_{ls} is interesting here because it is based on probabilistic ideas similar to our new method. However, the main problem with this measure is that C_{ls} is difficult to compute, even if the examined sets are polygons [9]. For this reason, the actual validation of this measure usually has to be done by sampling. The measure C_{mcs} is not very practical, because the $|\mathrm{mcs}(K)|$ is not straightforward to compute.

Sometimes too much robustness is undesirable and we want to have strong effects on the value of the measure, even if the object changes only slightly. There are also some theoretical examples [9] where the area based measures are not well suited to detecting convexity. In these cases it would be better to use so called boundary based measures. We define two of them, for simplicity just for polygons, as follows:

Definition 3 (Boundary based convexity measures): For a given planar shape K, where $K \subseteq \mathbf{R}^2$ is a compact and connected polygon, we define

$$C_{chp}(K) = \frac{\mathrm{Per_2(ch(K))}}{\mathrm{Per_2(K)}},$$

where $\mathrm{Per_2(K)}$ is the usual L_2 perimeter, see [9], of K and $\mathrm{ch}(K)$ is the convex hull of K. Further, let $R(K, \alpha)$ denote the minimal rectangle with edges parallel to the coordinate axes which contains the polygon K rotated by angle α. We now define

$$C_{poly}(K) = \min_{\alpha \in [0, 2\pi]} \frac{\mathrm{Per_2(R(K, \alpha))}}{\mathrm{Per_1(K)}},$$

where $\mathrm{Per_1(K)}$ is the L_1 perimeter, for definition see [9], of K.

In this paper, we present a new method for constructing convexity measures which is based on similar probabilistic ideas as C_{ls}, but uses points rather than lines. It turns out that there is an elegant and efficient way of evaluating the measure using the Fourier transform. There is also the possibility of applying the measure directly to gray-scale images, which eliminates the need to threshold a slightly noisy image to a strictly binary one. In the presented framework, our measure belongs to the area-based class, and it will be shown to fulfill all the requirements given in Definition 1. We begin with an introduction to the new measure and its properties, followed by a discussion of the actual implementation. Finally, we illustrate the behavior of the measure in several experiments, comparing it to other similar ones.

II. New convexity measure

In this section, we will define our new measure and illustrate its properties. Before continuing, we recall the relevant definitions and fix some assumptions on the examined sets.

Definition 4: A set $A \subseteq \mathbf{R}^2$ is called convex if for any points $x, y \in A$, the line segment $\{(1-t)x + ty\,;\, 0 \leq t \leq 1\}$ is contained in A. Equivalently, A is convex if and only if for

any points $x_1, \ldots, x_n \in A$*, any convex combination* $a_1x_1 + \ldots + a_nx_n$*, where* $a_j \geq 0$ *and* $a_1 + \ldots + a_n = 1$*, is in* A*.*

Let K denote the set to be studied. We restrict our attention to sets K which belong to

$$\mathscr{K} = \{K \subseteq \mathbf{R}^2 \,;\, K \text{ is compact, nonempty, and } K = \overline{\text{int}(K)}\},$$

where $\text{int}(A)$ is the interior and $\overline{A}$ is the closure of a set $A \subseteq \mathbf{R}^2$. With this restriction, we only simplify the discussion by excluding very large sets and sets of measure zero. In image processing, this results in no loss of generality, since digital images always satisfy these conditions.

A. Definition of the convexity measure

The new convexity measure is based on the simple idea of generating pairs of points from a set K, and then checking if certain points on the corresponding line segments belong to K. Here is the precise definition.

Definition 5: For $0 < \alpha < 1$ *and* $K \in \mathscr{K}$*, we define a convexity measure*

$$C_\alpha(K) = P(\alpha X + (1-\alpha)Y \in K),$$

where X *and* Y *are independent random variables drawn uniformly from* K*.*

In other words, $C_\alpha(K)$ measures the probability that a point which lies on the line segment between two random points from the set K, is also contained in this set. The parameter α determines the location of the point on this line segment. By varying α, we obtain infinitely many different measures. Definition 5 is somewhat similar to the definition of C_{ls}, but the major difference is that here we measure probabilities of points instead of whole line segments. This results in two substantial advantages. The first is that by varying α we will have a number of convexity measures. Secondly, we will show that there is a computationally efficient way of evaluating C_α, unlike in the case of C_{ls} which is usually difficult to compute. Definition 5 also gives rise to several generalizations of this measure, which we consider in more detail later. Note that C_α has the symmetry $C_\alpha(K) = C_{1-\alpha}(K)$, and hence $0 < \alpha \leq 1/2$ will give all the information.

B. Connection with MSA

At this point, we observe that C_α has a strong connection to the image transform called Multiscale Autoconvolution (MSA) [12] [13]. Let us now consider this connection in more detail.

Let $f \geq 0$ be an image intensity function in $L^1(\mathbf{R}^2) \cap L^2(\mathbf{R}^2)$, where $L^1(\mathbf{R}^2)$ is the space of absolutely integrable functions with the norm $\|f\|_{L^1} = \int_{\mathbf{R}^2} |f(x)|\, dx$ and $L^2(\mathbf{R}^2)$ is the space of square integrable functions with the norm $\|f\|_{L^2} = \left(\int_{\mathbf{R}^2} |f(x)|^2\, dx\right)^{1/2}$ [14]. We define a random variable

$$U_{\alpha,\beta} = \alpha X_1 + \beta X_2 + \gamma X_0,$$

where $\gamma = 1 - \alpha - \beta$ and (X_0, X_1, X_2) are independent random variables with values in $\mathbf{R}^2$ so that $P(X_j = x_j) = \frac{1}{\|f\|_{L^1}} f(x_j)$ (for notational convenience, we write $P(X = x)$ for the value of the probability density function of X at x). The MSA transform Mf of f is defined for $\alpha, \beta \in \mathbf{R}$ as the expected value of $f(U_{\alpha,\beta})$,

$$Mf(\alpha, \beta) = E[f(U_{\alpha,\beta})].$$

We now consider functions f which correspond to sets $K \in \mathscr{K}$. The characteristic function χ_K of the set K is defined as

$$\chi_K(x) = \begin{cases} 1 & \text{if } x \in K, \\ 0 & \text{otherwise.} \end{cases}$$

Now by setting $f = \chi_K$ in the MSA transform, we have

$$M\chi_K(\alpha, \beta) = \int_K P(U_{\alpha,\beta} = u)\, du = P(U_{\alpha,\beta} \in K). \tag{1}$$

In other words, $M\chi_K(\alpha, \beta)$ is the probability that $\alpha X_1 + \beta X_2 + \gamma X_0$ lies in K when the X_j are drawn uniformly from K. Since $\alpha + \beta + \gamma = 1$, one recognizes that if $\alpha, \beta, \gamma \geq 0$ then $M\chi_K(\alpha, \beta)$ is the probability that a convex combination of X_0, X_1, and X_2 lies in K. We now easily see that the measure C_α corresponds to the special case of this convex combination where $\beta = 1 - \alpha$. Thus $C_\alpha(K) = M\chi_K(\alpha, 1 - \alpha)$.

C. Explicit expressions for C_α

We proceed to derive different explicit expressions for the measure C_α. As it was noted that C_α is a special case of the MSA transform, we follow similar ideas to those in the derivation of MSA [13].

Let $f = \chi_K$ be the characteristic function of K as given above. Then for $0 < \alpha < 1$ we define a random variable

$$U_\alpha = \alpha X_1 + (1-\alpha)X_2,$$

where X_1 and X_2 are independent random variables with values in $\mathbf{R}^2$ so that $P(X_j = x_j) = \frac{1}{|K|} f(x_j)$. Now it can be easily shown that U_α has a probability density function given by

$$P(U_\alpha = u) = \frac{1}{|K|^2}(f_\alpha * f_{1-\alpha})(u), \tag{2}$$

where $f_a(x) = a^{-2} f(x/a)$ and $*$ denotes convolution. Using this notation, the value of C_α can be expressed as

$$C_\alpha(K) = P(\alpha X + (1-\alpha)Y \in K) = P(U_\alpha \in K) = \int_K P(U_\alpha = u)\, du. \tag{3}$$

Substituting (2) in (3) we can write the measure C_α in terms of the probability density function as

$$\begin{aligned} C_\alpha(K) &= \int_K \frac{1}{|K|^2}(f_\alpha * f_{1-\alpha})(u)\, du \\ &= \frac{1}{|K|^2}\int_{\mathbf{R}^2} f(u)(f_\alpha * f_{1-\alpha})(u)\, du \\ &= \frac{1}{|K|^2}\frac{1}{\alpha^2(1-\alpha)^2}\int_{\mathbf{R}^2}\int_{\mathbf{R}^2} f(u) f(\frac{u-x}{1-\alpha}) f(\frac{x}{\alpha})\, dx\, du. \end{aligned} \tag{4}$$

The double integral in (4) is computationally expensive to evaluate for large images, and it is important that one can use the Fourier transform with the convolution theorem and the Plancherel formula, $\int_{\mathbf{R}^2} f\bar{g} = \int_{\mathbf{R}^2} \hat{f}\bar{\hat{g}}$ (see [14]), to write (4) as a single integral,

$$C_\alpha(K) = \frac{1}{|K|^2}\int_{\mathbf{R}^2} \hat{f}(-\xi)\hat{f}(\alpha\xi)\hat{f}((1-\alpha)\xi)\, d\xi, \tag{5}$$

where $\hat{f}(\xi) = \int_{\mathbf{R}^2} e^{-2\pi i x\cdot\xi} f(x)\, dx$ is the Fourier transform of the image function $f = \chi_K$.

D. Properties of C_α

Here, we shall consider some important properties of the proposed measures, including the general requirements for convexity measures given in Definition 1.

Property 1: C_α satisfies the four requirements for a convexity measure:

1) $0 < C_\alpha(K) \leq 1$ for any $K \in \mathscr{K}$.
2) $C_\alpha(K) = 1$ for $K \in \mathscr{K}$ if and only if K is convex.
3) $\inf_{K \in \mathscr{K}} C_\alpha(K) = 0$.
4) The measure C_α is affine invariant: if $\mathscr{A}$ is an affine transformation, then $C_\alpha(\mathscr{A}(K)) = C_\alpha(K)$ for any $K \in \mathscr{K}$.

Proof: The properties are quite natural, and we will give here brief explanations of how to prove them. The details are in Appendix I.

1) Since $C_\alpha(K)$ is the probability of an event it is always ≤ 1. Any set $K \in \mathscr{K}$ contains a disk, so this probability is always > 0.
2) It is well known that K is convex if and only if the midpoint of any two points of K belongs to K. If $K \in \mathscr{K}$, then we also know that K is convex if and only if the midpoint of two points of K belongs to K with probability one. This is true for other points than the midpoint as well, so we get that K is convex if and only if $C_\alpha(K) = 1$.
3) Let $K = K_\delta = \{x \in \mathbf{R}^2 ; 1 - \delta \leq |x| \leq 1\}$ be a thin annulus. Then $C_\alpha(K_\delta) \to 0$ as $\delta \to 0$, showing that these thin annuli have an arbitrarily small convexity measure.
4) It is geometrically evident that the probability $C_\alpha(K)$ stays unchanged in an affine transformation of K.

■

Next we turn our attention to the continuity properties of the measure. First we consider the case where a binary image $f = \chi_K$ is disturbed with nonbinary noise h. We make the assumptions that $h \in L^1 \cap L^\infty$, $|h|$ is bounded by a constant R, and $\|h\|_{L^1} \leq \frac{1}{2}|K|$. Here L^∞ is the space of bounded functions with the norm $\|f\|_{L^\infty} = \sup_{x \in \mathbf{R}^2} |f(x)|$ [14]. In practice $\|h\|_{L^1}$ is much smaller than $|K|$, so the nonbinary image $f + h$ is close to the original binary image. We state the result in terms of the MSA transform $Mf(\alpha, \beta)$, which is the natural analog of C_α for nonbinary images (recall that $C_\alpha(K) = M\chi_K(\alpha, 1 - \alpha)$).

Property 2: Let $f = \chi_K$, $K \in \mathscr{K}$, and let $h \in L^1(\mathbf{R}^2) \cap L^\infty(\mathbf{R}^2)$ satisfy $\|h\|_{L^\infty} \leq R$ and

$\|h\|_{L^1} \leq \frac{1}{2}|K|$, where $R \geq 1$. Then there exists an absolute constant C_0 so that

$$|M(f+h)(\alpha,\beta) - Mf(\alpha,\beta)| \leq C_0 R \frac{\|h\|_{L^1}}{|K|}$$

for all $\alpha, \beta \in \mathbf{R}$.

The proof of Property 2 is given in Appendix II. This property justifies the idea that one may compute the convexity measure of a gray-scale image without first converting it to a binary one. The assumption is that the gray-scale image is a slightly noise disturbed model of a binary image, and the result shows that the convexity measures of the original binary image and the gray-scale image are almost the same. Note that for nonbinary images the convexity measure may be larger than one.

In practice one may have digital images of shapes where the boundary pixels have gray-scale values between 0 and 1, and in this case one may interpret these nonbinary values as the probability of a pixel belonging to the shape. Applying the convexity measure directly to this image removes the need for limiting the image to a strictly binary one.

The second version of the continuity property is given for binary images. If $A, B \subseteq \mathbf{R}^2$ we define the symmetric difference

$$A \triangle B = (A \setminus B) \cup (B \setminus A).$$

Let $K \in \mathscr{K}$, and suppose $K' \in \mathscr{K}$ satisfies $|K' \triangle K| \leq \frac{1}{2}|K|$. We think of K' as a disturbed version of K. The result states that the convexity measures of K and K' differ at most by $C_0 \frac{|K' \triangle K|}{|K|}$.

Property 3: Let $K, K' \in \mathscr{K}$ and suppose $|K' \triangle K| \leq \frac{1}{2}|K|$. Then

$$|C_\alpha(K') - C_\alpha(K)| \leq C_0 \frac{|K' \triangle K|}{|K|}$$

where C_0 is an absolute constant.

The proof of this property is given in Appendix III. Property 3 states that the value of C_α depends continuously on the measured set, and it gives an upper bound for the change in the value of the measure for a given distortion scale in K. This continuity is important because it ensures that sets differing only slightly will also have C_α values close to each other.

We conclude the theoretical discussion by giving inequalities which relate C_α to other area-based measures.

Property 4: If $0 < \alpha < 1$ and $K \in \mathscr{K}$, then

$$C_{mcs}(K)^2 \leq C_{\text{ls}}(K) \leq C_\alpha(K).$$

We will prove this in Appendix IV. This property gives some insight to the relations between the absolute convexity values of different measures. However, it does not give any information about the order in which the different measures rank objects.

E. Generalizations of C_α

When looking at (1) we observed that in the case where $\alpha, \beta, \gamma \geq 0$, the MSA transform of χ_K becomes the probability that a convex combination of three points from the set K also belongs to K. It was further noticed that if we take this convex combination so that it contains only two points, i.e. $\beta = 1-\alpha$ so that $\gamma = 0$, we have the C_α measure. However, using a convex combination of three or more points leads to a generalization of the C_α measure as follows:

Definition 6: For $K \in \mathscr{K}$ and $\alpha_1, \ldots, \alpha_n \geq 0$ such that $\alpha_1 + \ldots + \alpha_n = 1$, we define a convexity measure

$$C_{\alpha_1,\ldots,\alpha_n}(K) = P(\alpha_1 X_1 + \ldots + \alpha_n X_n \in K)$$

where $X_1, \ldots, X_n$ are independent random variables drawn uniformly from K.

Similarly to C_α it can be shown that also $C_{\alpha_1,\ldots,\alpha_n}$ can be computed through the Fourier transform of $f = \chi_K$ as

$$C_{\alpha_1,\ldots,\alpha_n}(K) = \frac{1}{|K|^n} \int_{\mathbf{R}^2} \hat{f}(-\xi)\hat{f}(\alpha_1\xi)\cdots\hat{f}(\alpha_n\xi)\, d\xi.$$

It is also possible to combine C_α with different α values to form new convexity measures. To be more precise, we present the following definition:

Definition 7: Let μ be a probability measure on the interval $(0,1)$. For $K \in \mathscr{K}$ we define a convexity measure

$$C_\mu(K) = \int_{(0,1)} C_\alpha(K)\, d\mu(\alpha).$$

Due to the symmetry $C_\alpha(K) = C_{1-\alpha}(K)$ it is enough to consider measures μ which are symmetric with respect to $\alpha = 1/2$. For such a measure one has

$$C_\mu(K) = 2\int_{(0,1/2)} C_\alpha(K)\, d\mu(\alpha) + C_{1/2}(K)\mu(\{1/2\}).$$

There are two main examples of suitable measures μ:

1) If $\mu = \sum_{j=1}^{k} a_j \delta_{\alpha_j}$ where $a_j \geq 0$, $\sum_{j=1}^{k} a_j = 1$, $0 < \alpha_1 < \ldots < \alpha_k < 1$, and δ_α is the Dirac measure at α, then
$$C_\mu(K) = \sum_{j=1}^{k} a_j C_{\alpha_j}(K).$$
In the special case $\mu = \delta_\alpha$ we obtain the original convexity measure C_α.
2) If $d\mu(\alpha) = w(\alpha)\, d\alpha$ where $w \geq 0$ and $\int_0^1 w(\alpha)\, d\alpha = 1$, then
$$C_\mu(K) = \int_0^1 C_\alpha(K) w(\alpha)\, d\alpha.$$

Thus $C_\mu(K)$ may be thought of as a weighted average, or integral average, of the $C_\alpha(K)$. The possibility of using different measures μ gives additional flexibility to the construction of convexity measures. We note that the properties derived for C_α in Section II-D are also valid for C_μ.

It is also straightforward to extend C_α and C_μ to cover sets in $\mathbf{R}^n$ where $n \geq 3$. This requires no changes in Definition 5, as long as we consider sets K belonging to

$$\mathscr{K} = \{K \subseteq \mathbf{R}^n\,;\, K \text{ is compact, nonempty, and } K = \overline{\operatorname{int}(K)}\}.$$

We also note that in this case the the expression in terms of the Fourier transform of $f = \chi_K$ extends similarly to

$$C_\alpha(K) = \frac{1}{|K|^2} \int_{\mathbf{R}^n} \hat{f}(-\xi)\hat{f}(\alpha\xi)\hat{f}((1-\alpha)\xi)\, d\xi,$$

where $\hat{f}(\xi) = \int_{\mathbf{R}^n} e^{-2\pi i x\cdot\xi} f(x)\, dx$. Such measures can be used to analyze the shape of three-dimensional objects, but as the main subject of this article is the two-dimensional case we will not discuss this in more detail here.

III. Implementation

For the implementation of C_α we have basically two possibilities. We can simply approximate the probability in Definition 5 by taking random pairs of points from the set K and testing whether the dividing point determined by α belongs to K. The other possibility is to evaluate the Fourier transform based expression (5). The first approach is straightforward to implement, and the only parameter is the number of pairs of points used in the approximation. The sampling, however, becomes more time consuming if we also allow gray-scale images approximating the binary one because in this case the examined lines are selected with different probabilities.

In the second approach, we need to discretize the integral (5) resulting in

$$C_\alpha(K) = \frac{1}{N_1 N_2}\frac{1}{|K|^2}\sum_{i=0}^{N_1N_2-1} \hat{f}(-w_i)\hat{f}(\alpha w_i)\hat{f}((1-\alpha)w_i), \tag{6}$$

where $f = \chi_K$ and w_i are $N_1 \times N_2$ points in a rectangular grid in $\mathbf{R}^2$. We may think of $\hat{f}$ as the discrete Fourier transform of a discrete function f, i.e. f is an $N_1 \times N_2$ matrix which represents the digital image. Formula (6) contains scaled versions of the continuous Fourier transform $\hat{f}$. We know that $\hat{f}(aw) = \widehat{f_a}(w)$ where $f_a(x) = \frac{1}{a^2}f(\frac{x}{a})$, so we may write (6) equivalently as

$$C_\alpha(K) = \frac{1}{N_1 N_2}\frac{1}{|K|^2}\sum_{i=0}^{N_1N_2-1} \hat{f}(-w_i)\widehat{f_\alpha}(w_i)\widehat{f_{1-\alpha}}(w_i). \tag{7}$$

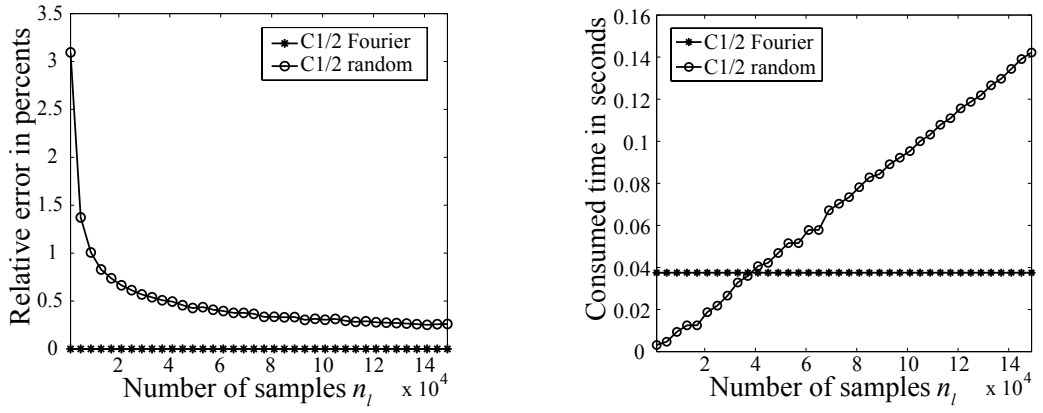
In the experiments, we mainly use the measure $C_{1/2}$, and in this case when $\alpha = 1/2$ expression (7) simplifies further to:

$$C_{1/2}(K) = \frac{1}{N_1 N_2}\frac{1}{|K|^2}\sum_{i=0}^{N_1N_2-1} \hat{f}(-w_i)\widehat{f_{1/2}}(w_i)^2. \tag{8}$$

Implementing (7) or (8) is quite straightforward, apart from the fact that we need to scale the original image f. We recall that since $0 < \alpha < 1$ there is only need for image decimation. In our implementation, this was done by dividing the image into equal sized regions and then summing the elements in them to form one element in the decimated image. This method was chosen because, in the probability sense, it preserves statistical relations between each part of the image. One should also keep in mind that the Fourier transform length must be long enough to avoid the wrap-around error, so if the original image f is an $M_1 \times M_2$ matrix, the transform length N_i must satisfy $N_i \geq (|\alpha| + |\beta| + |\gamma|)M_i - 2$. The Matlab algorithms we implemented and used to compute the C_α values can be retrieved from the URL: http://www.ee.oulu.fi/research/imag/msa/. As an example we give the full Matlab program for computing $C_{1/2}$ in Appendix V.

IV. Computational complexity

We next turn our attention to the computational complexity of convexity measures. Considering first the two implementations for C_α, it turns out that computing the measure using random sampling has asymptotic complexity $O(n_l)$ and the Fourier approach has the complexity $O(N^2 \log_2 N)$, where n_l is the number of randomly chosen pairs of points, and N is the side length of the image. The proportional efficiency between these implementations thus depends on the chosen n_l, which again determines the accuracy of the random sampling approach.



(a) (b)

Fig. 1. (a) Mean relative error and (b) computation time with different numbers of samples.

To get some insight into this matter, we performed a simple experiment involving Matlab implementations of both approaches. We had binary images I of size 240×240, consisting of three rectangular areas of pixels with value 1 on a background having pixel values 0. Each time image I was generated the side lengths of these rectangles were randomly taken from interval $[10, 60]$ and their relative positions were set so far from each other that in any case when two points are taken from different rectangles their midpoint lies in the background. For this set one can directly see that $C_{1/2}$ has the value

$$C_{1/2} = (\frac{A_1}{A})^2 + (\frac{A_2}{A})^2 + (\frac{A_3}{A})^2,$$

where A_1, A_2, A_3 are the areas of the three rectangles and $A = A_1 + A_2 + A_3$. We measured both computation time and relative error from the theoretical value, defined as $d = abs(\text{correct value} - \text{estimate})/(\text{correct value})$, with different numbers of sampling points n_l and averaging over 1000 samples of I. Figure 1 shows the results, and we can observe that at the point where both approaches work equally fast, the Fourier transform based evaluation gives clearly more accurate results. The point of equal computation time differs, of course, if we take a larger image than 240×240. However even by taking 140 000 samples, which was separately measured to correspond to equal computation time for a 600×600 image, we do not achieve a similar accuracy to the Fourier based approach. Taking also into account that the Fourier transform is easy to implement and it can be used directly for gray-scale approximations of binary sets, we

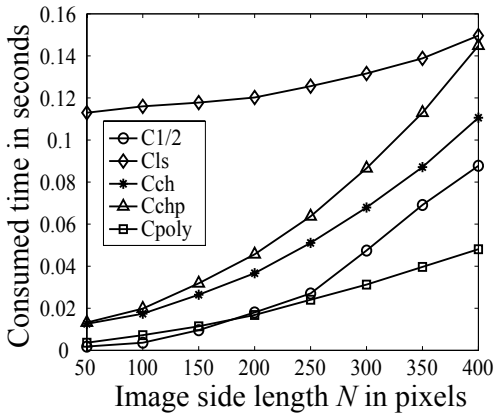


Fig. 2. Measured computation times with different convexity measures and image sizes.

saw no reason to use random sampling evaluation in the later experiments in this paper.

Considering the other convexity measures, the C_{ch} and C_{chp} have the same complexity as determining the convex hull of an object which is $O(n \log h)$ where n is the number of nonzero pixels in the examined image and h is the number of vertices of the convex hull [16]. According to [9], C_{poly} has a complexity of $O(n_p^2)$, where n_p is the number of vertices of the polygon. Finally the random sampling evaluation of C_{ls} results in the same complexity $O(n_l)$ as the random sampling approach of the C_α. However the computation is considerably slower than in the case of C_α, because it requires evaluation at a number of points on the examined line. The actual relative performances are not easily determined from the asymptotic complexities, especially when these depend on different properties of the measured sets. To get a rough idea, we ran an experiment where we measured the mean time of evaluating the convexity measures of 26 binary images, of size $N \times N$, of letters from A to Z. We illustrate the computation times with different image sizes N in Figure 2. For the C_{ls} measure, we used a random sampling approach with 10 000 samples. Looking at the results, it can be observed that computation times for all the measures are roughly in the same range, except for the C_{ls} which is clearly slowest at small image sizes. The exact performance depends on the considered sets as well as the implementations, but altogether the measures are quite fast to compute even for moderately large images.

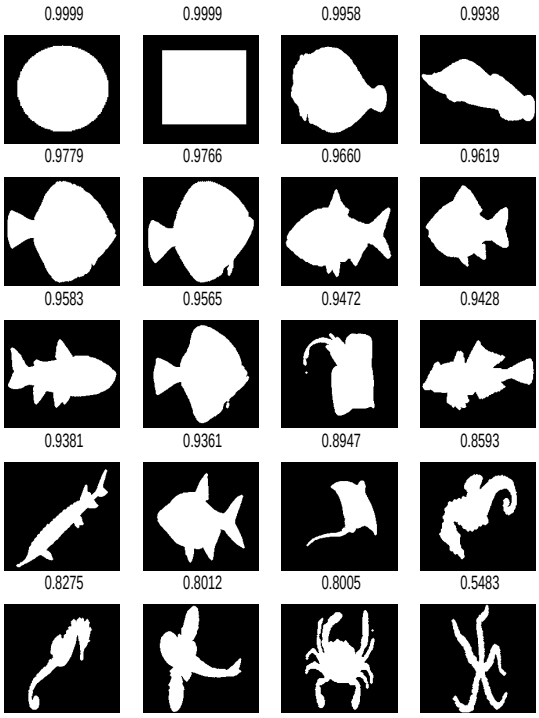


Fig. 3. 20 objects ranked according to their convexity. $C_{1/2}$ values are shown at the top of each image.

V. EXPERIMENTS

In this section, we illustrate the properties of our measure and compare it with the other measures defined in Section I. However, as already discussed, it should be kept in mind that the optimal behavior of a measure depends heavily on the application, and our experiments consider situations in which our measure could be useful. Also, in order to achieve a fairer comparison, we made a polygon approximation before applying C_{chp} or C_{poly} and used 50 000 samples when evaluating C_{ls}. We begin the experiments by a short qualitative evaluation. We took a set of 20 different shapes and classified them according to their convexity values computed using $C_{1/2}$. We illustrate the results in Figure 3, where the object having highest convexity is on the upper left corner. The point of this experiment was to indicate that at least intuitively the obtained results are sound.

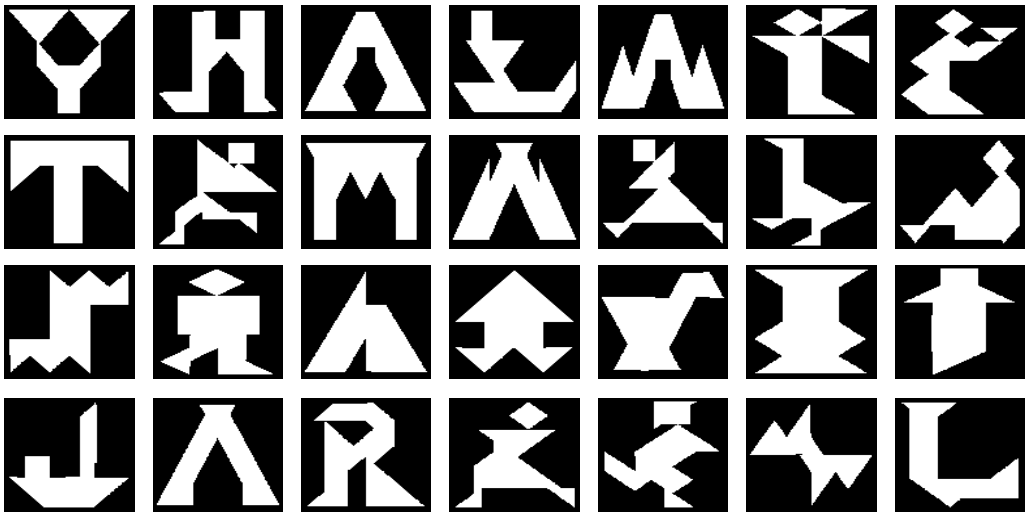


Fig. 4. Samples of tangram images.

A. Classifying and sorting binary objects

In this first experiment we assess the convexity measures in object sorting and recognition tasks in noisy conditions. As test objects we took 100 binary 128×128 images of tangrams, which are produced in an ancient Chinese game, as figures composed of seven pieces of given shape. Some of these are illustrated in Figure 4. The reason for using tangrams is that they give rise to a wide variety of shapes which have roughly the same area and perimeter. All the objects were different in shape and only one was convex, because it would make no sense to try to classify convex objects using convexity measures. The noise used to distort the images was generated in the following way. First, we added uniformly distributed binary noise to the images, where the strength D of the noise is the probability of a pixel to be inverted. Then we selected

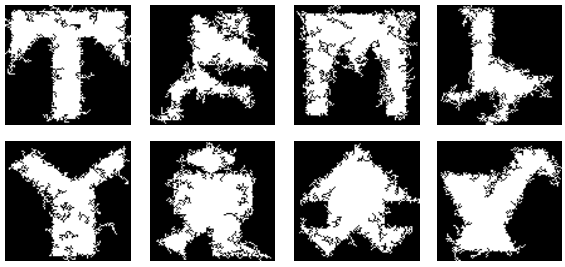
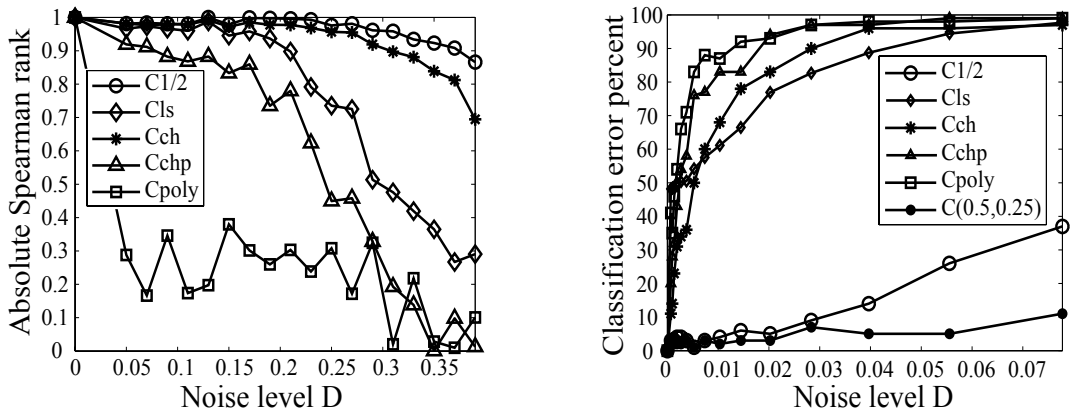


Fig. 5. Samples of noise distorted tangrams, using noise number 0.25.



(a) (b)

Fig. 6. Results of the experiments using tangram dataset, (a) the Spearman ranks and (b) the classification error rates.

the largest connected component, and further used connectivity analysis to remove all holes that were less than 20 pixels. Figure 5 depicts a few samples of the resulting noisy tangrams.

We computed the convexity values from the original and noise distorted versions of the tangram set with several different noise levels. According to these values, we sorted the tangrams in order of convexity at each noise level, and measured how these correlated with the convexity order of the original tangrams. As a measure for this correlation we used the Spearman rank [17], which is a way of estimating the correlation of two data groups possibly having very different scales. In this way the correlation is estimated only using the ranking data and not the absolute values. The resulting correlations are plotted in Figure 6(a), which clearly illustrates the noise tolerance of the area based measures compared to the boundary based ones. It can be further observed that the proposed $C_{1/2}$ measure outperforms the other measures, producing consistent rankings even under severe noise conditions.

In addition to sorting, another important use of shape descriptors is in object classification. To see how the measures perform in such a task we also recognized the noisy tangrams using the convexity values as features. A noisy tangram was classified to be the one from the originals having the closest convexity value. Since only one of the measured objects was actually convex, we expect that when changing the value α in C_α, the information that it gives from the shape also changes. In this case a combination of several C_α, with different values α, could result in

more discriminative information from the objects and thus improve recognition performance. To see this, we also computed $C_{1/4}$ values from the tangrams and used the measurement vector $[C_{1/2}, C_{1/4}]$ for the classification. The corresponding classification error rates are plotted in Figure 6(b). We see that already $C_{1/2}$ outperforms the other measures by a clear margin, but the combination of $C_{1/2}$ and $C_{1/4}$ improves this still further. The results obviously indicate that using several C_α values may improve the results.

B. Lesion classification

One of the issues in medical image analysis is to automatically identify prominent melanomas from images of skin lesions. This topic has recently been more broadly discussed in [1], where also a novel method for such detection is presented. It is known that one of the most important factors in discriminating benign melanocytic lesions from malignant melanomas is the shape of the lesion border. Using this fact in [1] they constructed a measure of the lesion border irregularity and classified those having very irregular border, in the context of this measure, as the most prominent melanomas. To evaluate the method they classified 40 lesion borders, some illustrated in Figure 7, and computed the Spearman rank correlation of the obtained result to a classification made by 14 human experts. The method achieved a rank of 0.88, which was better than the results obtained with the other techniques they considered.

The convexity measure is also very suitable for such shape measuring, which was also discovered in [10]. We wanted to see how the $C_{1/2}$, C_{ch}, C_{ls}, C_{chp}, and C_{poly} perform in such a task by performing the same experiment as in [10]. In this case we classified those having the lowest convexity as the most prominent melanomas. The Spearman ranks achieved with different convexity measures are illustrated in Table I. From these one can observe that boundary based measures are not suitable for this type of a problem as was already stated in [1], however

Fig. 7. 4 samples of the classified lesions.

TABLE I

ABSOLUTE SPEARMAN RANKS IN LESION CLASSIFICATION TEST.

$C_{1/2}$	C_{ch}	C_{ls}	C_{chp}	C_{poly}
0.95	0.90	0.84	0.37	0.42

the area based convexity measures work remarkably well. Already C_{ch} outperforms the results presented in [1], and $C_{1/2}$ does even better. The experiments done in [10] resulted in comparable performance, but evaluating the measures in [10] is not trivial and requires genetic algorithms. Taking into account also how simple for instance $C_{1/2}$ is to compute compared to the irregularity measures presented in [1], one can see the potential of these methods as a part of medical image analysis systems.

C. Pollen classification

Pollen allergies are very common and are usually treated with a medication that is used only during those times of the year when certain pollens are in the air. For this purpose pollen forecasts and measurements are essential. To take such measurements one has to collect pollens from the air and somehow recognize how many and which types of pollens there are in the atmosphere. In Albert-Ludwigs Universität in Freiburg, such recognition has been accomplished using microscope images of pollens collected from outside air [18]. One of the discriminating features between different pollen types is the object structure that appears in these images. In this experiment we wanted to see how our convexity measures work as descriptors for these structures and how they perform in classification. This recognition task is very difficult since real world pollens are not always geometrically transformed versions of each other, and in addition many types of disturbances appear during the collection and imaging processes.

We had a sample pollen image database from the Albert-Ludwigs Universität Freiburg containing a total of 1751 gray-scale images of 12 different types of pollens. In order to produce binary images describing the structure of pollens we applied two different procedures. In the first approach, we applied the Canny filter [19] and then dilated the obtained edge image using a circular element of 2 pixels diameter. In the second approach, we computed the absolute values of the image gradients, normalized them to the interval $[0, 1]$, and then thresholded them. The

thresholding was done using either the so-called ”hard” or ”soft” technique, which we define as follows:

$$f_h(x) = \begin{cases} 1 & \text{if } f(x) > t, \\ 0 & \text{otherwise.} \end{cases}$$

$$f_s(x) = 1 - e^{-((1/t)\log(2)^{(1/6)} f(x))^6},$$

where f_h is the result of hard thresholding, and f_s is the result of soft thresholding. The parameter t, in both approaches, is chosen using the Otsu algorithm [20] to be such that it minimizes the intraclass variance of the 0 and 1 pixels in f_h. In other words, the soft thresholding gives a gray-scale image that approximates the strict binarization f_h, and we presume that in the case of distortions, this approximation changes less than in the corresponding f_h. Therefore we expect a measure computed from f_s to behave more robustly compared to that computed from the corresponding f_h. We recall here that only C_α could be applied to the gray-scale image f_s and

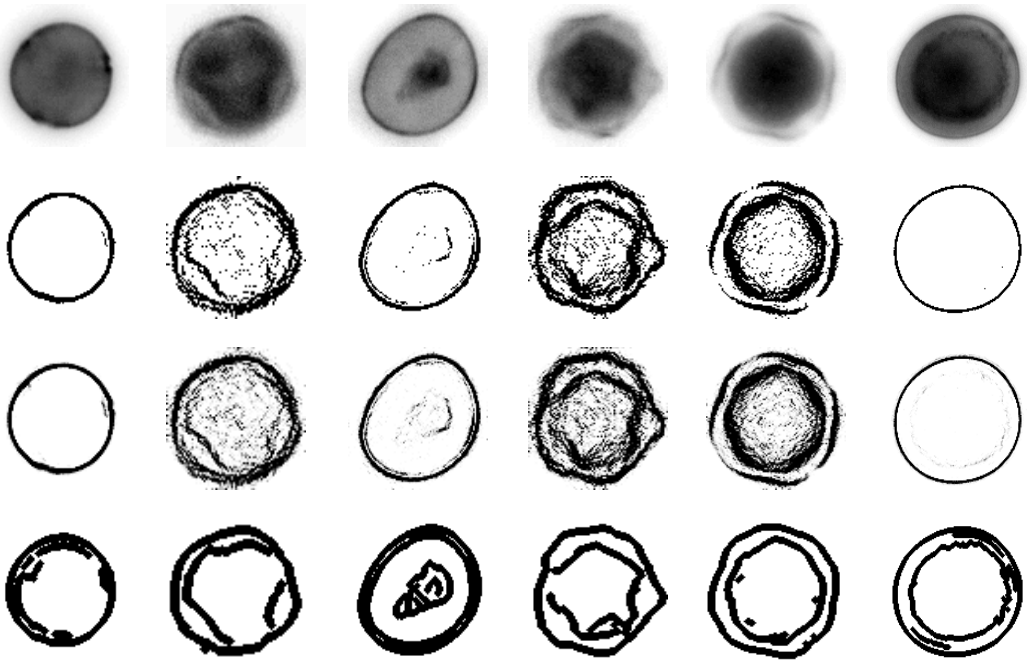


Fig. 8. Six samples of different pollen types. The first row has the original images, the second row has the hard thresholded images, the third row has the softly thresholded images, and the fourth row has the dilated edge images. The measured sets are denoted by black color.

these nonbinary values are interpreted as the probability of a pixel belonging to the shape, as discussed in Section II. In all cases, the convexity measures were directly applied to the achieved edge dilated images, or binarized gradient images without any further preprocessing. Figure 8 illustrates 6 different types of pollen and their thresholded gradient and dilated edge images.

The state-of-the-art pollen classification systems use around 20-60 features, so it was expected that a single convexity value does not have enough information to distinguish between the pollen classes. This fact was also verified in experiments where the correct classification rates with single measures remained around $20\% - 46\%$. However, as in the tangram experiment, we can take several C_α measures to construct measurement vectors which contain more discriminative information about the objects.

Recalling Section 4, due to the symmetry it is enough to consider the interval $0 < \alpha \leq 1/2$, but unfortunately there is no other theoretical result which would show how to select an optimal set of α values. It is also likely that the optimal choice would depend on the measured objects, and might be found using an evaluation process with training data. For our experiment, we decided to take somewhat arbitrarily four sets of α, having 1, 3, 5 and 10 values. The largest set A_3, having 10 values, was taken to be $A_3 = \{C_{1/(2^\beta)} \,|\, \beta = 1, \ldots, 10\}$. The set A_2, having 5 values, was taken from the the set A_3 simply by picking every second value. Similarly A_1 was created by taking every second value from A_2. Finally A_0 was taken to have only $C_{1/2}$. We made such a selection in order to have a situation where $A_0 \subset A_1 \subset A_2 \subset A_3$, making it possible to observe whether the addition of new features brings new information or increases classification perfomance.

To evaluate the features we computed the 10-fold cross validation errors for both edge images and gradient images, using the Support Vector Classifier [21] with the RBF kernel. The results achieved are presented in Table II. It is interesting to observe that already three simple shape descriptors contain significant discriminating information on the objects, and with 10 descriptors the classification accuracy is 89%. We can also observe that when comparing the two thresholdings the soft approach results in better performance, supporting the assumptions presented. The state-of-the-art pollen classification systems have 96% classification accuracy with the same data, but as they use plenty of gray-scale features and are optimized for such detection, we consider the results with only few convexity descriptors to be really promising. In addition, even if the classification results of $50 - 89\%$ with a few convexity values will not satisfy the

TABLE II

CLASSIFICATION RATES USING 10-FOLD CROSS VALIDATION FOR DIFFERENT α-SETS.

α-set	A_0	A_1	A_2	A_3
Gradient images, soft threshold	46.1 %	73.1 %	80.2 %	89.4 %
Gradient images, hard threshold	44.9 %	65.9 %	77.6 %	87.7 %
Edge images	29.3 %	52.9 %	63.2 %	72.3 %

needs of the application, one could use this easily computable information to complement other features.

VI. DISCUSSION

Different shape measurement applications set different requirements for the convexity measure. In some cases, strong reactivity to small defects in the object is desirable, and in other cases the image quality might be poor and one needs robustness against noise or other distortions. Due to these different objectives it is hard to design a measure that is applicable to all situations. If we consider the C_{chp}, and especially C_{poly} measures, one would expect that these boundary-based techniques work well when strong reactivity is desired and good quality images are available. On the other hand in the presence of noise these measures would clearly not be applicable. The experiments in Section V and in [9] further support this conclusion.

Comparing the other probability based approach C_{ls} to our method, there are clear advantages in using points rather than lines. The use of points gives rise to a variety of measures and offers an elegant way of evaluating them, which is lacking from C_{ls}. Also the results presented in Section V illustrate that the performance and the robustness of the point based approach is significantly better.

In this paper, we propose a new area based convexity measure and provide a comprehensive discussion about its properties. We show that the measure fulfills the four basic requirements given in Section I, and also that it depends continuously on examined set. This additional fact is important for applications where stability under noise is desired. We also gave an efficient procedure for computing the measure using the Fourier transform. A new feature of the measure is that it does not require an exact binary image, but can also be applied to a gray-scale image

which approximates the binary set. This property removes the need for strict binarization, and according to the experiments this seems to improve the robustness of the measure. Generally this feature might be useful when the objects are poorly segmented or suffer from gray-scale noise. Also, our formulation of the measure gives rise to a whole set of convexity measures, which can be used to extract more information from the sets under study. The results with tangrams and pollen classification support this idea. The existence of several measures gives new possibilities for choosing the proper convexity measures for a specific application.

VII. Conclusions

In this paper we presented a novel convexity measure for shape analysis. The method turns out to be easy to implement using the Fast Fourier Transform, and it can be applied directly to an image containing a binary set or a gray-scale approximation of it. An example program for computing the measure is included. The theory presented also provides a basis for constructing a whole set of convexity measures which will make it possible to extract more information about the object shape, and will give flexibility to adjust the method to the needs of applications. In addition, the measure can be extended in a straightforward way to cover also 3-dimensional objects. According to the experiments presented, our method provides robust results with respect to noise and in this sense outperforms the other measures included. We also demonstrated our novel measure in two real world classification tasks where it seemed to perform very efficiently. Taking into account how simple the proposed convexity measures are to compute, we expect them to be useful in many computer vision applications.

Acknowledgments

The authors would like to thank the Academy of Finland for providing funding for this research (project no. 102083). We are also grateful to Tangram Man (http://www.reijnhoudt.nl/ tangram) for offering his tangram database and Dr. Paul Rosin for providing his algorithms for computing C_{poly} and the lesion database for our experiments. Finally, we would like to thank Mr. Juho Kannala for helpful comments on the manuscript.

APPENDIX I
PROOF OF PROPERTY 1

We begin with a precise statement which characterizes convex sets in terms of the probability that a point on the line segment between two points from a set belongs to the same set.

Lemma 1.1: Let X and Y be independent random variables drawn uniformly from the set K.

(a) If $K \subseteq \mathbf{R}^2$ is a compact convex set, then $P(\alpha X+(1-\alpha)Y \in K) = 1$ whenever $\alpha \in (0,1)$.

(b) Let K be a compact subset of $\mathbf{R}^2$ which satisfies $K = \overline{\mathrm{int}(K)}$. If $P(\alpha X + (1-\alpha)Y \in K) = 1$ for some $\alpha \in (0,1)$, then K is convex.

Proof: (a) Since $X, Y \in K$ almost surely and K is convex, one has $\alpha X + (1-\alpha)Y \in K$ almost surely, so $P(\alpha X + (1-\alpha)Y \in K) = 1$.

(b) If $f = \chi_K$ then according to (4) we have

$$P(\alpha X + (1-\alpha)Y \in K) = \frac{1}{\alpha^2(1-\alpha)^2|K|^2} \int_{\mathbf{R}^2}\int_{\mathbf{R}^2} f(u) f(\frac{x}{\alpha}) f(\frac{u-x}{1-\alpha})\, dx\, du.$$

We make the change of variables $x = \alpha s, u = \alpha s + (1-\alpha)t$. The Jacobian of this change of variables is $\alpha^2(1-\alpha)^2$, and we obtain

$$\begin{aligned} P(\alpha X + (1-\alpha)Y \in K) =& \frac{1}{|K|^2} \int_{\mathbf{R}^2}\int_{\mathbf{R}^2} f(\alpha s + (1-\alpha)t) f(s) f(t)\, ds\, dt \\ =& \frac{1}{|K|^2} \int_K \int_K \chi_K(\alpha s + (1-\alpha)t)\, ds\, dt. \end{aligned}$$

Since $\chi_K \leq 1$ the last expression is always ≤ 1. But we had $P(\alpha X+(1-\alpha)Y \in K) = 1$, and this implies that there is a set $E \subseteq K^2 = K \times K$ which has measure zero, so that $\alpha s+(1-\alpha)t \in K$ whenever $(s,t) \in K^2 \setminus E$.

Take first $\mathbf{s} = (s,t) \in \mathrm{int}(K)^2$. We claim that there is a sequence in $K^2 \setminus E$ which converges to $\mathbf{s}$. For if not, then there is a small ball in $\mathrm{int}(K)^2$ with center $\mathbf{s}$ which would only contain points of E, and this contradicts the fact that $|E| = 0$. Hence $\mathbf{s}$ is the limit of some sequence in $K^2 \setminus E$, and we have $\alpha s' + (1-\alpha)t' \in K$ for any point $\mathbf{s}'$ in that sequence. Since K is closed we obtain $\alpha s + (1-\alpha)t \in K$ for any points s, t in $\mathrm{int}(K)$.

Now let s, t be any points of K. The assumption $K = \overline{\mathrm{int}(K)}$ implies that there is a sequence in $\mathrm{int}(K)^2$ which converges to (s,t). We have $\alpha s' + (1-\alpha)t' \in K$ for any point (s', t') in this sequence by the above argument, and again the fact that K is closed implies that $\alpha s+(1-\alpha)t \in K$. It follows that K is convex. ■

Proof: (of Property 1)

1) One has $C_\alpha(K) = P(\alpha X + (1-\alpha)Y \in K) \leq 1$. The condition $K \in \mathscr{K}$ implies that $\mathrm{int}(K)$ is nonempty, so there is a ball $B \subseteq K$. This shows that

$$C_\alpha(K) \geq P(\alpha X + (1-\alpha)Y \in K, X \in B, Y \in B)$$
$$= P(X \in B, Y \in B) = \frac{|B|^2}{|K|^2}.$$

2) This is a consequence of Lemma 1.1.

3) We take $K = K_\delta = \{x \in \mathbf{R}^2\,;\, 1-\delta \leq |x| \leq 1\}$. One has

$$C_\alpha(K) = P(\alpha X + (1-\alpha)Y \in K) = \int_{\mathbf{R}^2} P(\alpha X + (1-\alpha)Y \in K | Y = y)$$
$$\times\, P(Y = y)\, dy = \frac{1}{|K|}\int_K P(\alpha X + (1-\alpha)y \in K)\, dy.$$

Note that

$$P(\alpha X + (1-\alpha)y \in K) = P(X \in \frac{1}{\alpha}K - \frac{1-\alpha}{\alpha}y) = \frac{|K \cap (\frac{1}{\alpha}K - \frac{1-\alpha}{\alpha}y)|}{|K|}.$$

Using complex notation $y = re^{i\theta}$, rotational symmetry gives $|K \cap (\frac{1}{\alpha}K - \frac{1-\alpha}{\alpha}y)| = |K \cap (\frac{1}{\alpha}K + \frac{1-\alpha}{\alpha}r)|$. The last expression is the area of the intersection of two annuli, and a geometric argument shows that this area is maximal when $r = 1$. We obtain

$$C_\alpha(K) \leq \frac{|K \cap (\frac{1}{\alpha}K + \frac{1-\alpha}{\alpha})|}{|K|}.$$

The numerator of the right side is majorized by the area of a subsector of K with angle 2θ, where $\theta \to 0$ as $\delta \to 0$. This shows that $C_\alpha(K_\delta) \to 0$ as $\delta \to 0$ for fixed α.

4) This follows from the affine invariance of the MSA transform [13], since $\chi_{\mathscr{A}(K)} = \chi_K \circ \mathscr{A}^{-1}$.

■

APPENDIX II
PROOF OF PROPERTY 2

The following result will be used in the proof of Property 2.

Lemma 2.1: [14] Let $f \in L^1(\mathbf{R}^2)$ and $g \in L^p(\mathbf{R}^2)$, $p = 1$ or $p = \infty$. Then

$$\|f * g\|_{L^p} \leq \|f\|_{L^1}\|g\|_{L^p}.$$

Proof: [of Property 2] Fix $\alpha, \beta \in \mathbf{R}$, and let $\gamma = 1 - \alpha - \beta$. We assume first that all of α, β, γ are nonzero. Then one of $|\alpha|, |\beta|, |\gamma|$ is $\geq 1/3$, say $|\gamma| \geq 1/3$ (the same argument works if $|\alpha| \geq 1/3$ or $|\beta| \geq 1/3$). We write $I(p,q,r,s) = \int_{\mathbf{R}^2} p(u)(q_\alpha * r_\beta * s_\gamma)(u)\, du$ for $p, q, r, s \in L^1 \cap L^\infty$. Then Lemma 2.1 and the fact that $\|f_a\|_{L^1} = \|f\|_{L^1}$ for $a \neq 0$ imply

$$|I(p,q,r,s)| \leq \|p\|_{L^\infty} \|q\|_{L^1} \|r\|_{L^1} \|s\|_{L^1}, \tag{9}$$

$$|I(p,q,r,s)| \leq 9\|p\|_{L^1} \|q\|_{L^1} \|r\|_{L^1} \|s\|_{L^\infty}. \tag{10}$$

We now have

$$\begin{aligned} M(f+h)(\alpha,\beta) - Mf(\alpha,\beta) &= \frac{1}{\|f+h\|_{L^1}^3} I(f+h, f+h, f+h, f+h) - \frac{1}{\|f\|_{L^1}^3} I(f,f,f,f) \\ &= \frac{\|f\|_{L^1}^3 - \|f+h\|_{L^1}^3}{\|f\|_{L^1}^3 \|f+h\|_{L^1}^3} I(f,f,f,f) + \frac{1}{\|f+h\|_{L^1}^3} \Big[I(h,f,f,f) + I(f,h,f,f) + I(f,f,h,f) \\ &\qquad + I(f,f,f,h) + I(h,h,f,f) + \ldots + I(f,h,h,h) + I(h,h,h,h) \Big]. \end{aligned} \tag{11}$$

We need to estimate the expression on the right of (11). Since $\|f\|_{L^\infty} = 1$, (9) gives

$$|I(f,f,f,f)| \leq \|f\|_{L^1}^3.$$

The assumption $\|h\|_{L^1} \leq \frac{1}{2}|K| = \frac{1}{2}\|f\|_{L^1}$ implies $\|f+h\|_{L^1} \geq \frac{1}{2}\|f\|_{L^1}$. Then the absolute value of the first term in (11) is bounded by

$$\frac{8|\|f+h\|_{L^1}^3 - \|f\|_{L^1}^3|}{\|f\|_{L^1}^3}. \tag{12}$$

We have

$$\begin{aligned} \|f+h\|_{L^1}^3 - \|f\|_{L^1}^3 \leq (\|f\|_{L^1} + \|h\|_{L^1})^3 - \|f\|_{L^1}^3 &= 3\|f\|_{L^1}^2 \|h\|_{L^1} + 3\|f\|_{L^1} \|h\|_{L^1}^2 + \|h\|_{L^1}^3 \\ &\leq C_0 \|f\|_{L^1}^2 \|h\|_{L^1} \end{aligned}$$

and

$$\begin{aligned} \|f+h\|_{L^1}^3 - \|f\|_{L^1}^3 \geq (\|f\|_{L^1} - \|h\|_{L^1})^3 - \|f\|_{L^1}^3 &= -3\|f\|_{L^1}^2 \|h\|_{L^1} + 3\|f\|_{L^1} \|h\|_{L^1}^2 - \|h\|_{L^1}^3 \\ &\geq -C_0 \|f\|_{L^1}^2 \|h\|_{L^1} \end{aligned}$$

where C_0 denotes an absolute constant which is replaced by a larger such constant whenever needed. Now (12) is $\leq C_0 \frac{\|h\|_{L^1}}{\|f\|_{L^1}}$, which takes care of the first term in (11).

To continue, first note that $\frac{1}{\|f+h\|_{L^1}^3} \leq \frac{8}{\|f\|_{L^1}^3}$. Now (10) gives

$$|I(h,f,f,f)| \leq 9\|h\|_{L^1}\,\|f\|_{L^1}^2$$

and $\frac{1}{\|f+h\|_{L^1}^3}|I(h,f,f,f)| \leq C_0\frac{\|h\|_{L^1}}{\|f\|_{L^1}}$. For the other terms of the form $I(h,\,\cdot\,)$ we use (10) and also $\|h\|_{L^1} \leq \frac{1}{2}\|f\|_{L^1}$, and obtain that $|I(h,\,\cdot\,)| \leq 9(\max\{R,1\})\|f\|_{L^1}^2\,\|h\|_{L^1}$. Then

$$\frac{1}{\|f+h\|_{L^1}^3}|I(h,\,\cdot\,)| \leq C_0 R\frac{\|h\|_{L^1}}{\|f\|_{L^1}}$$

since $R \geq 1$. Finally, for the terms $I(f,\,\cdot\,)$ we use (9) and $\|h\|_{L^1} \leq \frac{1}{2}\|f\|_{L^1}$ and obtain

$$\frac{1}{\|f+h\|_{L^1}^3}|I(f,\,\cdot\,)| \leq C_0\frac{\|h\|_{L^1}}{\|f\|_{L^1}}.$$

This ends the proof for the case where $\alpha,\beta,\gamma \neq 0$.

If one of α,β,γ is zero, say $\gamma = 0$, then $\beta = 1-\alpha$ and

$$M(f+h)(\alpha,\beta) - Mf(\alpha,\beta) = \frac{1}{\|f+h\|_{L^1}^2}I(f+h,f+h,f+h) - \frac{1}{\|f\|_{L^1}^2}I(f,f,f)$$

where $I(p,q,r) = \int p(u)(q_\alpha * r_\beta)(u)\,du$. In this case one of $|\alpha|, |\beta|$ is $\geq 1/2$, and using this fact and obvious counterparts for (9) and (10) one proceeds in a similar way as above. This takes care of the case where one of α,β,γ is zero.

Finally, suppose two of α,β,γ are zero. We may assume $\alpha = \beta = 0$ and $\gamma = 1$, and then

$$\begin{aligned} M(f+h)(0,0) - Mf(0,0) &= \frac{\|f+h\|_{L^2}^2}{\|f+h\|_{L^1}} - \frac{\|f\|_{L^2}^2}{\|f\|_{L^1}} \\ &= \frac{(\|f+h\|_{L^2}^2 - \|f\|_{L^2}^2)\|f\|_{L^1} + (\|f\|_{L^1} - \|f+h\|_{L^1})\|f\|_{L^2}^2}{\|f\|_{L^1}\|f+h\|_{L^1}}. \end{aligned}$$

We note that $\|f+h\|_{L^1} \geq \frac{1}{2}\|f\|_{L^1}$, $|\|f+h\|_{L^1} - \|f\|_{L^1}| \leq \|h\|_{L^1}$, and

$$\|f+h\|_{L^2}^2 - \|f\|_{L^2}^2 = \int_{\mathbf{R}^2}((f+h)^2 - f^2)\,dx = \int_{\mathbf{R}^2}(2fh+h^2)\,dx$$

so that $|\|f+h\|_{L^2}^2 - \|f\|_{L^2}^2| \leq (2\|f\|_{L^\infty} + \|h\|_{L^\infty})\|h\|_{L^1} \leq 3R\|h\|_{L^1}$. Further, $\|f\|_{L^2}^2 \leq \|f\|_{L^1}$. Using these facts we obtain

$$|M(f+h)(0,0) - Mf(0,0)| \leq (6R+2)\frac{\|h\|_{L^1}}{\|f\|_{L^1}}.$$

The last expression is bounded by $8R\frac{\|h\|_{L^1}}{\|f\|_{L^1}}$, which ends the proof. ■

APPENDIX III
PROOF OF PROPERTY 3

Proof: We let $f = \chi_K$ and $h = \chi_{K'} - \chi_K$, so that $f + h = \chi_{K'}$. Then $|h| \leq 1$ and $\|h\|_{L^1} = |K' \setminus K| + |K \setminus K'| = |K' \triangle K|$, so h satisfies the conditions of Property 2 with $R = 1$. Thus

$$|M\chi_{K'}(\alpha, 1-\alpha) - M\chi_K(\alpha, 1-\alpha)| \leq C_0 \frac{|K' \triangle K|}{|K|}.$$

The left hand side is the same as $|C_\alpha(K') - C_\alpha(K)|$, so we obtain

$$|C_\alpha(K') - C_\alpha(K)| \leq C_0 \frac{|K' \triangle K|}{|K|}.$$

■

APPENDIX IV
PROOF OF PROPERTY 4

Proof: If A is any convex subset of K, one has

$$\begin{aligned} C_{\text{ls}}(K) &= P([X,Y] \subseteq K) \geq P([X,Y] \subseteq K, X \in A, Y \in A) \\ &= P(X \in A, Y \in A) = \frac{|A|^2}{|K|^2}. \end{aligned}$$

Taking the supremum over all convex subsets of K gives

$$C_{\text{ls}}(K) \geq \frac{|\text{mcs}(K)|^2}{|K|^2} = C_{mcs}(K)^2.$$

Also, for fixed α,

$$C_\alpha(K) = P(\alpha X + (1-\alpha)Y \in K) \geq P([X,Y] \subseteq K) = C_{\text{ls}}(K).$$

■

APPENDIX V
MATLAB PROGRAM FOR COMPUTING $C_{1/2}$

The input f in the algorithm is a binary image, or a gray-scale image which is close to a binary image. We note that the second and third lines are just to make the image dimensions even numbers.

```
function F=convexity(f)
[m,n]=size(f);
f=[[f,zeros(m,rem(n,2))];zeros(rem(m,2),n+rem(n,2))];
[m,n]=size(f);
f0=sum(f(:));
f1=f(1:2:m,1:2:n)+f(2:2:m,1:2:n)+f(1:2:m,2:2:n)+f(2:2:m,2:2:n);
F=real(sum(sum(conj(fft2(f,m,n)).*(fft2(f1,m,n).^2))))/(n*m*f0^2);
```

REFERENCES

[1] T. Lee, D. McLean, and M. Atkins, "A new irregularity measure for cutaneous lesions." *Medical Image Analysis*, vol. 7, no. 1, pp. 47–64, 2003.

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

[3] S. Hyde, *The language of shape*. Elsevier, 1997.

[4] F. Valentine, *Convex sets*. San Fransisco, CA: McGraw-Hill, 1964.

[5] A. Held and K. Abe, "On approximate convexity," *Pattern Recognition Letters*, vol. 15, pp. 611–618, 1994.

[6] A. T. Popov, "Fuzzy morphology and fuzzy convexity measures," in *Proceedings of 13th International Conference on Pattern Recognition*, vol. 2, pp. 611–614, Vienna, Austria, 1996.

[7] L. Boxer, "Computing deviations from convexity in polygons," *Pattern Recognition Letters*, vol. 14, pp. 163–167, 1993.

[8] H. I. Stern, "Polygonal entropy: A convexity measure," *Pattern Recognition Letters*, vol. 10, pp. 229–235, 1989.

[9] J. Zunic and P. L. Rosin, "A new convexity measure for polygons," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 7, pp. 923–934, 2004.

[10] P. L. Rosin and C. L. Mumford, "A symmetric convexity measure," in *Proceedings of 17th International Conference on Pattern Recognition*, vol. 4, pp. 11–14, Cambridge, England, 2004.

[11] E. Rahtu, M. Salo, and J. Heikkilä, "Convexity recognition using Multi-Scale Autoconvolution," in *Proceedings of 17th International Conference on Pattern Recognition*, vol. 1, pp. 692–695, Cambridge, England, 2004.

[12] J. Heikkilä, “Multi-Scale Autoconvolution for affine invariant pattern recognition,” in *Proceedings of 16th International Conference on Pattern Recognition*, pp. 119–122, Québec, Canada, 2002.

[13] E. Rahtu, M. Salo, and J. Heikkilä, “Affine invariant pattern recognition using Multiscale Autoconvolution,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 6, pp. 908–918, June 2005.

[14] W. Rudin, *Real and complex analysis*. McGraw-Hill, 1987.

[15] M. Frigo and S. G. Johnson, “FFTW: An adaptive software architecture for the FFT,” in *Proceedings of IEEE International Conference on Acoustics Speech and Signal Processing*, vol. 3. IEEE, pp. 1381–1384, 1998.

[16] T. H. Cormen, C. E. Leiserson, R. L. Rivest, and C. Stein, *Introduction to algorithms*, Second Edition. MIT Press and McGraw-Hill, 2001.

[17] E. L. Lehmann and H. J. M. D’Abrera, *Nonparametrics: statistical methods based on ranks, rev. ed.*, Prentice Hall, 1998.

[18] O. Ronneberger, E. Schultz, and H. Burkhardt, “Automated pollen recognition using 3D volume images from fluorescence microscopy,” *Aerobiologia*, vol. 18, pp. 107–115, 2005.

[19] J. Canny, “A computational approach to edge detection,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 8, pp. 679–698, 1986.

[20] N. Otsu, “A threshold selection method from gray-level histograms,” *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 9, no. 1, pp. 62–66, 1979.

[21] V. Vapnik, *The nature of statistical learning theory*. Springer N.Y., 1995.

Esa Rahtu received his MSc degree in electrical engineering in 2004 from the University of Oulu, Finland. Currently, he is a researcher in the Machine Vision Group at the University of Oulu. His research interests include geometric invariants and object classification.

Mikko Salo received his MSc and PhLic degrees in mathematics in 2001 and 2003 from the University of Oulu, Finland, and received the Doctor of Philosophy degree in applied mathematics in 2004 from the University of Helsinki, Finland. He has been a visiting researcher at the University of Washington in Seattle and the Mathematical Sciences Research Institute in Berkeley. Currently he is a researcher in the Department of Mathematics and Statistics / Rolf Nevanlinna Institute at the University of Helsinki. His research interests include inverse problems, partial differential equations, and pattern recognition.

Janne Heikkilä received his MSc degree in electrical engineering in 1993, and the Doctor of Technology degree in information engineering in 1998 from the University of Oulu, Finland. From 1998 to 2000, he was a postdoctoral fellow in the Machine Vision and Media Processing Unit at the University of Oulu. Since 2000 he has been working as an acting professor of signal processing engineering, and an acting professor of information engineering. From 2002 to 2004, he was a professor of digital video processing at the University of Oulu. His research interests include geometric camera calibration, 3-D vision, geometric invariants, image registration, and motion estimation in video coding.