

# Fourier Descriptors for Broken Shapes

Christoph Dalitz · Christian Brandt · Steffen Goebbels · David Kolanus

Received: 15. July 2013 / Accepted: 18. September 2013

**Abstract** Fourier descriptors are powerful features for the recognition of two-dimensional connected shapes. In this article, we propose a method to define Fourier descriptors even for broken shapes, i.e., shapes that can have more than one contour. The method is based on the convex hull of the shape and the distance to the closest actual contour point along the convex hull. We define different invariant Fourier descriptors for this three-dimensional representation of a two-dimensional shape and compare them on different data sets. The recognition rates are comparable to normal Fourier descriptors while the new scheme at the same time offers the option to also deal with broken contours. We also discuss and evaluate different normalisation schemes that make the descriptors invariant under scale and rotation.

**Keywords** Image Processing · Shape Description · Object Recognition

## 1 Introduction

*Shape descriptors* are numbers that are computed from a two-dimensional shape. In some cases, the set of numbers is complete in the sense that the original shape can be reconstructed from the shape descriptors [1, 2], but even in these situations, only a subset of the shape descriptors is typically used in practical applications. The shape descriptors can thus be considered as an approximative description of the shape such that shape similarity somehow corresponds to similarity of the shape descriptors. Consequently, they can be used for object recognition and object similarity detection.

---

C. Dalitz, C. Brandt, S. Goebbels, and D. Kolanus  
Hochschule Niederrhein, Institute for Pattern Recognition,  
Reinarzstr. 49, 47805 Krefeld, Germany  
E-mail: christoph.dalitz@hsnr.de

There are two main categories of shape descriptors: *volume-based* and *contour-based*. Volume based descriptors use all pixels of the object and include descriptors like geometric moments [3] or Zernike moments [4]. Contour based descriptors compute the descriptors only from the shape boundary and include descriptors like curvature scale space [5] or Fourier descriptors [6]. Which approach is better depends on the application, especially whether the internal content of the shape or the boundary is more important.

The term *Fourier descriptors* covers a wide variety of shape descriptors, which have in common that they compute the discrete Fourier transform of some representation of a closed contour. They vary in the representation (or “signature” [6]) of the contour and in the additional manipulations to achieve invariance properties under certain geometric transformations. Typically, invariance under translation, scaling, and rotation is achieved, but there are also Fourier descriptors that are invariant to shearing [7]. As Fourier descriptors require the extraction of a closed contour from the shape, they are restricted to connected shapes and cannot be applied to possibly broken objects.

To overcome this restriction, we define a new *three-dimensional* contour signature function, based upon the convex hull of the shape and the distance of the convex hull to the closest point of the shape. As the convex hull is defined for arbitrary shapes, including broken shapes, the new shape signature no longer requires connectivity of the shape. Based upon this shape signature, we derive different invariant Fourier descriptors and compare their performance on different data sets.

This paper is organised as follows: in Sec. 2 we give an overview over different Fourier descriptors described in the literature for closed contours of unbroken shapes. In Sec. 3 we describe our new method for a contour rep-

representation of broken shapes and define different methods to obtain invariant Fourier descriptors from this representation. Secs. 4 and 5 contain the results of a comparative evaluation of the different Fourier descriptors and the conclusions drawn therefrom.

## 2 Fourier descriptors

Every connected object has a closed contour that can be represented as a sequence of the pixel coordinates  $x(t), y(t)$  where  $t = 0, \dots, N-1$ . A popular algorithm for contour extraction can be found in [8]. The coordinates can be considered to be sampling values

$$(x(t), y(t)) = f\left(\frac{2\pi}{N}t\right) \quad (1)$$

of a continuous, closed curve  $f : [0, 2\pi] \rightarrow \mathbb{R}^2$  such that  $f$  can be extended continuously to a  $2\pi$ -periodic function. When the function  $f$  (or, more generally, any signature function  $g(x(t), y(t))$  derived from the coordinates) is expanded into a Fourier series, a fixed number of discrete Fourier coefficients approximately represents the contour shape. This allows for data reduction.

This is the basic idea underlying all Fourier descriptors suggested in the literature. They vary however in the contour representation  $g(x(t), y(t))$  used as the starting point for the Fourier expansion. The representations typically fall into one of the following categories:

- Complex representation:  $x(t) + j \cdot y(t) \in \mathbb{C}$
- Multidimensional representation:  $(x(t), y(t)) \in \mathbb{R}^2$
- Scalar representation:  $(x(t), y(t)) \mapsto g(t) \in \mathbb{R}$ .

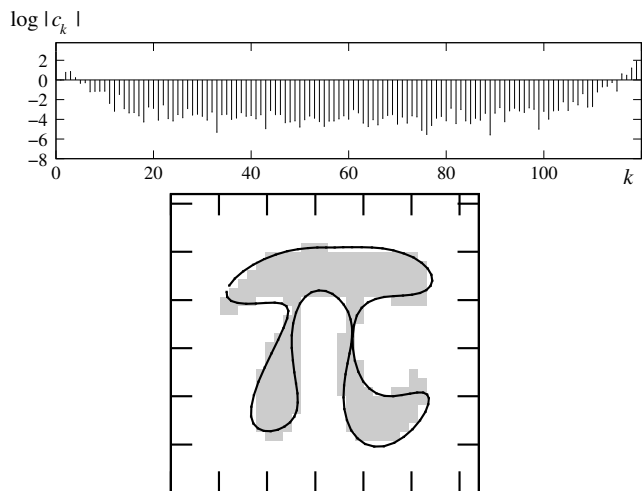
Depending on the contour representation, the resulting Fourier coefficients will behave differently under the geometric transformations scaling, rotation, and start point shift. When the contour points are transformed by any of these operations, the Fourier coefficients change according to simple rules, which can be used to define invariant descriptors.

In the following subsections, we give an overview over the three categories and the normalisation approaches proposed in the literature to achieve invariance under these geometric transformations.

### 2.1 Complex contour representation

When the two dimensional plane is interpreted as a complex plane, the contour is represented by a sequence of complex numbers  $z(t) = x(t) + j \cdot y(t)$  which has the discrete Fourier expansion ( $t = 0, \dots, N-1$ )

$$z(t) = \sum_{k=0}^{N-1} c_k \exp(j2\pi kt/N) \quad (2)$$



**Fig. 1** Absolute values of the Fourier coefficients  $c_k$  and the reconstruction of a contour from only the coefficients  $k \leq 8$  and  $k \geq N-8$ .

with discrete Fourier coefficients

$$c_k = \frac{1}{N} \sum_{t=0}^{N-1} z(t) \exp(-j2\pi kt/N) \quad (3)$$

When the coefficients  $c_k$  are interpreted as numerical approximations of the Fourier coefficients  $\hat{f}(k)$  of the continuous curve  $f(\tau) = x(\tau N/2\pi) + jy(\tau N/2\pi)$  in (1)

$$\hat{f}(k) := \frac{1}{2\pi} \int_0^{2\pi} f(\tau) \exp(-jk\tau) d\tau, \quad k \in \mathbb{Z}$$

then the connection between  $\hat{f}(k)$  and the discrete Fourier coefficients (3) is given by

$$\begin{aligned} \hat{f}(k) &\approx c_k && \text{for } 0 \leq k < \frac{N}{2} \\ \hat{f}(-k) &\approx c_{N-k} && \text{for } 1 \leq k < \frac{N}{2} \end{aligned} \quad (4)$$

According to the Riemann-Lebesgue lemma [9, p. 45], it is

$$\lim_{|k| \rightarrow \infty} \hat{f}(k) = 0$$

so that coefficients for large values of  $|k|$  are small and only describe less important details. Cutting off higher frequencies  $|k|$  in (4) is thus equivalent to omitting coefficients in the middle of the vector  $(c_0, \dots, c_{N-1})$ . An example can be seen in Fig. 1.

The zeroth coefficient  $c_0$  is the centre of gravity of the contour. As the smallest period of the contour curve  $f$  is the length  $2\pi$  of the parameter interval, we can assume that at least  $c_1 \neq 0$  or  $c_{N-1} \neq 0$ . Which of these two coefficients actually is guaranteed to be nonzero depends on the orientation of the contour path: for Pavlidis' algorithm [8], e.g., it is  $c_1 \neq 0$ .

In general, it is however not guaranteed that *both* coefficients are nonzero, as can be seen from the unit circle  $f(t) := (\cos(t), \sin(t))$ , which has coefficients  $c_1 = 1$ ,  $c_{N-1} = 0$ ,  $N \geq 3$ .

Based on elementary properties of the discrete Fourier transform, simple rules for the change of the coefficients  $c_k$  under translation, scale and rotation immediately follow. Let  $c_k$  be a coefficient calculated before any of the following operations. Then the geometric operations have the following effects:

*Translation:* Adding the same complex number  $u$  to all points  $z(t)$  leads to new coefficients  $(c_0 + u, c_1, \dots, c_{N-1})$ .

*Scale:* Multiplying all points  $z(t)$  with the same real factor  $d > 0$  leads to new coefficients  $d \cdot c_k$ .

*Rotation:* Rotation in the complex plane is the same as multiplying all points  $z(t)$  with a factor  $\exp(j\varphi)$ , where  $\varphi$  is the angle of rotation. This leads to new coefficients  $\exp(j\varphi)c_k$ .

*Start point shift:* Starting the contour at a different point results in a cyclic shift of vector  $(z(t))_{t=0}^{N-1}$ . If the index shift is  $m$ , then the new coefficients are  $\exp(jkm\frac{2\pi}{N})c_k$ .

To achieve translation invariance, the first coefficient  $c_0$  can be discarded because it is the only one that depends on translation. For scale invariance, all coefficients can be divided by the absolute value of a non-zero coefficient  $|c_r|$ ,  $r > 0$ . Usually, a fixed coefficient  $r$  is chosen, e.g.  $r = 1$ , but our experiments in Sec. 4.3 have shown that it is better to always choose the coefficient  $c_r$  with the largest absolute value.

Since  $|\exp(j\varphi)| = |\exp(jkm\frac{2\pi}{N})| = 1$ , a simple approach to obtain a rotation and shift invariant descriptor is to completely drop the phase information and to only use the absolute values of the Fourier coefficients. This approach was used, e.g., by Zhang & Lu in their comparative study [6]. The resulting absolute value descriptors are

$$|l_k| := \frac{|c_k|}{|c_r|} \quad \text{for } k = 1, N-1, 2, N-2, \dots \quad (5)$$

It is also possible to define invariant Fourier descriptors that still keep the phase information, as already observed by Dimov & Laskov [10]. Let  $c_r \neq 0$  and  $c_s \neq 0$ ,  $r \neq s$ , be two non-zero coefficients with polar angles  $\alpha_r = \arg(c_r)$  and  $\alpha_s = \arg(c_s)$ . Rotation invariance is achieved by multiplying each coefficient with  $\exp(-j\alpha_r)$ , and shift invariance is established by replacing the polar angle  $\alpha_k = \arg(c_k)$  with  $s\alpha_k - k\alpha_s$ . Combining these phase normalisations yields the invari-

ant descriptors

$$l_k := \frac{|c_k| \exp(j(s-r)[\alpha_k - \alpha_r])}{|c_r| \exp(j(k-r)[\alpha_s - \alpha_r])} \quad (6)$$

$$= \frac{|c_k|}{|c_r|} \exp(j[(s-r)\alpha_k + (k-s)\alpha_r + (r-k)\alpha_s])$$

for  $k = 1, N-1, 2, N-2, \dots$ . The two normalisation coefficients  $c_r$  and  $c_s$  should be chosen as the coefficients with the largest and second largest absolute value, respectively<sup>1</sup>.

Granlund [11] proposed different descriptors as

$$d_k := \frac{c_{1+k}c_{N+1-k}}{c_1^2} \quad \text{for } k = 2, \dots, N-2 \quad (7)$$

These descriptors include the phase information, but because of  $d_k = d_{N-k}$ , there is a considerable loss of information compared to (6). Granlund also defined the  $(N-1)(N-2)$  (sic!) descriptors  $c_{1+k}^i c_{N+1-i}^k / c_1^{i+k}$ , but these have a lot of redundancy and it is not clear how to select a small subset therefrom.

## 2.2 Multidimensional contour representation

Instead of interpreting the contour coordinates as complex numbers, the  $x$  and  $y$  coordinate can alternatively be transformed separately:

$$c_k^{(x)} = \frac{1}{N} \sum_{t=0}^{N-1} x(t) \exp(-j2\pi kt/N) \quad (8)$$

$$c_k^{(y)} = \frac{1}{N} \sum_{t=0}^{N-1} y(t) \exp(-j2\pi kt/N),$$

or, split into into real and imaginary part:

$$a_k^{(x)} = \frac{1}{N} \sum_{t=0}^{N-1} x(t) \cos(2\pi kt/N) \quad (9)$$

$$b_k^{(x)} = \frac{1}{N} \sum_{t=0}^{N-1} x(t) \sin(2\pi kt/N)$$

$$a_k^{(y)} = \frac{1}{N} \sum_{t=0}^{N-1} y(t) \cos(2\pi kt/N)$$

$$b_k^{(y)} = \frac{1}{N} \sum_{t=0}^{N-1} y(t) \sin(2\pi kt/N).$$

Half of these components are redundant because  $b_0^{(x)} = b_0^{(y)} = 0$  and, for  $0 < k \leq N/2$ , it is

$$a_k^{(x)} = a_{N-k}^{(x)} \quad b_k^{(x)} = -b_{N-k}^{(x)} \quad (10)$$

$$a_k^{(y)} = a_{N-k}^{(y)} \quad b_k^{(y)} = -b_{N-k}^{(y)}$$

<sup>1</sup> When only one coefficient is non zero, this can be chosen as  $c_r$  and no phase normalisation is necessary.

The inverse formula then reads

$$\begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = \begin{pmatrix} a_0^{(x)} \\ a_0^{(y)} \end{pmatrix} + 2 \sum_{k=1}^{\lceil \frac{N-1}{2} \rceil} \begin{pmatrix} a_k^{(x)} & b_k^{(x)} \\ a_k^{(y)} & b_k^{(y)} \end{pmatrix} \cdot \begin{pmatrix} \cos 2\pi kt/N \\ \sin 2\pi kt/N \end{pmatrix} \quad (11)$$

where  $\lceil \frac{N-1}{2} \rceil$  is the smallest integer  $i$  with  $i \geq \frac{N-1}{2}$ . For the sake of simplicity let  $N$  be an odd number throughout the rest of this section.

The real coefficients (9) of the multidimensional representation are connected to the complex coefficients (3) by

$$c_k = c_k^{(x)} + j c_k^{(y)} = a_k^{(x)} + b_k^{(y)} + j[a_k^{(y)} - b_k^{(x)}]$$

In contrast to  $c_k$ , higher frequencies directly correspond to higher indices  $k$  of the real coefficients  $a_k^{(x/y)}, b_k^{(x/y)}$  because of the symmetry relations (10).

Kuhl & Giardina [12] interpreted each summand

$$\begin{pmatrix} x_k(t) \\ y_k(t) \end{pmatrix} = 2 \begin{pmatrix} a_k^{(x)} & b_k^{(x)} \\ a_k^{(y)} & b_k^{(y)} \end{pmatrix} \cdot \begin{pmatrix} \cos 2\pi kt/N \\ \sin 2\pi kt/N \end{pmatrix}$$

in (11) as a parameterisation with parameter  $t \in [0, 2\pi]$  of an ellipse that visualises the  $k$ -th Fourier coefficients. Therefore they called the resulting Fourier descriptors *elliptic features*. Based upon this idea, Lin & Hwang [13] proposed the following translation, rotation and shift invariant (but not scale invariant) Fourier descriptors:

$$I_k := [a_k^{(x)}]^2 + [b_k^{(x)}]^2 + [a_k^{(y)}]^2 + [b_k^{(y)}]^2 \quad (12)$$

$$J_k := \det \begin{pmatrix} a_k^{(x)} & b_k^{(x)} \\ a_k^{(y)} & b_k^{(y)} \end{pmatrix} = a_k^{(x)} b_k^{(y)} - b_k^{(x)} a_k^{(y)}$$

$$\begin{aligned} K_k := & \operatorname{sgn} \left( (a_k^{(x)} a_k^{(y)} + b_k^{(x)} b_k^{(y)}) \left( |c_k^{(y)}|^2 - |c_k^{(x)}|^2 \right) \right. \\ & \left. + (a_i^{(x)} a_i^{(y)} + b_i^{(x)} b_i^{(y)}) \left( |c_k^{(x)}|^2 - |c_k^{(y)}|^2 \right) \right) \\ & \times \left[ |c_i^{(x)}|^2 |c_k^{(x)}|^2 + |c_i^{(y)}|^2 |c_k^{(y)}|^2 \right] \\ & + 2 (a_i^{(x)} a_i^{(y)} + b_i^{(x)} b_i^{(y)}) (a_k^{(x)} a_k^{(y)} + b_k^{(x)} b_k^{(y)}) \end{aligned}$$

where  $i \in 1, \dots, (N-1)/2$  is a fixed index and  $\operatorname{sgn}$  denotes the signum function. The properties of the real Fourier coefficients imply

$$I_{N-k} = I_k, \quad J_{N-k} = J_k, \quad K_{N-k} = K_k$$

To make these features also scale invariant, they additionally need to be divided by a normalisation factor, e.g.  $I_1$ , which leads to the invariant descriptors  $I_k/I_1$ ,  $J_k/I_1$ , and  $K_k/I_1^2$ .

A nice point about the features (12) is that they have a geometric interpretation:  $I_k$  is the sum of two semi-axis lengths of the  $k$ -th ellipse,  $|J_k|$  is proportional to the area of the  $k$ -th ellipse, and  $K_k$  contains the phase difference between ellipses  $i$  and  $k$  for the fixed  $i$ . Compared to the  $N-2$  complex features (6), the  $\frac{3}{2}(N-1)$  real elliptic Fourier features contain however considerably less information about the shape. Lin & Hwang tried to compensate this with additional features, which were not rotation invariant, however.

An interesting generalisation of Fourier descriptors from two-dimensional curves to  $n$ -dimensional closed curves was made by Badreldin et al. [14]. They transformed each component separately and then built a vector containing  $l_2$ -norms of all Fourier coefficients of a given index. In two dimensions this is equivalent to the descriptors by Shridhar & Badreldin [15] ( $k = 0, \dots, (N-1)/2$ ):

$$\sqrt{|c_k^{(x)}|^2 + |c_k^{(y)}|^2} = \sqrt{[a_k^{(x)}]^2 + [b_k^{(x)}]^2 + [a_k^{(y)}]^2 + [b_k^{(y)}]^2} \quad (13)$$

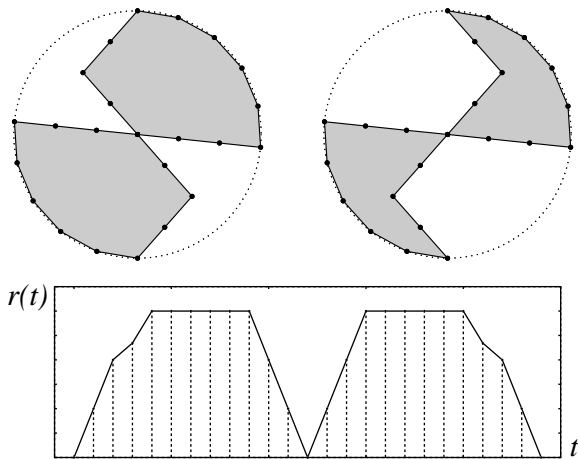
Rotation and startpoint shift invariance is obtained because absolute values are used, and translation invariance results from discarding  $k=0$ . For scale invariance, they additionally need to be divided by a normalisation factor, e.g.  $\sqrt{|c_1^{(x)}|^2 + |c_1^{(y)}|^2}$ . The Fourier descriptors (13) discard however a considerable portion of the shape information: not only phase information is lost but also  $x$ - and  $y$ -components are not coupled. Therefore, for example, a shift of  $x$ -coordinates without a change to  $y$ -coordinates cannot be detected.

### 2.3 Scalar contour representation

A two-dimensional contour can also be represented in one dimension by mapping it to a one-dimensional *signature function*:  $(x(t), y(t)) \mapsto f(t)$ . The signature function  $f$  can already be invariant under translation, scaling and rotation, like Zahn & Roskies' *cumulative angular function* [16], or the invariance normalisation can be applied after the Fourier transform. Mapping two dimensions onto one generally leads to some loss of shape information (see Fig. 2), but the hope is that the essential features are still captured for most shapes. For an overview of possible signature functions, see [6].

In the comparative study [6], the *centroid distance* performed best. Let  $(x_0, y_0) := \frac{1}{N} \sum_{k=0}^{N-1} (x_k, y_k)$  be the centre of gravity of the contour. Then the centroid distance is defined as

$$\begin{aligned} r(t) := & |(x(t), y(t)) - (x_0, y_0)| \\ = & \sqrt{(x(t) - x_0)^2 + (y(t) - y_0)^2} \end{aligned} \quad (14)$$



**Fig. 2** Two different shapes (grey) with the same signature function “centroid distance”  $r(t)$  defined in Eq. (14).

These values are already rotation and translation invariant. Let  $(c_k)_{k=0}^{N-1}$  be the (complex) Fourier transform of  $(r(t))_{t=0}^{N-1}$ , i. e.

$$c_k = \frac{1}{N} \sum_{t=0}^{N-1} r(t) \exp\left(-j \frac{2\pi kt}{N}\right) \quad (15)$$

Descriptors  $R_k$  that are also scale and start point shift invariant can be obtained with the phase normalisation

$$R_k = \exp(j s \alpha_k - j k \alpha_s) \cdot \frac{|c_k|}{|c_0|} \quad (16)$$

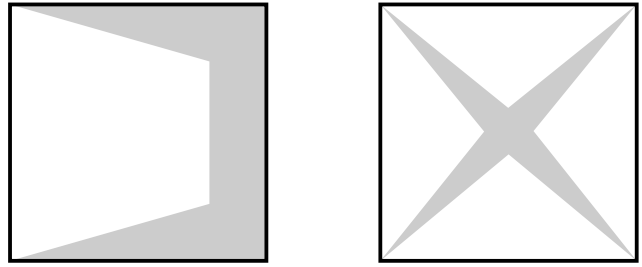
where  $\alpha_k = \arg(c_k)$  denotes the polar angle of coefficient  $c_k$ , and  $\alpha_s$  is the polar angle of the coefficient  $c_s$ ,  $s \geq 1$  with the second largest absolute value. Note that  $|c_0| = c_0$  is always the largest absolute value because

$$\begin{aligned} c_0 &= \frac{1}{N} \sum_t r(t) = \frac{1}{N} \sum_t \left| r(t) \exp\left(-j \frac{2\pi kt}{N}\right) \right| \\ &\geq \left| \frac{1}{N} \sum_t r(t) \exp\left(-j \frac{2\pi kt}{N}\right) \right| = |c_k| \text{ for all } k \end{aligned}$$

An alternative, simpler normalisation is to discard the phase information and use the absolute value  $|R_k|$ . This normalisation was used by Zhang & Lu.

### 3 Application to broken shapes

All Fourier descriptors described in Sec. 2 start from a closed contour description of the shape and are therefore not applicable when the shape is broken, i.e. consists of more than one connected component. In this section we first present a method to describe the contour of an arbitrary (broken or unbroken) shape by a periodic *three-dimensional* curve and then derive different Fourier descriptors for this curve which are invariant under translation, scale, rotation, and start point shift.



**Fig. 3** Two different shapes (grey) with the same convex hull (solid black).

#### 3.1 Contour representation of broken shapes

A simple solution to circumvent the problem of broken shapes would be to replace the shape parts with a *single* closed curve that contains all parts, and to compute the Fourier descriptors from this curve instead. An obvious candidate for such a curve is the *convex hull*, i.e. the smallest convex polygon that contains all points of the shape. There are efficient algorithms for computing the convex hull from a set of points [17]. As can be seen in Fig. 3, replacing a contour with its convex hull loses a considerable amount of information because very different shapes can have the same convex hull. To encode more shape information, we therefore compute for each point  $(x, y)$  on the convex hull its closest Euclidean distance  $d$  to the shape  $S$ :

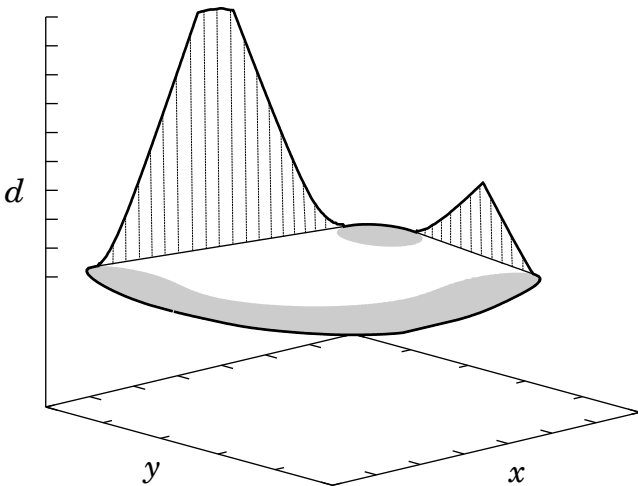
$$d = \min\{|(x, y) - (u, v)| \text{ with } (u, v) \in S\} \quad (17)$$

Instead of a two-dimensional contour  $(x(t), y(t))$ , we then obtain a three-dimensional parametric curve  $(x(t), y(t), d(t))$  representing the shape, as shown in Fig. 4.

When implementing an algorithm for computing the contour representation  $(x(t), y(t), d(t))$ , two questions occur: how the convex hull should be sampled and how the distances  $d(t)$  can be efficiently computed. The vertices of the convex hull polygon can be obtained, e.g., with Graham’s scan algorithm [17]. These vertices are obvious sampling points, but their distance can be arbitrary, so that the edges need to be sampled. As the image sampling distance is one pixel, it is natural to compute the edge length  $l$  and to add  $\lfloor l - 1 \rfloor$  equidistant sampling points on each edge.

To compute the distance  $d(t)$  for each sampling point  $x(t), y(t)$ , two efficient approaches are possible:

- Compute the distance transform image [18] of the original shape and approximate  $d(t)$  by linear interpolation of the distance image at the real point  $x(t), y(t)$ .
- Store all shape contour points in a kd-tree [19] and compute (17) for each sampling point  $x(t), y(t)$  with a nearest neighbour search in the kd-tree.



**Fig. 4** Representation of a broken shape (grey) by a three-dimensional curve  $(x(t), y(t), d(t))$ .  $(x, y)$  are the coordinates of the convex hull and  $d$  is the distance between convex hull and shape.

To estimate the runtime complexity of both algorithms, let us first observe that a shape with an  $n \times n$  bounding box has  $O(n^2)$  volume pixels, but only  $O(n)$  contour points. As the fastest algorithms for computing the distance transform require two runs over all image pixels [18], the first algorithm requires  $O(n^2)$  operations to compute all contour distances. The second algorithm requires  $O(n \log n)$  operations for building the kd-tree and  $O(n \log n)$  operations for querying all nearest neighbours, resulting in a total runtime of  $O(n \log n)$ . The second approach is thus faster, and we have implemented it with the kd-tree library shipped with the Gamera framework [19].

### 3.2 Broken shape Fourier descriptors

For the derivation of invariant Fourier descriptors for the three dimensional point sequence  $(x(t), y(t), d(t))_{t=0}^{N-1}$ , we propose three different approaches. Our first Fourier descriptor is built upon the techniques of Sec. 2.1. It builds a complex number by taking the centroid distance  $r(t) := |(x(t), y(t)) - (x_0, y_0)|$ , see Eq. (14), as real part and the distance  $d(t)$  as imaginary part. The sequence  $(r(t) + jd(t))_{t=0}^{N-1}$  is already invariant under translation and rotation. Scale and start point shift invariance of the Fourier coefficients

$$c_k := \frac{1}{N} \sum_{t=0}^{N-1} [r(t) + jd(t)] \exp(-j2\pi kt/N) \quad (18)$$

is either achieved with the phase normalisation (compare Eq. (16))

$$A_k := \exp(js\alpha_k - jk\alpha_s) \frac{|c_k|}{|c_r|} \quad (19)$$

or, simply, by using the absolute values  $|A_k|$ , where  $\alpha_k = \arg(c_k)$  denotes the polar angle of coefficient  $c_k$ , and  $c_r$  and  $c_s$  are the two coefficients with the largest and second largest absolute value ( $r \geq 0, 0 < s < N/2$ ), and  $\alpha_s = \arg(c_s)$  is the polar angle of  $c_s$ . For a fixed number of  $n$  descriptors, the first  $n$  values in the sequence  $A_0, A_{N-1}, A_1, A_{N-2}, \dots$  should be selected.

The second Fourier descriptor under investigation follows the multidimensional approach by Badreldin et al. as described in Sec. 2.2. Let  $c_k^{(x)}$ ,  $c_k^{(y)}$ , and  $c_k^{(d)}$  be the complex Fourier coefficients of the three dimensions  $x(t)$ ,  $y(t)$ , and  $d(t)$  according to (8). Invariant Fourier descriptors are then obtained as ( $k = 1, 2, \dots, \lceil \frac{N-1}{2} \rceil$ )

$$B_k := \frac{\sqrt{|c_k^{(x)}|^2 + |c_k^{(y)}|^2 + |c_k^{(d)}|^2}}{\sqrt{|c_1^{(x)}|^2 + |c_1^{(y)}|^2}} \quad (20)$$

For a fixed number of  $n$  descriptors, the first  $n$  values  $B_1, B_2, \dots, B_n$  should be selected.

The third descriptor uses the scalar representation  $r(t) - d(t)$  that already is invariant under translation and rotation. It is an approximation to the local radius of the shape and leads to the Fourier coefficients

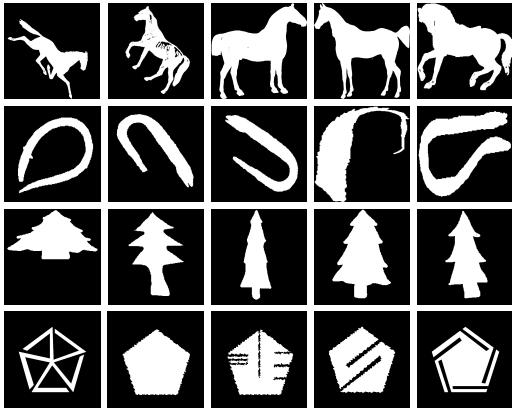
$$c_k = \frac{1}{N} \sum_{t=0}^{N-1} [r(t) - d(t)] \exp(-j2\pi kt/N) \quad (21)$$

As  $r(t) - d(t)$  are real values, the Fourier coefficients at  $k$  and  $n - k$  are complex conjugates:  $c_k = c_{N-k}^*$ ,  $1 \leq k < N$ . Therefore, only values for  $0 \leq k \leq \lceil \frac{N-1}{2} \rceil$  are relevant. Again, the coefficients (21) can be made scale and startpoint shift invariant either with the phase normalisation

$$C_k := \exp(js\alpha_k - jk\alpha_s) \frac{|c_k|}{|c_r|} \quad (22)$$

name	symbol	definition
complex position	$l_k$	Eq. (6)
Granlund	$d_k$	Eq. (7)
Elliptic	$I_k, J_k, K_k$	Eq. (12)
real position	–	Eq. (13)
centroid distance	$R_k$	Eq. (16)
broken A	$A_k$	Eq. (19)
broken B	$B_k$	Eq. (20)
broken C	$C_k$	Eq. (22)

**Table 1** Names and symbols used for the Fourier descriptors in the present study.



**Fig. 5** Example shapes from the MPEG-7 CE-1 part B data set. Samples in each row belong to the same class.

or, simply, by using the absolute values  $|C_k|$ , where  $\alpha_k = \arg(c_k)$  denotes the polar angle of coefficient  $c_k$ , and  $c_r$  and  $c_s$  are the two coefficients with the largest and second largest absolute value ( $r \geq 0, 0 < s < N/2$ ). For a fixed number of  $n$  descriptors, the first  $n$  values  $C_0, C_1, C_2, \dots$  should be selected.

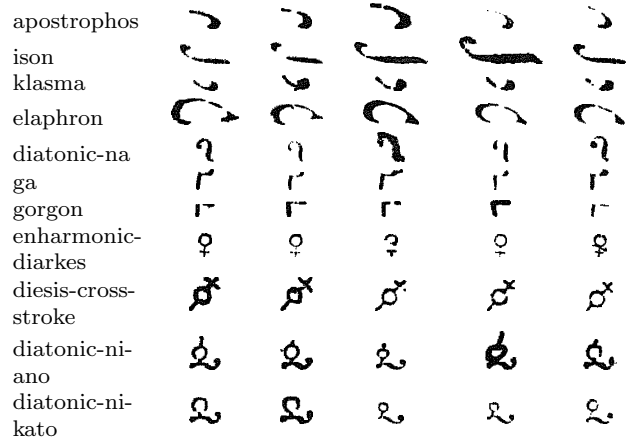
## 4 Evaluation

We have evaluated our new Fourier descriptors on two different data sets, the MPEG-7 database of unbroken shapes and a new real world data set with broken shapes from scans of 19th century chant books in the Eastern neumatic notation [20]. Both data sets are described in detail in Sec. 4.2. Apart from a performance comparison of the broken shape descriptors (Sec. 4.5), we have also investigated the effect of different normalisation schemes (Sec. 4.3) and the number of descriptors needed for similarity based retrieval (Sec. 4.4).

We have implemented all Fourier descriptors as a toolkit for the Gamera framework for document analysis and recognition<sup>2</sup> [21]. The toolkit is published under a free license together with the new data set of broken neumes in the ‘‘Addons’’ section of the Gamera website<sup>3</sup>. For convenience, a brief summary of all Fourier descriptors under investigation is given in Table 1.

### 4.1 Performance measures

As evaluation criteria for shape based image retrieval, we have used two different performance measures, the precision/recall curve and the leave-one-out error rate of a  $k$ -nearest neighbour ( $k$ -NN) classification. For a single query image belonging to class  $\omega$ , *precision* and



**Fig. 6** Unrotated example shapes from our newly created NEUMES data set. Samples in each row belong to the same class.

*recall* are defined as follows: let  $n_\omega$  be the number of all images of class  $\omega$ , and let  $k_\omega$  be the number of images of class  $\omega$  among the  $k$  nearest neighbours of the query image; then  $k_\omega/k$  is the *precision* and  $k_\omega/n_\omega$  is the *recall* for this query. The precision of all test images is averaged to yield a single precision value. Typically  $k$  is not fixed, but the precision is measured for a given recall rate. When the recall is increased, the precision will generally decrease, but less so for better similarity measures. The decrease of the precision/recall curve can thus serve as a performance measure for similarity based retrieval.

To evaluate the classification performance of the different Fourier descriptors, a natural criterion is the *cross-validation* or *leave-one-out* error rate of a  $k$ NN classifier because it is an unbiased estimator of the expected error rate [22]. A *kNN classifier* assigns a test sample to the majority class among its  $k$  nearest training samples. The leave-one-out error rate is the average error rate when each sample is classified with a  $k$ NN classifier that has been trained with the remaining  $n-1$  samples, thereby yielding a single performance measure.

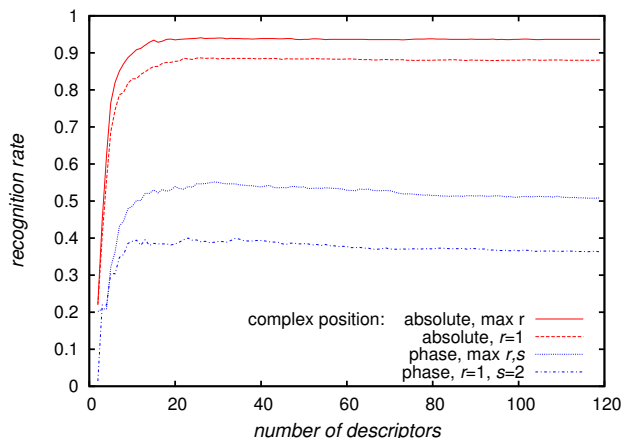
### 4.2 Data sets

A data set that has already been used for the evaluation of different Fourier descriptors in the study [6] is part B of the MPEG-7 CE-Shape-1 database<sup>4</sup> [23]. It consists of 1400 shapes that have been classified into 70 classes with 20 similar items in each class. Fig. 5 shows sample shapes from this data set. As pointed out by the authors of the data set, a 100% retrieval rate is impossible because some shapes are more similar to the shapes

<sup>2</sup> <http://gamera.sf.net/>

<sup>3</sup> <http://gamera.informatik.hsnr.de/addons/fd/>

<sup>4</sup> <http://www.dabi.temple.edu/~shape/MPEG7/>



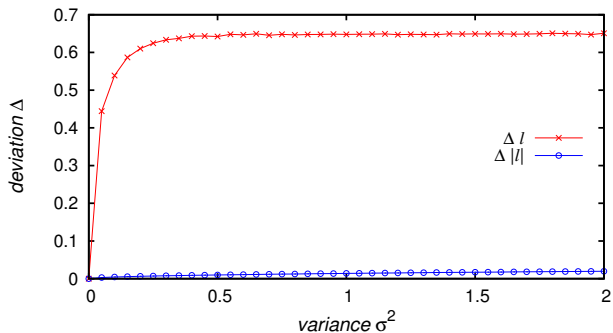
**Fig. 7** Impact of different normalisations on the *complex position* Fourier descriptor from Eq. (6). Recognition rates have been measured by leave-one-out with a kNN classifier ( $k = 1$ ) on the MPEG-7 data set.

from different classes than to their own class, so that “it is not possible to group them into the same class using only shape knowledge” [23]. In some images, there are noise pixels which form additional small random shapes. In order to ignore this noise, we have computed the contour of the largest connected component for each image only.

The MPEG-7 data set does not contain any broken shapes, and thus allowed for a performance comparison of the new descriptors with the ordinary Fourier descriptors described in Sec. 2. To also test the new descriptors from Sec. 3.2 on actual broken shapes, we have created a data set of broken glyphs from the four 19th century music prints in Byzantine neume notation that have also been used in [20] (sources HA-1825, HS-1825, AM-1847, and MP1-1850). This “NEUMES” data set consists of 640 images out of 40 different classes with 16 items in each class. Due to varying print quality, some glyphs are connected while others are randomly broken into up to eight fragments. As can be seen in Fig. 6, some neumes are mirrored or elongated versions of different neumes. It is thus important that the shape descriptors used for discrimination are not invariant to axial mirroring or arbitrary affine transformations. The sample images in Fig. 6 are not rotated; to make rotational invariance of the shape descriptors mandatory, we have rotated the 16 items in each class in steps of  $22.5^\circ$ .

### 4.3 Normalisation schemes

The Fourier descriptors from Sec. 2 can be normalised (i.e., made invariant) in different ways. There are generally two degrees of freedom:



**Fig. 8** Impact of random disturbances on the *complex position* Fourier descriptors  $l_k$ . Disturbances have been simulated as Gaussian random noise with variance  $\sigma^2$ .

- Phase normalisation versus absolute values
- The index choices  $s$  and  $r$  for the normalisation coefficients  $c_r$  and  $c_s$

Fig. 7 shows the effect of the different normalisation schemes on the leave-one-out recognition rate on the MPEG-7 data set for the *complex position* Fourier descriptor. Both for the absolute values and the phase normalised descriptors, it is better to normalise not with a fixed coefficient, but with the coefficient with the largest absolute value (which may vary from shape to shape). This normalisation limits the numeric range of the descriptors to a fixed interval, a feature normalisation scheme that is known to improve the recognition rate in many cases [24].

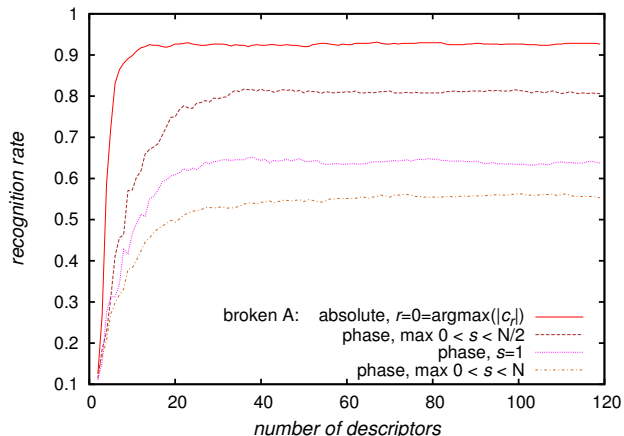
The observation that the phase normalisation performs poorer than the absolute values is surprising, however, because the phase normalised descriptors carry information that is lost in the absolute values. It turned out, however, that the phase angles of the descriptors are much less robust with respect to small changes in the contour coordinates. To demonstrate this phenomenon, we did a small Monte Carlo experiment. We added normally distributed random noise independently to the  $x$  and  $y$  coordinates of a sample contour and measured the deviation  $\Delta$  of the resulting descriptors  $\tilde{l}_k$  and  $|\tilde{l}_k|$  as

$$\Delta l := \frac{1}{L} \sum_{k=1}^m |l_k - \tilde{l}_k| + |l_{N-k} - \tilde{l}_{N-k}|$$

$$\Delta |l| := \frac{1}{L} \sum_{k=1}^m ||l_k| - |\tilde{l}_k|| + ||l_{N-k}| - |\tilde{l}_{N-k}||$$

where  $l_k$  are the undisturbed descriptors and  $L := \sum_{k=1}^m |l_k| + |l_{N-k}|$ . Fig. 8 shows these deviations, averaged over 10 000 random experiments, as a function of the variance  $\sigma^2$  of the random noise. The phase normalisation obviously is much less robust. The same phenomenon already occurs for the Fourier coefficients  $c_k$





**Fig. 9** Impact of different normalisations on the *broken A* Fourier descriptor from Eq. (19). Recognition rates have been measured by leave-one-out with a kNN classifier ( $k = 3$ ) on the NEUMES data set. Note that the curves for  $r = 0$  and  $r = \text{argmax}\{|c_r|\}$  are identical.

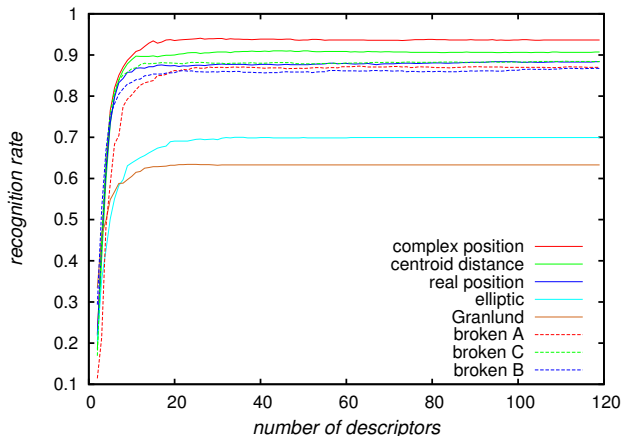
due to  $||c_k| - |\tilde{c}_k|| \leq |c_k - \tilde{c}_k|$ , and this is amplified by the phase normalisation (6) because the phase angles are multiplied with large integer values ( $s - r$ ,  $k - s$ , and  $r - k$ ).

Further evidence for the instability of the phase angles can be derived from Fig. 9, which shows the recognition rates for different normalisations of the Fourier descriptor *broken A* on the NEUMES data set. When the phase normalisation coefficient  $c_s$  is chosen as the largest coefficient for  $0 < s < N$ , this often results in a high value  $s \approx N - 1$ . This amplifies the phase angle error of  $\alpha_k = \arg c_k$  because  $\alpha_k$  is multiplied with  $s$  in the phase normalisation (19), thereby even resulting in a negative effect on the recognition rate compared to a fixed normalisation with  $s = 1$ . When the maximum coefficient  $c_s$  is only searched for small  $s$  (in most cases this led to  $s = 2$ ), the recognition rate is considerably better. Nevertheless, in any case, the absolute values performed yet better.

We therefore conclude that it is generally better to use the absolute values instead of the phase normalised coefficients, and that the scale invariance normalisation should be done with the coefficient  $c_r$  with the largest absolute value rather than with fixed  $r$ .

#### 4.4 Number of descriptors

Figs. 7 & 9 show that a small number of Fourier descriptors is sufficient for shape retrieval. In our experiments, this behaviour was universal, as can be seen in Fig. 10: using more than 20 descriptors generally does not increase the recognition rate any further. In our experiments in the following subsection, we have there-

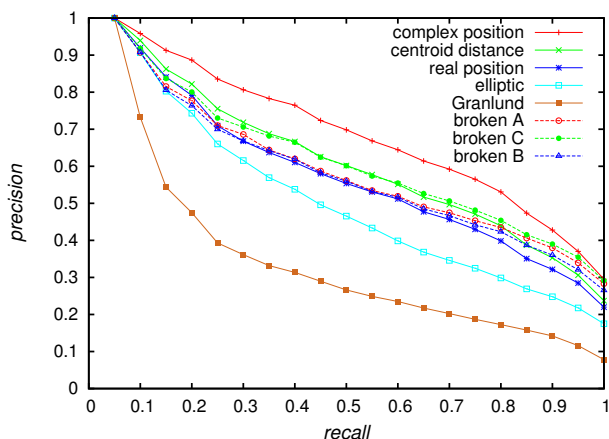


**Fig. 10** Leave-one-out recognition rates of a kNN classifier ( $k = 1$ ) on the MPEG-7 data set. All descriptors have been normalised with the largest coefficient  $c_r$  and the absolute values have been taken.

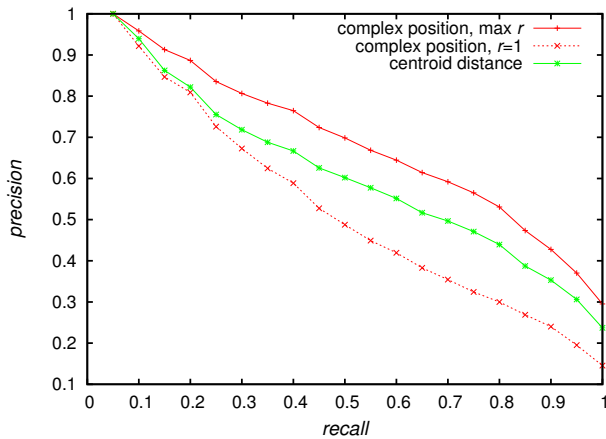
fore limited the number of descriptors to 60, to be on the safe side. It is interesting to note that these numbers, which are derived from the leave-one-out recognition rates, are much lower than the numbers derived by Zhang & Lu from the absolute magnitude of the descriptors [6]. The reason for this difference is that a criterion based on the magnitude does not take the discriminative power of the coefficients into account.

#### 4.5 Descriptor comparison

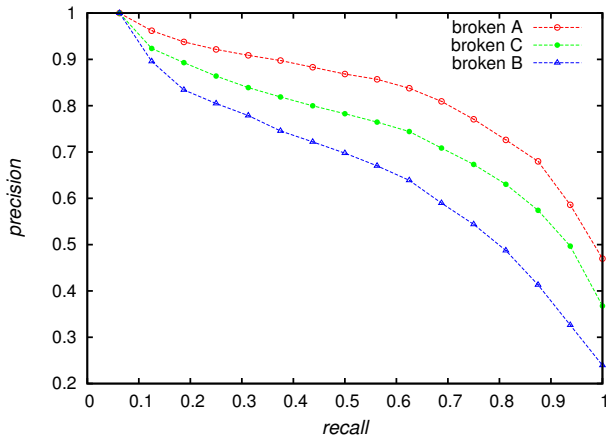
Fig. 11 shows the precision/recall curves for all investigated Fourier descriptors on the MPEG-7 data set. To all descriptors, the recommendations from the preceding subsections have been applied, i.e. they have been normalised with the largest coefficient and by taking



**Fig. 11** Precision/recall curves for all Fourier descriptors on the MPEG-7 data set. In all cases, 60 descriptors have been used.



**Fig. 12** Comparison between *complex position* and *centroid distance*. The *complex position* Fourier descriptor is only better than the *centroid distance* when the normalisation coefficient  $c_r$  is chosen as the largest coefficient (“max  $r$ ”).



**Fig. 13** Precision/recall curves for the broken shape Fourier descriptors on the NEUMES data set. In all cases, 60 descriptors have been used.

the absolute value, and the first 60 descriptors have been used.

The best performing Fourier descriptor was the *complex position*, which seems to be a contradiction to the experiments by Zhang & Lu [6], who found the *centroid distance* to be best performing. This discrepancy can however be explained with the different choice of the normalisation coefficient  $c_r$ , as shown in Fig. 12: when the *complex position* Fourier descriptor is normalised with a fixed coefficient, e.g.  $r = 1$ , it performs poorer than the *centroid distance*, as was the case in the study by Zhang & Lu.

On the MPEG-7 data set, the broken shape Fourier descriptors did not perform as well as the best single shape Fourier descriptor (*complex position*), but the precision/recall curve of our *broken C* descriptor is almost identical to the *centroid distance* Fourier de-

scriptor, with *broken A* and *broken B* performing only slightly poorer. That the *broken C* and *centroid distance* descriptors behave very similar is hardly surprising, because for single closed shapes the signature  $r(t) - d(t)$  is simply an approximation of the signature (14).

On the NEUMES data set, only our new Fourier descriptors are applicable, and the resulting precision/recall curves are shown in Fig. 13. On this data set, there is a more distinct difference between the three broken Fourier descriptors. Actually, the ranking is in ascending order of the information loss of the descriptor: mapping the complex number  $r + jd$  (*broken A*) onto the real number  $r - d$  (*broken C*) loses some information, and the *broken B* descriptor loses even more shape information because it decouples the  $x$ ,  $y$ , and  $d$  coordinates. On the MPEG-7 data set, this has a smaller impact on the recognition performance because the shapes within a single class vary considerably. On the NEUMES data set, in contrast, this is of importance because detailed shape information is required for class discrimination; see e.g. the two bottom rows in Fig. 6.

## 5 Conclusions

The new Fourier descriptors for broken shapes have shown retrieval performances that were comparable to common closed contour shape descriptors like the “*centroid distance*” Fourier descriptor. As the new descriptors have the benefit of being applicable to arbitrary shapes (connected or broken), they can serve as a general replacement for other Fourier descriptors lacking this flexibility.

Our experiments have shown that it is generally better to use the absolute values rather than the phase normalised descriptors and that scale invariance normalisation should be done with the largest coefficient, rather than with a fixed coefficient. For practical applications on real world data, we would recommend to use the “*broken A*” Fourier descriptor.

## Competing interests

The authors declare that they have no competing interests.

## Acknowledgements

We are grateful to Georgios K. Michalakis for providing us with scans of Byzantine chant books that we have used for extracting the NEUMES test data set.

## References

1. Crimmins, T.: A complete set of Fourier descriptors for two-dimensional shapes. *Systems, Man and Cybernetics, IEEE Transactions on* **12**, 848–855 (1982)
2. Ghorbel, F., Derrode, S., Mezhoud, R., Bannour, T., Dhahbi, S.: Image reconstruction from a complete set of similarity invariants extracted from complex moments. *Pattern Recognition Letters* **27**, 1361–1369 (2006)
3. Gonzalez, R., Woods, R.: *Digital Image Processing*, 2nd edn. Prentice-Hall, New Jersey (2002)
4. Khotanzad, A., Hong, Y.H.: Invariant image recognition by Zernike moments. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **12**, 489–497 (1990)
5. Mokhtarian, F., Abbasi, S., Kittler, J.: Robust and efficient shape indexing through curvature scale space. In: *British Machine Vision Conference*, pp. 53–62 (1996)
6. Zhang, D., Lu, G.: Study and evaluation of different Fourier methods for image retrieval. *Image and Vision Computing* **23**, 33–49 (2005)
7. Arbter, K., Snyder, W., Burkhardt, H., Hirzinger, G.: Application of affine-invariant Fourier descriptors to recognition of 3-D objects. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **12**, 640–647 (1990)
8. Pavlidis, T.: *Algorithms for Graphics and Image Processing*. Springer, New York (1982)
9. Zygmund, A.: *Trigonometric Series*, 3rd edn. Cambridge University Press (2002)
10. Dimov, D., Laskov, L.: Invariant Fourier descriptors representation of medieval Byzantine neume notation. In: *Joint COST 2101 & 2102 International Conference on Biometric ID Management and Multimodal Communication*, pp. 192–199 (2009)
11. Granlund, G.: Fourier preprocessing for hand print character recognition. *IEEE Transactions on Computers* **21**, 195–201 (1972)
12. Kuhl, F., Giardina, C.: Elliptic Fourier features of a closed contour. *Computer Graphics and Image Processing* **18**, 236–258 (1982)
13. Lin, C.S., Hwang, C.L.: New forms of shape invariants from elliptic Fourier descriptors. *Pattern Recognition* **20**, 535–545 (1987)
14. Badreldin, A., Wong, A., Prasad, T., Ismail, M.: Shape descriptors for n-dimensional curves and trajectories. In: *International Conference on Cybernetics and Society*, pp. 713–717 (1980)
15. Shridhar, M., Badreldin, A.: High accuracy character recognition using Fourier and topological descriptors. *Pattern Recognition* **17**, 515–524 (1984)
16. Zahn, C., Roskies, R.: Fourier descriptors for plane closed curves. *IEEE Transactions on Computers* **C-21**, 269–281 (1972)
17. Cormen, T., Stein, C., Leiserson, C., Rivest, R.: *Introduction to algorithms*, 2nd edn. MIT Press (2001)
18. Bailey, D.: An efficient euclidean distance transform. In: *10th International conference on Combinatorial Image Analysis (IWCI)*, pp. 394–408 (2004)
19. Dalitz, C.: Kd-trees for document layout analysis. In: *Document Image Analysis with the Gamera Framework, Schriftenreihe des Fachbereichs Elektrotechnik und Informatik, Hochschule Niederrhein*, vol. 8, pp. 39–52. Shaker-Verlag (2009)
20. Dalitz, C., Michalakis, G., Pranzas, C.: Optical recognition of psaltic byzantine chant notation. *International Journal of Document Analysis and Recognition* **11**, 143–158 (2008)
21. Droettboom, M., MacMillan, K., Fujinaga, I.: The gamera framework for building custom recognition systems. In: *Symposium on Document Image Understanding Technologies*, pp. 275–286 (2003)
22. Raudys, S., Jain, A.: Small sample size effects in statistical pattern recognition: recommendations for practitioners. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* **13**, 252–264 (1991)
23. Latecki, L., Lakamper, R., Eckhardt, T.: Shape descriptors for non-rigid shapes with a single closed contour. In: *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 424–429 (2000)
24. Guan, H., Antani, S., Long, L., Thoma, G.: Comparative study of shape retrieval using feature fusion approaches. In: *Computer-Based Medical Systems (CBMS), 2010 IEEE 23rd International Symposium on*, pp. 226–231 (2010)