



Phase preserving Fourier descriptor for shape-based image retrieval



Emir Sokic*, Samim Konjicija

^a Faculty of Electrical Engineering, University of Sarajevo, Zmaja od Bosne bb, 71000 Sarajevo, Bosnia and Herzegovina

ARTICLE INFO

Available online 22 November 2015

Keywords:

Content based image retrieval
Fourier descriptors
Phase
Nominal orientation
Pseudomirror points

ABSTRACT

Shape is one of the most important discriminative elements for the content based image retrieval and the most challenging for quantification and description. Fourier descriptors are a very efficient shape description method used in shape-based image retrieval tasks. In order to achieve invariance under rotation and starting point change, most Fourier descriptor implementations disregard the phase of Fourier coefficients, consequently losing valuable information about the shape. This paper proposes a novel method of extracting Fourier descriptors that preserve the phase of Fourier coefficients. We introduce specific points, called *pseudomirror* points, and use them as a shape orientation reference. They facilitate the extraction of phase-preserving Fourier descriptors which are invariant under translation, scaling, rotation and starting point change. The proposed descriptor was tested on four popular benchmarking datasets: MPEG7 CE-1 Set B, Swedish leaf, ETH-80 and Kimia99 datasets. Performance and computational complexity measures indicate that the proposed method outperforms other state-of-the-art phase-based Fourier descriptors. In addition, it outperforms other state-of-the-art magnitude-based Fourier descriptors, and many non-Fourier based shape description methods in terms of performance – complexity ratio.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

The shape of the presented objects on images is an important feature for image understanding. Thus, shape is widely used as a discriminative element in the field of content-based image retrieval (CBIR). In many applications, shape captures most of the perceptual information of the observed objects on images, while color and texture can often be omitted without affecting retrieval performance. Unfortunately, shape may be subject to significant changes, such as deformation, scaling, changes in orientation, noise, and partial concealment. Hence, accurate

description of the shape remains a challenging technical problem.

A variety of shape description techniques have been developed over the years [1]. Best shape descriptors are typically described using the following attributes: compact, easy to compute, informative, discriminative, tolerant to geometric transformations, efficient. It is very hard to satisfy all these requirements. The aim of many researchers in this field is to improve descriptor performance and reduce computational costs.

Shape description methods usually belong to one of the following four classes: (1) global, (2) local, (3) combined global and local methods, and (4) post-processing/learning shape similarity methods.

Global methods capture the object's global shape information and are relatively fast to compute and

* Corresponding author. Tel.: +387 33 250 700.
E-mail address: esokic@etf.unsa.ba (E. Sokic).

compare. Although they are robust to noise, they exhibit poor performance when it comes to discriminating occluded shapes or do partial matching. In contrast, local techniques precisely represent local shape features. However, they are sensitive to noise and often more computationally complex than global techniques. The choice between local and global descriptor is context-dependent. Recently, a palette of so-called “rich” (as called by Wang et al. [2]) descriptors has been proposed. These descriptors combine both local and global shape characteristics. Therefore, they obtain good retrieval rates on popular benchmark datasets, at a cost of more complicated descriptor computation and matching. The last group of shape descriptors consists of those that use either training sets and machine learning techniques to “learn” shape similarity, or use information about other shape retrievals in order to increase the retrieval rate.

In this paper, Phase-including Fourier Descriptor (PIFD) is introduced. It is a novel global, spectrally hierarchical, information-preserving, contour-based shape description technique. Development of PIFD is inspired by the work on magnitude-based Fourier descriptor (FD) by Zhang and Lu [3–5] and by a recent work of Sokic and Konjicija [6]. We use the term magnitude-based FDs to denote FDs which use only the magnitude of Fourier coefficients. Magnitude-based FD is established as very compact, efficient and effective global shape descriptors. We propose to improve the retrieval performance of FD even more, while attaining the same computational complexity. This is obtained by preserving the phase of the Fourier coefficients. In order to achieve invariance of the Fourier coefficients under rotation and starting point change, we propose to use so-called *pseudomirror* points for determining the nominal orientation of the shape. Obtained experimental results demonstrate that PIFD is superior to other Fourier descriptor based methods in terms of retrieval performance and discrimination ability. Although being a global technique, experiments confirm that PIFD is also comparable to other state-of-the-art shape description methods, while having much lower computational complexity.

The main contributions of this paper are

- introduction of a novel method for determining nominal shape orientation(s),
- development of a versatile phase-preserving Fourier descriptor which is invariant under translation, rotation, scaling, starting point change and optionally mirror transformations,
- development of a shape descriptor with one of the best retrieval performance- computational complexity ratios.

The rest of the paper is organized as follows. A brief review of related work is given in Section 2. In Section 3, Phase-including Fourier Descriptor (PIFD) is introduced. Experimental results are demonstrated and discussed in Section 4. Concluding remarks and future research directions are given in the last section.

2. Related work

Shape description methods may be classified into two groups: region based and contour based [1,7]. Region based techniques use the boundary of the shape as well as the interior of the shape, while contour based techniques take into account only the contour of the shape. Contour based approaches are generally more compact, faster and sometimes perform better than region based methods. On the other hand, contour based methods find it difficult to identify shapes which consist of disjoint parts, such as trademarks, logos and characters [8].

The contour of the shape is commonly described using shape signatures. They are one-dimensional functions which capture most of the perceptual features of the shape [9]. Shape signatures are sensitive to noise and distortions and often dependent on rotation, translation and scaling. To overcome these problems, different transformations are conducted over shape signatures.

Fourier descriptors (FD) are obtained by applying the discrete Fourier transform (DFT) over a shape signature [3]. By disregarding phase information, descriptors become invariant under rotation, translation, scale and change of the starting point of the contour. They also show good retrieval accuracy, compactness, insensitivity to noise and have a hierarchical representation in spectral domain.

Magnitude-based Fourier descriptors have been derived from several shape signatures: Complex coordinates [6], Centroid/Radial distance, Tangent angle [5], Curvature function, Area function, Triangle-area representation [10], Triangular centroid area, Chord length [4], Polar coordinates, Farthest point distance [11], Perimeter area function [12], Improved arc-height function [13], Rectangle centroid distance [14] and many others.

Apart from being a global shape description technique and having hierarchical representation only in spectral domain, there are essentially two main disadvantages of magnitude-based Fourier descriptors.

Fourier descriptors are not information-preserving, which means that the original shape cannot be reconstructed from descriptor coefficients. Therefore, they are not suitable for shape evolution problems or shape retrieval tasks where rotation invariance is not desirable (e.g. traffic signs recognition [15]). However, information-preserving property is not always a required attribute of a CBIR system, hence the usability of the descriptor is not significantly limited. Moreover, this drawback is compensated by their compact notation and simple matching.

The most important drawback of magnitude-based Fourier descriptors is that they disregard phase information in order to obtain invariance under rotation and starting point change. Using this simplistic approach, valuable information contained in shape description is lost. To illustrate this fact, two completely different shapes with equal magnitudes of Fourier coefficients are depicted in Fig. 1. Interestingly, Oppenheim and Lim [16] in 1981. showed that phase contains a lot of valuable information about the shape and that even, with specific initial assumptions, images may be reconstructed using the

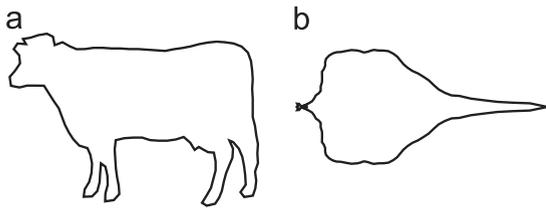


Fig. 1. Reconstruction of shapes with equal FD magnitudes: (a) shape reconstructed with the original phase content, (b) shape reconstructed with different phase content.

phase information only. Nevertheless, there exist only few papers that considered preserving the phase of the Fourier coefficients.

Arbter et al. [17] proposed to exploit the phase of FD with their Affine-Invariant Fourier descriptor (AFD), which is inherently invariant under affine transformations. The invariance of AFD to starting point change is essentially deduced from the estimate of the starting point shift, by using phase differences of selected harmonics. A low order estimate is actually based on the first and second harmonic, while for higher orders an appropriate Diophantine equation needs to be solved.

There were several authors that also used the first harmonic for phase normalization, such as Persoon and Fu [18] (similar approach to Kuhl and Giardina [19]), Li et al. [20] and Fieguth et al. [21]. Li et al. [20] used a two-dimensional Fourier descriptor, but also exploited first harmonic's phase for normalization. Fieguth et al. [21] proposed a generalization, in a sense that the desired rotation and starting point change invariance are achieved using the phase of the most dominant harmonic. In practice, this usually leads to phase normalization using the first or second harmonic, since the latter have the largest magnitude.

In their work, Bartolini et al. [22] used the Fourier coefficients in order to normalize the contour scale and position. They proposed to normalize the phase of the harmonics using the nominal phase of the first order harmonic as a reference. Instead of matching Fourier descriptors in spectral domain, Bartolini et al. suggested to use Constrained Dynamic Time Warping (CDTW) of the contours in spatial domain, whereas Fourier descriptors are only used for contour normalization. CDTW is a more computationally demanding similarity measure, but unlike the Euclidean distance, it allows constrained stretching of the contour.

It is noticeable that the usage of the first harmonic's phase is a relatively common method of phase normalization. Unfortunately, if the first harmonic has a neglectable magnitude or noise is present, these methods usually fail. Normalization with the first harmonic is an ill-conditioned task, because negligible errors in the computation of first order harmonic's phase may cause large variations of phase of higher orders normalized harmonics. As Larsson et al. noted in [15], using first order harmonics for phase estimation depends on the type of the first harmonic locus. For the elliptic harmonic locus this method performs well, whereas for a circular harmonic locus it might accidentally create arbitrary orientation.

Furthermore, very thin and lengthy structures are more or less invisible to the first order harmonic, but have a huge impact on the spatial orientation estimation.

Instead of normalizing the phase, Larsson et al. preserved the phase of FD without previous normalization and used cross-correlation for shape matching [15]. Our method of preserving phase is similar to the work of Larsson et al. [15], but instead of using L_2 -norm during scale normalization, we have adopted L_1 -norm. Moreover, we use a modified City-block distance in spectral domain instead of cross-correlation in spatial domain. The method presented in [15] is based on finding the best possible match between the two shapes using every possible rotation (shift) of one shape. This is not an optimal procedure, since it may affect the discriminability among shapes. For example, certain shapes from the same class may be de-rotated for different angles in order to align with another query shape. Therefore, shapes from the same class are compared to others using different orientations. This nonuniformity leads to a lower discriminability among shapes, and consequently to lower performance of the descriptor. Instead, we introduce a novel method to determine nominal orientation(s) of the shape. These orientations are inherent to the shape, embed some shape semantics in Fourier descriptors and consequently contribute to better discriminability.

Apart from FD-based methods, other important contour based techniques are multiscale-based [23,24], Curvature Scale Space (CSS) approaches [25,26], Multiscale Convexity Concavity (MCC) [27], and Curve Edit Distance [28].

Popular structural methods that recently appeared are Shape tree [29], Hierarchical Procrustes Matching (HPM) [30], Contour Flexibility [31], Shape Context [32] followed by Inner-Distance Shape Context [33] and Height functions [2]. These methods are based on "rich" shape descriptors. At a cost of a higher computational complexity, they achieve exceptionally good results on a popular MPEG7 CE-1 Set B dataset [34]. In addition, most of these techniques require complex matching schemes and often use non-symmetrical similarity measures.

Recently, a number of context-sensitive methods appeared in CBIR. Context-sensitive or contextual similarity methods [35–38] try to learn pairwise similarities among the database (context) and use this context to infer semantic similarities between shapes against the database. Methods such as Label Propagation (LP), Locally Constrained Diffusion Process (LCDP), Meta Similarity (MS), Contextual Dissimilarity Measure (CDM) are developed, to name just a few. Using these techniques, Donoser and Bischof [39] obtained a first time ever 100% Bull-Eye score on MPEG7 CE-1 Set B. CBIR systems based on these methods achieve the best retrieval performance results, but they require either additional specific knowledge of the dataset, or prior training of the retrieval system.

3. Phase-Including Fourier Descriptor (PIFD)

The shapes that are analyzed in this paper are outline shapes, which can be described as single plane closed (discrete) curves. First, in preprocessing stage, the

coordinates of the shape boundary are extracted from the image. In order to apply a shape description method, the contour of the shape is re-sampled with a fixed number of points (N) using equal arc-length sampling [5].

For subsequent analysis, it will be assumed that a shape contour is given with N boundary points $P_n = (x_n, y_n)$ where $n = 0, 1, \dots, N-1$. The contour points $P_n = (x_n, y_n)$ will be represented in the form of complex numbers:

$$Z_n = x_n + jy_n, \quad (1)$$

for which the Discrete Fourier Transform may be computed as in the following equation:

$$a_k = \frac{1}{N} \sum_{n=0}^{N-1} Z_n e^{-j2\pi nk/N}, \quad (2)$$

where $k = 0, 1, \dots, N-1$. Fourier coefficients a_k are used to derive Fourier descriptors. The coefficients a_k must be additionally transformed to be invariant under translation, rotation, scale and starting point change.

It can be easily shown that invariance under rotation and starting point change is obtained using only the magnitude of the Fourier coefficients, while the invariance under translation is achieved by disregarding the DC component (coefficient a_0). In order to introduce scale invariance, all earlier implementations ([5,11–13,22] and many others), have used $Sc = |a_1|$ as the scale normalization coefficient. Sokic and Konjicija [6] have shown that using a different scaling coefficient:

$$Sc = \sum_{i=1}^{N-1} |a_i|, \quad (3)$$

yields a better retrieval performance. A scale normalization approach given in [6] is similar to the one presented in [15]. The authors in [6] propose to use L_1 -norm of the Fourier descriptor (Eq. (3)) for scale normalization and City-block distance as similarity measure, whereas the authors in [15] propose to use L_2 -norm for scale normalization and Euclidean distance as a metric. Experimental results showed that using L_1 -norm and City-block distance yields a more efficient descriptor, therefore we adopted the scaling coefficient given with (3) in the further analysis.

Although Normalized Complex Coordinates (NCC) FD presented in [6] have promising performance and compactness vs. computational time ratio, they suffer from the same problems of all magnitude-based FD – the information contained in the phase is discarded.

Therefore, we start with the following translation and scaling invariant descriptor:

$$\mathbf{F}_{TS} = \left\{ \frac{a_{-M/2}}{Sc}, \dots, \frac{a_{-1}}{Sc}, \frac{a_1}{Sc}, \frac{a_2}{Sc}, \dots, \frac{a_{M/2}}{Sc} \right\}, \quad (4)$$

where a_i and Sc are computed using (2) and (3), and M is the number of Fourier coefficients needed for representation. Number M is usually small (≤ 30) and independent of the number of points N . Note that the DFT is a periodic sequence with period N ($a_{-m} = a_{N-m}$).

Descriptor given with (4) is invariant under translation and scaling, which is easy to demonstrate as in [6]. Unlike NCC FD (4), the \mathbf{F}_{TS} descriptor is information preserving, so

the initial shape could be hierarchically reconstructed for a different number of Fourier coefficients (M).

In order to use information contained in the phase of Fourier coefficients, one must obtain their invariance under starting point and orientation change. This could be easily achieved if a nominal orientation and the starting point of the shape contour are known. Methods based on using landmarks in spatial domain (such as the axis of least inertia [40] or the point with maximum radius [19], to name just a few) are methods sensitive to noise, shape deformations, spurious peaks etc. In addition, approaches based on aligning shapes using the first harmonic (see Section 2) showed to perform well on objects with dominant 1st harmonic and elliptic locus, but rather poorly on other shape types.

In the following subsection, we introduce a novel method for determining the nominal orientation of the shape, which achieves the desired phase invariance.

3.1. Nominal orientation based on pseudomirror points

It is clear that the derived descriptor \mathbf{F}_{TS} given by (4) is not invariant under rotation and starting point change. Without the loss of generality, it may be assumed that the rotation of the shape is always conducted around the centroid of the shape: $Z_C = (1/N) \sum_{n=0}^{N-1} Z_n$. The Fourier coefficients of the shape with the starting point P_m , rotated for an angle ϕ , are given by

$$a_k^{(new)} = e^{i\phi} e^{j2\pi km/N} a_k^{(old)}, \quad (5)$$

where $k \in M_{ind}$, $M_{ind} = \{-M/2, -M/2+1, \dots, -2, -1, 1, 2, \dots, M/2-1, M/2\}$. The set M_{ind} contains all indices of the M low frequency Fourier coefficients. The coefficient a_0 contains no relevant information about the shape other than the position of shape centroid. In order to obtain the invariance under translation the coefficient a_0 is disregarded, thus the set M_{ind} does not contain zero.

It is noticeable from relation (5) that shape rotation affects the phase of all harmonics equally with a constant amount (ϕ), while the change of the starting point manifests as de-rotating the k th harmonic for the angle of $(2\pi/N)km$. Since the phase is computed as 2π -modulus, it is impossible to uniquely identify the unknowns m and ϕ just by observing $a_k^{(new)}$ for $|k| > 1$.

On the other hand, if a starting point is known, this point could be set to have a specific predetermined argument by de-rotating the shape (e.g. zero). This actually means that for the shape with the predefined starting point, rotation invariance of FD is constructed without much effort. An example of shape with several arbitrary orientations and indicated starting points is given in Fig. 2. For each chosen starting point P_m , the contour is rotated by an angle $-\arg(Z_m)$, so that the starting point is always located on the real axis. Once a starting point is uniquely determined for a shape, then the shape becomes invariant under starting point change and rotation, thus the Fourier descriptor given in (4) may be used for shape description. One simplified approach would be the derivation of Fourier descriptor for each shape point considered to be a starting point. Consequently, matching of the shapes may

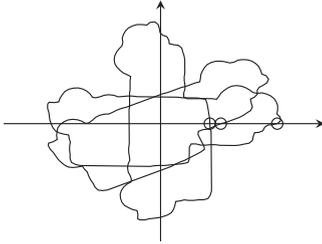


Fig. 2. Several arbitrary orientations of the “Classic car” shape. Starting points are marked with symbol \circ .

be achieved by finding the best match between the FD vector from one shape and all other (rotated) FDs from the other shape. Nevertheless, this would lead to a large dimensionality of the descriptor and to a computationally expensive descriptor matching scheme. Hence, in order to obtain effectiveness and efficiency while preserving compactness, the vertical dimension of the descriptor needs to be reduced.

What may not be noticed at first glance, is that reducing the number of possible starting points would actually improve the intra-class similarity and reduce the cross-class similarity. If one shape is compared to all possible rotations of another shape, it will yield the best possible match between shapes, even if they do not semantically correlate on these nominal orientations.

We propose the method for finding point(s) with specific geometrical and shape discriminative meaning. For each contour point chosen as a starting point, the contour is de-rotated around the centroid of the shape, so that the starting point lies on the positive real axis. This contour will be marked C1. Then another, “mirrored” contour is created. The “mirrored” contour C2 is derived by conjugating all C1-contour points and changing the direction of encircling. The whole process is illustrated in Fig. 3.

Let P_m be arbitrarily chosen as a starting point. For the original contour C1 and for the “mirrored” contour C2, FDs $a_k^{(1)}$ and $a_k^{(2)}$ are derived, respectively. If a_k are the coefficients corresponding to the nominal orientation and nominal starting point, then the coefficients extracted from contour C1 are

$$a_{k(m)}^{(1)} = e^{-j\arg(Z_m)} e^{j2\pi mk/N} a_k. \quad (6)$$

The corresponding points of the mirrored contour C2 are

$$C2 = e^{-j\arg(\bar{Z}_m)} \{\bar{Z}_m, \bar{Z}_{m-1}, \dots, \bar{Z}_0, \bar{Z}_{N-1}, \bar{Z}_{N-2}, \dots, \bar{Z}_{m+1}\}, \quad (7)$$

where \bar{Z} denotes conjugate complex value of Z . It can be shown that

$$\begin{aligned} a_{k(m)}^{(2)} &= e^{-j\arg(\bar{Z}_m)} \left[\frac{1}{N} \sum_{n=0}^m \bar{Z}_{m-n} e^{-j2\pi nk/N} \right. \\ &\quad \left. + \frac{1}{N} \sum_{n=m+1}^{N-1} \bar{Z}_{N+m-n} e^{-j2\pi nk/N} \right] \\ &= e^{-j\arg(\bar{Z}_m)} \left[e^{-j2\pi mk/N} \frac{1}{N} \sum_{r=0}^{N-1} \bar{Z}_r e^{j2\pi rk/N} \right] \\ &= e^{-j\arg(\bar{Z}_m)} e^{-j2\pi mk/N} \bar{a}_k \\ &= e^{j\arg(Z_m)} e^{-j2\pi mk/N} \bar{a}_k \\ &= \bar{a}_{k(m)}^{(1)}. \end{aligned} \quad (8)$$

We introduce the following objective function for determining the starting points:

$$f(m) = \sum_{k=1}^{N-1} |a_{k(m)}^{(1)} - a_{k(m)}^{(2)}|. \quad (9)$$

Objective function given by (9) describes the similarity of the original and mirrored shape in terms of phase-preserving Fourier descriptors. It is easy to show that

$$f(m) = 2 \sum_{k=1}^{N-1} \left| \text{Im} \{ a_{k(m)}^{(1)} \} \right|, \quad (10)$$

where $\text{Im}\{\}$ denotes imaginary part. We propose to use a point P_m that corresponds to the global minimum of the objective function $f(m)$ as a starting point. Apart from the global minimum, the function $f(m)$ may have several local minima. All of these local minima should also be considered as potential starting points. They will be called *pseudomirror* points.

Typical representation of $f(m)$ is given in Fig. 4(a). In accordance with the given shape, several local minima and several corresponding pseudomirror points can be found. The proposed method does not search for symmetry/mirror axis but rather for points. Each pseudomirror point usually has its own complementary (“mirrored”) version. Thus, there are usually at least four points that could be used as the starting points for the phase-preserving FD.

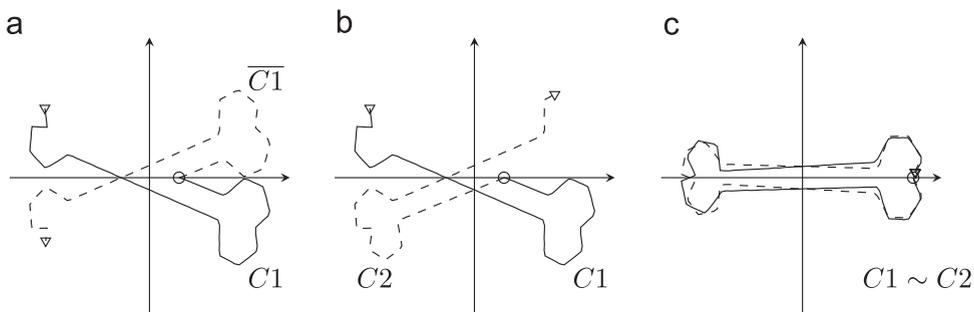


Fig. 3. Determining pseudomirror points: (a) original and “mirrored” contour for an arbitrary starting point are depicted (part of the closed contour is omitted for clarity), (b) original and “mirrored” contour with the same direction (part of the closed contour is omitted for clarity), (c) one of the derived nominal orientations for which C1 and C2 are very similar (potential pseudomirror point). Symbol \circ marks the starting point, while the triangle symbol Δ marks the end point.

Experiments have shown that these points are typically located $N/4$ points apart along the contour, but due to specific contour shapes, transformations, noise or inadequate sampling, the relative positions of these points may differ. The corresponding nominal orientations of the shape for the determined pseudomirror points in Fig. 4 (a) are depicted in Fig. 4(b).

Pseudomirror points are inherent to the shape and may hold some shape semantics. First, in the case of symmetrical objects, these points determine a line which coincides with the axis of symmetry. If there are many axes of symmetry, there will be many local minima of the function $f(m)$, hence several pseudomirror points may be determined. However, pseudomirror points may be determined in the pre-processing stage instead of the matching stage, which results in a simpler shape matching procedure.

We recommend that the number of chosen pseudomirror points is bounded by a parameter Q during descriptor extraction. If the number of local minima is larger than Q , then the local minima with the smallest values are chosen as pseudomirror points. However, if the number of local minima is smaller than Q , then we propose that the points with the lowest value of $f(m)$ are appended to this set. They are usually located near the local minima. Their inclusion in this set improves retrieval performance for a coarser sampling of the contour.

Before proceeding to the extraction algorithm of the phase-preserving FD, there are some technical implementation issues that need to be addressed. As illustrated in Fig. 4(a) and (b), when using a smaller number of contour points N , the gradient of $f(m)$ computed in points close to the centroid tends to be large, so the exact minimum of $f(m)$ could be located between m and $m+1$, leading to smaller errors. This means that for a smaller number of points, perfect “alignment” may not be achievable. A simple workaround is to extrapolate and sub-sample the discrete contour to a finer discretization scheme in a neighborhood of m , therefore the function $\arg(Z_m)$ can be computed on a finer grid. This allows us to search for a more exact local pseudomirror point within a range $[m-S, m+S]$, where P_m is assumed to be a pseudomirror point and S is a predefined integer parameter (usually not larger than $S=2$). The described procedure has two advantages: it enables computing a more exact local pseudomirror point and it facilitates establishing a hierarchical computation scheme, if complexity needs to be further reduced. The function $m \rightarrow \arg(Z_m)$ is approximated in this local neighborhood by a function (linear, cubic, spline or similar) $m_{ss} \rightarrow g(m_{ss})$, where $m_{ss} \in [m-S, m+S]$. The effectiveness of sub-sampling is demonstrated in Fig. 5.

Experimental results show that another small modification of the objective function can improve retrieval

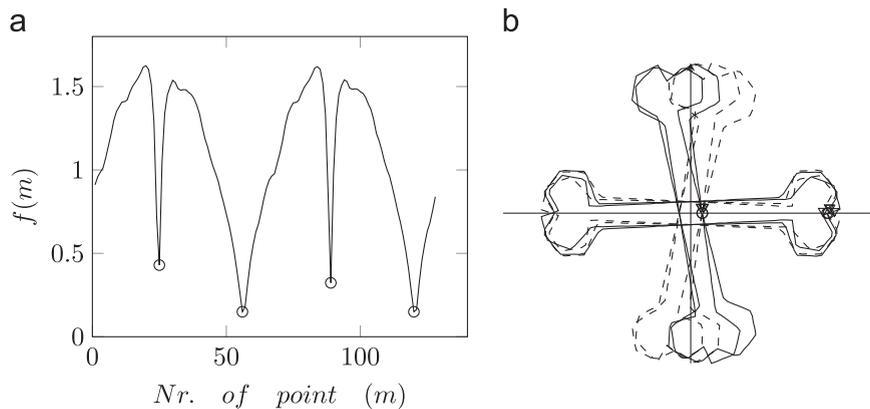


Fig. 4. Determining local minima of $f(m)$: (a) objective function, (b) corresponding pseudomirror point based shape orientations.

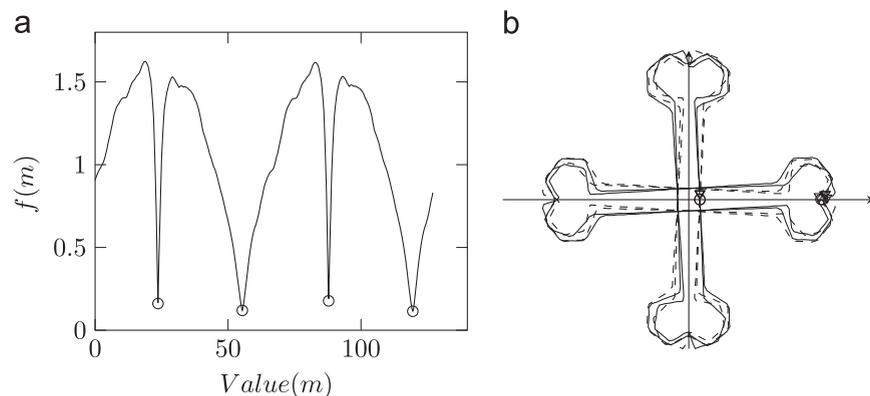


Fig. 5. Determining local minima of $f(m)$ using sub-sampling: (a) objective function, (b) pseudomirror point based shape orientations.

accuracy. It is well-known that low frequency Fourier coefficients used for shape description usually have much larger magnitude than Fourier coefficients that correspond to higher frequencies. Thus, it is possible to sort the magnitude values of a_k in descending order as

$$|a_{K_1}| \geq |a_{K_2}| \geq \dots \geq |a_{K_R}| \geq \dots \geq |a_{K_M}|, \quad (11)$$

where $K_i \in M_{ind}$. Harmonics with larger magnitude contribute to the coarse description of the shape, while the ones with smaller magnitude usually contain subtle intra-class variations, details and noise. Therefore, if a smaller number of coefficients M are used for computation of $f(m)$ (given by relation (10)), the normalization may be less accurate. In contrast, if all of the coefficients are used, then the alignment of different shapes from the same class could also be less accurate, because of the details present. Experiments on all datasets have shown that the computation of the objective function $f(m)$ using only first R largest magnitude harmonics:

$$f(m) = 2 \sum_{k \in R_{ind}} |Im\{a_{k(m)}\}|, \quad (12)$$

where R is determined as the maximal integer value for which the inequality $\sum_{i=1}^R |a_{K_i}| \leq 0.95 \sum_{i=1}^M |a_{K_i}|$ is satisfied, improves retrieval performance while reducing computational time. Thus, only 10–15 harmonics are usually needed to determine the pseudomirror points.

3.2. PIFD extraction procedure

The PIFD extraction procedure may be summarized through the following steps:

1. Each point P_m of the shape is considered to be the starting point of the shape. The shape is de-rotated so that the starting point is always located on the real axis.
2. Fourier descriptors $a_{k(m)}$ and objective function $f(m)$ are computed for every starting point in the previous step.
3. Best Q local minima of the objective function $f(m)$ are found. These minima define potential pseudomirror points.
4. (Optionally) Function $f(m)$ is subsampled near the potential pseudomirror points and finer pseudomirror points are computed.
5. Fourier descriptors computed in finer pseudomirror points are used to form the PIFD descriptor.

Now, a more formal description of the PIFD extraction procedure shall be given.

1. The initial contour of the shape is denoted as $\{Z_0^in, Z_1^in, \dots, Z_{N-1}^in\}$. In order to increase efficiency and avoid repetitive normalization under translation and scale, the initial contour is firstly translated to the origin and scaled in space domain using relation $\{Z_0, Z_1, \dots, Z_{N-1}\} = (\{Z_0^in, Z_1^in, \dots, Z_{N-1}^in\} - Z_C) / Sc$ where $Sc = \sum_{i=1}^{N-1} |a_i|$ is the scaling coefficient, $a_k = (1/N) \sum_{n=0}^{N-1} Z_n^in e^{-j2\pi nk/N}$ are Fourier coefficients of the initial shape and $Z_C = (1/N) \sum_{n=0}^{N-1} Z_n^in$ is the centroid (or DC component) of the shape.
2. The Fourier coefficients are sorted in descending order $|a_{K_1}| \geq |a_{K_2}| \geq \dots \geq |a_{K_R}| \geq \dots \geq |a_{K_M}|$, where $K_i \in M_{ind}$.

Maximal R that satisfies $\sum_{i=1}^R |a_{K_i}| \leq 0.95 \sum_{i=1}^M |a_{K_i}|$ is found. An array of indices is collected in a set $R_{ind} = \{K_1, K_2, \dots, K_R\}$. The set R_{ind} is used only for computing the objective function $f(m)$ in Step 5 and Step 8.

3. The first point of the contour $\{Z_0, Z_1, \dots, Z_{N-1}\}$ has to be de-rotated to zero argument, therefore the contour $e^{-jarg(Z_0)} \{Z_0, Z_1, \dots, Z_{N-1}\}$ is formed. The Fourier transform of the resulting contour is computed as $a_{k(0)} = e^{-jarg(Z_0)} (1/N) \sum_{n=0}^{N-1} Z_n e^{-j2\pi nk/N} = e^{-jarg(Z_0)} (Sc \cdot a_k)$ (for $k \in M_{ind}$)
4. For each $m \in \{1, 2, \dots, N-1\}$ the initial contour is shifted (changing the starting point) and the contour is de-rotated so that the first contour point P_m has the argument of zero degrees. Using this procedure a contour $e^{-jarg(Z_m)} \{Z_m, Z_{m+1}, \dots, Z_{N-1}, Z_0, Z_1, \dots, Z_{m-1}\}$ is obtained. Fourier transform does not have to be computed again, but $a_{k(m)}$ may be computed by coefficient multiplication $a_{k(m)} = e^{-jarg(Z_m)} e^{j2\pi mk/N} e^{jarg(Z_0)} a_{k(0)}$, for $k \in M_{ind}$. This signifies that for each contour point Z_m used as starting point, a descriptor $a_{k(m)}$ is computed.
5. Then, for each $m \in \{0, 1, 2, \dots, N-1\}$ function $f(m) = 2 \sum_{k \in R_{ind}} |Im\{a_{k(m)}\}|$ is computed. Note that the $f(m)$ is now computed on a smaller set $R_{ind} = \{K_1, K_2, \dots, K_R\}$, instead of $M_{ind} = \{-M/2, -M/2+1, \dots, -2, -1, 1, 2, \dots, M/2-1, M/2\}$ proposed with relation (10).
6. Let Q be defined as a number of chosen pseudomirror points. The goal is to find Q best local minima of $f(m)$ using the following procedure:
 - (a) sort pairs $(m, f(m))$ by value of $f(m)$ in ascending order into two sets:
 - (i) Set f_{LM} which contains the points that correspond to the local minima, obtaining corresponding indices $\{m_{j0}, m_{j1}, \dots, m_{j(LM-1)}\}$ (m_{j0} being the index of the global minimum, LM is the number of local minima),
 - (ii) (optionally) Set f_{NON-LM} which contains the rest of the points that are not local minima, but sorted in ascending order: $\{m_{l0}, m_{l1}, \dots, m_{l(N-1-LM)}\}$.
 - (b) choose the first $\min(Q, LM)$ pseudomirror points P_q from set f_{LM} where $q \in \{m_{j0}, m_{j1}, \dots, m_{j(\min(Q, LM)-1)}\}$
 - (c) optionally: if one wants to keep a predefined descriptor size and $Q > LM$, append the rest of the pseudomirror points from first $Q - LM$ points from set f_{NON-LM} and form pseudomirror points P_q , where $q \in \{m_{j0}, m_{j1}, \dots, m_{j(LM-1)}, m_{l0}, m_{l1}, \dots, m_{l(Q-LM-1)}\}$. This is a useful feature which improves retrieval accuracy, especially for a coarser sampling of the contour (when N is small).
7. The resulting P_q are potential starting (pseudomirror) points.
8. For each pseudomirror point $P_{ms} \in \{P_q\}$ found in the previous steps, function $arg(Z_{ms})$ is interpolated on the interval $[m_s - S, m_s + S]$ with $g(m_s)$, where S is a positive integer (usually $S = 1$ or $S = 2$) and $g(m_s)$ is a linear or cubic interpolation function. Now the coefficients are computed as:

$$a_{k(m)} = e^{-jg(m)} e^{j2\pi mk/N} e^{jarg(Z_0)} a_{k(0)} \quad (13)$$

and the global minimum of the function $f(m) = 2 \sum_{k \in R_{ind}} |Im\{a_{k(m)}\}|$ is found using discretization

Table 1
Similarity measures.

Mirror non-invariant	$d(P_1, P_2) = \min_{(i_1, i_2)} \sum P_1(i_1, :) - P_2(i_2, :) $
Mirror invariant	$d(P_1, P_2) = \min_{(i_1, i_2)} \left[\min \left\{ \sum P_1(i_1, :) - P_2(i_2, :) , \sum P_1(i_1, :) - \overline{P_2(i_2, :)} \right\} \right]$

structure $m = m_s + r\Delta m$ (where $r = -S/\Delta m, \dots, S/\Delta m$). These minima are located in m_s^{fine} .

9. For each discrete pseudomirror candidate point m_s , a finer pseudomirror candidate point $m_s^{fine} \in [m_s - S, m_s + S]$ is found. Instead of a set of points P_q , the corresponding P_q^{fine} is determined, hence deriving the corresponding descriptors as in (13).
10. Finally, the descriptor is formed as:

$$PIFD_{Q \times M}(S, \Delta m) = \begin{bmatrix} \{a_{k(q_0^{fine})}\} \\ \{a_{k(q_1^{fine})}\} \\ \vdots \\ \{a_{k(q_{(Q-1)}^{fine})}\} \end{bmatrix}_{q_i \in q, k \in M_{ind}} \quad (14)$$

It is interesting to note that PIFD is actually a generalization of many other Fourier-based methods. If $Q = 1$, the modulo of PIFD is equal to NCC FD ($|PIFD_{Q=1}| \rightarrow NCC$) [6]. Also, the minimum of $f(m)$ is often close to $\min_{(m)} \{ |Im\{a_{k(m)}^{(1)}\}| \}$, where K is the dominant harmonic (harmonic with the largest magnitude). For shapes with a dominant first harmonic, solution of $f(m) = 2|Im\{a_{1(m)}^{(1)}\}| = 0$ (a special case of (10)) may be used to align shapes using only the first harmonic, resulting in a normalization method that is widely adopted in [19,22] and many others.

3.3. Similarity measures

It is clear that the relative positions of the proposed pseudomirror points are mirror-invariant, which means that the points will be positioned at the same locations on the shape, even if the shape contour is mirrored over an arbitrary axis. One of the advantages of PIFD is that it is simple to introduce mirror invariance in shape description. It is easily noticeable from relation (8) that mirrored shapes have conjugated PIFDs.

Suppose that two different shapes are described by PIFDs P_1 and P_2 . We propose two similarity measures, given in Table 1. The distance is essentially the minimum of all City-block distances computed between corresponding FDs.

3.4. Computational complexity

As for the computational complexity of the PIFD descriptor, there are two elements to consider. The first is the computational complexity of the descriptor extraction and the second is computation of the similarity measure.

Computational complexity which corresponds to the PIFD extraction procedure is given by $O(N \log N) + O(N \log N) + O(N \cdot M) + O(N \cdot M^2) + O(N \log N) + O(S/\Delta m \cdot N \cdot M \cdot Q)$, which is equivalent to $O(N \log N)$. The overall complexity is actually determined by the computational complexity of FFT

and the sorting algorithms. As in the previous sections, variables N, M, Q and $S/\Delta m$ denote the number of contour points, the number of Fourier coefficients chosen for representation, the number of pseudomirror points and the number of subsampled points around a pseudomirror point, respectively.

The computation of the similarity measures is equal to determining the minimum of previously computed Q^2 or $2 \cdot Q^2$ City-block distances (for mirror non-invariant or mirror invariant similarity measure, respectively). Since Q is usually fixed within the interval [4,20], the computational complexity of the similarity measure is equivalent to the computational complexity of City-block distance. Although the computational time is increased to a limited extent, the computational complexity of the proposed similarity measure remains linear ($O(N)$).

4. Experimental results and discussion

The presentation of experimental results is organized as follows.

First, the functionality and importance of pseudomirror points were presented and discussed. Pseudomirror points were compared to competitive methods for determining nominal orientation and starting point of a shape contour.

In order to illustrate the effectiveness and efficiency of the phase-preserving Fourier descriptor based on pseudomirror points (PIFD), the retrieval performance and computational complexity were analyzed.

In the paper, four image databases were used for performance analysis:

- MPEG7 CE-1 Set B database [34],
- Swedish Leaf database [41], binarized and preprocessed by Xu et al. [13],
- ETH-80 dataset [42],
- Kimia99 dataset [43].

The experiments on the MPEG7 CE-1 Set B database discuss four aspects:

1. the influence of the different parameters on PIFD performance,
2. PIFD performance against phase-preserving Fourier descriptors,
3. PIFD performance against magnitude-based Fourier descriptors,
4. PIFD performance against other non-Fourier based descriptors.

Computational complexity of the algorithms is analyzed in two different stages: descriptor extraction and descriptor matching. It is important to note that almost all Fourier

descriptor implementations share the same computational complexity as PIFD. Therefore, PIFD is compared to competitive Fourier-based methods using only retrieval performance measures. In contrast, other state-of-the-art shape retrieval methods which are not based on the Fourier transform have different computational complexity. Thus, the trade-off between accuracy and computational complexity is emphasized when PIFD is compared with other non-Fourier based shape descriptors.

The experimental results and suitable discussions are given in the following subsections of the paper.

4.1. Pseudomirror points as orientation landmarks

In order to demonstrate the usefulness of pseudomirror points for determining nominal shape orientation, two illustrations are provided.

Fig. 6 presents how nominal orientation and starting points are determined for different shapes belonging to different classes (“Glass”, “Cow”, “Fork” and “Jar”, all extracted from MPEG7 CE-1 Set B). Up to eight pseudomirror points are extracted from each shape. In order to increase the visibility of different pseudomirror points, they were divided into three groups denoted with \square , \circ and \times

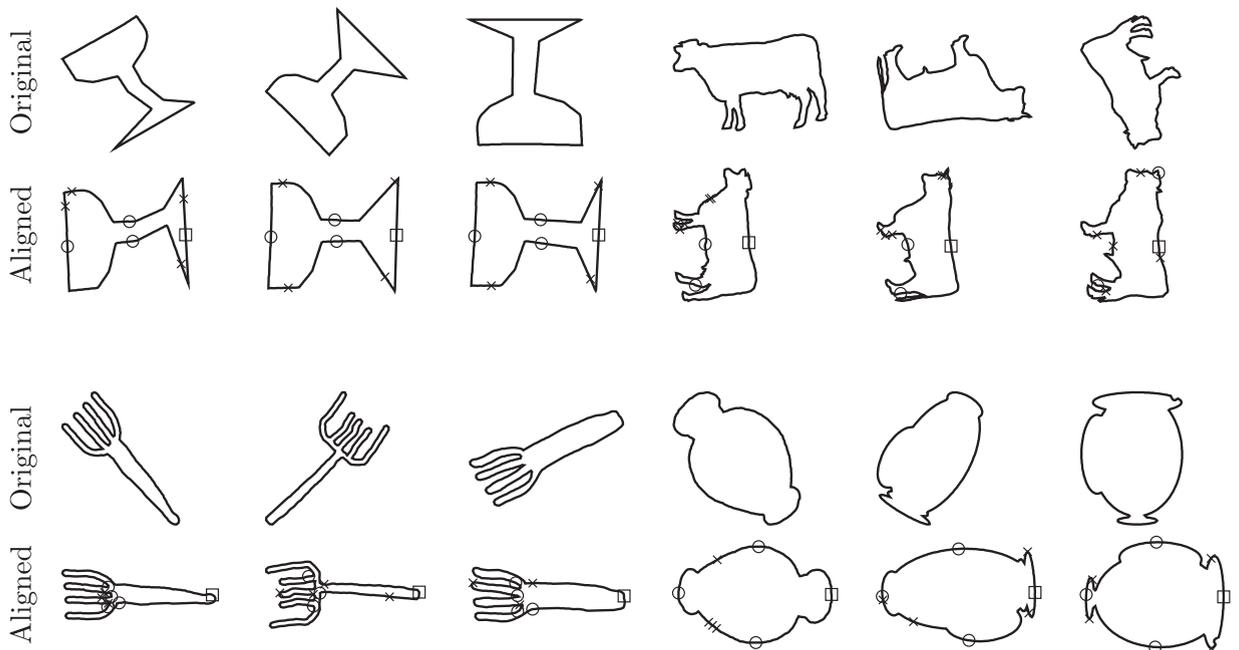


Fig. 6. The capability of pseudomirror points to determine nominal orientation and starting point of the shape contour. The pseudomirror points are marked using symbols \square (most important point – global minimum), \circ (less important points – local minima) and \times (least important points – local minima).

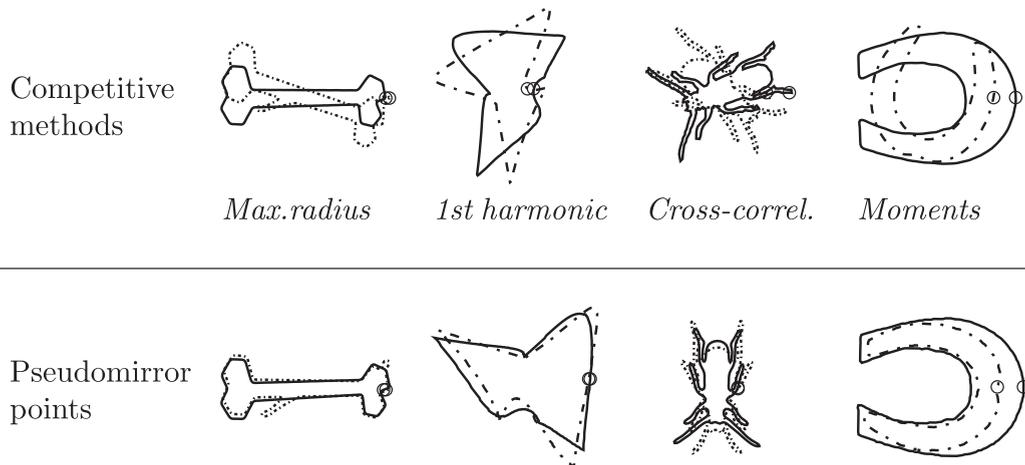


Fig. 7. Comparison of competitive methods for determining nominal orientation and starting points (point with maximal radius, phase of first harmonic, cross-correlation, moments) with pseudomirror points. The determined starting points are marked with a circle symbol.

\times , respectively. The point that corresponds to the absolute minimum of the cost function $f(m)$ is marked with a square symbol. The pseudomirror points that correspond to the low-valued local minima, but do not represent a global minimum, are marked with a circle symbol. The rest of the determined pseudomirror points, that corresponds to less important higher-valued local minima, are marked with an x-mark. It is noticeable from the results in Fig. 6 that pseudomirror points are able to properly align different shapes and locate a common starting point of the contour. In addition, the same groups of pseudomirror points tend to appear in similar locations within a single shape class. Therefore, affine and non-rigid transformations of the shape may affect the position of several pseudomirror points on the shape contour, but usually the rest of the pseudomirror points preserve their relative position. The robustness of pseudomirror points is based on the fact that often these preserved contour points can be used for accurate shape alignment.

The superiority of pseudomirror points to other competitive methods for determining nominal orientation is presented in Fig. 7. A suitable pair of similar shapes that needed to be aligned was chosen to point out the disadvantages of the competitive methods. The results of the alignment obtained using the point of maximal radius [19] and our proposed method of the “Bone” shape are illustrated in the first column of Fig. 7. Non-rigid transformations of the shape contour (e.g. arisen by inadequate contour extraction) significantly affect the position of the centroid and the point of maximal radius, therefore the nominal orientation of the shape is incorrectly determined. On the other hand, pseudomirror points exhibit enough robustness under limited shape deformations. Methods that use alignment schemes based on the phase on first harmonic [19,18,22] fail in cases of noise, circular harmonic locus or first harmonics with neglectable magnitude. An example of such case is presented in the second column in Fig. 7, where pseudomirror points clearly achieve a more accurate alignment than first-harmonic based method. As stated earlier, cross-correlation based techniques [15] attempt to find the best possible concurrence among shapes during matching stage. This may result in an inadequate nominal shape orientation, as shown in the third row in Fig. 7. Finally, orientation methods based on moments or the axis of least inertia, unlike the pseudomirror point based orientation, may fail

in cases of rounded or horseshoe-like objects, as presented in the last row in Fig. 7.

4.2. MPEG-7 CE-1 Set B database

Initial experiments were conducted on the MPEG7 CE-1 Set B shapes dataset [34]. Representative elements of MPEG7 CE-1 Set B dataset are illustrated in Fig. 9(a). MPEG7 CE-1 Set B consists of 1400 shapes representing real life objects, classified into 70 classes with 20 similar shapes for each class. This database is widely used for shape retrieval testing purposes. It includes rotation, scaling, skew, stretching, deflection, indentation and articulation of shapes which may be used to test robustness of the shape descriptors.

Typical measure of retrieval performance in CBIR are the precision and recall (PR) diagrams [5,11–13]. Precision is defined as the ratio of the number of the relevant shapes to the total number of retrieved shapes, while recall is defined as the ratio of the number of retrieved relevant shapes to the total number of relevant shapes in the entire database. Average precision for all recall values for all query shapes in the database is computed and presented by PR diagrams. If an image retrieval algorithm has better precision than another one, for same recall values, it is generally considered better.

Another widely adopted performance measure is Bulls-Eye score [29,30,33]. Unlike PR diagrams, Bulls-Eye score is a scalar, so the comparison of retrieval performance between different algorithms is simpler. Bulls-Eye score is defined as the percentage of relevant results in the first $2 \cdot K$ retrieved results of a query, where K is the number of shapes in the query class. Average Bulls-Eye score is computed after all elements in the dataset have been used as a query. Thus, the best possible rate is 100%. As opposed to PR diagrams which demonstrate precision across all recall values, Bulls-Eye scores favor algorithms which provide higher precision for top retrieval results.

4.2.1. Evaluation of PIFD descriptor

In order to investigate the scalability of PIFD, retrieval performance on MPEG-7 dataset was determined for different values of the following parameters: number of points (N), number of coefficients (M) and number of pseudomirror points (Q). During the experiments, one parameter was variable, while the other two were fixed.

Table 2

Bulls-Eye scores achieved on MPEG-7 dataset, computed for different PIFD parameters.

N	64	128	256	512	1024					
($M=16, Q=8$)	81.24	81.47	81.50	81.51	81.58					
M	6	8	10	12	16	18	20	24	32	
($N=256, Q=8$)	76.04	78.52	79.99	80.70	81.50	81.90	81.83	81.60	81.50	
Q	1	4	6	8	10	12	16	20	24	32
($N=256, M=18$)	55.57	81.32	81.79	81.90	81.90	81.93	81.94	81.91	81.91	81.60

Table 3
Bulls-Eye score for MPEG-7 dataset.

Method	Bulls Eye score	CC extracting	CC matching
Height functions+SC [2]	90.35	$O(N^2)$	$O(N^3)$
Locally affine invariant descriptors [44]	89.62	$O(N^2)$	$O(N^3)$
Shape tree [29]	87.70	$O(N^2)$	$O(N^4)$
HPM [30]	86.35	$O(N^2)$	$O(N^3)$
IDSC+DP [33]	85.40	$O(N^3)$	$O(N^3)$
Planar graph cuts [45]	85.00	$O(N \log N)$	$O(N^2 \log N)$
MCC+SC [27]	84.93	$O(N^2)$	$O(N^3)$
PIFD	82.03	$O(N \log N)$	$O(N)$
Generative Model [46]	80.03	$O(N^2)$	$O(N^2)$
MCSS [23]	78.8	$O(N)$	$O(N)$
Curve Edit [28]	78.17	$O(N)$	$O(N^2 \log N)$
Shape Context [32]	76.51	$O(N^2)$	$O(N^2)$
Visual parts [34]	76.45	$O(N \log N)$	$O(N^2)$
NCC* [6]	75.75	$O(N \log N)$	$O(N)$
CSS [25,34]	75.44	$O(N^2)$	$O(N)$
CPAF* [12]	74.47	$O(N \log N)$	$O(N)$
IARH* [13]	73.52	$O(N \log N)$	$O(N)$
CRFD* [15]	72.57	$O(N \log N)$	$O(N \log N)$
FPD* [11]	65.54	$O(N^2)$	$O(N)$
WARP* [22]	58.50	$O(N \log N)$	$O(N^2)$
FHAFD* [18–21]	51.34	$O(N \log N)$	$O(N)$
AFD* [17]	41.08	$O(N \log N)$	$O(N)$

All the retrieval results were compared using Bulls-Eye score. The obtained results are given in Table 2.

For all the experiments the variation of the remaining parameters (S , Δm , interpolation type) does not affect the results by more than 0.1%, as long as $S \geq 1$ and $\Delta m \leq N/512$.

The maximal Bulls-Eye score achieved by PIFD on the MPEG7 CE-1 Set B dataset was 81.98 for $N = 512$, $M = 18$ and $Q = 8$. Using 1024 points it can reach 82.03.

Although capturing more details with finer sampling clearly improves shape recognition, it is clear that the number of points does not significantly affect the retrieval performance, since interpolation around pseudomirror points is used. Satisfactory results are obtained with $N = 256$ points, while the best results in terms of performance vs. computational cost are obtained with 512 points. The points are chosen to be a power of two, so that FFT could be computed.

Like all Fourier descriptor based techniques, the descriptor is compact and only few low-frequency components are enough for adequate discrimination between shapes. This implies that M does not need to be large. The lowest coefficients provide the rough estimate of the shape and its principal characteristics, while the higher harmonics carry information about details and, eventually, noise. If a very large value for M is chosen, more details are included in the descriptor. These details help distinguishing among classes, but introduce errors when some intra-class variations (or noise) are present. Thus, the value of $M = 18$ has proved to be optimal on the MPEG7 CE-1 Set B.

It is noticeable that the best results are obtained for 6–16 pseudomirror points Q . Increasing the number of pseudomirror points negatively affects retrieval performance. This

occurs due to the fact that a larger number of pseudomirror points diminish the discriminability between objects that do not share nominal orientation. Eight pseudomirror points ($Q = 8$) should achieve optimal results on most datasets.

PIFD is a parameter based descriptor, which means that tuning parameters for different datasets may provide better results. Nevertheless, parameters $N = 512$, $M = 20$ and $Q = 8$ will obtain promising results for almost all applications, while keeping the descriptor compact and efficient. In all the following experiments, except in Table 3, parameters $N = 512$, $M = 18$ and $Q = 8$ are used. The Bulls-Eye score presented in Table 3 is achieved with $N = 1024$.

4.2.2. Evaluation of PIFD descriptor against other phase-preserving Fourier descriptors

It has been shown in the previous sections that PIFD is actually a phase-preserving Fourier descriptor. Therefore, PIFD is first evaluated against other implemented phase-based Fourier shape description methods: The Correlation based FD by Larsson et al. (CRFD) [15], WARP FD proposed by Bartolini et al. (WARP) [22], The Affine FD proposed by Arbter et al. (AFD) [17] and FHAFD (First Harmonic Aligned FD) – actually PIFD with $Q = 1$ and aligned only using the phase of the first order harmonic (as proposed by [18–21]).

In order to compare algorithms in a fair environment, all phase-preserving Fourier-based techniques use the same number of points ($N = 512$) and Fourier coefficients ($M = 18$). The PR diagrams for MPEG7 CE-1 Set B are given in Fig. 8(a) and Bulls-Eye scores as a part of Table 3.

As it may be seen, PIFD outperforms other comparable techniques by a large margin. It is important to note that all other methods share the same computational complexity in extraction stage as PIFD, while CRFD and WARP have higher computational complexity in the matching stage.

To the best of the authors' knowledge, PIFD is the first FD implementation for which the phase was effectively and directly used in shape-based image retrieval, while obtaining worth-mentioning results. Moreover, it is the first time these descriptors (CRFD, WARP, AFD) were tested on the whole MPEG7 CE-1 Set B.

It is interesting to note that CRFD performs relatively well on the MPEG7 CE-1 Set B, but unlike our proposed method it implicitly uses correlation in spatial domain. This procedure, due to possible contour transformations, may more often lead to false matching and inadequate comparisons.

The inadequate scaling coefficient and first harmonic alignment used by the WARP method cause large variation of starting point positions. Consequently, the descriptor retrieval performance is significantly reduced.

4.2.3. Evaluation of PIFD descriptor against other magnitude-based Fourier descriptors

In contrast to the phase-preserving Fourier descriptors, magnitude-based Fourier descriptors are considerably more popular. This may be explained by two facts: magnitude-based FD achieve the desired invariance without much difficulty (by simply ignoring the phase content) and magnitude-based FD usually exhibit higher retrieval performance.

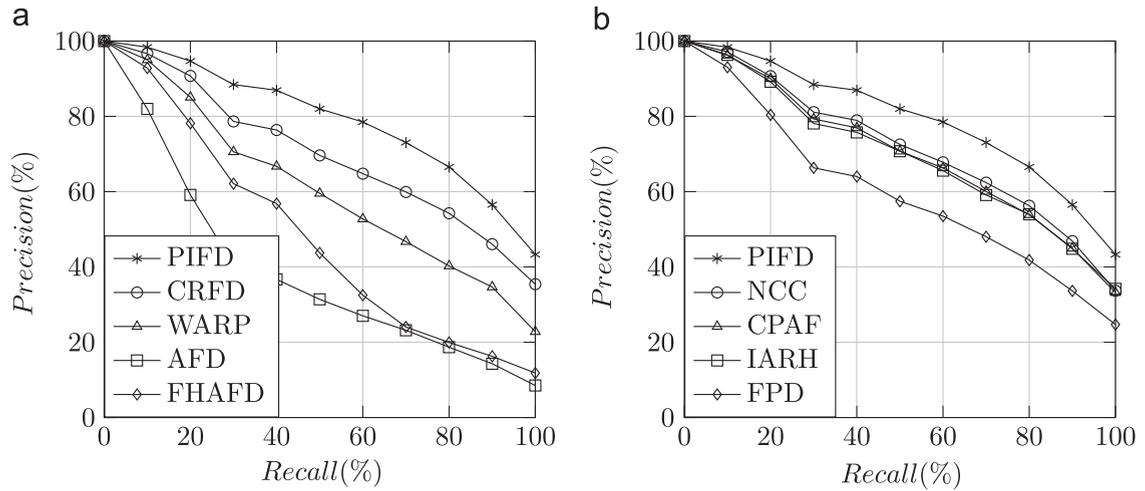


Fig. 8. PR diagrams obtained on MPEG-7 dataset (a) PIFD compared to phase-preserving FDs, (b) PIFD compared to magnitude-based FDs.

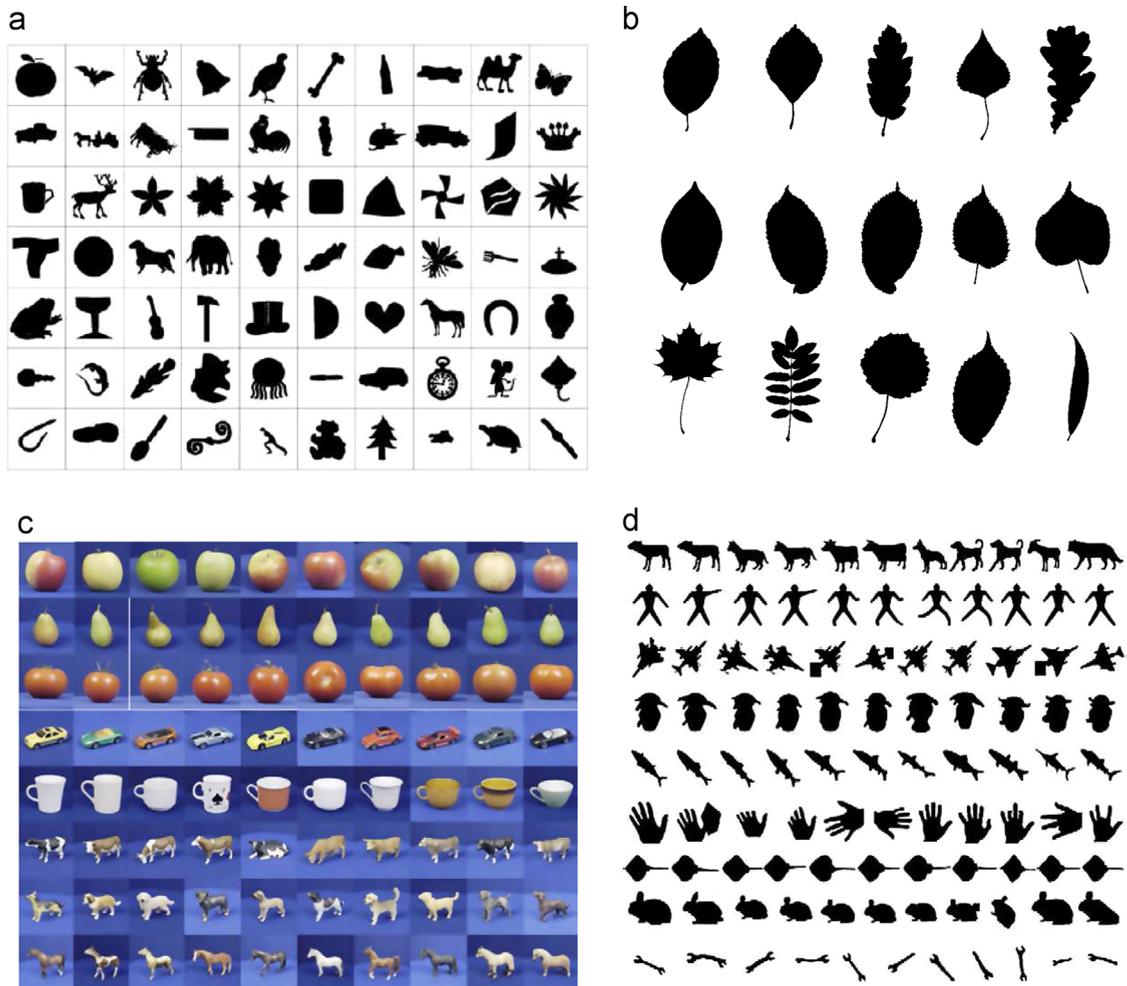


Fig. 9. (a) MPEG7 CE-1 Set B dataset representative shapes (70 classes with 20 variations per class), (b) Leaf dataset representative shapes (15 classes with 75 variations per class), (c) ETH-80 dataset (8 classes with 10×41 variation per class), (d) Kimia99 dataset (9 classes with 11 variations per class).

To illustrate this fact, we compared PIFD with our implementation of several best performing state-of-the-art Fourier descriptors based solely on magnitude of Fourier coefficients, using the following shape signatures: Furthest point distance (FPD) [11], Improved arc-height function (IARH) [13], Combined perimeter area function (CPAF) [12] and Normalized complex coordinates (NCC) [6].

All magnitude-based Fourier descriptors are extracted from the same number of points ($N = 512$) and have the same number of Fourier coefficients ($M = 18$) as PIFD. The PR diagrams for MPEG7 CE-1 Set B are given in Fig. 8 (b) and Bulls-Eye scores are given as a part of Table 3.

PIFD clearly outperforms all other magnitude-based FDs, without any additional increase in computational complexity. It is notable that the inclusion of phase in PIFD descriptor significantly improves the retrieval

performance and discriminability of the existing magnitude-based Fourier descriptor.

Experimental results demonstrate that many phase-preserving Fourier descriptors under-perform regular magnitude-based FDs. This is due to the fact that total exclusion of the phase introduces less error than an inappropriate phase normalization procedure.

4.2.4. Evaluation of PIFD descriptor against other state-of-the-art shape description methods

More than a dozen of widely known non-FD-based state-of-the-art shape description methods have been chosen and compared with PIFD on the MPEG7 CE-1 Set B. It is not justified to expect that a global, relatively simple shape description method such as PIFD, yields good results as other more complex and application-specific algorithms. Nevertheless, PIFD has one of the best performance vs. computational complexity ratios of all compared algorithms.

Some of the best Bulls-Eye scores for non-supervised shape description techniques reported in the literature are given in Table 3. Asterisk indicates that we have implemented the method ourselves. Our proposed method does not excel on MPEG7 CE-1 Set B, but has comparable performance with other shape descriptors. Nevertheless, all the methods presented in Table 3 that outperform PIFD have higher computational complexity and “rich” shape descriptions. Most of them are able to perform partial shape matching. In addition, many descriptors that underperformed PIFD also have higher computational complexity. In the 2nd and 3rd columns of Table 3, extraction and matching computational complexity of over 20 state-of-the-art methods are listed. It is clear that our method has the best performance vs. computational complexity ratio.

It is important to note that the Bulls-Eye scores in Table 3, as well as other performance scores presented in the following sections of the paper are achieved with the optimal setups of the corresponding methods reported by the authors. In order to reach the best possible performance scores, most of the reported results are obtained using different number of contour points, number of parameters or descriptor size. Therefore, the Bulls-Eye score achieved by PIFD on MPEG-7 CE-1 Set B corresponds to the following parameters: $N = 1024$, $M = 18$ and $Q = 8$. A reduced number of contour points ($N = 512$) is used on the other datasets.

Table 4

Classification score for Leaf dataset.

Method	Score
TSLA [48]	96.53
Shape tree [29]	96.28
PIFD	95.47
MDS+SC+DP [33]	95.33
NCC* [6]	94.53
IDSC+DP [33]	94.13
sPACT [49]	90.77
Fourier descriptors (result from [33])	89.60

Table 5

Performance score for ETH-80 dataset.

Method	Score
Decision tree (multi-cue method) [42]	93.02
PIFD	91.64
Symbolic representation [50]	90.28
Height functions+SC [2]	89.73
Height functions [2]	88.72
IDSC+DP [33]	88.11
NCC* [6]	87.01
MDS+SC+DP [33]	86.80
SC+DP [42]	86.40
SC greedy [42]	86.40
PCA masks [42]	83.41
PCA gray [42]	82.99
Mag-lap [42]	82.23
Dx-Dy [42]	79.79
Color histogram [42]	64.85

Table 6

Performance score for Kimia99 dataset.

Method	1st	2nd	3rd	4th	5th	6th	7th	8th	9th	10th
Height functions [2]	99	99	99	99	98	99	99	96	95	88
Shape tree [29]	99	99	99	99	99	99	99	97	93	86
IDSC+DP [33]	99	99	99	98	98	97	97	98	94	79
Shock Graph [43]	99	99	99	98	98	97	96	95	93	82
Path similarity [51]	99	99	99	99	96	97	95	93	89	73
MDS+SC+DP [33]	99	98	98	98	97	99	97	96	97	85
PIFD	97	96	96	97	88	87	83	74	68	57
Generative Model [46]	99	97	99	98	96	96	94	83	75	48
Shape Context [32] (from [43])	97	91	88	85	84	77	75	66	56	37

4.3. Leaf dataset

The Swedish leaf dataset is challenging because of its high inter-species similarity. Leaf database consists of 15 species of leaves, with 75 leaves per species, illustrated in Fig. 9(b). Contour-based differences between species are very small. It is notable in Fig. 9(b) that the 7th and 8th class are almost identical.

According to the available papers based on Leaf dataset, Bulls-Eye score is rarely used to evaluate descriptor performance (only in [47]). Instead, many authors demonstrate the classification ability of their proposed descriptors by randomly selecting 25 training images from each species (25×15 images in total) and classify the remaining 50×15 images using the 1-nearest neighbor (1NN) approach. The percentage of correct classifications is used as a performance measure. This parameter better describes the discriminative ability of the descriptor.

Score obtained by PIFD is compared to other performance scores available in the literature and presented in Table 4.

Note that our method performs comparably to the best reported methods on this dataset. It is interesting that PIFD even outperforms IDSC based methods presented in [33], although the computational complexity of PIFD is much lower ($O(N \log N)$ compared to $O(N^3)$). Because of its local hierarchical structure IDSC performs better than PIFD on MPEG7 CE-1 Set B, but the Leaf dataset results indicate that, in contrast, PIFD has promising classification abilities. It is also interesting to note that PIFD performance is comparable to the performance of TSLA [48], although the latter is specifically developed for leaf classification. Nevertheless, the improvement of PIFD over NCC on Leaf dataset is not substantial. This could be explained by the fact that leaves' contours are usually smooth and contour changes (tops, peduncles) usually appear in similar places, hence the difference in phase of Fourier coefficients is not that important for shape discrimination.

4.4. ETH-80 dataset

The ETH-80 data set [42] contains 10 categories of 8 objects. Each object has 41 color images captured from different viewpoint. Thus, there are 3280 images in this database. Some of the instances for every objects are given in Fig. 9(c). Only the contours of the shapes are used during classification. As proposed in [42], the descriptors are tested using *leave one object out cross-validation*. This means that each shape in the dataset is removed from the dataset and afterwards the removed shape is classified using all others as training shapes. Recognition is considered successful if the object is assigned to the correct category label. As presented in Table 5, our method is outperformed only by Decision tree [42], which is a multi-class method where the color information is exploited during classification. Moreover, it outperforms many methods with considerably higher computational complexity (such as Height functions and IDSC).

4.5. Kimia99 dataset

This dataset contains 11 images in 9 categories, which are presented in Fig. 9(d). Because of the existing occlusions, missing parts and articulation, this dataset is particularly difficult for shape-based retrieval. Performance score on Kimia99 dataset is computed in the following way: every shape is considered as a query, and the retrieval result is summarized as the number of top1 to top10 closest matches in the same class, excluding the query shape itself. The best possible result is 99 for each of the ranking. The results are presented in Table 6.

It is clear that PIFD does not perform well on the Kimia99 dataset as it does on other, mostly curvature-based shape datasets. PIFD is essentially a global contour-based shape descriptor, hence it is unable to perform partial matching, recognize articulations of a silhouette or compare heavily occluded shapes. However, it still exhibits solid performance and outperforms more complex shape descriptors (e.g. Generative Models [46] or Shape Context [32]). Other Fourier-based shape description methods fail to achieve any significant performance on this dataset.

5. Conclusion and future work

As we presented in the paper, PIFD has proved to be a versatile and flexible shape descriptor that performs well on many different datasets. PIFD is characterized by simple extraction and matching, which makes it convenient for usage in real-time applications. Additionally, it has promising classification abilities.

Pseudomirror points proved to be valuable for determining the nominal shape orientation. They may be used to improve shape description methods wherever starting point and rotation invariance are needed. Combined with the normalization of scale and translation based on Fourier coefficients, they may be used to normalize the contour in order to apply other more complex shape description techniques.

Experimental results pointed out the future directions for the improvement of PIFD. Although PIFD outperforms many other state-of-the-art shape description methods on curvature based datasets such as ETH-80 or Leaf dataset, it fails in scenarios that contain articulations, large artifacts or missing parts. This explains the lower average retrieval performance on MPEG-7 CE-1 Set B and Kimia99 datasets, where it was outperformed by hierarchical structural-based methods and methods that allow partial matching. Similar to many other contour based descriptors, PIFD fails in region-based shape-retrieval tasks.

As a part of future research, PIFD should be improved in two directions: it should exploit certain hierarchical structure in spatial domain in order to perform partial (local) matching of the shape and it should be extended to a region based descriptor, perhaps similarly as Zhang proposed to extend his FD into 2D Generic FD [52].

References

- [1] S. Loncaric, A survey of shape analysis techniques, *Pattern Recognit.* 31 (8) (1998) 983–1001.
- [2] J. Wang, X. Bai, X. You, W. Liu, L.J. Latecki, Shape matching and classification using height functions, *Pattern Recognit. Lett.* 33 (2) (2012) 134–143.
- [3] D. Zhang, G. Lu, Study and evaluation of different Fourier methods for image retrieval, *Image Vis. Comput.* 23 (2005) 33–49. January.
- [4] D. Zhang, G. Lu, A comparative study of Fourier descriptors for shape representation and retrieval, In: *Proceedings of the Fifth Asian Conference on Computer Vision*, January, Citeseer, 2002, pp. 646–651.
- [5] D. Zhang, G. Lu, A comparative study on shape retrieval using Fourier descriptors with different shape signatures, In: *Proceedings of International Conference on Intelligent Multimedia and Distance Education (ICIMADE01)*, 2001, pp. 1–9.
- [6] E. Sokic, S. Konjicija, Novel Fourier descriptor based on complex coordinates shape signature, In: *2014 12th International Workshop on Content-Based Multimedia Indexing (CBMI)*, IEEE, 2014, pp. 1–4.
- [7] T. Sikora, The MPEG-7 visual standard for content description—an overview, *IEEE Circuits Syst. Video Technol.* 11 (6) (2001) 696–702.
- [8] W.Y. Kim, Y.S. Kim, A region-based shape descriptor using Zernike moments, *Signal Process.: Image Commun.* 16 (1) (2000) 95–102, [http://dx.doi.org/10.1016/S0923-5965\(00\)00019-9](http://dx.doi.org/10.1016/S0923-5965(00)00019-9).
- [9] R.B. Yadav, N.K. Nishchal, A.K. Gupta, V.K. Rastogi, Retrieval and classification of shape-based objects using Fourier, generic Fourier, and wavelet-Fourier descriptors technique: a comparative study, *Opt. Lasers Eng.* 45 (6) (2007) 695–708.
- [10] N. Alajlan, I. El Rube, M. Kamel, G. Freeman, Shape retrieval using triangle-area representation and dynamic space warping, *Pattern Recognit.* 40 (7) (2007) 1911–1920.
- [11] A. El-ghazal, O. Basir, S. Belkasim, Farthest point distance: a new shape signature for Fourier descriptors, *Signal Process.: Image Commun.* 24 (August) (2009) 572–586.
- [12] B. Wang, Shape retrieval using combined Fourier features, *Opt. Commun.* 284 (14) (2011) 3504–3508.
- [13] G.-Q. Xu, Z.-C. Mu, B.-F. Nan, Shape retrieval using improved arc-height function, In: *2012 International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR)*, IEEE, 2012, pp. 73–77.
- [14] S. Eini, A. Chalechale, Shape retrieval using smallest rectangle centroid distance, *Int. J. Signal Process. Image Process. Pattern Recognit.* 6 (4) .
- [15] F. Larsson, M. Felsberg, P.-E. Forssen, Correlating Fourier descriptors of local patches for road sign recognition, *Comput. Vis. IET* 5 (4) (2011) 244–254.
- [16] A.V. Oppenheim, J.S. Lim, The importance of phase in signals, *Proc. IEEE* 69 (5) (1981) 529–541.
- [17] K. Arbter, W. Snyder, H. Burkhardt, G. Hirzinger, Application of affine-invariant Fourier descriptors to recognition of 3-D objects, *IEEE Trans. Pattern Anal. Mach. Intell.* 12 (7) (1990) 640–647.
- [18] E. Persoon, K.S. Fu, Shape discrimination using Fourier descriptors, *IEEE Trans. Pattern Anal. Mach. Intell.* 8 (March) (1986) 388–397.
- [19] F.P. Kuhl, C.R. Giardina, Elliptic Fourier features of a closed contour, *Comput. Graph. Image Process.* 18 (3) (1982) 236–258, [http://dx.doi.org/10.1016/0146-664X\(82\)90034-X](http://dx.doi.org/10.1016/0146-664X(82)90034-X).
- [20] S. Li, M.-C. Lee, D. Adjeroh, Effective invariant features for shape-based image retrieval, *J. Am. Soc. Inf. Sci. Technol.* 56 (May) (2005) 729–740.
- [21] P. Fieguth, P. Bloore, A. Domsa, Phase-based methods for Fourier shape matching, in: *2004 Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, (ICASSP'04)*, vol. 3, IEEE, 2004, pp. iii–425.
- [22] I. Bartolini, P. Ciaccia, M. Patella, WARP: accurate retrieval of shapes using phase of Fourier descriptors and time warping distance, *IEEE Trans. Pattern Anal. Mach. Intell.* 27 (January) (2005) 142–147.
- [23] A. Jalba, M. Wilkinson, Shape representation and recognition through morphological curvature scale spaces, *IEEE Image Process.* 15 (2) (2006) 331–341.
- [24] I. El Rube, N. Alajlan, M. Kamel, M. Ahmed, G. Freeman, Efficient multiscale shape-based representation and retrieval, in: *Image Analysis and Recognition*, Springer, Berlin Heidelberg, 2005, pp. 415–422.
- [25] F. Mokhtarian, S. Abbasi, J. Kittler, Efficient and robust retrieval by shape content through curvature scale space, *Ser. Softw. Eng. Knowl. Eng.* 8 (1997) 51–58.
- [26] S. Abbasi, F. Mokhtarian, J. Kittler, Curvature scale space image in shape similarity retrieval, *Multim. Syst.* 7 (November) (1999) 467–476.
- [27] T. Adamek, N.E. O'Connor, A multiscale representation method for nonrigid shapes with a single closed contour, *IEEE Trans. Circuits Syst. Video Technol.* 14 (5) (2004) 742–753.
- [28] T.B. Sebastian, P.N. Klein, B.B. Kimia, On aligning curves, *IEEE Trans. Pattern Anal. Mach. Intell.* 25 (1) (2003) 116–125.
- [29] P.F. Felzenszwalb, J.D. Schwartz, Hierarchical matching of deformable shapes, In: *2007 IEEE Conference on Computer Vision and Pattern Recognition, CVPR'07*, IEEE, 2007, pp. 1–8.
- [30] G. McNeill, S. Vijayakumar, Hierarchical procrustes matching for shape retrieval, In: *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, IEEE, 2006, pp. 885–894.
- [31] C. Xu, J. Liu, X. Tang, 2d shape matching by contour flexibility, *IEEE Trans. Pattern Anal. Mach. Intell.* 31 (1) (2009) 180–186.
- [32] S. Belongie, J. Malik, J. Puzicha, Shape matching and object recognition using shape contexts, *IEEE Trans. Pattern Anal. Mach. Intell.* 24 (4) (2002) 509–522.
- [33] H. Ling, D.W. Jacobs, Shape classification using the inner-distance, *IEEE Trans. Pattern Anal. Mach. Intell.* 29 (2) (2007) 286–299.
- [34] L.J. Latecki, R. Lakämper, U. Eckhardt, Shape descriptors for non-rigid shapes with a single closed contour, in: *2000 Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, IEEE, 2000, pp. 424–429.
- [35] X. Yang, S. Koknar-Tezel, L.J. Latecki, Locally constrained diffusion process on locally densified distance spaces with applications to shape retrieval, in: *2009 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009*, IEEE, 2009, pp. 357–364.
- [36] X. Yang, X. Bai, L.J. Latecki, Z. Tu, Improving shape retrieval by learning graph transduction, in: *Computer Vision-ECCV 2008*, Springer, Berlin Heidelberg, 2008, pp. 788–801.
- [37] X. Bai, B. Wang, C. Yao, W. Liu, Z. Tu, Co-transduction for shape retrieval, *IEEE Trans. Image Process.* 21 (5) (2012) 2747–2757.
- [38] P. Kontschieder, M. Donoser, H. Bischof, Beyond pairwise shape similarity analysis, in: *Computer Vision-ACCV 2009*, Springer, Berlin Heidelberg, 2010, pp. 655–666.
- [39] M. Donoser, H. Bischof, Diffusion processes for retrieval revisited, In: *2013 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, IEEE, 2013, pp. 1320–1327.
- [40] Y. Rui, A.C. She, T.S. Huang, Modified Fourier descriptors for shape representation—a practical approach, In: *Proceedings of First International Workshop on Image Databases and Multi Media Search*, 1996, pp. 22–23.
- [41] O. Söderkvist, Computer vision classification of leaves from swedish trees.
- [42] B. Leibe, B. Schiele, Analyzing appearance and contour based methods for object categorization, In: *2003 Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, IEEE, 2003, pp. ii–409.
- [43] T.B. Sebastian, P.N. Klein, B.B. Kimia, Recognition of shapes by editing their shock graphs, *IEEE Trans. Pattern Anal. Mach. Intell.* 26 (May) (2004) 550–571.
- [44] Z. Wang, M. Liang, Locally affine invariant descriptors for shape matching and retrieval, *IEEE Signal Process. Lett.* 17 (9) (2010) 803–806.
- [45] F.R. Schmidt, E. Toppe, D. Cremers, Efficient planar graph cuts with applications in computer vision, In: *2009 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009*, IEEE, 2009, pp. 351–356.
- [46] Z. Tu, A.L. Yuille, Shape matching and recognition—using generative models and informative features, In: *Computer Vision-ECCV 2004*, Springer, Berlin Heidelberg, 2004, pp. 195–209.
- [47] A. Temlyakov, P. Dalal, J. Waggoner, D. Salvi, S. Wang, Shape and image retrieval by organizing instances using population cues, In: *2013 IEEE Workshop on Applications of Computer Vision (WACV)*, IEEE, 2013, pp. 303–308.
- [48] S. Mouine, I. Yahiaoui, A. Verroust-Blondet, A shape-based approach for leaf classification using multiscaletriangular representation, In: *Proceedings of the 3rd ACM Conference on International Conference on Multimedia Retrieval*, ACM, New York, 2013, pp. 127–134.
- [49] J. Wu, J.M. Rehg, Where am i: place instance and category recognition using spatial PACT, In: *2008 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2008*, IEEE, 2008, pp. 1–8.
- [50] M.R. Daliri, V. Torre, Robust symbolic representation for shape recognition and retrieval, *Pattern Recognit.* 41 (5) (2008) 1782–1798.
- [51] X. Bai, L.J. Latecki, Path similarity skeleton graph matching, *IEEE Trans. Pattern Anal. Mach. Intell.* 30 (7) (2008) 1282–1292.
- [52] D. Zhang, G. Lu, Shape-based image retrieval using generic Fourier descriptor, *Signal Process.: Image Commun.* 17 (10) (2002) 825–848.