# Using the Time Warping Distance for Fourier-Based Shape Retrieval

Ilaria Bartolini, Paolo Ciaccia, Marco Patella
*DEIS - IEIIT/BO-CNR*, University of Bologna, Italy
e-mail: {ibartolini, pciaccia, mpatella}@deis.unibo.it

**Abstract**

Modern content-based image retrieval systems use different low-level features, like color, texture, and shape, to search in large image databases for those images which are perceptually similar to a given query image. Effective and efficient retrieval by shape similarity is still an open issue, despite the high importance of the shape feature in describing the content of an image. In this paper, we propose a new approach (based on the Discrete Fourier Transform) for assessing the shape similarity between two objects. The use of Fourier coefficients allows a compact representation of shape boundaries which is robust to noise and can be easily made independent of translation, scaling, rotation and changes in the starting point used to describe each boundary. Since the Euclidean distance, which is used by almost all Fourier-based approaches to shape-matching, is not effective whenever a phase shifting exists between the two boundaries to be compared, we propose the use of the (Dynamic) Time Warping distance to compare shape descriptors, allowing a (limited) elastic stretching of the time axis to accommodate possible phase differences. A comparative experimental analysis conducted on a real data set shows the superior effectiveness of our approach with respect to existing Fourier-based techniques.

## 1 Introduction

Large image databases are increasingly used in many multimedia areas like crime prevention, architectural and engineering design, fashion, medical diagnosis, journalism and advertising, geographical information systems. The efficient and effective retrieval of images by their content has, consequently, gained a considerable attention. The classical approach to solve the image retrieval problem is to characterize the image content by way of a set of features, then, at query time, only those images are retrieved whose features are most similar (as assessed by means of a similarity function) to those extracted from the query image [FSN+95, PPS96, SC96, ABP99, CTB+99, NRS99, SWS+00, WLW01]. Even if features like color and texture, able to effectively represent the content of images, have been proven

to be quite easy to extract and to require few parameters to be represented [HSE$^+$95, SO95, MM96], the same it is not true for the shape property [Pav78, RSH96, MM99, BDelBP00]. Shape representation has, in fact, proven to be a difficult problem [Mum87] because capturing the shape content is a more complex task than representing color and texture features. For this reason, shape characterization still represents a challenging topic for the multimedia scientific research community.

When designing a shape-based retrieval system two primary issues have to be considered:

**Shape representation.** How can an object be represented in terms of its shape properties?

**Similarity measure.** Given the representation of two shapes, how they should be compared?

Since we are dealing with large image databases, the shape representation and the similarity measure should satisfy three basic requirements:

**Compactness:** A significant compactness of the shape representation with respect to the original shape boundary is required, since the reconstruction of the original shape from its representation is not required.

**Indexability:** Since we suppose the cardinality of the database to be very large (in the order of millions of objects), the naïve solution of comparing (the representation of) the query object with all the database objects is not feasible. To avoid this, the shape representation and the similarity measure should allow the objects to be indexed using a suitable structure, e.g. with metric access methods like the M-tree [CPZ97] which are already profitably used in several other image and multimedia applications [HSE$^+$95, BDelBP00].

**Robustness to noise:** Imaging conditions, e.g. perspective or discretization transformations resulting from changing the viewing angle, and distortion introduced by the segmentation and the shape extraction processes can slightly alter the overall shape of an object. Both the shape representation and the similarity measure should be robust to such noise, in the sense that the similarity between a given shape and the same shape with added noise should be high.

**Invariance to transformations:** The overall shape representation should be invariant to a number of transformations. In particular, we are interested in translation, scaling, and rotation transformation. Moreover, since the extraction process typically produces a boundary which is considered as a discrete-time signal, the representation should also be invariant with respect to the starting point, i.e. the initial point of the boundary parameterization.

Shape representation methods fall in two major categories [Del99]: The transformational approach and the feature vector technique. Even if, in the general case, the former case has proved to be more robust to noise and distortion than the latter [Del99], since its shape similarity model is closer to human perception, the choice of a particular representation is driven by application needs, like characteristics of the shapes being analyzed, robustness again noise, possibility of indexing, and invariance properties.

The transformational model is typically used in the pattern recognition literature [Wid73]. With this approach a shape is regarded as a template and it is deformed in order to improve its match with a target image, i.e. the original image is deformed until a fitting with the target image is found. The overall dissimilarity between two shapes is taken as the cost for transforming one shape into the other. This solution has been largely used in image retrieval [DelBP97], to find correspondences between points of two different shapes, or for classification purposes. The main disadvantage of this approach, however, is that it does not support indexing, due to the fact that the method used to assess similarity usually does not satisfy metric postulates, like symmetry and triangle inequality [BDelBP00], and consequently it is not suited for large image databases.

On the other hand, the feature vector approach is widely employed in information retrieval and allows effective indexing. In detail, a shape is represented as a numerical vector using a parametric *internal* (or *global*) method, where the region enclosed by the object contour is represented [FSN+95], or a parametric *external* method, where the external *boundary* of the object is represented [RSH96, MM99, BDelBP00]. Since the parametric external representation is less sensible to edge noise with respect to the internal one, it is thus largely used in the field of content-based image retrieval [Del99]. In external methods, a finite set of points is extracted from each boundary, representing the vertices of a polygon that approximates the shape of the object. The polygon can then be seen as a discrete-time signal, each vertex corresponding to a signal sample. This allows the application to objects boundaries of existing techniques for signal processing (either in the time or in the frequency domain). Several schemas have been proposed to represent such complex contour: All of them use autoregressive models [SKO92] or Fourier analysis [ZR72] to extract descriptors from the boundary of each object. An experimental comparison of shape representation models based on these two techniques is presented in [KSP95] where the superior effectiveness of Fourier-based method with respect to autoregressive-based solutions, especially for noisy images, is proven. Moreover, the preservation of invariance properties (i.e. scale-, translation-, rotation-, and starting point-invariance) is easily obtained by switching to the frequency domain.

At present time, however, shape-based retrieval systems following the Fourier-based approach are hindered by the problem of following too simplistic ways to ensure the invariance properties and/or to determine the sim-

ilarity value between two shapes, since they discard a part of the information about Fourier coefficients. Moreover, almost all the existing techniques use the Euclidean distance to compare the extracted descriptors. This does not allow phase shiftings of subsequences of the shape, thus similar shapes that are not aligned along the time axis lead to counter-intuitive high distance values [Keo02].

In this paper we present a new shape matching approach based on the Discrete Fourier Transform. We show how a transformation-invariant representation of the shape boundary can be derived from Fourier coefficients, using both the magnitude and the phase information, since, in our view, phase is as much important as magnitude, carrying a significant part of the information about the shape. Furthermore, to overcome the limits of the Euclidean metric in accommodating phase differences between similar shapes, we propose the use of the *Dynamic Time Warping* distance [BC94] to compare the (modified) Fourier descriptors. Even if the DTW distance does not satisfy the metric postulates, recent results [Keo02] demonstrate how the Dynamic Time Warping distance can be indexed, thus the indexability requirement is satisfied. Experiments conducted on a real data set of shapes of marine animals demonstrate the robustness to noise of our approach and its superior effectiveness with respect to existing Fourier-based approaches.

The rest of the paper is organized as follows. Section 2 reports the state of the art of Fourier-based shape retrieval systems. In Section 3 we describe the details of the approach we propose, formally demonstrating how the invariance properties are satisfied. In Section 4 the experimental setup used to prove the effectiveness and the efficiency of our shape matching algorithm is described. An experimental comparison between our approach and existing methods is given in Section 5. Finally, Section 6 concludes the paper, pointing out interesting directions for future research.

## 2  Related Work

In the following, we report the state of the art for external shape-based representations, together with the distance used for comparing shape descriptors. We will primarily focus on Fourier-based methods because of the their proved effectiveness [KSP95].

The use of the Discrete Fourier Transform (DFT) [PTVF92] for the retrieval of shapes dates back to the early years of computer science [Cos60, ZR72, PF77]. The analysis in the frequency domain allows to achieve the invariance properties with simple calculations on the Fourier coefficients. Basically, existing Fourier-based methods work as follows:

1. The shape boundary is parameterized to obtain a discrete-time periodic signal, where the signal period equals the number $N$ of points in

the boundary.

2. Since the number of boundary points differs from one object to another, and it is usually large, the signal parameterization is re-sampled to include only a limited number, fixed for all objects, of interest points, e.g. by including only contour points where the curvature has a singular value. If the number of interest points is appropriately chosen, this step also allows to efficiently compute the DFT, e.g. using the Fast Fourier Transform (FFT).

3. The DFT is performed on the re-sampled signal, obtaining a (reduced) number of Fourier coefficients.

4. The so-obtained coefficients are then modified to achieve the desired invariance properties. For example, in order to obtain the translation- and the size-invariance, the first (DC) Fourier coefficient is discarded and the amplitude of other coefficients is normalized by dividing them by the first non-zero coefficient amplitude.

5. Finally, at query time, the modified coefficients of the query object are compared against the coefficients of database shapes using a distance function to find the shapes more similar to the query.

Differences between Fourier-based techniques, thus, arise in: (i) The parameterization of the original signal; (ii) How the parameterized signal is re-sampled to reduce its length; (iii) How the Fourier coefficients are modified in order to achieve the rotation- and starting point-invariance; (iv) The distance function used to compare the coefficients.

In general, the following different parameterizations are considered as functions of the curvilinear position $l$ of each point along the boundary:

***Complex coordinate* function:** The parameterization assigns to each point of the boundary, with coordinates $x(l)$ and $y(l)$, a complex value $z(l) = x(l) + j\,y(l)$, with $j = \sqrt{-1}$. This function was considered in [KSP95, RSH96, MM99, ZL02, RM02].

***Centroid distance* function:** Here the parameterization consists of the distances of boundary points from the centroid of the object, $r(l) = \sqrt{(x(l) - x_c)^2 + (y(l) - y_c)^2}$, where $x_c = 1/N \sum_{l=0}^{N-1} x(l)$ and $y_c = 1/N \sum_{l=0}^{N-1} y(l)$. This parameterization is translation-invariant and was used in [KSP95, MM99, ZL02]. The major drawback of this approach resides in the fact that the phase information of the shape boundary is discarded, since it is immediate to derive that $r(l) = |z(l)|$ except for a translation of the space origin in the shape centroid.

***Tangent angle* function:** Here, each point of the boundary is represented by the value of the tangent angle in that point, indicating the change

5

in angular direction of the shape boundary. This function was first introduced in [ZR72] and was also used in [PF77, ZL02]. The drawback of this parameterization is that it is very sensitive to discretization noise, since noisy contours introduce significant variations in the tangent angle value, however it is translation- and scale-invariant.

*Curvature* **function:** In this case, for each point of the boundary the change in the tangent angle function is assessed, computed as follows:

$$K(l) = \arctan\left(\frac{y(l) - y(l-w)}{x(l) - x(l-w)}\right) - \arctan\left(\frac{y(l-1) - y(l-1-w)}{x(l-1) - x(l-1-w)}\right)$$
(1)

where the smoothing factor $w > 1$ is used to reduce the influence of the discretization noise in computing the differentiation of the angle function. This parameterization was used in [KSP95, MM99, ZL02] and is translation-, scale-, and rotation-invariant, but, as the tangent angle parameterization, suffers from discretization noise sensitivity .

*Area* **function:** For each point, the area of the triangle formed by the point itself, the previous point and the centroid of the object boundary is considered. Such parameterization was considered in [ZL02] and is translation- and rotation-invariant.

Most of the approaches reduce the length of the boundary signals by maintaining only a fixed number $M$ of points. This is usually carried out by sampling the parameterized signal every $N/M$ steps [KSP95, RSH96, MM99]. However, this could lead to miss significant boundary points, e.g. points where the boundary tangent angle changes sharply (see differences in the tails of the two ray images in Figure 2). A possible solution to this problem is to maintain the $M$ points having highest curvature values, where the curvature in each point is defined as in Equation 1. However, this approach is prone to discretization noise, as Figure 3 shows (note the different representation of the ray nose and tail). Other approaches [ZL01, ZL02, RM02] compute the DFT on the original signal and only keep a limited number of Fourier coefficients. This approach does not allow the use of Fast DFT algorithms, since the whole signal should be used to compute the DFT, but, as shown in Figure 4, points extracted from similar images will be very similar. We investigate the effectiveness of the different approaches in Section 5.

In order to achieve the translation-invariance for the complex parameterization, one has to simply discard the DC Fourier coefficient. To obtain the scale-invariance for the complex, the centroid distance, and the area functions, the amplitude of the Fourier coefficients has to be normalized, e.g. by dividing all the modules by the first non-zero module. Changes in the rotation of the shape or in the starting point, however, influence the phase of the
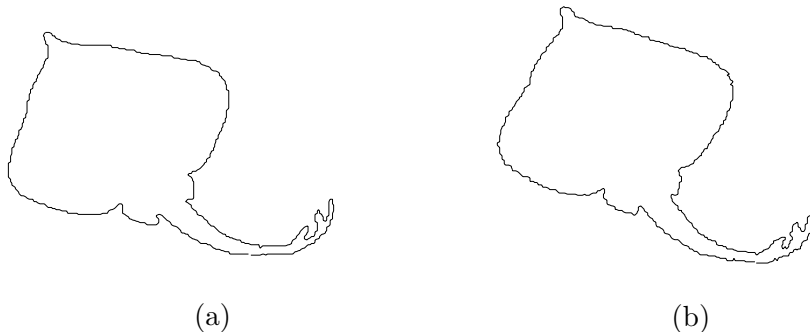
(a)             (b)

Figure 1: Ray image: (a) Original image; (b) image rotated by 10 degrees.
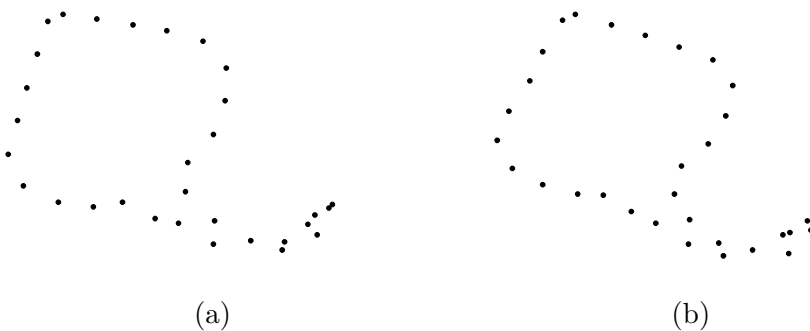


(a)             (b)

Figure 2: Ray image sampled with $M = 32$: (a) Original image; (b) rotated image.

Fourier coefficients: Most of the approaches [KSP95, MM99, ZL02, RM02] simply consider the (normalized) modules of the Fourier coefficients, thus discarding the phase information. However, since the phase of the Fourier coefficients contains a significant fraction of the information of the original signal, by simply discarding it a part of the information useful to retrieve shapes similar to the query is potentially lost. The only exception to this simplistic approach is presented in [RSH96], where the phase information is maintained and the starting point-invariance is resolved by modifying the phase coefficients. However, the amplitude and the phase coefficients are not integrated to obtain a single distance between modified signals, but Fourier coefficients of the two shapes to be matched are directly compared obtaining two different distance measures, $d_{ampl}$ and $d_{phase}$ (for example, in [RSH96] such distances are obtained as the variance over all the ratios of corresponding amplitude coefficients, $d_{ampl}$, and the variance over all the
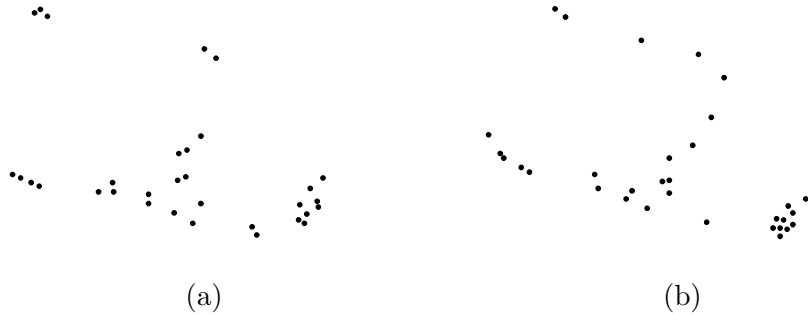
(a)                                           (b)

Figure 3: Maximum curvature points for ray images: (a) Original image; (b) rotated image.



(a)                                           (b)

Figure 4: Points obtained by computing the Inverse DFT using only the the first $M = 32$ Fourier coefficients for ray images: (a) Original image; (b) rotated image. The original image is super-imposed to show differences.

shifts of corresponding phase coefficients, $d_{phase}$). In order to obtain a single distance value, used to rank shape descriptors, the two distance values are then linearly combined using a weight value: $d = (1 - \lambda) \, d_{ampl} + \lambda \, d_{phase}$. Now, the issue of finding a good value for $\lambda \in [0, 1]$ is raised; such problem cannot, however, solved by recurring to intuition, since $\lambda$ has no perceptual meaning, and an empirical value is suggested in [RSH96].[1]

As to the distance function used to compare (modified) Fourier coefficients, all the approaches except [RSH96] use the Euclidean distance. As we pointed out in the introduction, however, this does not allow an elastic shifting of the time scale to allow phase differences between similar shapes.

Another technique similar to Fourier-based methods is based on the curvature scale-space (CSS) representation of the boundary [MM92], and it

---

[1]It has to be noted that the value of the weight $\lambda$ has to be fixed in advance and cannot be modified at query time; otherwise, indexability has to be given up.

is included in the MPEG-7 standard [Bob01]. In this case, the curvature parameterization expressed by Equation 1 is smoothed by applying a 1-D Gaussian kernel of width $\sigma$. As a result of the smoothing, by increasing the value of $\sigma$, the contour evolves to become a convex curve with no curvature zero-crossings. For each value of $\sigma$, the CSS image of a boundary is implicitly defined as the position of the points where the smoothed curvature has a value of zero (note that, since the boundary is convex, the number of such points is even for all values of $\sigma$). The CSS descriptor consists of the peaks, i.e. the maxima, of the CSS image. Such image is invariant to translation, scale, and rotation transformations. However, since changing the starting point introduces an horizontal shift in the CSS representation, the invariance is obtained by shifting the whole CSS image until the position of the highest peak is in the origin. When comparing two CSS images, the peaks of the two CSS images are matched and the similarity between two shapes is measured by summing all the peak differences. The matching algorithm, however, is very complex, since the number of peaks of the two images to be compared can be very different and an alignment between peaks sometimes cannot be found. Moreover, since CSS descriptors only capture local features such as the location and the degree of concavity (or convexity) of segments of the shape boundary, global features are not considered.

## 3 Proposed Approach

Our approach is based on the *complex coordinate* parameterization introduced in Section 2, thus we consider the boundary of an object $O$ as a discrete-time complex periodical signal, $z(l) = x(l) + j\,y(l)$, with period $N$, where $N$ is the total shape length, i.e. the number of points included in the object boundary. We extract, from each shape boundary, a feature descriptor $Z = f(z)$ representing the original signal in a compact way. The requirements for such descriptor are the following:

**Robustness:** The representation should be robust with respect to possible noise, e.g. the *spatial discretization* noise.

**Invariance to a group of transformations:** A measure is invariant to a group of transformations $G$ if, for every $g \in G$, it is $f(z) = f(g(z))$. In particular, the transformations we are interested in are: Scaling, shifting, rotation, and changes in the starting point used to define the boundary.

**Compactness:** Since each descriptor will be indexed in the database, it is important to keep its size limited to the maximum possible extent.

**Extraction efficiency:** The computation of $Z$ from the original signal $z$ should be performed quickly. This requirement is not strict, since it is

9

only performed once at population time for each database shape and once for each query.

One of the simplest and most used ways (e.g. see [Cos60, ZR72, PF77, KSP95, RSH96, MM99, ZL02]) to accomplish the above requirements is to map the original signal $z$ in the frequency domain by way of the Discrete Fourier Transform [PTVF92]. The DFT $Z(m)$ of a signal $z(l)$ is defined as

$$Z(m) = \sum_{l=0}^{N-1} z(l)e^{-j\frac{2\pi lm}{N}} = R(m)e^{j\Theta(m)} \qquad m \in \mathbb{Z} \qquad (2)$$

where $R(m)$ and $\Theta(m)$ are the module and the phase of each Fourier coefficient, respectively. It has to be noted that $Z(m)$ is also periodic, with period $N$.

In order to guarantee compactness and robustness to noise, we propose to only use the low frequency Fourier coefficients. Denoting with $M$ the number of coefficients to maintain, this equals to discard all the coefficients except for the first $M$. Thus we obtain

$$Z(m) = \sum_{l=0}^{N-1} z(l)e^{-j\frac{2\pi lm}{N}} \qquad m = -M/2, \ldots, -1, 0, 1, \ldots, M/2 - 1 \quad (3)$$

In general, the computation of each Fourier coefficient is in $O(N)$, thus the overall complexity for the extraction of the $Z$ descriptors is in $O(M \cdot N)$. Note that we cannot use the Fast Fourier Transform [PTVF92], whose complexity for extracting *all* the $N$ coefficients is in $O(N \log N)$, since the FFT requires $N$ to be a power of 2, and this is not true in general.[2]

As for the value of $M$, we have to trade-off the accuracy in representing the original signal, which increases with $M$, with the compactness and the extraction efficiency, which decrease with $M$. To help in the choice for a good value of $M$, it is useful to analyze the spectral characteristics of the data set at hand. For example, in Figure 5, we plot the contribution to the total signal energy for the first $M$ Fourier coefficients (we omitted the 0-frequency coefficient, since it only gives information about the position of the shape), obtained by averaging results over all the FISHES data set described in Section 5.

From the graph we see that, when using the first 16 coefficients, we keep the 83.7% of the energy of the original signal. For $M = 32$ the energy value increases to 89.8%. Increasing the value of $M$ to 64, we obtain the 93.4% of the total energy. In order to obtain the 99% of the total energy, we have to increase $M$ to the value of 512, which is far too high if we want to keep

---

[2]Other fast DFT algorithms exist [PTVF92]; however, since we are only interested in the first $M \ll N$ coefficients, our experiments (see Section 5) show that the described approach is quick enough for the scenario we are interested in.
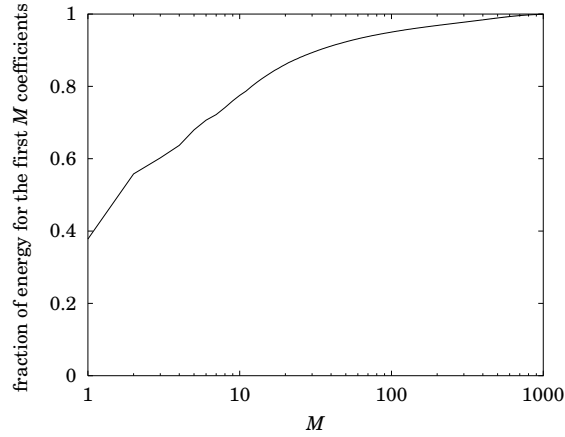
Figure 5: Total energy for Fourier coefficients. The graph shows the fraction of the total signal energy obtained when maintaining only the first $M$ Fourier coefficients.

extraction costs low. Therefore, for the FISHES data set, we can conclude that $M$ should be chosen in the interval $[16 \div 64]$. We will investigate in Section 5 the effect of $M$ on the effectiveness of the method for the FISHES data set.

By using the Fourier coefficients, we are also able to guarantee the translation-, scaling-, rotation-, and starting point-invariance, as it will be shown in the following.

## 3.1 Translation Invariance

Consider the boundary $z'(l)$ obtained from the signal $z(l)$ by simply shifting each point by a constant value $\overline{z}$, $z'(l) = z(l) + \overline{z}$ (see Figure 6).
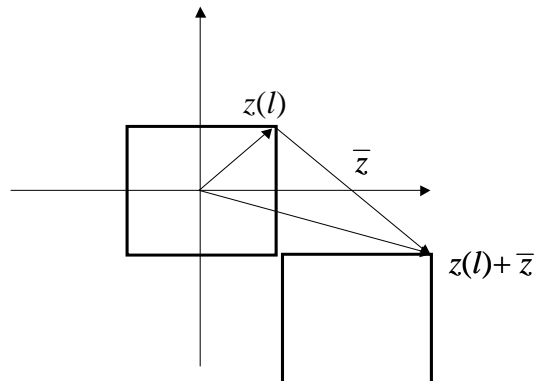


Figure 6: Geometrical interpretation of the translation operation.

11

The DFT $Z'$ of the shifted signal is:

$$Z'(m) = \sum_{l=0}^{N-1} z'(l)e^{-j\frac{2\pi lm}{N}} = \sum_{l=0}^{N-1} (z(l) + \overline{z})e^{-j\frac{2\pi lm}{N}} =$$
$$= \sum_{l=0}^{N-1} z(l)e^{-j\frac{2\pi lm}{N}} + \sum_{l=0}^{N-1} \overline{z}e^{-j\frac{2\pi lm}{N}} = Z(m) + \overline{z}\sum_{l=0}^{N-1} e^{-j\frac{2\pi lm}{N}} = \quad (4)$$
$$= Z(m) + \overline{z}\chi_0(m)$$

where $\chi_0(m) = 0$, $\forall m \neq 0$, and $\chi_0(0) = 1$. Thus, the translation introduces a variation in the descriptor only in the DC coefficient. The shape descriptor, thus, should simply discard the zero frequency coefficient in order to achieve the translation invariance. This is reasonable, since the DC coefficient represents the mean of the original signal $z(l)$, thus it gives information only about the shape centroid, i.e. on the spatial location of the object, and not about its shape.

## 3.2   Scale Invariance

Consider a zero-mean signal $z(l)$:[3] Let $z'(l)$ be a boundary obtained from $z(l)$ by scaling each point by a constant factor $\beta \in \Re^+$, $z'(l) = \beta z(l)$ (see Figure 7).
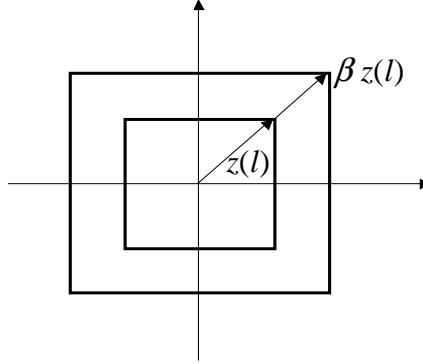


Figure 7: Geometrical interpretation of the scale transformation.

The Fourier coefficients $Z'$ of the new scaled signal are obtained as:

$$Z'(m) = \sum_{l=0}^{N-1} z'(l)e^{-j\frac{2\pi lm}{N}} = \sum_{l=0}^{N-1} \beta z(l)e^{-j\frac{2\pi lm}{N}} = \beta Z(m) \quad (5)$$

Each coefficient, thus, is scaled by the value of $\beta$. To achieve scale-invariance, we can easily normalize the module of all coefficients, e.g. by dividing them by the first module $Z(\overline{m}) \neq 0$, the arguments remaining untouched.

---

[3]If the boundary signal has mean different from zero, we can simply apply the translation invariance presented in Section 3.1.

## 3.3 Rotation Invariance

Consider the zero-mean signal $z(l) = r(l)e^{j\theta(l)}$, where with $r(l)$ and $\theta(l)$ we denote module and phase of each sample, respectively. Let $z'(l)$ be a boundary obtained from $z(l)$ by rotating each point by a constant factor $\theta$, $z'(l) = r(l)e^{j(\theta(l)+\bar{\theta})} = r(l)e^{j\theta(l)}e^{j\bar{\theta}} = z(l)e^{j\bar{\theta}}$ (see Figure 8).
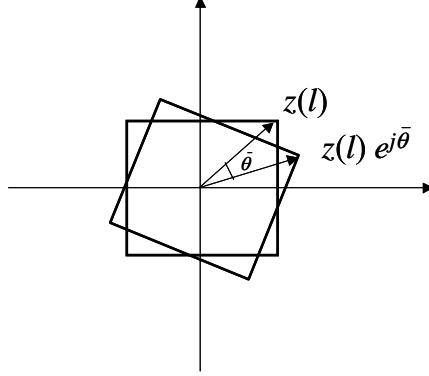


Figure 8: Geometrical interpretation of the rotation transformation.

The corresponding DFT is:

$$Z'(m) = \sum_{l=0}^{N-1} z'(l)e^{-j\frac{2\pi lm}{N}} = \sum_{l=0}^{N-1} z(l)e^{j\bar{\theta}}e^{-j\frac{2\pi lm}{N}} = Z(m)e^{j\bar{\theta}} \qquad (6)$$

It is clear that object rotation only changes the argument of the coefficients (the modules remaining untouched). Thus, to obtain rotation invariance, it is sufficient to subtract to all the arguments a constant value, e.g. the argument of the first coefficient, $\Theta(1) = arg(Z(1))$.[4]

## 3.4 Starting Point Invariance

Changing the starting point used in the definition of the boundary sequence, corresponds to a shifting in the time domain, since the signal $z(l)$ is periodic. The new boundary $z'(l)$ can therefore be obtained as $z'(l) = z(l - l_0)$, where $l_0$ is the index of the new starting point in the original signal (see Figure 9).

The corresponding Fourier coefficients are obtained as:

$$\begin{aligned} Z'(m) &= \sum_{l=0}^{N-1} z'(l)e^{-j\frac{2\pi lm}{N}} = \sum_{l=0}^{N-1} z(l - l_0)e^{-j\frac{2\pi lm}{N}} = \\ &= \sum_{l=0}^{N-1} z(l - l_0)e^{-j\frac{2\pi(l-l_0)m}{N}}e^{-j\frac{2\pi l_0 m}{N}} = Z(m)e^{-j\frac{2\pi l_0 m}{N}} \end{aligned} \qquad (7)$$

---

[4]Remember that the DC coefficient has been already discarded to achieve translation-invariance.
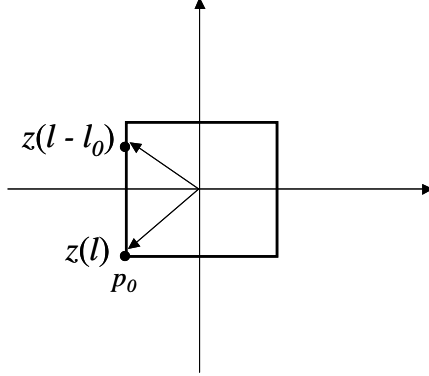
Figure 9: Geometrical interpretation of the starting point transformation.

A shift in the time domain thus introduces a rotation in the Fourier coefficients which is linear in the frequency value. To get rid of this factor, we should subtract, to the arguments of all the DFT coefficients, a term which is linear in $m$. Omitting the straightforward algebra, we obtain that such term is equal to $m\frac{\Theta(-1)-\Theta(1)}{2} - \frac{\Theta(1)+\Theta(-1)}{2}$.

Summarizing, the invariant Fourier coefficients $\widehat{Z}(m) = \widehat{R}(m)e^{j\widehat{\Theta}(m)}$ can be obtained from the DFT coefficients as follows (without loss of generality, suppose that $R(1) = |Z(1)| \neq 0$):

$$\widehat{Z}(0) = 0$$

$$\widehat{R}(m) = \frac{R(m)}{R(1)} \qquad m = -M/2, \ldots, -1, 1, \ldots, M/2 - 1 \tag{8}$$

$$\widehat{\Theta}(m) = \Theta(m) - \frac{\Theta(1) + \Theta(-1)}{2} + m\frac{\Theta(-1) - \Theta(1)}{2} \qquad m = -M/2, \ldots, -1, 1, \ldots, M/2 - 1$$

We are now able to obtain a modified signal $\widehat{z}(l)$, which is scale-, translation-, rotation-, and starting point-invariant, by simply performing the Inverse DFT on the $\widehat{Z}(m)$ coefficients of Equation 8:

$$\widehat{z}(l) = \frac{1}{M} \sum_{m=-M/2}^{M/2-1} \widehat{Z}(m)e^{j\frac{2\pi lm}{M}} \qquad l = 0, \ldots, M \tag{9}$$

For each shape in the database, the corresponding modified signal $\widehat{z}(l)$ is stored and used, at query time, to find the shapes which are more similar to the given query. To this end, a (dis-)similarity measure $d$ is used to compare the modified query signal $\widehat{z_q}$ with the database descriptors. Formally, if we denote with $\mathcal{U}$ the space of all possible signals, we have to find a function $d : \mathcal{U} \times \mathcal{U} \to \Re_0^+$ having the following characteristics:

**Computation efficiency:** Since the query resolution is done on-line, the computation of $d$ between any two descriptors should be performed quickly.

14

**Metric property:** The use of index structures, advocated in Section 1 to reduce query resolution times, requires that the distance used to compare the shapes should be a metric, i.e. $d$ should be symmetric, $d(z_1, z_2) = d(z_2, z_1)$, and should also satisfy the triangle inequality: $d(z_1, z_2) \leq d(z_1, z_3) + d(z_3, z_2)$, for all $z_1, z_2, z_3 \in \mathcal{U}$.

**Stability:** If two shape boundaries are perceptually similar, the distance between the relative signals should be low. If two signals have a small distance between them, the boundaries they represent should be perceptually similar.

The usual way to compare fixed-length vectors is to use the Euclidean distance $L_2$, representing the energy of the difference between the two signals $z$ and $z'$:

$$L_2(z(l), z'(l)) = \sqrt{\sum_{l=0}^{M} |z(l) - z'(l)|^2}$$

The metric property is obviously satisfied, as well as the computation efficiency. Moreover, we do not need to perform the IDFT, since the Parseval's theorem guarantees that $L_2(\widehat{Z}(m), \widehat{Z}'(m)) = L_2(\widehat{z}(l), \widehat{z}'(l))$, thus the distance between the modified Fourier descriptors (see Equations 8) equals the distance between the modified signals.

In recent times, however, the use of other distance measures has been considered, since the Euclidean distance does not allow elastic shiftings of the time axis, e.g. to accommodate signals which are similar, but where a phase difference between some samples exists, as shown in Figure 10. To this end, we consider the use of a (Dynamic) Time Warping distance [BC94].
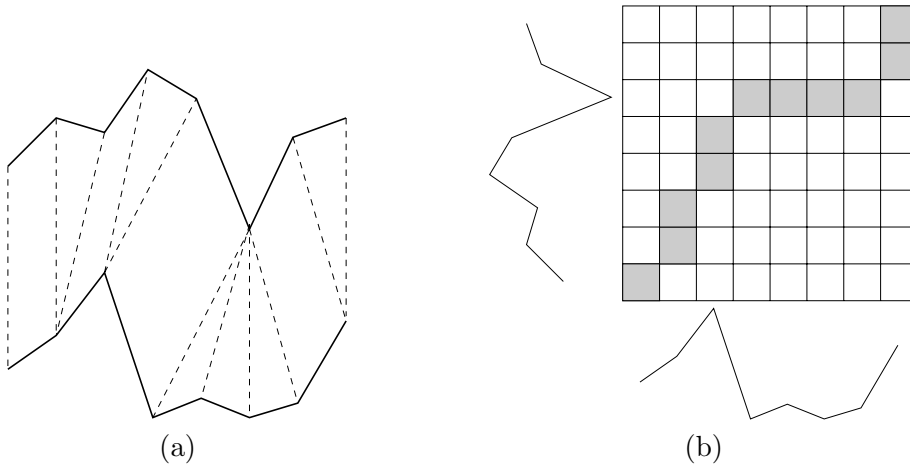


Figure 10: An example of the Time Warping distance on two real signals (a) and the alignment matrix for computing it (b).

**Definition 3.1 (Time Warping Distance)** *Given two sequences $z(l)$ and $z'(l)$ of length $M$, construct an $M \times M$ matrix, where each element $(i, j)$ corresponds to the alignment of the $i$-th sample of $z$ with the $j$-th sample of $z'$. A warping path $\mathcal{W}$ is a contiguous sequence of matrix cells defining a mapping between $z$ and $z'$. More precisely, $\mathcal{W}$ is a sequence $\mathcal{W}(k)$, $k = 0, \dots, K_{\mathcal{W}} - 1$, where $\mathcal{W}(k) = (i_k, j_k)$, and $M \leq K_{\mathcal{W}} < 2M$. The warping path should satisfy the following constraints:*

**Monotonicity:** *Cells in $\mathcal{W}$ should be monotonically ordered, i.e. for consecutive pairs $\mathcal{W}(k) = (i_k, j_k)$ and $\mathcal{W}(k + 1) = (i_{k+1}, j_{k+1})$, it is $i_{k+1} - i_k \geq 0$ and $j_{k+1} - j_k \geq 0$, with at least one inequality being strict.*

**Continuity:** *Steps in the path are constrained to neighboring cells, i.e. for consecutive pairs $\mathcal{W}(k) = (i_k, j_k)$ and $\mathcal{W}(k + 1) = (i_{k+1}, j_{k+1})$, it is $i_{k+1} - i_k \leq 1$ and $j_{k+1} - j_k \leq 1$.*

**Boundary condition:** *$\mathcal{W}(0) = (0, 0)$ and $\mathcal{W}(K) = (M - 1, M - 1)$, thus the first and last samples of the two sequences should be reciprocally aligned.*

*For each path $\mathcal{W}$, the alignment cost $d_{\mathcal{W}}$ can be defined as the (square root of the) sum of distances between aligned samples:*

$$d_{\mathcal{W}} = \sqrt{\sum_{k=0}^{K_{\mathcal{W}}-1} \delta(z(i_k), z'(j_k))^2} \tag{10}$$

*where $\delta$ is an arbitrary distance function between samples (in the following we will use the Euclidean distance between the two samples, $\delta(z(i), z'(j)) = |z(i) - z'(j)|$).*

*The Time Warping distance $d_{tw}(z, z')$ is defined as the minimum cost taken over all the possible paths:*

$$d_{tw}(z, z') = \min_{\mathcal{W}} \{d_{\mathcal{W}}\} = \min_{\mathcal{W}} \left\{ \sqrt{\sum_{k=0}^{K_{\mathcal{W}}-1} \delta(z(i_k), z'(j_k))^2} \right\} \tag{11}$$

$\square$

Although an exponential number of possible paths exist, the Time Warping distance can be computed using a dynamic programming algorithm in $O(M^2)$ time. The algorithm is based on the following recurrence relation, defining the cumulative cost $d(i, j)$ for each cell in the matrix:

$$d(i, j) = \delta(z(i), z'(j))^2 + \min\{d(i - 1, j - 1), d(i, j - 1), d(i - 1, j)\}$$

Thus, we obtain the cost for matching $z(i)$ and $z'(j)$ as the sum of the distance between the samples and the minimum cost for matching previous samples. The overall Time Warping distance is obtained as the (square root of the) cost for the last cell in the matrix, $d_{tw}(z, z') = \sqrt{d(M-1, M-1)}$.

In the vast majority of applications of the Time Warping distance, the warping path is constrained by limiting how far it can deviate from the diagonal. The most frequently used constraint is the Sakoe-Chiba band [SC78] (see Figure 11), which limits the deviation from the diagonal up to a value $\omega$, which is called the window length; thus, we have the additional constraint $|i_k - j_k| \le \omega$, $k = 0, \ldots, K_{\mathcal{W}} - 1$.[5] The motivation for this constraint is twofold: First, it limits the complexity of computing $d_{tw}$ to $O(M \times \omega)$; second, and more important, it prevents the creation of pathological paths, aligning samples which are far away from each other.
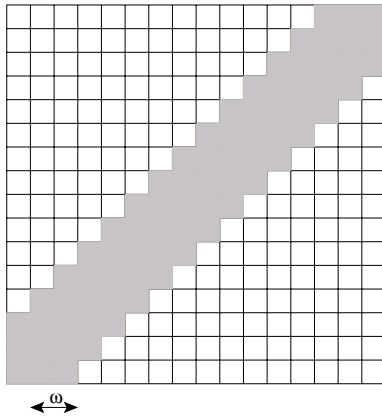


Figure 11: An illustration of the Sakoe-Chiba band for constraining the warping path for the computation of $d_{tw}$. In this example, the window length $\omega$ is 2.

Even if the Time Warping distance has a number of desirable properties, it suffers from the problem of not being a metric, since $d_{tw}$ does not satisfy the triangle inequality [YJF98]. However, indexing without losing relevant objects is still possible as long as we can find a metric $d_I$ which lower-bounds $d$, i.e. $d_I(z_1, z_2) \le d(z_1, z_2)$, $\forall z_1, z_2 \in \mathcal{U}$ [CP02]. In [Keo02] a lower-bounding metric distance for $d_{tw}$ is proposed under the conditions that the two sequences to be compared have equal length and that a global constraint on alignment, like the Sakoe-Chiba band, is used. Since this is exactly our case, we conclude that such lower-bounding distance can be used for indexing shape descriptors.

Compared to existing solution for modifying Fourier descriptors presented in Section 2, our technique nicely combines both amplitude- and

---

[5]It is worth noting that by setting $\omega = 0$ we obtain the Euclidean distance. Therefore, it is also $d_{tw}(z, z') \le L_2(z, z')$, for any value of $\omega$.

phase-related invariance, without losing important phase information or recurring to artificial weights. Moreover, to the best of our knowledge, this is the first application of the Time Warping distance to the retrieval of shapes.

# 4   Experimental Evaluation

This section presents the goals and the setup for the experimental tests of Section 5.

We implemented the shape-matching method presented in Section 3 and tested its performance, shown in Section 5, in order to (i) realize a tuning of the method itself and (ii) to compare the presented solution with other techniques described in Section 2. In details, we are interested in providing an answer to the following questions:

- Which is the best choice for the number of Fourier coefficients $M$ to be retained? Is the use of the Time Warping distance really increasing the retrieval effectiveness?

- Which is the discriminant capacity of our approach in terms of retrieval results, i.e. is it able to retrieve, as first results, only shapes that are perceptually similar to the query?

- Is our approach robust to spatial discretization noise? Is it invariant to transformations?

- What is the complexity of our approach for the extraction of the Fourier coefficients as compared to other techniques?

- Can we assert that, using both the magnitude and the phase information to ensure the rotation-invariance, results are better compared to those obtained by discarding the phase data?

**Data Set.**   To test the effectiveness of our method, we used the data provided by [AMK], consisting of 1100 text files each containing the $x$ and $y$ coordinates of boundary points of an object, each object representing a marine animal. To better test our approach in term of robustness to noise and rotation invariance, we have added to the data set other 100 images as follows: 30 pictures were randomly chosen from the starting data set and were differently rotated (e.g. by 10, 30, 90, 180, 270 degrees, etc.) obtaining about 1200 images, that we refer as the FISHES data set. Note that, while a rotation of a multiple of $\pi/2$ produce a signals without spatial discretization noise with respect to the original image, a free rotation introduces noise to the original signal due to the spatial discretization.

Since assessing the similarity of two shapes is a subjective task, we decided to use the feedback of a number of volunteers to classify the FISHES

data set. In details, using a set of 30 randomly chosen query images, we asked to each user to select, for each image, the most similar pictures in the data set. As a result, we obtained a characterization of the FISHES images based on 10 *categories* (for a total of about 300 images) where queries are (randomly) extracted from. Other images were associated to a big *compound* class, called "Misc" (for a total number of about 900 images), that we used to further complicate the retrieval process in testing the discriminant capacity of our approach. Each of the 1200 images in the FISHES data set is thus annotated with a category: "Seahorses" (10 images), "Seamoths" (6), "Sharks" (63), "Soles" (72), "Tonguefishes" (24), "Crustaceans" (4), "Eels" (31), "U-Eels" (25), "Pipefishes" (21), "Rays" (46), and "Misc" (860). For each query image, any image in the same category is considered *relevant* whereas all other images are considered *irrelevant*. Table 1 shows some examples of images together with the cardinality of their relative category.

| image | category | cardinality | image | category | cardinality |
|---|---|---|---|---|---|
|  | Rays | 46 |  | Seahorses | 10 |
|  | Sharks | 63 |  | Pipefishes | 21 |
|  | U-Eels | 25 |  | Soles | 72 |

Table 1: Examples of query images extracted from some defined classes.

**Workload.**  The workload used for our experiments consists of 30 query images randomly chosen from the set of 300 FISHES images representing the 10 extracted categories of the dataset.

**Metrics.**  To measure the effectiveness of our shape-matching algorithm, we consider classical *precision* and *recall* metrics [Sal89], averaged over the set of processed queries. For a given number $k$ of retrieved objects, the precision $P$ is the number of retrieved relevant objects over $k$, and the recall $R$ is the number of retrieved relevant objects over the total number of relevant objects (in our case, the number of images in the category of the query).

19

The formal definition is:

$$P = \frac{|relevant \bigcap retrieved|}{|retrieved|} \tag{12}$$

$$R = \frac{|relevant \bigcap retrieved|}{|relevant|} \tag{13}$$

where $|\cdot|$ indicates the cardinality of a set.

As reported above, the cardinality of each category varies from one class to another (e.g. category "Crustaceans" has only 4 examples, whereas in the "Rays" category there are 46 candidates). To overcome this disparity, we show in our results the value of the precision at a fixed recall level.

From an efficiency point of view, we compare extraction and distance times: The extraction time represents the average time needed to extract the Fourier coefficients from a boundary during the feature extraction phase, whereas the distance time represents the average time needed to compute a distance between two shapes during the retrieval process.

## 5    Experimental Results

This section presents experimental results obtained applying the implemented shape-matching algorithm to the FISHES data set. The software has been developed in C++ under the Windows NT 4.0 operating system and its performance has been tested on a 450-Mhz Pentium II PC with 256 MBytes of RAM.

Section 5.1 presents results concerning the effect of the number of Fourier coefficients $M$. In Section 5.2 results obtained using the Dynamic Time Warping distance are compared to those obtained with the simple Euclidean distance. Section 5.3 investigates the discriminant capacity of our solution during the retrieval process together with its robustness to spatial discretization noise and its invariance to transformations. Section 5.4 compares our approach with other existing Fourier-based techniques. Finally, Section 5.5 summarizes experimental results.

### 5.1    Effect of the Number of Fourier Coefficients $M$

The first question we answer concerns which is the best choice for the parameter $M$, i.e. the number of Fourier coefficients extracted from the outer boundary of each shape. As stressed in Section 3, the problem is to find a good trade-off between an accurate description of the contour and a compact representation together with an efficient extraction of it. In particular, we suggested that a good value for $M$ should be chosen in the interval $[16 \div 64]$. Towards this goal, Table 2 reports the average time needed to extract the feature vector of a single image and to compute the distance between a query

image and an image belonging the FISHES data set for different values of $M$. It should be noted that, as anticipated in Section 3, the time complexity of the extraction time is in $O(M)$, whereas the distance computation time is $O(M^2)$, since we always choose $\omega = M/10$, this being the most typical value used in the literature [SC78, Keo02].

| M | extraction time (s) | distance time (s) |
|---|---|---|
| 16 | 0.032 | $0.060 \times 10^{-3}$ |
| 32 | 0.065 | $0.198 \times 10^{-3}$ |
| 64 | 0.129 | $0.500 \times 10^{-3}$ |

Table 2: Time needed for extracting Fourier coefficients and for computing the DTW distance for 16, 32, and 64 Fourier coefficients.

To determine in which measure this time difference is translated in term of effectiveness of results, Figure 12 shows the precision-recall (P/R) graph for different values of $M$. We can observe that the P/R curve for $M = 32$ is really close to that obtained with $M = 64$. On the other hand, results for $M = 16$ are almost always 10% less than other values. This means that $M = 32$ represents a good trade-off between effectiveness of result and efficiency of storage overhead and computation.
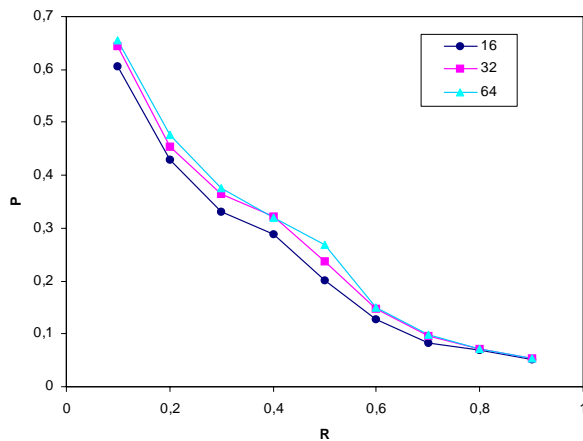


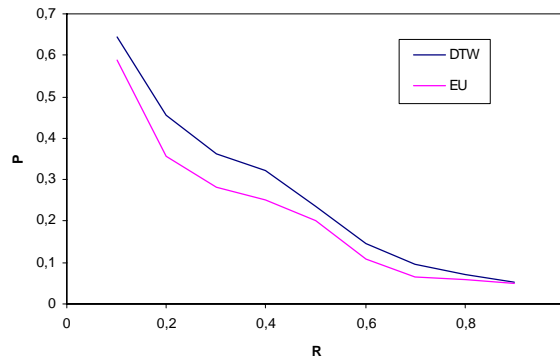Figure 12: Precision-recall graphs for different values of $M$.

## 5.2 Effect of the Distance Function

The second point we investigate is the effect of the distance function used to compare Fourier coefficients in terms of both effectiveness and efficiency.

To this end, in Figure 13 (a) we compare the distance computation time of the Euclidean distance (EU) and the Dynamic TIme Warping (DTW) distance for different values of $M$. Figure 13 (b) then shows the averaged precision-recall graph computed on the complete query workload. It is clear that, even if the DTW distance function is more complex than the Euclidean metric, the difference in effectiveness obtained during the retrieval process is more than compensating the difference in computation costs. Thus, we conclude that the DTW distance function improves the quality of results over the Euclidean distance.
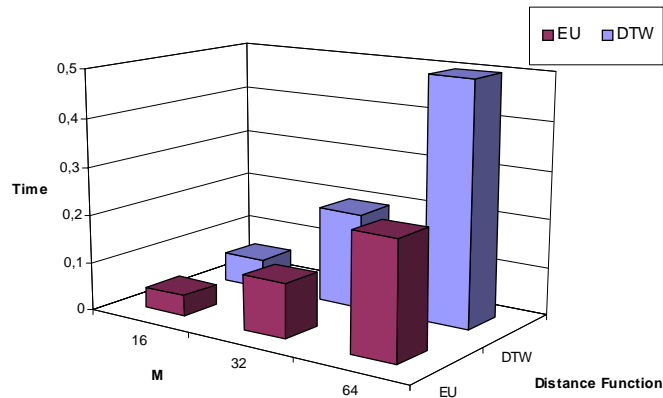


(a)



(b)

Figure 13: Euclidean vs. DTW distance function: (a) Distance computation time for different values of $M$; (b) precision-recall graph for $M = 32$.

To better show the superior retrieval effectiveness of the DTW distance, we report in Figure 14 (a) and (b), respectively, the precision-recall graphs computed over two single query images. This allows us to demonstrate the

different level of difficulty in answering a query image depending on the category of the query itself.
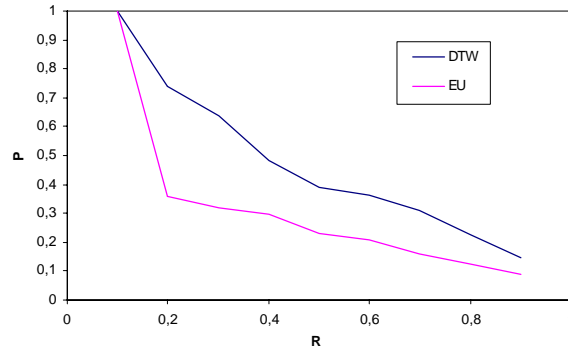


(a)
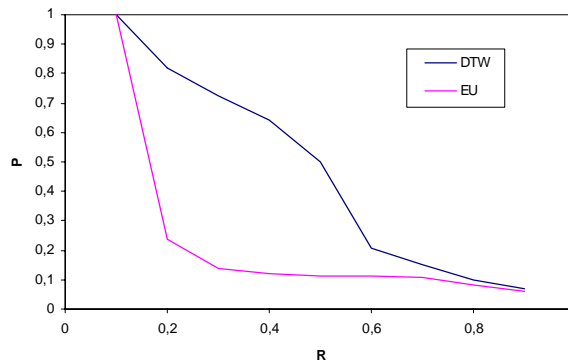


(b)

Figure 14: Precision-recall graphs for the two query images "Soles" (a) and "Rays" (b).

Finally, in Figure 15 we report a visual comparison of results obtained using the DTW and the Euclidean distances.

## 5.3 Discriminating Capacity, Robustness to Noise, Invariance to Transformations

This section demonstrates that our shape-matching algorithm has a very good discriminating power, allowing the retrieval of an high fraction of relevant objects. Furthermore, we show that it is robust to noise and invariant to transformations. To this end, we report in Figure 16 five query examples where the first image of each row is the query image and the others are the most similar shapes found by the algorithm from the FISHES data set.

We can observe that in all cases reported in Figure 16 our shape-matching algorithm is able to retrieve both the images rotated by a multiple of $\pi/2$ (i.e. signals without spatial discretization noise with respect to the query one, as observed in Section 4) and the freely rotated pictures, that present noise with respect to the query. Moreover, results show that, for each query, our approach has the capacity of discriminating among the whole FISHES data set the images belonging to the same category of the image query, returning, as first results of the search, only relevant objects.

## 5.4  Comparisons with Other Techniques

This section reports comparison results between our shape-matching algorithm and the other techniques described in Section 2. The first point we investigate concerns the best way to extract the Fourier coefficients from a boundary. As reported in Section 2, there are basically two possible methods to do this: Apply the inverse DFT approach to all the boundary points (that we refer as IDFT in the following) by computing the first 32 coefficients, or extract the $M$ points with maximum curvature from the original boundary and apply the Fourier transform to the so-obtained $M$ points [MM99] (referred as MAXC). Figures 17 (a) and 18 (a) show results obtained applying our method to two distinct query images (the first image of each row). In the same figures, row (b) reports the results obtained for the same query images using MAXC. We can observe that, in all cases, our method is much more effective than MAXC (the trend is also confirmed for images in other categories, that are not shown here for brevity). This means that our approach is more precise in finding "interesting" points of a shape. On the other hand, Table 3 reports the needed time to extract the $M$ Fourier coefficients for both the methods. We can observe that our approach requires about three times more time than MAXC, but, as observed in Section 3, this can be considered an acceptable trade-off, since the feature extraction process is only performed once at the population time for each database shape and once for each query image.

| method | extraction time (s) |
|--------|---------------------|
| IDFT   | 0.065               |
| MAXC   | 0.022               |

Table 3: Extraction time for IDFT and MAXC. Results are obtained with $M = 32$.

Our last experiment show that the conjunctive use of magnitude and phase allows to further improve the effectiveness of results. To this end, row (c) in Figures 17 and 18 show results for the same query images when we apply IDFT by dropping the phase information of each Fourier coefficient

(we call this technique IDFTM). Experimental results confirm that, also in this case, simply ignoring the phase reduces the retrieval effectiveness, because the phase carries some part of the information about the shape. We can in fact observe that, even if result for IDFTM are good compared to ours, it happens that some returned images do not belong to the query image category.

## 5.5   Summary of Experiments

In previous sections we evaluated our shape-matching algorithm on the FISHES data set for different values of $M$ and two distinct distance functions. In particular, when we use 32 Fourier coefficients with the DTW distance function, we obtain a good trade-off between efficiency of the method and effectiveness of query results. In summary, our method has proven to be robust to noise and invariant to transformations. We also showed the strong discriminating capacity of our approach in retrieving shapes relevant to the query. Furthermore, the superior effectiveness of our approach with respect to other solutions presented in Section 2 was shown.

# 6   Conclusions

In this paper we presented an effective Fourier-based approach for shape matching, able to preserve the invariance properties by way of an opportune adjustment of both magnitude and phase of Fourier coefficients. As to the distance function used to compare coefficients extracted from shape boundaries, we proposed the use of the Dynamic Time Warping distance in place of the commonly used Euclidean distance. The benefit of using the DTW distance is that it allows a (limited) elastic stretching of the time axis to accommodate for phase shiftings existing between the two boundaries to be matched. Even if, in its general form, the DTW does not satisfy the metric postulates [YJF98], thus preventing the use of indices to speed-up the retrieval phase, by exploiting recent results on metric access structures [CP02] and on the DTW itself [Keo02], we can still guarantee the indexability of Fourier coefficients extracted from large data sets. The effectiveness of our approach in retrieving only images which are relevant to the query has been demonstrated on a real world data set. A comparative analysis on the same data set has also proven the superiority of our approach with respect to other existing Fourier-based shape retrieval techniques. We are currently integrating the approach proposed here into our region-based image reteieval system, WINDSURF [ABP99], and we plan to conduct further experiments on the retrieval efficiency, when an index like the M-tree [CPZ97] is used.

# References

[ABP99]    Stefania Ardizzoni, Ilaria Bartolini, and Marco Patella. Wind-surf: Region-based image retrieval using wavelets. In *Proceedings of the 1st International Workshop on Similarity Search (IWOSS'99)*, pages 167–173, Florence, Italy, September 1999. IEEE Computer Society.

[AMK]      Sadegh Abbasi, Farzin Mokhtarian, and Josef Kittler. SQUID Demo dataset 1,500. http://www.ee.surrey.ac.uk/Research/VSSP/imagedb/demo.html.

[BC94]     Donald J. Berndt and James Clifford. Using dynamic time warping to find patterns in time series. In *Advances in Knowledge Discovery in Databases: Papers from the 1994 AAAI Workshop*, pages 359–370, Seattle, WA, July 1994. AAAI Press.

[BDelBP00] Stefano Berretti, Alberto Del Bimbo, and Pietro Pala. Retrieval by shape similarity with perceptual distance and effective indexing. *IEEE Transaction on Multimedia*, 2(4):225–239, December 2000.

[Bob01]    Miroslaw Bober. MPEG-7 visual shape descriptors. *IEEE Transactions on Circuits and Systems for Video Technology*, 11(6):716–719, June 2001.

[Cos60]    R.L. Cosgriff. Identification of shape. Technical Report 820-11, Ohio State University, Columbus, OH, December 1960.

[CP02]     Paolo Ciaccia and Marco Patella. Searching in metric spaces with user-defined and approximate distances. *ACM Transactions on Database Systems*, 27(4):398–437, December 2002.

[CPZ97]    Paolo Ciaccia, Marco Patella, and Pavel Zezula. M-tree: An efficient access method for similarity search in metric spaces. In *Proceedings of the 23rd International Conference on Very Large Data Bases (VLDB'97)*, pages 426–435, Athens, Greece, August 1997. Morgan Kaufmann.

[CTB+99]   Chad Carson, Megan Thomas, Serge Belongie, Joseph M. Hellerstein, and Jitendra Malik. Blobworld: A system for region-based image indexing and retrieval. In *Proceedings of the 3rd International Conference on Visual Information Systems VISUAL'99*, pages 509–516, Amsterdam, The Netherlands, June 1999. http://elib.cs.berkeley.edu/photos/blobworld/.

[Del99]        Alberto Del Bimbo. *Visual Information Retrieval*. Morgan Kaufmann, San Francisco, CA, 1999.

[DelBP97]     Alberto Del Bimbo and Pietro Pala. Visual image retrieval by elastic matching of user sketches. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(2):121–132, February 1997.

[FSN⁺95]    Myron Flickner, Harpreet Sawhney, Wayne Niblack, Jonathan Ashley, Qian Huang, Byron Dom, Monika Gorkani, Jim Hafner, Denis Lee, Dragutin Petkovic, David Steele, and Peter Yanker. Query by image and video content: The QBIC system. *IEEE Computer*, 28(9):23–32, September 1995.
`http://wwwqbic.almaden.ibm.com/`.

[HSE⁺95]    Jim Hafner, Harpreet Sawhney, Will Equitz, Myron Flickner, and Wayne Niblack. Efficient color histogram indexing for quadratic form distance functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(7):729–736, July 1995.

[Keo02]      Eamonn J. Keogh. Exact indexing of dynamic time warping. In *Proceedings of the 28th International Conference on Very Large Data Bases (VLDB 2002)*, pages 406–417, Hong Kong SAR, China, September 2002. Morgan Kaufmann.

[KSP95]     Hannu Kauppinen, Tapio Seppänen, and Matti Pietikäinen. An experimental comparison of autoregressive and Fourier-based descriptors in 2D shape classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17(2):201–207, February 1995.

[MM92]      Farzin Mokhtarian and Alan K. Mackworth. A theory of multiscale, curvature-based shape representation for planar curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(8):789–805, August 1992.

[MM96]      B.S. Manjunath and Wei-Ying Ma. Texture features for browsing and retrieval of image data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(8):837–842, August 1996.

[MM99]      Wei-Ying Ma and B. S. Manjunath. NeTra: A toolbox for navigating large image databases. *Multimedia Systems*, 7(3):184–198, May 1999.
`http://www-iplab.ece.ucsb.edu/netra/Netra.html`.

[Mum87]     D. Mumford. The problem of robust shape descriptor. In *Proceedings of 1st International Conference on Computer Vision*, pages 602–606, London, England, June 1987.

[NRS99]     Apostol Natsev, Rajeev Rastogi, and Kyuseok Shim. WALRUS: A similarity retrieval algorithm for image databases. In *Proceedings 1999 ACM SIGMOD International Conference on Management of Data*, pages 396–405, Philadelphia, PA, June 1999. ACM Press.

[Pav78]     Theodosios Pavlidis. A review of algorithms for shape analysis. *Computer Vision, Graphics and Image Processing*, 7:243–258, 1978.

[PF77]      Eric Persoon and King-Sun Fu. Shape discrimination using Fourier descriptors. *IEEE Transactions on Systems, Man, and Cybernetics*, 7(3):170–179, March 1977.

[PPS96]     Alex Pentland, Rosalind W. Picard, and Stan Sclaroff. Photobook: Content-based manipulation of image databases. In Borko Furht, editor, *Multimedia Tools and Applications*, chapter 2, pages 43–80. Kluwer Academic, 1996. http://web.media.mit.edu/~tpminka/photobook/.

[PTVF92]    William H. Press, Saul A. Teukolsky, William T. Vetterling, and Brian P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, Cambridge, NY, 1992.

[RM02]      Davood Rafiei and Alberto O. Mendelzon. Efficient retrieval of similar shapes. *The VLDB Journal*, 11(1):17–27, August 2002.

[RSH96]     Yong Rui, Alfred C. She, and Thomas S. Huang. Modified Fourier descriptors for shape representation – A practical approach. In *Proceedings of the 1st International Workshop on Image Databases and Multi Media Search*, pages 22–23, Amsterdam, The Netherlands, August 1996.

[Sal89]     Gerard Salton. *Automatic Text Processing: The Transformation, Analysis, and Retrieval of Information by Computer*. Addison-Wesley, Reading, MA, 1989.

[SC78]      Hiroaki Sakoe and Seibi Chiba. A dynamic programming algorithm optimization for spoken word recognition. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 26(1):43–49, February 1978.

[SC96]     John R. Smith and Shih-Fu Chang. VisualSEEk: A fully au-
           tomated content-based image query system. In *Proceedings of
           the 4th ACM International Conference on Multimedia*, pages
           87–98, Boston, MA, November 1996. ACM Press.
           `http://www.ctr.columbia.edu/VisualSEEk/`.

[SKO92]    Iwao Sekita, Takio Kurita, and Nobuyuki Otsu. Complex au-
           toregressive model for shape recognition. *IEEE Transactions
           on Pattern Analysis and Machine Intelligence*, 14(4):489–496,
           April 1992.

[SO95]     Markus Stricker and Markus Orengo. Similarity of color im-
           ages. In *Storage and Retrieval for Image and Video Databases
           SPIE*, volume 2420, pages 381–392, San Jose, CA, February
           1995.

[SWS$^+$00]  Arnold W. M. Smeulders, Marcel Worring, Simone Santini,
           Amarnath Gupta, and Ramesh Jain. Content-based image re-
           trieval at the end of the early years. *IEEE Transactions on
           Pattern Analysis and Machine Intelligence*, 22(12):1349–1380,
           December 2000.

[Wid73]    B. Widrow. The rubber mask technique-II - Pattern storage
           and recognition. *Pattern Recognition*, 5:199–211, 1973.

[WLW01]    James Ze Wang, Jia Li, and Gio Wiederhold. SIMPLIcity:
           Semantics-sensitive Integrated Matching for Picture LIbraries.
           *IEEE Transactions on Pattern Analysis and Machine Intelli-
           gence*, 23(9):947–963, September 2001.
           `http://wang.ist.psu.edu/cgi-bin/zwang/regionsearch_show.cgi`.

[YJF98]    Byoung-Kee Yi, H. V. Jagadish, and Christos Faloutsos. Ef-
           ficient retrieval of similar time sequences under time warping.
           In *Proceedings of the 14th International Conference on Data
           Engineering (ICDE 1998)*, pages 201–208, Orlando, FL, 1998.
           IEEE Computer Society.

[ZL01]     Dengsheng Zhang and Guojun Lu. A comparison of shape
           retrieval using Fourier descriptors and short-time fourier de-
           scriptors. In *Proceedings of 2nd IEEE Pacific Rim Conference
           on Multimedia*, pages 855–860, Bejing, China, October 2001.
           Springer.

[ZL02]     Dengsheng Zhang and Guojun Lu. A comparative study of
           Fourier descriptors for shape representation and retrieval. In

*Proceedings of the 5th Asian Conference on Computer Vision (ACCV02)*, pages 646–651, Melbourne, Australia, January 2002.

[ZR72]    Charles T. Zahn and Ralph Z. Roskies. Fourier descriptors for plane closed curves. *IEEE Transactions on Computers*, 21(3):269–281, March 1972.
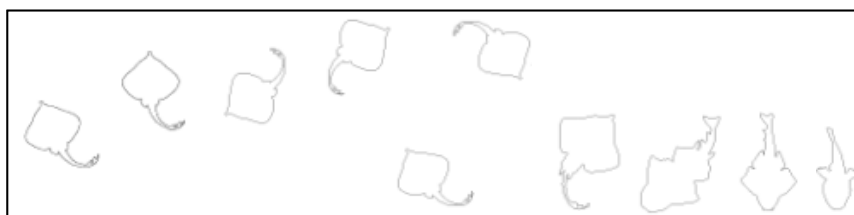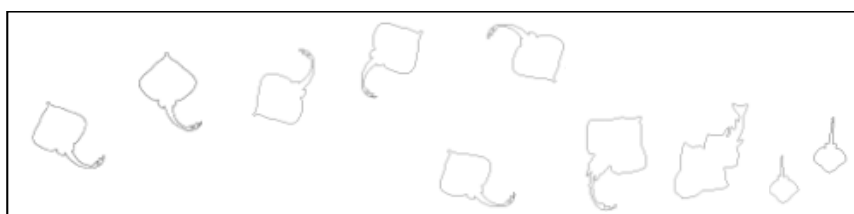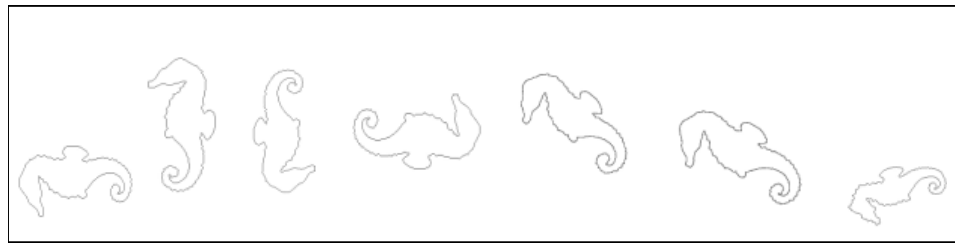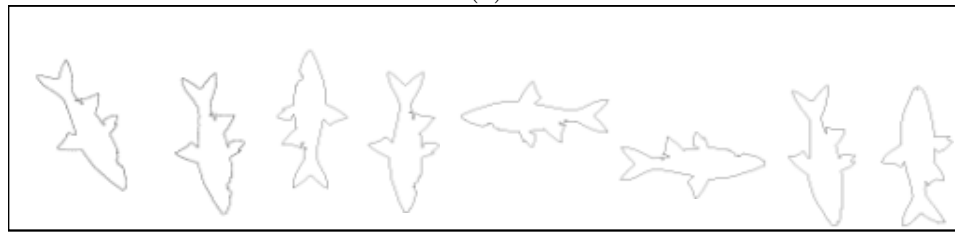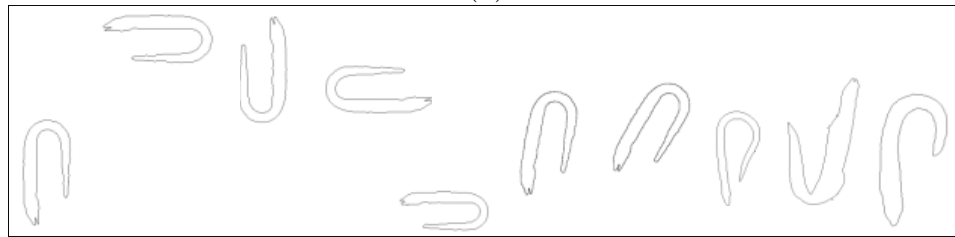
(a)
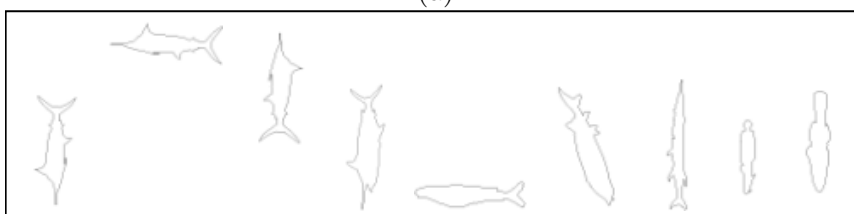

(b)


(c)

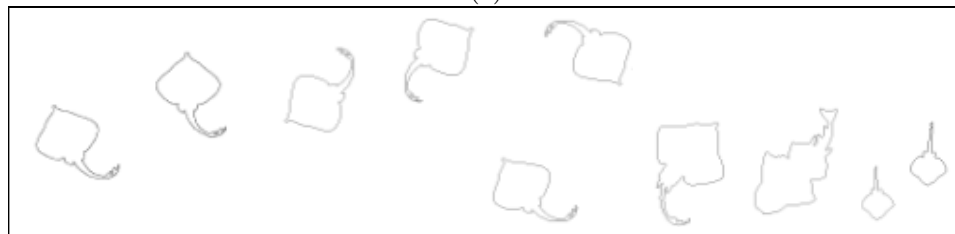Figure 15: Comparison results between DTW (first row of each example) and EU (second row) for some query images.
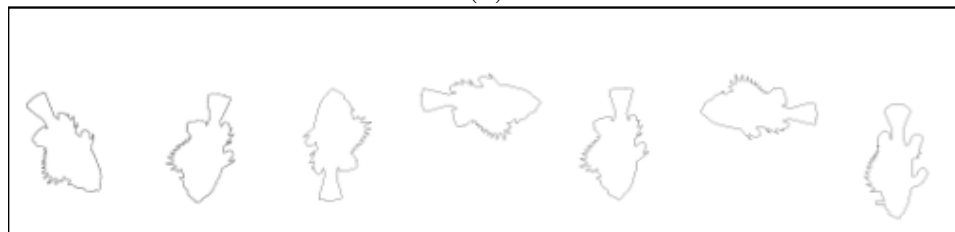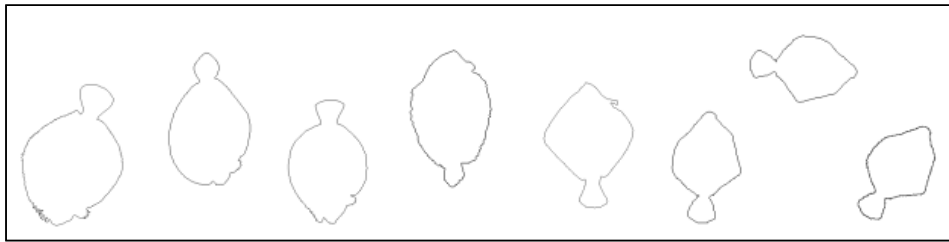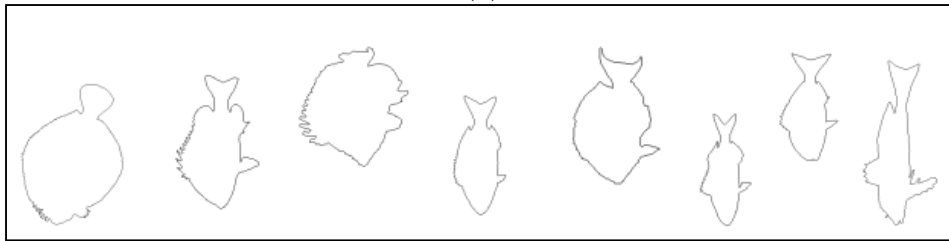
(a)



(b)



(c)



(d)



(e)

Figure 16: Results of retrieval evaluation for queries extracted from the categories "Seahorses" (a), "Sharks" (b), "U-Eels" (c), "Rays" (d), and "Tonguefishes" (e), respectively.
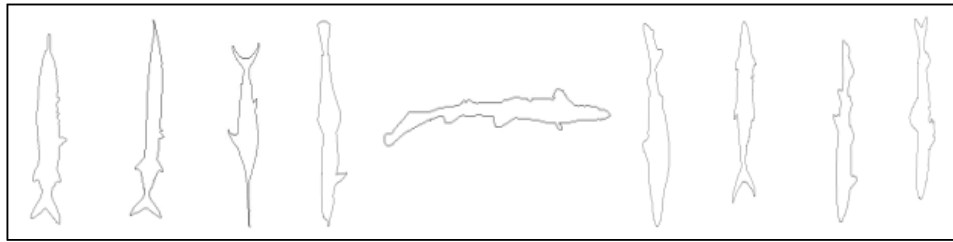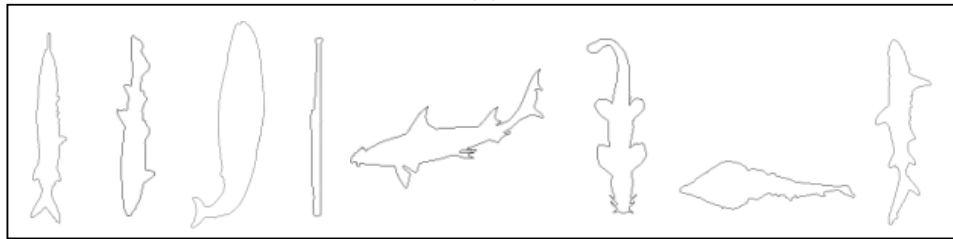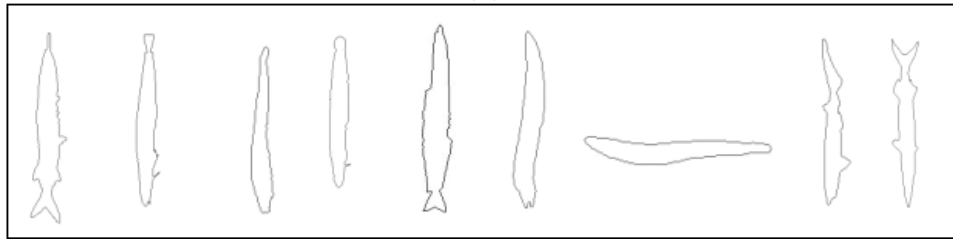
(a)



(b)
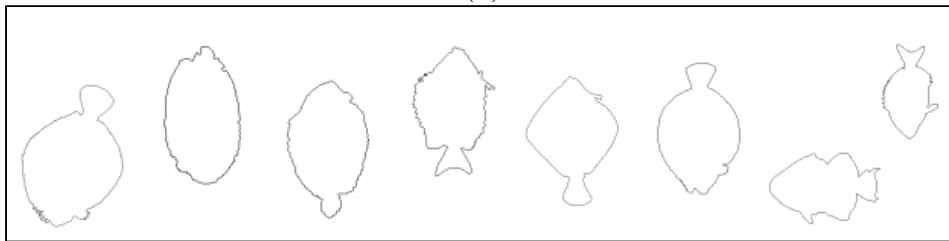


(c)

Figure 17: Results of retrieval evaluation for the "Sole" query using: IDFT (a), MAXC (b), and IDFTM (c). Experimental results are obtained with $M = 32$ and using the DTW distance function.

(a)



(b)



(c)

Figure 18: Results of retrieval evaluation for the "Pipefish" query using: IDFT (a), MAXC (b), and IDFTM (c). Experimental results are obtained with $M = 32$ and using the DTW distance function.