

# WARP: Accurate Retrieval of Shapes Using Phase of Fourier Descriptors and Time Warping Distance

Ilaria Bartolini, Paolo Ciaccia, *Member, IEEE*, and Marco Patella

**Abstract**—Effective and efficient retrieval of similar shapes from large image databases is still a challenging problem in spite of the high relevance that shape information can have in describing image contents. In this paper, we propose a novel Fourier-based approach, called WARP, for matching and retrieving similar shapes. The unique characteristics of WARP are the exploitation of the phase of Fourier coefficients and the use of the Dynamic Time Warping (DTW) distance to compare shape descriptors. While phase information provides a more accurate description of object boundaries than using only the amplitude of Fourier coefficients, the DTW distance permits us to accurately match images even in the presence of (limited) phase shiftings. In terms of classical precision/recall measures, we experimentally demonstrate that WARP can gain, say, up to 35 percent in precision at a 20 percent recall level with respect to Fourier-based techniques that use neither phase nor DTW distance.

**Index Terms**—Shape matching, Dynamic Time Warping distance, discrete Fourier transform.

## 1 INTRODUCTION

LARGE image databases are increasingly used in many application areas like crime prevention, architectural and engineering design, fashion, medical diagnosis, journalism, advertising, and land analysis. This has motivated growing research interest on efficient and effective methods enabling the retrieval of images on the basis of their content from large databases. With respect to other features, like color and texture, shape is much more effective in semantically characterizing the content of an image [1], [2], [3], [4]. However, properly extracting and representing shape information are still challenging tasks. In particular, even when accurate object boundaries are obtainable (this is the case when some domain knowledge is available or when images represent simple objects), the problem of representing them so as to allow the implementation of efficient and effective matching and retrieval methods is still not solved in a satisfactory way. The scenario is further complicated when invariance, with respect to a number of possible transformations, such as scaling, shifting, and rotation, is required.

Effective and efficient retrieval of images based on their shape content calls for a set of basic, often contrasting, requirements: *Compactness* and *simplicity* of shape descriptors are necessary for minimizing the storage overhead and the extraction time; for an effective retrieval, the shape descriptors should be *robust to noise* and *invariant to transformations* (namely, translation, scaling, and rotation); finally, in order to avoid a sequential scan of the whole (large) database, shape descriptors should be *indexable*, e.g., by using metric trees, like the M-tree [5], that are already profitably applied in several other image and multimedia applications [4], [6].

Among the different approaches that are available for the representation of shape information, those based on the Discrete Fourier Transform (DFT) describe the outside contour by means of a limited number of coefficients in the frequency domain. It is well-recognized nowadays that Fourier-based approaches are able to

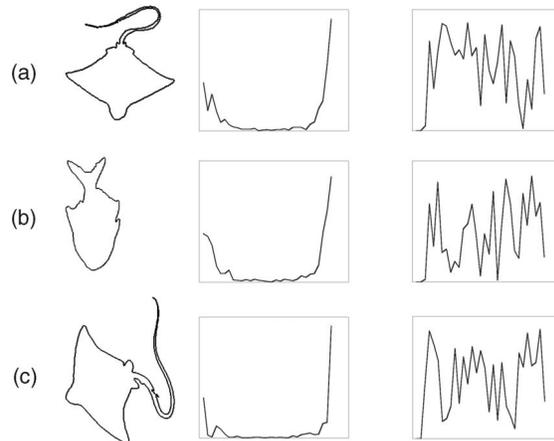


Fig. 1. Three different images, their relative magnitude spectra (i.e., the amplitude of the DFT coefficients), and their phase spectra.

provide all of the above-mentioned requirements, obtaining good effectiveness levels [7] along with efficiency in retrieval and indexability [8]. It has to be observed that, since DFT coefficients also carry information about the size, the orientation, and the position of the object, they have to be properly normalized in order to achieve invariance with respect to the desired transformations.

In this paper, we propose a novel Fourier-based approach for shape retrieval, called WARP, that extends previous methods with two innovative characteristics: 1) the preservation of *phase* information and 2) the use of the *Dynamic Time Warping* (DTW) distance to compare the shape descriptors. In terms of classical precision/recall measures, WARP can achieve up to 35 percent precision gain at a 20 percent recall level with respect to state-of-the-art Fourier-based techniques that use neither phase nor DTW distance and, at the same time, allows efficient indexing of shape information.

**Phase.** The rationale for maintaining phase information is based on the observation that current Fourier-based techniques discard the phase of DFT coefficients with the purpose of achieving rotation invariance as well as independence from the starting point of the parameterization [3], [7], [8], [9]. This is a consequence of the fact that rotating an object boundary or changing the starting point introduces a phase shifting in the DFT coefficients. The drawbacks of this simplistic approach are made evident in Fig. 1, where, for each image, the amplitude and phase diagrams of DFT coefficients are shown. Although Fig. 1a is perceptually more similar to Fig. 1c than to Fig. 1b, if we just consider the amplitude spectra and compare them using the Euclidean distance, Fig. 1a turns out to be closer to Fig. 1b than to Fig. 1c.

We resolve the apparent contradiction of preserving phase without giving up rotation and starting point invariances by deriving appropriate “compensation” terms that are added to the original phases, thus yielding a modified phase spectrum.

**Dynamic Time Warping.** As for the distance used to compare shape descriptors, Fourier-based techniques typically adopt the Euclidean distance. However, as our experiments also demonstrate, the Euclidean metric does not tolerate phase shiftings of subsequences of the shape, thus, similar shapes that are not perfectly aligned along the time axis lead to counter intuitive high distance values. To overcome the limits of the Euclidean metric, we consider using the Dynamic Time Warping distance [10] to compare the shape descriptors, thus allowing elastic deformations of objects’ boundaries to be matched. The DTW distance has already proven to yield superior performance for the retrieval of time series [11], [12] and we show that this is also the case for shape matching.

The rest of the paper is organized as follows: Section 2 presents the details of our WARP approach, formally demonstrating how the invariance properties are obtained. Experimental results are shown in Section 3, where we also compare WARP with other approaches. Section 4 concludes the paper.

• The authors are with DEIS, University of Bologna and IEIIT-BO/CNR, Viale Risorgimento 2, 40136-Bologna, Italy.  
E-mail: {ibartolini, pciaccia, mpatella}@deis.unibo.it.

Manuscript received 20 Dec. 2002; revised 10 Oct. 2003; accepted 7 July 2004.  
Recommended for acceptance by C. Schmid.

For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 118023.

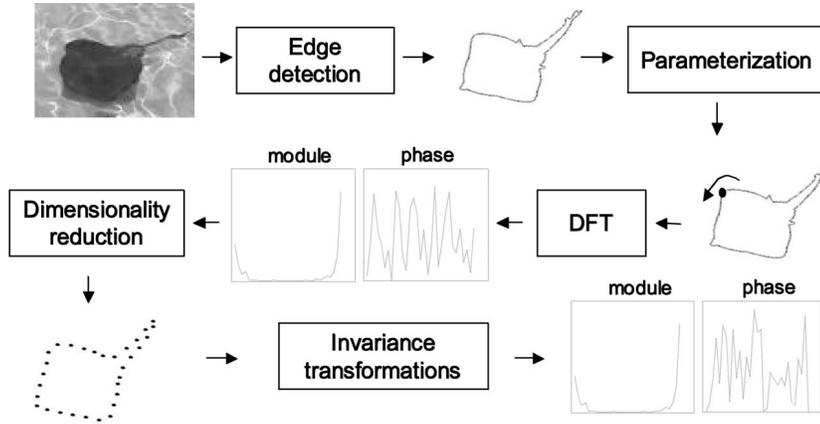


Fig. 2. WARP: Basic steps.

## 2 THE WARP METHOD

Fig. 2 presents a general overview of our method, called WARP. (WARP stands for “WARP: Accurate Retrieval based on Phase.”) Starting from a boundary description obtained through a shape extraction algorithm, we first parameterize the boundary to obtain a complex discrete-time periodic signal whose period  $N$  equals the number of points in the object boundary. Then, the DFT of the signal is computed. In order to reduce the size of shape descriptors, in the subsequent step (“dimensionality reduction”), only the first  $M$  low-frequency DFT coefficients are retained. Finally, these  $M$  coefficients are modified so as to achieve the invariances we are interested in and are then stored in the database.

At query time (see Fig. 3), the shape descriptor of the query object is obtained in the same way. To compare it with a database object, the Inverse DFT is applied to the two descriptors and the two reconstructed signals are then compared using the DTW distance.

### 2.1 Boundary Representation

As in [2], [3], [7], [8], [9], we consider the boundary of an object as a discrete-time complex periodical signal,  $\mathbf{z} = \langle z_0, \dots, z_{N-1} \rangle$ , where  $z_l = x_l + j y_l$  ( $j = \sqrt{-1}$ ) and  $x_l$  and  $y_l$  are the real coordinate values of the  $l$ th sampled point ( $l = 0, \dots, N - 1$ ). The  $\mathbf{z}$  signal is then mapped to the frequency domain by way of the Discrete Fourier Transform:

$$Z_m = \sum_{l=0}^{N-1} z_l e^{-j \frac{2\pi m l}{N}} = R_m e^{j \Theta_m} \quad (1)$$

$$m = -N/2, \dots, -1, 0, 1, \dots, N/2 - 1,$$

where  $R_m$  and  $\Theta_m$  are the module and the phase of the  $m$ th DFT coefficient, respectively.

### 2.2 Dimensionality Reduction

In order to guarantee compactness and robustness to noise, we use only low frequency DFT coefficients as in [8], [9]. In particular, we keep only the  $M$  ( $M \ll N$ ) coefficients whose frequency is closer to 0, i.e., those obtained for  $m \in [-M/2, M/2 - 1]$ .

The choice of a suitable value for  $M$  has to trade off the accuracy in representing the original signal with the compactness and the extraction efficiency.<sup>1</sup> To this end, we consider the spectral characteristics of the data set at hand, in particular, the *energy* of the signal retained by the  $M$  coefficients, defined as  $E(M) = \sum_{m=-M/2}^{M/2-1} |Z_m|^2$ . The reason for excluding the DC (zero frequency) coefficient is that it only provides information about object position and says nothing about the actual object shape (see also Section 2.3).

As an example, Fig. 4 plots the average value of  $E(M)/E(N)$  for the FISHES data set described in Section 3. From the graph, we see that, when choosing  $M \in [16, 64]$ , the retained energy varies from 83.7 percent to 93.4 percent of the total. On the other hand, to obtain, say,  $E(M)/E(N) = 0.99$ ,  $M$  should be equal to 512, which is far too high.

Experiments on the FISHES data set, described in [13], show that the effectiveness of WARP using  $M = 16$  is much lower than that obtained for  $M = 32$ , whereas increasing the value of  $M$  to 64 does not lead to further improvements, only inflating extraction and distance times. Thus, in the experiments of Section 3, we will use  $M = 32$ .

### 2.3 The Shape Descriptor

As anticipated in Section 1, a change in the position of the object, in its size, orientation, or the initial point used to parameterize the boundary, introduces a modification in the DFT coefficients. In order to guarantee the translation, scaling, rotation, and starting point invariances, we have to modify the DFT coefficients accordingly. In Table 1, we show how *Normalized DFT* (N-DFT) coefficients  $\tilde{Z}_m = \hat{R}_m e^{j \hat{\Theta}_m}$  satisfying the invariance requirements can be obtained.<sup>2</sup>

If one wants to achieve both rotation and starting point invariances, the two corresponding modifications have to be integrated, leading to  $\hat{\Theta}_m = \Theta_m - \frac{\Theta_{-1} + \Theta_1}{2} + m \frac{\Theta_{-1} - \Theta_1}{2}$ . The translation and scaling invariances are obtained as in [2], [3], [7], [8], [9], whereas, for rotation and starting point invariances, the derivation is

1. The complexity of computing  $M$  DFT coefficients is  $O(M \cdot N)$ . Since we do not require  $N$  to be a power of 2, we cannot make use of the Fast Fourier Transform (FFT), which would reduce the complexity to  $O(N \log N)$ .

2. Note that the WARP method for phase normalization is similar, although simpler, to that proposed in [14], which is unnecessarily complex for our goals.

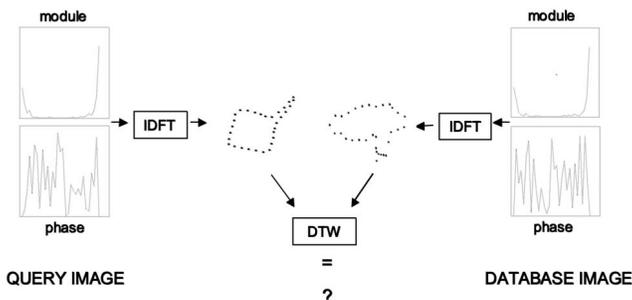


Fig. 3. WARP: Assessing the similarity between two shape descriptors.

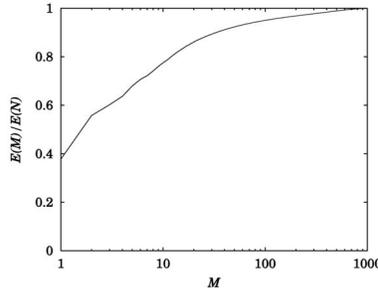


Fig. 4. Fraction of energy associated to the first  $M$  DFT coefficients.

as follows (see also [13]): Consider the origin-centered signals  $\mathbf{z}$  and  $\mathbf{z}'$ , where  $\mathbf{z}'$  is obtained from  $\mathbf{z}$  by rotating each point counter-clockwise by a constant factor  $\bar{\theta}$  and by shifting the starting point by  $l_0$  positions, i.e.,  $z'_i = z_{i-l_0} e^{j\bar{\theta}}$ . The corresponding DFT coefficients are:

$$Z'_m = Z_m e^{j\bar{\theta}} e^{-j\frac{2\pi l_0 m}{N}} = R_m e^{j(\Theta_m + \bar{\theta} - \frac{2\pi l_0 m}{N})}, \quad (2)$$

thus, it is  $\Theta'_m = \Theta_m + \bar{\theta} - \frac{2\pi l_0 m}{N}$ , in particular  $\Theta'_1 = \Theta_1 + \bar{\theta} - \frac{2\pi l_0}{N}$  and  $\Theta'_{-1} = \Theta_{-1} + \bar{\theta} + \frac{2\pi l_0}{N}$ . By referring to Table 1, it is obtained:

$$\begin{aligned} \hat{\Theta}'_m &= \Theta'_m - \frac{\Theta'_{-1} + \Theta'_1}{2} + m \frac{\Theta'_{-1} - \Theta'_1}{2} \\ &= \Theta_m + \bar{\theta} - \frac{2\pi l_0 m}{N} - \frac{\Theta_{-1} + \Theta_1}{2} - \bar{\theta} + m \frac{\Theta_{-1} - \Theta_1}{2} + \frac{2\pi l_0 m}{N} \\ &= \Theta_m - \frac{\Theta_{-1} + \Theta_1}{2} + m \frac{\Theta_{-1} - \Theta_1}{2} = \hat{\Theta}_m. \end{aligned} \quad (3)$$

By simply performing the Inverse DFT on N-DFT coefficients, we then obtain a modified (normalized) signal  $\hat{\mathbf{z}}$ , which satisfies all the requested invariances:

$$\hat{z}_l = \frac{1}{M} \sum_{m=-M/2}^{M/2-1} \hat{Z}_m e^{j\frac{2\pi l m}{M}} \quad l = 0, \dots, M-1. \quad (4)$$

## 2.4 Comparing Shape Descriptors

The standard approach [3], [7], [8], [9] to compare (normalized, in our case) signals  $\hat{\mathbf{z}}$  and  $\hat{\mathbf{z}'}$  makes use of the Euclidean distance,  $L_2(\hat{\mathbf{z}}, \hat{\mathbf{z}'}) = \sqrt{\sum_{i=0}^{M-1} |\hat{z}_i - \hat{z}'_i|^2}$ . This, by exploiting the Parseval Theorem, can be replaced by  $L_2(\hat{\mathbf{Z}}, \hat{\mathbf{Z}'})$ , thus without the need to perform the Inverse DFT on the shape descriptors.

The Euclidean distance has the major drawback of not considering elastic shiftings of the time axis, which is a necessity with signals which are similar although partially unaligned at some sample points, as exemplified in Fig. 5a. To obviate this problem, in WARP we consider the use of the Dynamic Time Warping (DTW) distance [10]:

**Definition 2.1 (DTW Distance).** Given two sequences,  $\hat{\mathbf{z}} = \langle \hat{z}_0, \hat{z}_1, \dots, \hat{z}_{M-1} \rangle$  and  $\hat{\mathbf{z}'} = \langle \hat{z}'_0, \hat{z}'_1, \dots, \hat{z}'_{M-1} \rangle$  of length  $M$ , let us denote with  $\text{tail}(\hat{\mathbf{z}}) = \langle \hat{z}_1, \dots, \hat{z}_{M-1} \rangle$  the sequence obtained from  $\hat{\mathbf{z}}$  by removing its first element. The Dynamic Time Warping (DTW) distance  $d_{tw}(\hat{\mathbf{z}}, \hat{\mathbf{z}'})$  is recursively defined as follows:

$$\begin{aligned} d_{tw}(\langle \rangle, \langle \rangle) &= 0 \\ d_{tw}(\hat{\mathbf{z}}, \langle \rangle) &= d_{tw}(\langle \rangle, \hat{\mathbf{z}'}) = \infty \\ d_{tw}(\hat{\mathbf{z}}, \hat{\mathbf{z}'}) &= \sqrt{\delta(\hat{z}_0, \hat{z}'_0) + \min \begin{cases} d_{tw}(\hat{\mathbf{z}}, \text{tail}(\hat{\mathbf{z}'})) \\ d_{tw}(\text{tail}(\hat{\mathbf{z}}), \hat{\mathbf{z}'}) \\ d_{tw}(\text{tail}(\hat{\mathbf{z}}), \text{tail}(\hat{\mathbf{z}'})) \end{cases}}, \end{aligned} \quad (5)$$

where  $\delta$  is an arbitrary distance function between samples (in the following, we will use the (squared) Euclidean distance between the two samples,  $\delta(\hat{z}_i, \hat{z}'_j) = |\hat{z}_i - \hat{z}'_j|^2$ ).

TABLE 1  
The N-DFT WARP Coefficients

Invariance	Modified coefficient
Translation	$\hat{Z}_0 = 0$
Scale	$\hat{R}_m = R_m / R_1$
Rotation	$\hat{\Theta}_m = \Theta_m - \frac{\Theta_{-1} + \Theta_1}{2}$
Starting point	$\hat{\Theta}_m = \Theta_m + m \frac{\Theta_{-1} - \Theta_1}{2}$

In the vast majority of applications of the DTW 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 [15] (see Fig. 5d), which limits the deviation from the diagonal up to a value  $\omega$ , which is called the window length.<sup>3</sup> The motivation for this constraint is two-fold: First, it limits the complexity of computing  $d_{tw}$  from  $O(M^2)$  to  $O(M \times \omega)$ ; second, and more importantly, it prevents the creation of pathological paths, aligning samples which are far away from each other.

Even if the DTW distance has a number of desirable properties, it suffers from the problem of not being a metric since it does not satisfy the triangle inequality [11]. However, as demonstrated in [16], indexing without losing relevant objects is still possible as long as we can find a metric  $d_I$  which lower-bounds  $d_{tw}$ , i.e.,  $d_I(\hat{\mathbf{z}}, \hat{\mathbf{z}'}) \leq d_{tw}(\hat{\mathbf{z}}, \hat{\mathbf{z}'})$ ,  $\forall \hat{\mathbf{z}}, \hat{\mathbf{z}'}$ . In [12], 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 a lower-bounding distance can be used in conjunction with a metric access method, like the QIC-M-tree [16], for indexing shape descriptors.

## 3 EXPERIMENTAL EVALUATION

We implemented the WARP method in C++ under Windows NT 4.0. For performance evaluation, we used the FISHES data set provided by [17], consisting of 1,100 text files containing the coordinates of boundary points of an object with each object representing a marine animal. The length of each boundary varied from 256 to 1,653 points.

We manually classified each image based on 10 semantic categories ("Seahorses" (5 images), "Seamoths" (6), "Sharks" (58), "Soles" (52), "Tonguefishes" (19), "Crustaceans" (4), "Eels" (26), "U-Eels" (25), "Pipefishes" (16), and "Rays" (41)). Images not belonging to any category were assigned to a default class (848). Table 2 shows some images in the data set, along with their category.

The query workload used in our experiments consists of 30 query images chosen from the 10 semantic categories. For evaluation purposes, any image in the same category of the query is considered *relevant* to that query, whereas all other database images are considered *irrelevant*. To measure the retrieval effectiveness, we considered classical *precision* (P) and *recall* (R) metrics averaged over the set of processed queries.

### 3.1 Parameterization and Dimensionality Reduction

In our first experiment, we compare the parameterization and dimensionality reduction approach of WARP to the one based on the extraction from the boundary of the  $M$  points having highest curvature values, hereafter called MAXC [3]. We do not consider methods that reduce dimensionality by simply resampling the original boundary every  $N/M$  points [2], [3], [7] since this easily leads to missing significant shape details. Results in Fig. 6a clearly show that WARP consistently outperforms MAXC, that improvement in precision varies between 65 percent and 220 percent up to

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

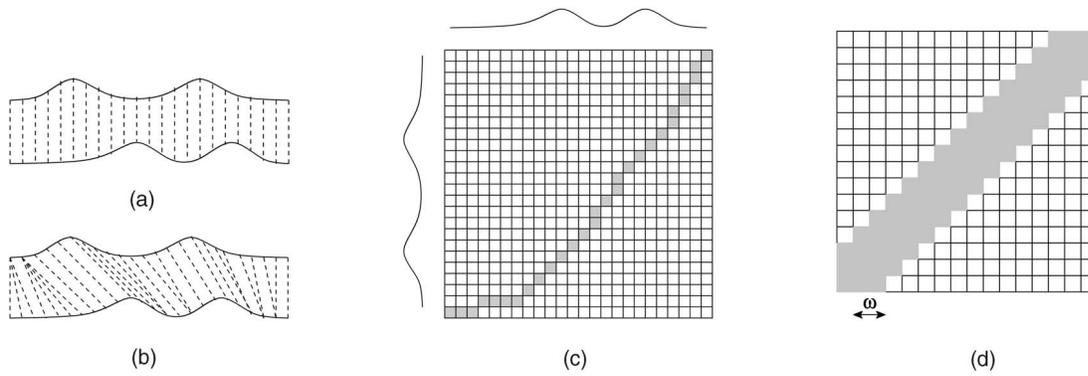


Fig. 5. An example of the (a) Euclidean and (b) Dynamic Time Warping distance on two real signals. (c) The alignment matrix for computing the warping path. (d) An illustration of the Sakoe-Chiba band for constraining the warping path. In this example, the window length,  $\omega$ , is 2.

TABLE 2  
Sample Images from Some of the 10 Semantic Categories

image	category	card.	image	category	card.	image	category	card.
	Rays	41		Seahorses	5		Sharks	58
	Pipefishes	16		U-Eels	25		Soles	52

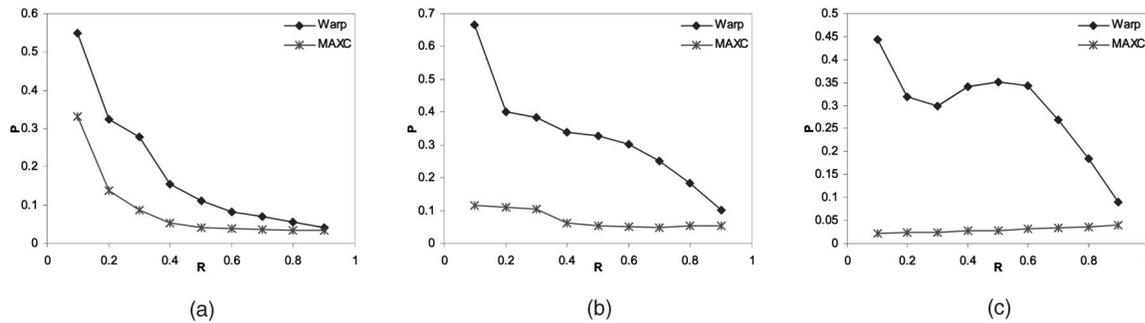


Fig. 6. P/R graphs comparing WARP and MAXC. (a) Whole data set. (b) Query image in the category "Sole." (c) Query image in the category "Ray."

$R = 0.7$ , and that it is always positive even for higher recall levels. A similar trend can be also observed on specific query images, as shown in Fig. 6b and Fig. 6c (see also Fig. 7 for visualization of retrieval results).

### 3.2 Effect of Phase Information and Distance Function

We now turn to considering the effects of maintaining phase information and using the DTW distance. To this end, in Fig. 8a we compare WARP with a method, termed NoPhase-EU, that discards the phase of DFT coefficients and uses the Euclidean distance to compare shape descriptors. Note that this is, with slight variations, the method used by almost all of the existing Fourier-based shape retrieval systems [3], [7], [8], [9]. As Fig. 8a clearly shows, WARP consistently performs better than NoPhase-EU, with a gain in precision that reaches 39 percent at  $R = 0.7$ .

In order to better understand the separate contribution of phase information and DTW distance, in Fig. 8a, we also show results for a variant of WARP, called WARP-EU, which retains phase

information but applies the Euclidean distance. From the graphs, it can be observed that only using phase leads to a swinging behavior, depending on the level of recall. This demonstrates that phase alone cannot guarantee adequate performance levels since phase shiftings cannot be captured by the Euclidean distance, thus it is only the *combined* use of phase and DTW distance that justifies the better performance of WARP. We end this section by showing, in Fig. 8b and Fig. 8c, P/R graphs for two specific queries (see Fig. 7 for visualization of retrieval results).

### 3.3 Invariance to Rotation

In order to better test robustness to noise and rotation, in the following experiment, we consider the FISHES data set extended with another 100 images, obtained by rotating 30 randomly chosen images by different angles (e.g., 10, 30, 90, 180, and 270 degrees). Note that, while a rotation of a multiple of  $\pi/2$  produces a signal without noise with respect to the original image, other rotation angles introduce noise due to spatial discretization.

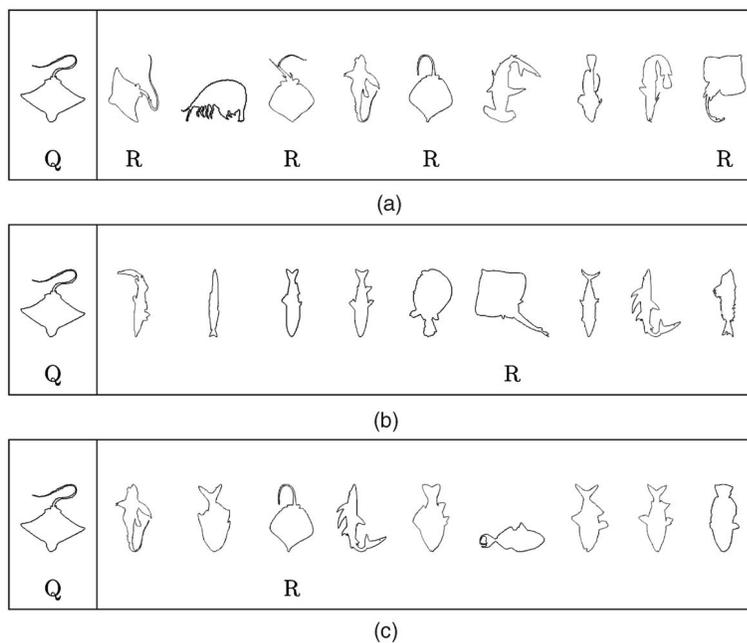


Fig. 7. Results for the "Ray" query. (a) WARP. (b) MAXC. (c) NoPhase-EU. Relevant images have label "R."

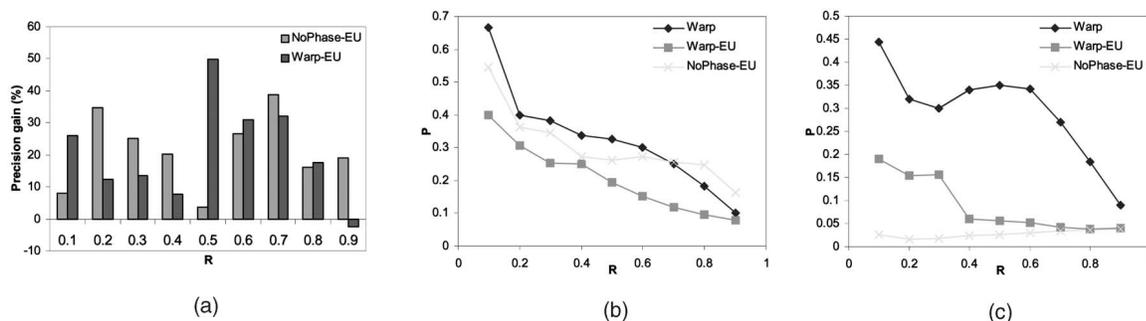


Fig. 8. (a) Precision gain of WARP over NoPhase-EU and Warp-EU. (b) P/R graph for the query image "Sole." (c) P/R graph for the query image "Ray."

Fig. 9 shows two examples where the first image of each row is the query image and the others are the most similar images retrieved by WARP. As can be observed, WARP is able to retrieve images rotated by a multiple of  $\pi/2$  as well as those freely rotated.

### 3.4 Comparison with the MPEG-7 Standard

We compare the results of WARP on the MPEG-7 data set for the Core Experiment CE-Shape-1, part B, illustrated in [18], with those obtained by the curvature scale-space (CSS) approach [19], which is the standard for MPEG-7 shape representation [20].

The CSS method adopts a curvature parameterization of a boundary that is smoothed by applying several Gaussian kernels, each with a different width  $\sigma$ . The CSS image of a boundary is then defined as the locus of points  $(l, \sigma)$  for which the (smoothed) curvature has zero value and the CSS descriptor consists of the peaks (maxima) of the CSS image. The CSS descriptor is invariant to translation, scale, and rotation transformations and can be made invariant with respect to the starting point by shifting the whole descriptor so as to have the highest peak in the origin. When comparing two CSS descriptors, the peaks are matched and the overall distance is obtained by summing all of the peak differences.

For comparison purposes, we use the data set from [21], consisting of 1,400 images divided in 70 shape classes of 20 images each. Each image is used as a query and the relevant images are

those found in the first 40 results; thus, precision is always one half of the recall.

According to [18], on this data set, CSS (method P320 in [18]) attains an average precision of 37.72 percent, whereas the precision of WARP is 29.25 percent (note that precision cannot be more than 50 percent in this experiment). Although CSS outperforms WARP in terms of effectiveness, this does not come without a price. Indeed, CSS shape descriptors are indexed, as described in [22], by using their *aspect ratio*, i.e., the ratio between the higher CSS peak and the contour length. If the aspect ratio of an image differs more than a user-specified threshold value from that of the query, the image is filtered out, thus without computing the exact match between the two CSS descriptors. This is an *approximate* query processing algorithm since it can easily lead to *false dismissals* (i.e., best-matching images may be filtered out). From this we can conclude that CSS is not suitable if one wants to find the *exact* best matches of a query image on a large database. Further, since the method requires the user to supply a value for the threshold, it is difficult to calibrate.

Finally, it should be noted that other techniques [23], [24], [25] are able to attain results similar to, or even slightly better than, those obtained by WARP for the MPEG-7 data set. However, such techniques cannot be efficiently indexed, thus they are only suitable for shape recognition in small-size databases.

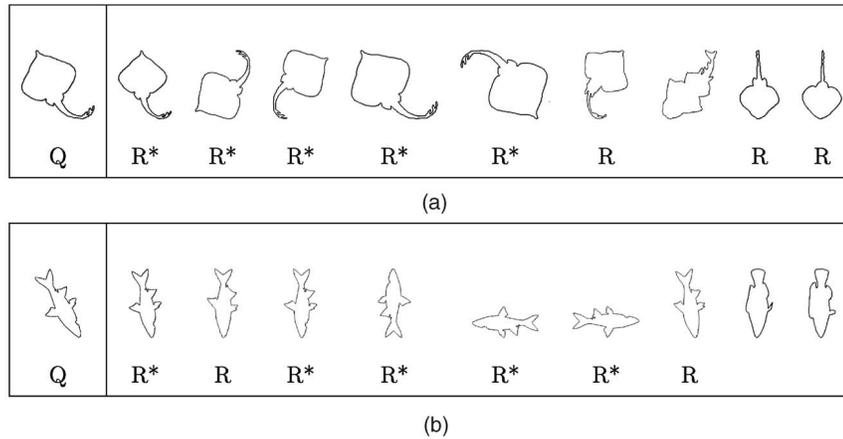


Fig. 9. Results of WARP for queries extracted from the categories (a) "Rays" and (b) "Sharks," respectively. Relevant images have label "R." Rotated versions of the query image are labeled "\*".

#### 4 CONCLUSIONS

In this paper, we have presented a novel Fourier-based method for shape matching and retrieval, called WARP, which is characterized by the combined preservation of phase information and use of the Dynamic Time Warping (DTW) distance in place of the more common Euclidean distance. We have detailed how invariance properties, such as scaling, shifting, rotation, and starting point, can be obtained by properly normalizing DFT coefficients and have experimentally demonstrated that WARP consistently outperforms state-of-the-art Fourier-based methods, as substantiated by an average improvement of 35 percent in retrieval accuracy at a 20 percent recall level.

Even if the DTW distance does not satisfy the metric postulates [11], thus preventing the use of indices to speed-up the retrieval phase, by exploiting recent results on metric access structures [16] and on the DTW distance itself [12], we can still guarantee the indexability of DFT coefficients extracted from large data sets.

We are now integrating WARP into our image retrieval system, Windsurf [26], and we plan to conduct thorough experimentation on retrieval efficiency when an index like the QIC-M-tree [16] is used.

#### ACKNOWLEDGMENTS

This work was supported by the European Commission under the IST-2001-33058 Thematic Network PANDA "PAtterns for Next-generation DATAbase Systems" (2001-04). The authors would like to thank the anonymous reviewers for the many useful observations that greatly contributed to improve the overall quality of the paper.

#### REFERENCES

- [1] T. Pavlidis, "A Review of Algorithms for Shape Analysis," *Computer Vision, Graphics, and Image Processing*, vol. 7, pp. 243-258, 1978.
- [2] Y. Rui, A.C. She, and T.S. Huang, "Modified Fourier Descriptors for Shape Representation—A Practical Approach," *Proc. First Int'l Workshop Image Databases and Multi Media Search*, pp. 22-23, Aug. 1996.
- [3] W.-Y. Ma and B.S. Manjunath, "NeTra: A Toolbox for Navigating Large Image Databases," *Multimedia Systems*, vol. 7, no. 3, pp. 184-198, <http://vision.ece.ucsb.edu/netra/>, May, 1999.
- [4] S. Berretti, A. Del Bimbo, and P. Pala, "Retrieval by Shape Similarity with Perceptual Distance and Effective Indexing," *IEEE Trans. Multimedia*, vol. 2, no. 4, pp. 225-239, Dec. 2000.
- [5] P. Ciaccia, M. Patella, and P. Zezula, "M-Tree: An Efficient Access Method for Similarity Search in Metric Spaces," *Proc. 23rd Int'l Conf. Very Large Data Bases (VLDB '97)*, pp. 426-435, Aug. 1997.
- [6] A.H.H. Ngu, Q.Z. Sheng, D.Q. Huynh, and R. Lei, "Combining Multi-Visual Features for Efficient Indexing in a Large Image Database," *The Very Large Data Bases J.*, vol. 9, no. 1, pp. 279-293, Mar. 2001.
- [7] H. Kauppinen, T. Seppänen, and M. Pietikäinen, "An Experimental Comparison of Autoregressive and Fourier-Based Descriptors in 2D Shape Classification," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 17, no. 2, pp. 201-207, Feb. 1995.
- [8] D. Rafiei and A.O. Mendelzon, "Efficient Retrieval of Similar Shapes," *The Very Large Data Bases J.*, vol. 11, no. 1, pp. 17-27, Aug. 2002.
- [9] D. Zhang and G. Lu, "A Comparative Study of Fourier Descriptors for Shape Representation and Retrieval," *Proc. Fifth Asian Conf. Computer Vision (ACCV '02)*, pp. 646-651, Jan. 2002.
- [10] D.J. Berndt and J. Clifford, "Using Dynamic Time Warping to Find Patterns in Time Series," *Advances in Knowledge Discovery in Databases: Papers from the 1994 AAAI Workshop*, pp. 359-370, July 1994.
- [11] B.-K. Yi, H.V. Jagadish, and C. Faloutsos, "Efficient Retrieval of Similar Time Sequences under Time Warping," *Proc. 14th Int'l Conf. Data Eng. (ICDE '98)*, pp. 201-208, Feb. 1998.
- [12] E.J. Keogh, "Exact Indexing of Dynamic Time Warping," *Proc. 28th Int'l Conf. Very Large Data Bases (VLDB '02)*, pp. 406-417, Sept. 2002.
- [13] I. Bartolini, P. Ciaccia, and M. Patella, "Using the Time Warping Distance for Fourier-Based Shape Retrieval," Technical Report IEHT-BO-03-02, IEHT-BO/CNR, 2002.
- [14] T.P. Wallace and P.A. Wintz, "An Efficient Three-Dimensional Aircraft Recognition Algorithm Using Normalized Fourier Descriptors," *Computer Graphics and Image Processing*, vol. 13, pp. 99-126, 1980.
- [15] H. Sakoe and S. Chiba, "A Dynamic Programming Algorithm Optimization for Spoken Word Recognition," *IEEE Trans. Acoustics, Speech, and Signal Processing*, vol. 26, no. 1, pp. 43-49, Feb. 1978.
- [16] P. Ciaccia and M. Patella, "Searching in Metric Spaces with User-Defined and Approximate Distances," *ACM Trans. Database Systems*, vol. 27, no. 4, pp. 398-437, Dec. 2002.
- [17] S. Abbasi, F. Mokhtarian, and J. Kittler, "SQUID Demo Dataset 1,500," <http://www.ee.surrey.ac.uk/Research/VSSP/imagedb/demo.html>, 1997.
- [18] L.J. Latecki, R. Lakämper, and U. Eckhardt, "Shape Descriptors for Non-Rigid Shapes with a Single Closed Contour," *Proc. IEEE Int'l Conf. Computer Vision and Pattern Recognition (CVPR '00)*, pp. 1424-1429, June 2000.
- [19] F. Mokhtarian and A.K. Mackworth, "A Theory of Multiscale, Curvature-Based Shape Representation for Planar Curves," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 14, no. 8, pp. 789-805, Aug. 1992.
- [20] M. Bober, "MPEG-7 Visual Shape Descriptors," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 11, no. 6, pp. 716-719, June 2001.
- [21] L.J. Latecki, "Shape Data for the MPEG-7 Core Experiment CE-Shape-1," <http://www.cis.temple.edu/~latecki/TestData/mpeg7shapeB.tar.gz>, 2002.
- [22] F. Mokhtarian, S. Abbasi, and J. Kittler, "Robust and Efficient Shape Indexing through Curvature Scale Space," *Proc. 1996 British Machine Vision Conf.*, pp. 53-62, Sept. 1996.
- [23] L.J. Latecki and R. Lakämper, "Shape Similarity Measure Based on Correspondence of Visual Parts," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1185-1190, Oct. 2000.
- [24] S. Belongie, J. Malik, and J. Puzicha, "Shape Matching and Object Recognition Using Shape Contexts," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 509-522, Apr. 2002.
- [25] C. Grigorescu and N. Petkov, "Distance Sets for Shape Filters and Shape Recognition," *IEEE Trans. Image Processing*, vol. 12, no. 10, pp. 1274-1286, Oct. 2003.
- [26] S. Ardizzoni, I. Bartolini, and M. Patella, "Windsurf: Region-Based Image Retrieval Using Wavelets," *Proc. First Int'l Workshop Similarity Search (IWSS '99)*, pp. 167-173, Sept. 1999.