

# Shape Retrieval Using Normalized Fourier Descriptors Based Signatures and Cyclic Dynamic Time Warping

Andrés Marzal, Vicente Palazón, and Guillermo Peris\*

Dept. Llenguatges i Sistemes Informàtics,  
Universitat Jaume I de Castelló, Spain  
{amarzal, palazon, peris}@lsi.uji.es

**Abstract.** The WARP system defines a dissimilarity measure between shapes described by their contours which is based on Dynamic Time Warping of Fourier Descriptors based signatures. These signatures are invariant to translation, scaling, rotation, and selection of the starting point. However, identical shapes present ambiguous signatures and similar shapes may yield significantly different signatures. Differences affect rotation and starting-point of the signatures, which may lead to poor performance in classification and shape retrieval tasks. We propose a different signature method to provide true rotation invariance and a Cyclic Dynamic Time Warping dissimilarity measure to achieve true starting-point invariance in shape comparisons.

## 1 Introduction

Content-based image retrieval is being increasingly demanded in many applications: digital libraries, broadcast media selection, multimedia editing, etc. [7]. In order to be effective in classification and retrieval tasks, shape descriptions, combined with (dis)similarity measures, must be robust to noise and invariant to transformations such as translation, scaling, and rotation.

Recently, Bartolini *et al.* have proposed a new Discrete Fourier Transform based approach to represent and compare shapes: the WARP System [1]. The normalized, low-frequency Fourier Descriptors (FDs) (including phase information) are used to reconstruct the original shape. We will refer to the reconstructed shape with the term *signature*. The signature is a good approximation to the original shape and contains a small number of points. Moreover, it is a sequence of complex values with a canonical starting point, which makes it amenable to be compared to other signatures by means of standard sequence comparison methods. The WARP system uses Dynamic Time Warping (DTW) in order to compare sequences [6]. In [1], some experiments on the SQUID Demo and MPEG-7 CE-Shape-1 databases show that the WARP system outperforms other indexable curvature-based shape descriptors and FDs-based signatures that do not take into account phase information.

---

\* This work has been supported by the Spanish *Ministerio de Ciencia y Tecnología* and FEDER under grant TIC2002-02684.

The WARP system presents two drawbacks: (1) reconstructing the shape contour from normalized FDs produces signatures with an ambiguity modulo a rotation of  $\pi$  radians [2] (which also affects the starting-point selection); and (2) perceptually similar shapes may have significantly different signatures (in orientation and starting point selection), which leads to poor performance of DTW-based comparisons. In order to solve these problems, we propose (1) a different encoding of the shape contour (which is based on the derivative of the reconstructed contour), and (2) to compare derivative-based signatures by means of a Cyclic Dynamic Time Warping dissimilarity measure.

The paper is organized as follows: In Sect. 2, some notation is introduced. In Sect. 3, the WARP system is reviewed and the observed drawbacks are pointed out. A simple improvement to the WARP system which provides better rotation invariance and a Cyclic Dynamic Time Warping procedure that provides starting point invariance when comparing signatures are presented in Sect. 4. In Sect. 5, experimental results on image retrievals tasks for the SQUID Demo and MPEG-7 CE-Shape-1 databases compare the different methods. Finally, some conclusions are presented in Sect. 6.

## 2 Notation

Shapes can be coded as a cyclic sequence of points along the contour. A cyclic sequence can be viewed as the set of sequences obtained by cyclically shifting a representative sequence (i.e., by choosing different starting points).

Let  $\mathbb{C}^*$  be the closure of  $\mathbb{C}$ , the field of complex numbers, under a concatenation operator and let  $a = a_0a_1 \dots a_{m-1} \in \mathbb{C}^*$  be a sequence of  $m$  points (complex values) describing a (counter-clockwise) contour<sup>1</sup>. A *cyclic shift*  $\sigma$  of  $a$  is a mapping  $\sigma : \mathbb{C}^* \rightarrow \mathbb{C}^*$  defined as  $\sigma(a_0a_1 \dots a_{m-1}) = a_1a_2 \dots a_{m-1}a_0$ . Let  $\sigma^k$  denote the composition of  $k$  cyclic shifts and let  $\sigma^0$  denote the identity. Two sequences  $a$  and  $\hat{a}$  are cyclically equivalent if  $a = \sigma^k(\hat{a})$  for some integer  $k$ . A cyclic sequence is an equivalence class  $[a] = \{\sigma^k(a) : 0 \leq k < m\}$ . Any of its members is a representative (non-cyclic) sequence.

## 3 The WARP System

Dynamic Time Warping (DTW) of sequences of 2D points describing shapes is sensitive to changes in position, scale, orientation of contours and to selection of their starting points. Therefore, DTW does not lead to good dissimilarity measures when the original, cyclic sequences describing shapes are used. The WARP images retrieval system [1] is based on the DTW-based comparison of compact, normalized signatures of shapes. These signatures are obtained by applying the Inverse Discrete Fourier Transform (IDFT) to the shape's Fourier Descriptors (FDs) after a normalization procedure.

---

<sup>1</sup> Note that  $a_0a_1 \dots a_{m-1}$  does not denote the product of  $m$  complex numbers, but their concatenation to form a sequence.

The Discrete Fourier Transform (DFT) of a sequence  $a = a_0 a_1 \dots a_{m-1}$  is an ordered set of complex values  $A = (A_{-m/2}, \dots, A_{-1}, A_0, A_1, \dots, A_{m/2-1})$  where  $A_i = \sum_{0 \leq k < m} a_k e^{-j2\pi ki/m}$  and  $j = \sqrt{-1}$ . These coefficients are the FDs and model the contour of a shape as a composition of ellipses revolving at different frequencies [2]. The main ellipse is centered at the contour centroid,  $A_0$ , and translation of the contour only affects this descriptor. Scaling by a factor  $\alpha$  scales the FDs by  $\alpha$ . Rotating the shape by an angle  $\theta$  yields a phase shift of  $\theta$  in the FDs. The cyclic shift  $\sigma^k(a)$  produces a linear phase shift of  $2\pi ki/m$  to  $A_i$ .

The  $A_0$  descriptor can be set to 0 in order to provide invariance to translation. Let us consider the polar representation of the descriptors:  $A_i = r_i e^{j\theta_i}$ . The value of  $A_1$  is the length of the main axis of the basic (low frequency) ellipse; therefore, dividing all the descriptors by  $r_1$  provides invariance to scale. Invariance to rotation can be obtained by subtracting  $(\theta_{-1} + \theta_1)/2$  (the orientation of the basic ellipse) to each  $\theta_i$ . Invariance with respect to the starting point can be achieved by adding  $i(\theta_{-1} - \theta_1)/2$  to each  $\theta_i$ . In principle, the shape can be reconstructed to a canonical form (invariant to translation, scaling, rotation, and starting point) by computing the IDFT. Noise in the contour can be reduced by taking only  $M \ll m$  low frequency components before computing the IDFT. The WARP system only uses the  $M = 32$  lower frequency FDs before computing the IDFT. The resulting shape is a more compact, canonical representation of the original one: a *signature*.

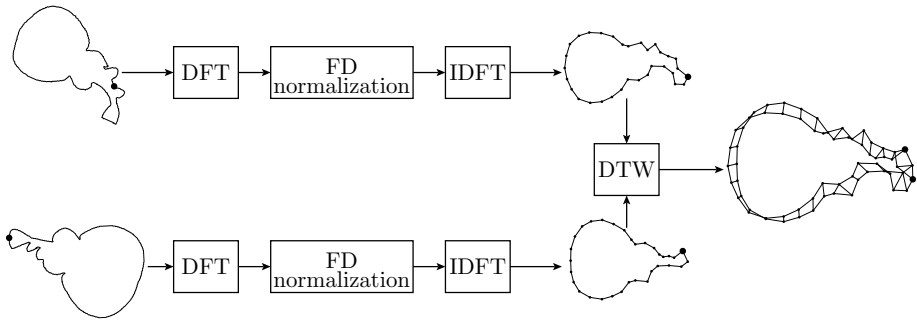
Let  $a = a_0 a_1 \dots a_{m-1}$  and  $b = b_0 b_1 \dots b_{n-1}$  be two sequences. An *alignment* between  $a$  and  $b$  is a sequence of pairs  $(i_0, j_0), (i_1, j_1), \dots, (i_{k-1}, j_{k-1})$  such that (a)  $0 \leq i_\ell < m$  and  $0 \leq j_\ell < n$  for  $0 \leq \ell < k$ ; (b)  $0 \leq i_{\ell+1} - i_\ell \leq 1$  and  $0 \leq j_{\ell+1} - j_\ell \leq 1$  for  $0 \leq \ell < k-1$ ; and (c)  $(i_\ell, j_\ell) \neq (i_{\ell+1}, j_{\ell+1})$  for  $0 \leq \ell < k-1$ . The pair  $(i_\ell, j_\ell)$  is said to *align*  $a_{i_\ell}$  with  $b_{j_\ell}$ . The weight of an alignment is defined as  $\sum_{0 \leq \ell < k} \delta(a_{i_\ell}, b_{j_\ell})$ , where  $\delta$  is a “local dissimilarity” function that the WARP system defines as  $\delta(a_i, b_j) = |a_i - b_j|^2$ . An optimal alignment is an alignment of minimum weight.

The DTW dissimilarity measure  $D(a, b)$  is defined as  $\sqrt{d(m-1, n-1)}$ , where  $d(m-1, n-1)$  is the weight of an optimal alignment and is defined as<sup>2</sup>

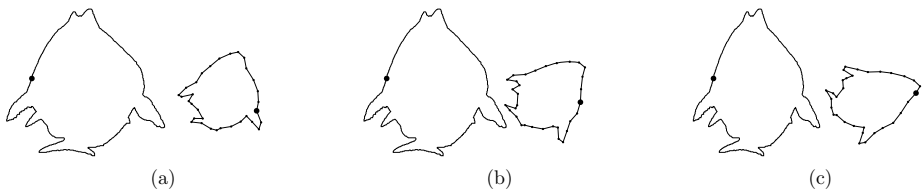
$$d(i, j) = \begin{cases} \delta(a_0, b_0), & \text{if } i = j = 0; \\ d(i-1, j) + \delta(a_i, b_0), & \text{if } i > 0 \text{ and } j = 0; \\ d(i, j-1) + \delta(a_0, b_j), & \text{if } i = 0 \text{ and } j > 0; \\ \min \begin{cases} d(i-1, j-1), \\ d(i-1, j), \\ d(i, j-1) \end{cases} + \delta(a_i, b_j), & \text{if } i > 0 \text{ and } j > 0. \end{cases} \quad (1)$$

This equation can be solved by Dynamic Programming in  $O(mn)$  time: the problem is reduced to the computation of an optimal path in the *warping graph*, a weighted, acyclic graph with  $O(mn)$  arcs. Fig. 1 depicts the complete WARP

<sup>2</sup> The recursive equation in [1, page 144] contains a typo: the square root should be applied only to  $d(m-1, n-1)$ , and not to all  $d(i, j)$ .



**Fig. 1.** The WARP system: shapes are compared by means of DTW on the IDFT of normalized FDs



**Fig. 2.** (a) Original shape and its normalized version. (b) The same shape compressed in the  $X$  axis and its normalized version, which has a different rotation and starting point. (c) A bit more compressed shape and its normalized version, which is also different.

comparison procedure. The DTW computation in the WARP system is  $O(M^2)$ , where  $M \ll m, n$ , since comparisons are performed on signatures.

### 3.1 Drawbacks of the WARP System

It should be noted that subtracting  $(\theta_{-1} + \theta_1)/2$  to the orientation of all FDs only provides rotation invariance modulo  $\pi$  radians [2]. The WARP system does not consider this ambiguity. Anyway, let us consider that the rotation ambiguity is not present. The basic idea of the WARP system is that, after normalization, all shapes have a canonical version with a “standard” centroid, scale, rotation, and starting point and thus, can be compared by means of the DTW dissimilarity measure. But this is a flawed reasoning: invariance is only achieved for different translations, scalings, rotations, and starting points *of the same shape*. Different shapes (even similar ones) may differ substantially in their normalized orientation and starting point. Fig. 2 shows three perceptually similar figures (in fact, the second and third ones have been obtained from the first one by slightly compressing the horizontal axis) whose normalized version are significantly different in terms of *orientation and starting point*. This problem appears frequently in shapes whose basic ellipse is almost a circle. Invariance to rotation and starting point election should be provided by a different method.

In the next section, we present an alternative signature which provides better rotation invariance for similar shapes and a dissimilarity measure which is not affected by the starting point of the signature.

## 4 Cyclic Dynamic Time Warping: A Rotation and Starting-Point Invariance Comparison

We have seen that the signature of similar shapes may present different orientations (Fig. 2). True rotation invariance can be obtained by taking the derivative of the normalized shape, i.e., replacing  $a'_i$  by  $a'_i - a'_{(i-1) \bmod M}$ . We need this derivative signature to use the dissimilarity measure that is detailed next.

When two signatures have “equivalent” starting points, DTW provides a good dissimilarity measure. However, we have seen that similar shapes can present very different starting points. It is useful to consider the problem under the framework of cyclic alignments, i.e., alignments between cyclic sequences.

Let  $[a] = [a_0 a_1 \dots a_{m-1}]$  and  $[b] = [b_0 b_1 \dots b_{n-1}]$  be two cyclic sequences. A *cyclic alignment* between  $[a]$  and  $[b]$  is a sequence of pairs  $(i_0, j_0), (i_1, j_1), \dots, (i_{k-1}, j_{k-1})$  such that, for  $0 \leq \ell < k$ , (a)  $0 \leq i_\ell < m$  and  $0 \leq j_\ell < n$ ; (b)  $0 \leq i_{(\ell+1) \bmod m} - i_\ell \leq 1$  and  $0 \leq j_{(\ell+1) \bmod n} - j_\ell \leq 1$ ; and (c)  $(i_\ell, j_\ell) \neq (i_{(\ell+1) \bmod m}, j_{(\ell+1) \bmod n})$ . The *weight of a cyclic alignment*  $(i_0, j_0), (i_1, j_1), \dots, (i_{k-1}, j_{k-1})$  is defined as  $\sum_{0 \leq \ell < k} \delta(a_{i_\ell}, b_{j_\ell})$ , where  $\delta$  is the local dissimilarity measure. An optimal cyclic alignment is a cyclic alignment of minimum weight.

The Cyclic Dynamic Time Warping (CDTW) dissimilarity measure  $\hat{D}([a], [b])$  is defined as the square root of the weight of the optimal cyclic alignment between  $a$  and  $b$ . First, we are going to show that the optimal cyclic alignment can be defined in terms of alignments between non-cyclic sequences, i.e., in terms of  $D(\cdot, \cdot)$ ; then, we will present an efficient procedure to compute it.

**Lemma 1.** *If  $m, n > 1$  and  $(i_0, j_0), (i_1, j_1), \dots, (i_{k-1}, j_{k-1})$  is an optimal alignment between two sequences  $a_0 a_1 \dots a_{m-1}$  and  $b_0 b_1 \dots b_{n-1}$ , there is at least one  $\ell$  such that  $i_\ell \neq i_{(\ell+1) \bmod m}$  and  $j_\ell \neq j_{(\ell+1) \bmod n}$ .*

*Proof:* Any alignment including  $(i_\ell, j_\ell), (i_\ell + 1, j_\ell)$ , and  $(i_\ell + 1, j_\ell + 1)$  can be “improved” by removing  $(i_\ell + 1, j_\ell)$ , since  $\delta(a_{i_\ell+1}, b_{j_\ell}) \geq 0$ . Analogously, any alignment including  $(i_\ell, j_\ell), (i_\ell, j_\ell + 1)$ , and  $(i_\ell + 1, j_\ell + 1)$  can be “improved” by removing  $(i_\ell, j_\ell + 1)$ .  $\square$

**Lemma 2.** *The CDTW dissimilarity between  $[a] = [a_0 a_1 \dots a_{m-1}]$  and  $[b] = [b_0 b_1 \dots b_{n-1}]$ ,  $\hat{D}([a], [b])$ , can be computed as  $\min_{0 \leq k < m} \min_{0 \leq \ell < n} D(\sigma^k(a), \sigma^\ell(b))$ .*

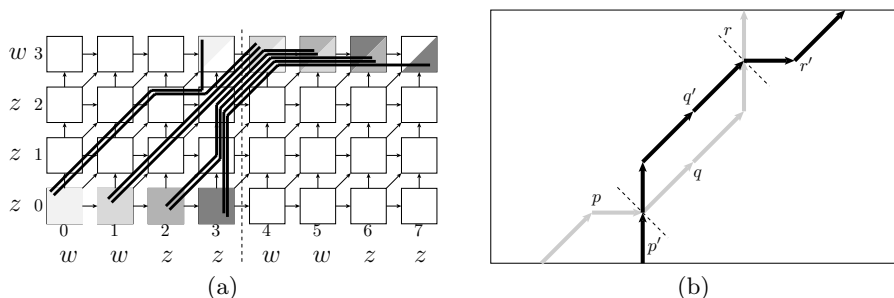
*Proof:* Trivial when  $m = 1$  or  $n = 1$ . Let us consider that  $m, n > 1$  and let  $(i_0, j_0), (i_1, j_1), \dots, (i_{k-1}, j_{k-1})$  be an optimal alignment. Let  $\ell$  be an index such that  $i_\ell \neq i_{(\ell+1) \bmod m}$  and  $j_\ell \neq j_{(\ell+1) \bmod n}$  (by Lemma 1). The weight of this cyclic alignment is  $D(\sigma^{(i_\ell+1) \bmod m}(a), \sigma^{(j_\ell+1) \bmod n}(b))$ , which is considered by the double minimization.  $\square$

According to Lemma 2, the value of  $\hat{D}([a], [b])$  can be trivially computed in  $O(m^2n^2)$  time by solving  $mn$  recurrences like equation (1). Maes showed in [4] that the Cyclic Edit Distance (CED), a related dissimilarity measure, can be computed in  $O(m^2n)$  time by performing cyclic shifts only on one of the sequences. This observation finally led to a  $O(mn \lg m)$  time algorithm. Is it possible to perform cyclic shifts on only one of the sequences when computing the CDTW? The answer is no: in general,  $\hat{D}([a], [b])$  is neither  $\min_{0 \leq k < m} D(\sigma^k(a), b)$  nor  $\min_{0 \leq k < n} D(a, \sigma^k(b))$ , as the following counter-example shows: let  $z$  and  $w$  be two complex numbers such that  $\delta(z, w) = 1$ ; the value of  $\hat{D}([zwz], [wzw])$  is 0, since  $D(zzw, zww) = 0$ , but  $D(zwz, wz w) = 3$  and  $D(wzz, wz w) = D(zzw, wz w) = D(zwz, zw w) = D(zwz, wwz) = 1$ . Therefore, an equivalent of Maes' algorithm for the CED computation cannot be directly applied to CDTW dissimilarity computation.

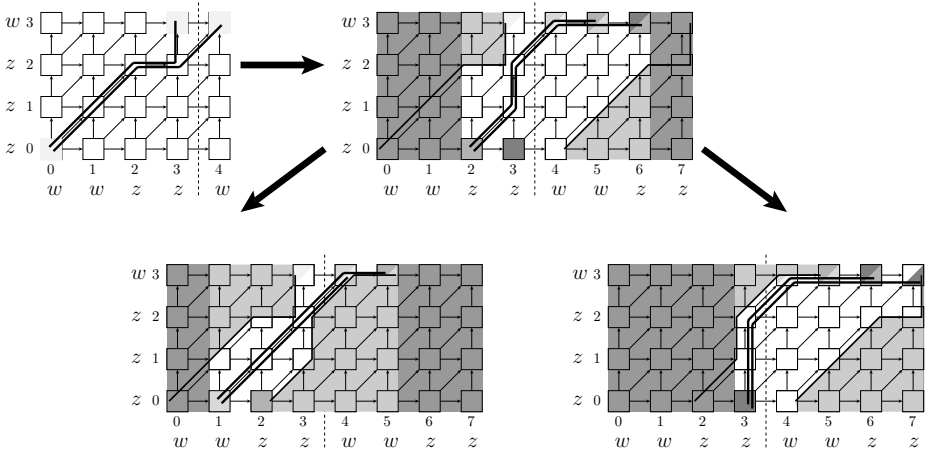
**Theorem 1.** *The CDTW dissimilarity between cyclic sequences  $[a]$  and  $[b]$  can be computed as  $\hat{D}([a], [b]) = \min_{0 \leq k < m} (\min(D(\sigma^k(a), b), D(\sigma^k(a)a_k, b)))$ .*

*Proof:* Each alignment induces a segmentation on  $a$  and a segmentation on  $b$ . All the values in a segment are aligned with the same value of the other cyclic sequence (Lemma 1). There is a problem when  $b_{n-p-1}, b_{n-p}, \dots, b_{n-1}$  and  $b_0, b_1, \dots, b_q$ , for some  $p, q \geq 0$ , should belong to the same segment of  $b$ . In that case, the optimal path cannot be obtained by simply shifting  $a$ , since  $b_{n-1}$  must be aligned with the last value of  $\sigma^k(a)$  and  $b_0$  must be aligned with its first value, i.e., they cannot belong to the same segment. The sequence  $\sigma^k(a)a_k$  allows to align  $b_{n-p}b_{n-p+1} \dots b_n$  and  $b_0b_1 \dots b_q$  with the first value of  $\sigma^k(a)$ , since  $a_k$  also appears at the end of  $\sigma^k(a)a_k$ .  $\square$

The value of  $D(\sigma^k(a), b)$  and  $D(\sigma^k(a)a_k, b)$ , for each  $k$ , can be obtained by computing shortest paths in an *extended warping graph* similar to the extended



**Fig. 3.** (a) Extended warping graph for  $a = wz z z$  and  $b = z z z w$ , where  $z$  and  $w$  are complex numbers such that  $\delta(z, w) = 1$ . Arcs ending at node  $(i, j)$  are weighted  $\delta(a_i, b_j)$ . The optimal alignment for  $[a]$  and  $[b]$  is the minimum weight path starting from any colored node in the lower row and ending at a node containing the same color in the upper row (all path candidates are shown with thick lines). (b) Optimal crossing paths can be avoided: if the weight of the subpath  $q$  is greater than the weight of the subpath  $q'$ , the black path can be improved by traversing  $q'$  instead of  $q$ .



**Fig. 4.** Divide-and-Conquer procedure to compute the CDTW dissimilarity between the sequences of Fig. 3. First, the optimal alignment (path) between  $a$  and  $b$  and between  $\sigma^0(a)a_0$  and  $b$  is computed. The first optimal path is used as a left and right frontier in the extended graph: only the white region must be explored to compute the optimal alignment between  $\sigma^2(a)$  and  $b$  and between  $\sigma^2(a)a_2$  and  $b$ . This idea is applied recursively to the computation of the other optimal alignments, but using also the optimal alignment between  $\sigma^2(a)$  and  $b$  as a new left or right frontier.

edit graph defined by Maes [4] (see Fig. 3 (a)). Since the non-crossing property of edit paths also holds for alignment paths (see Fig. 3 (b)), the Divide-and-Conquer approach proposed by Maes can be applied to CDTW. The reader is addressed to [4] to obtain a complete description of the Divide-and-Conquer procedure, which is depicted in Fig. 4. It should be taken into account that, unlike in Maes' algorithm, the optimal path starting at  $(k, 0)$  can finish either at node  $(k + m - 1, n - 1)$  or  $(k + m, n - 1)$ .

When applied to signatures, the running time of the algorithm is  $O(M^2 \log M)$ : each recursive step divides the search space in two halves and all recursive operations at the same recursion level require total  $O(M^2)$  time.

## 5 Experiments

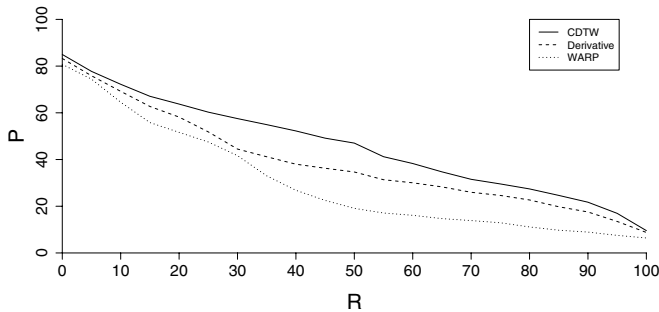
In [1], the WARP system was tested on a labeled version of the SQUID Demo database and the MPEG-7 Core Experiment CE-Shape-1 (part B). We have performed comparative experiments with the same test sets.

The SQUID Demo database consists of 1100 contours of marine species and is used as a demonstration application of the Shape Queries Using Image Databases system [5]. The original database does not divide the contours into classes. Bartolini *et al.* manually classified 252 images into 10 semantic categories<sup>3</sup>. They

<sup>3</sup> Seahorses (5 images), seamounts (6), sharks (58), soles (52), tonguefishes (19), crustaceans (4), eels (26), u-eels (25), pipefishes (16), and rays (41).

conducted some precision (P) versus recall (R) shape retrieval experiments with 30 query images from the 10 semantic categories. For each query, images in the same category were considered relevant and all the others were considered irrelevant. Since we do not know which query images were used, we have run queries on the 252 labeled shapes.

Fig. 5 shows the precision/recall graph for 3 retrieval procedures: (i) **WARP**: the standard WARP system; (ii) **Derivative**: derivative of the reconstructed contour as a shape signature and DTW-based comparison; (iii) **CDTW**: derivative of the reconstructed contour and comparison by means of the Cyclic Dynamic Time Warping dissimilarity. It can be seen that the two methods proposed in this work improve the WARP results. The signature based on the derivative provides a better precision/recall curve, thus confirming that the WARP system is sensitive to variation of orientation in the signatures of similar (but not identical) shapes. There is also a significant difference between CDTW-comparison and the other methods.



**Fig. 5.** Precision/Recall results on the SQUID Demo database

In [1], the WARP system was also compared to a Curvature Scale Space (CSS) based image retrieval on the same MPEG-7 experiment presented in [3]. A CSS-based query system obtained an average precision of 37.72% (the maximum precision attainable in that experiment is 50%) and the WARP system obtained a 29.25% average precision. Bartolini *et al.* explain in [1] that the CSS system is an approximate query processing algorithm that can easily lead to false dismissals (filtering out best-matching images) by discarding shapes with an aspect ratio greater than a user threshold. Other techniques with similar or slightly better results are not suitable for efficient indexing and, thus, can only be used in small-size databases. Using the derivative-based signature, the average precision is 31.29%. The precision raises to 34.17% when the CDTW is used.

## 6 Conclusions

In this work, we have critically studied the WARP system, detected some drawbacks, and presented several ways to improve its precision/recall behavior on



shape-based image retrieval tasks: (a) using the original signatures derivative, (b) using the signature derivative with a CDTW comparison. Proposal (a) provides better results than the WARP system and proposal (b) offers the best precision/recall.

The CDTW dissimilarity has been defined and an algorithm to compute it in  $O(M^2 \log M)$  for two signatures of length  $M$  has been presented. We have shown that the Cyclic Edit Distance algorithm presented by Maes cannot be directly extended to CDTW: two conventional DTW dissimilarities must be computed for each cyclic shift of one sequence. Fortunately, one of these dissimilarities can be obtained as a subproduct of the computation of the other.

## Acknowledgments

The authors wish to thank S. Abbasi, F. Mokhtarian, and J. Kittler for making the SQUID database publicly available and to I. Bartolini, P. Ciaccia, and M. Patella for providing their labeled version of the SQUID database.

## References

1. I. Bartolini, P. Ciaccia, and M. Patella. WARP: Accurate Retrieval of Shapes Using Phase of Fourier Descriptors and Time Warping Distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(1):142–147, 2005.
2. A. Folkers and H. Samet. Content-based Image Retrieval Using Fourier Descriptors on a Logo Database. In *Proc of the 16th Int Conf on Pattern Recognition*, pages 521–524, 2002.
3. J. Latecki, R. Lakämper, and U. Eckhardt. Shape descriptors for non-rigid shapes with a single closed contour. In *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pages 424–429, 2000.
4. M. Maes. On a Cyclic String-to-String Correction Problem. *Information Processing Letters*, 35:73–78, 1990.
5. F. Mokhtarian, J. Kittler, and S. Abbasi. Shape queries using image databases. <http://www.ee.surrey.ac.uk/Research/VSSP/imagenet/demo.html>.
6. D. Sankoff and J. Kruskal, editors. *Time warps, string edits, and macromolecules: the theory and practice of sequence comparison*. Addison-Wesley, Reading, MA, 1983.
7. T. Sikora. The mpeg-7 visual standard for content description – an overview. *IEEE Transactions on Circuits and Systems for Video Technology*, 11(6):696–702, 2001.