A Heuristic Based on the Intrinsic Dimensionality for Reducing the Number of Cyclic DTW Comparisons in Shape Classification and Retrieval Using AESA

Vicente Palazón-González and Andrés Marzal*
{palazon,amarzal}@lsi.uji.es

Dept. Llenguatges i Sistemes Informàtics and Institute of New Imaging Technologies.
Universitat Jaume I de Castelló. Spain.

Abstract. Cyclic Dynamic Time Warping (CDTW) is a good dissimilarity of shape descriptors of high dimensionality based on contours, but it is computationally expensive. For this reason, to perform recognition tasks, a method to reduce the number of comparisons and avoid an exhaustive search is convenient. The Approximate and Eliminate Search Algorithm (AESA) is a relevant indexing method because of its drastic reduction of comparisons, however, this algorithm requires a metric distance and that is not the case of CDTW. In this paper, we introduce a heuristic based on the intrinsic dimensionality that allows to use CDTW and AESA together in classification and retrieval tasks over these shape descriptors. Experimental results show that, for descriptors of high dimensionality, our proposal is optimal in practice and significantly outperforms an exhaustive search, which is the only alternative for them and CDTW in these tasks.

Keywords: Cyclic strings, cyclic sequences, cyclic dynamic time warping, shape classification, shape retrieval, intrinsic dimensionality, metric spaces, AESA.

1 Introduction

Shape classification and retrieval are very important problems with applications in several areas such as industry, medicine, biometrics and even entertainment.

Among the methods to solve this problem the ones related to Dynamic Time Warping (DTW) [1] and descriptors of the contour with sequences of components of several dimensions have had a significant presence [2–8]. In general, these shape descriptors aim to have information from all of the contour with respect to each point, that is the reason for their large size (see Figure 1 for an example of the shape descriptor used in [6]). These methods offer very competitive results because of their full description and the properties that DTW has as a dissimilarity (DTW is able to align parts instead of points and it is robust with elastic deformations). Nevertheless, this combination has a high computational cost. Besides, the problem of the starting point invariance appears, i.e., where we have to start the comparison in the sequence. Although there are many

* Work partially supported by the Spanish Government (TIN2010-18958 and *Consolider Ingenio 2010* CSD2007-00018), the Generalitat Valenciana (*Prometeo*/2010/028), and *Fundació Caixa Castelló-Bancaixa* (P11B2009-48).

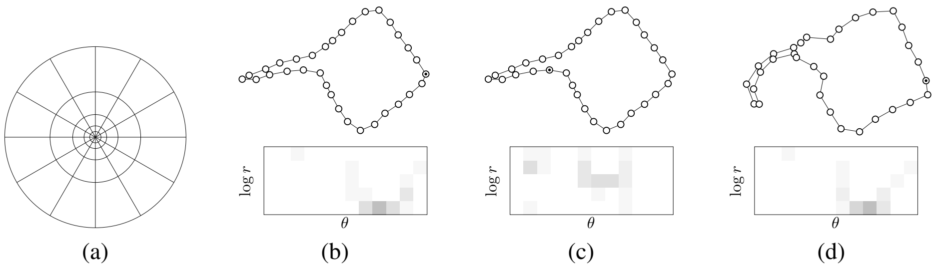

Fig. 1. Shape context computation. Given a set of landmark points from the contour, for each point is defined a histogram of the relative coordinates of the remaining points. (a) Diagram of log-polar histogram bins used in computing the shape contexts. Five bins for $\log r$ and 12 bins for θ, 60 dimensions. (b) Landmark points and the corresponding histogram of the point marked by a black dot. (c) The same shape but using a different point. (d) Another shape similar to (b).

heuristic methods to obtain this invariance, they are not suitable in most of the domains. Therefore, the literature accepts that to obtain a good starting point we must make the comparison between every possible starting point of the sequence [2, 3, 9, 4]. Hence the necessity to use cyclic sequences and then CDTW (Cyclic DTW) arises.

In [10], an algorithm is proposed to calculate the CDTW in time $O(n^2 \log n)$ (being n the size of the sequences). Although this algorithm considerably reduces the cost, with the shape descriptors mentioned before, the local distance or dissimilarity [10] between the components of the sequence has too much weight on the final cost, due to its dimensionality. Thus, in recognition tasks to use solutions that avoid the computation of CDTW over all the prototypes of the database is necessary, i.e., to avoid an exhaustive search.

In [9], the authors, using a method similar to their previous work with DTW [11], try to speed up the CDTW as well. In this work, they do not use the algorithm of [10], but they make clusters of sequences based on their similarity, treating every possible starting point as a different sequence and using indexing methods with lower bounds of these clusters. This solution seems to be suitable just for shape descriptors with only one dimension (such as the curvature) and not for much more dimensions [2–8]. For instance, in [5], 60 dimensions are required for each point (Figure 1). Another problem is that it cannot use more sophisticated local distances (in CDTW between elements of the shape descriptors) such as χ^2 [6], due to their lower bound.

AESA [12, 13] is characterised by a drastic reduction of the computation of distances. It is then specially interesting when the distance has a high cost and that is precisely our case. However, CDTW is not a metric because it does not satisfy the triangular inequality, which is an indispensable property for using AESA. In [14–17], the authors used AESA to speed up a speech recognition task based on DTW with good results in spite of not satisfying this property. In the current work, we improve their heuristic adding an important factor: the intrinsic dimensionality [18]. As far as we know the heuristic presented here is the only alternative to an exhaustive search in the context of shape classification and retrieval with cyclic sequences of high dimensionality.

The paper is organized as follows: The next section describes how the intrinsic dimensionality is affected in the search of nearest neighbours. In Section 3, the triangular inequality is related to the intrinsic dimensionality and how we can use AESA due to this relation is explained, that is to say, we present our heuristic. In Section 4, we show experiments to validate our proposal. Finally, conclusions are formulated in Section 5.

2 On the Intrinsic Dimensionality and Nearest Neighbours

Indexing methods based on metrics do not necessarily work with all databases and all metrics. Their efficiency is affected by the distribution of distances of the database. From this distribution we can obtain the intrinsic dimensionality. According to [18], given a database D and a metric m, the intrinsic dimensionality, ϱ, is: $\varrho(D, m) = \frac{\mu^2}{2\sigma^2}$, where μ and σ^2 are the mean and the variance of the distribution of distances.

In [18], it is shown, in an analytical and experimental way, that all the algorithms based on metrics degrade in a systematic way as the dimensionality increases, i.e., the computational cost is getting close to the one of an exhaustive search.

We can observe that the intrinsic dimensionality increases because of the two next reasons: the variance decreases and/or the mean of the distribution of distances increases. In Figure 2, we can see two distributions of distances showing a low and high intrinsic dimensionality. Two extreme cases, where both variance and mean vary. If the variance decreases, it means that the most distances have similar value, then we are going to have less information for pruning (in the case of AESA, bounds are going to be worse). On the other hand, if the mean increases, to obtain the nearest neighbours we will have to explore more prototypes (in AESA, we will take more time to find a prototype for a good pruning).

But, in our problem, what determines the distribution of distances?, i.e., what provokes that ϱ increases?. We can consider two causes. One is the sequence or the shape descriptor, especially affecting the number of points and the number of dimensions for each point. For instance, the BAS descriptor [2] uses 4 dimensions for each point and the shape contexts [5] 60 dimensions for each point. The second cause is the distance for comparing the sequences. Even though, if we set as the distance the CDTW, the local distance gains importance, which for the BAS descriptor is the euclidean distance and for the shape contexts is χ^2 [5].

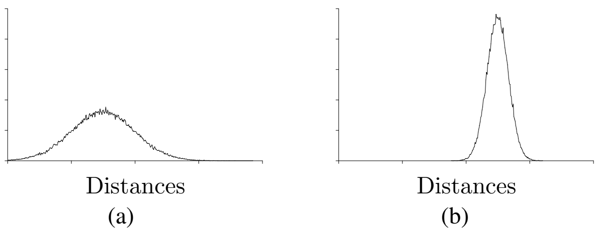

Fig. 2. Synthetic example of two distributions of distances. (a) With a low intrinsic dimensionality. (b) With a high intrinsic dimensionality.

3 Improving the Heuristic with the Intrinsic Dimensionality

The only problematic property for DTW to be a metric, is the triangular inequality:

$$d(x, z) \leq d(x, y) + d(y, z),$$

since it is possible to find counterexamples where DTW does not satisfy it [14, 19] (thus, CDTW is not a metric either). The correction of algorithms such as AESA (Figure 3) depends on having a metric distance and then it has to satisfy this property.

In [14–17], a study was performed with a task of speech recognition with isolated words using DTW. They aimed to see how not to satisfy the triangular inequality by DTW affected in samples of the real world. These samples were speech frames that were represented by sequences of components of eight dimensions. In [15], in 15 millions of triplets there were no cases where the triangular inequality was violated. In [19], the authors made experiments with synthetic time series (sequences of one dimension) of three types: *white-noise*, *random-walk* and *cylinder-bell-funnel*. The most problematic was *random-walk* where 20% of triplets violated the triangular inequality.

To see how many triplets x, y, z violate the triangular inequality we can use the next formula:

$$H = d(x, y) + d(y, z) - d(x, z). \tag{1}$$

All the triplets that have an H less than zero do not satisfy the triangular inequality. In [14], distributions (or histograms) of the frequencies of triplets for each H are shown. These distributions seem to have a gaussian form and when $H = 0$ the frequency is very low.

In (1), we can observe that the distribution of H has a relation to the distribution of distances (Section 2). That is to say, H is a composition of three random variables with the same distribution (the distribution of distances). The greater the mean, μ, of the distribution of distances, the greater the value of H of most of triplets, therefore, there will be more positive values because we will be adding two distances of the same distribution and subtracting another one of the same distribution too. In the case of the variance, σ^2, a similar thing will happen but with a lower variance, since the distances will be similar, and then, there will be more values of H that are greater or equal to zero. Therefore, we can say that, when the intrinsic dimensionality, ϱ, is greater, we will very probably find a lower number of triplets, x, y, z, that violate the triangular inequality. In practice, and in the case of CDTW, this statement shows that it will be easier to find triplets that violate triangular inequality in sets of sequences whose components have one dimension, like the curvature descriptor, than in sequences with 60 dimensions, like the shape contexts descriptor [5]. Thus, we can apply AESA with greater chances of success the greater the dimensionality of our cyclic sequences.

In our experiments with real world data (Section 4) we obtained few cases that violate the triangular inequality. However, in the curvature descriptor it arrives to almost a 3%. For the other types of descriptor the amount is very low as we expected and, given the characteristics of AESA (Figure 3), the recognition rates are not going to be significantly affected in practice. The fact that the intrinsic dimensionality increases is good for the triangular inequality but not for AESA, since it degrades the search [18], as we mentioned before. Even so, as we will see in the next section the results are satisfactory both in time and in classification and retrieval rates.

Fig. 3. AESA. In our case the distance d is the CDTW.

```
Input: P: prototypes, x: sample to classify, D ∈ ℝ^{|P|×|P|}: distances between prototypes
Output: nn ∈ P: nearest neighbour
begin
    for p ∈ P do
        G[p] = 0
    nn = unknown; d_nn = ∞; s = any element from P
    while |P| > 0 do
        d_s = d(x, s); P = P − s
        if d_s < d_nn then
            nn = s; d_nn = d_s
        next = unknown; gmin = ∞
        for p ∈ P do
            G[p] = max(G[p], |D[s, p] − d_s|)        // lower bound based on the
                                                     // triangular inequality
            if G[p] > d_nn then
                P = P − p
            else
                if G[p] < gmin then
                    gmin = G[p]; next = p
        s = next
end
```

4 Experiments

In order to assess the behaviour of our proposal, we performed experiments on an Intel i7 2.66GHz machine running under linux 3.2.0. The real world databases used were the MPEG-7 Core Experiment CE-Shape-1 (part B) [20] and the Silhouette database [21]. The shape descriptors were: curvature (as an example descriptor of one dimension for each point), BAS [2] (four dimensiones) and the shape contexts (SC) [5, 6] (60 dimensions). The results achieved with these descriptors, and in particular the ones with the shape contexts, can be applied to other ones of similar characteristics from the bibliography [3, 4, 6–8]. We also used a synthetic corpus of sequences of several dimensions (1, 5, 10, 20 and 60). We generated 1000 sequences (for each number of dimensions) with a random walk (for each dimension of the sequence) defined by $x_i = x_{i-1} + N(0, 1)$ and $x_1 = 0$ as in [19].

In the following, we will observe how the intrinsic dimensionality affects CDTW to satisfy the triangular inequality property. Subsequently, speeding up results are shown

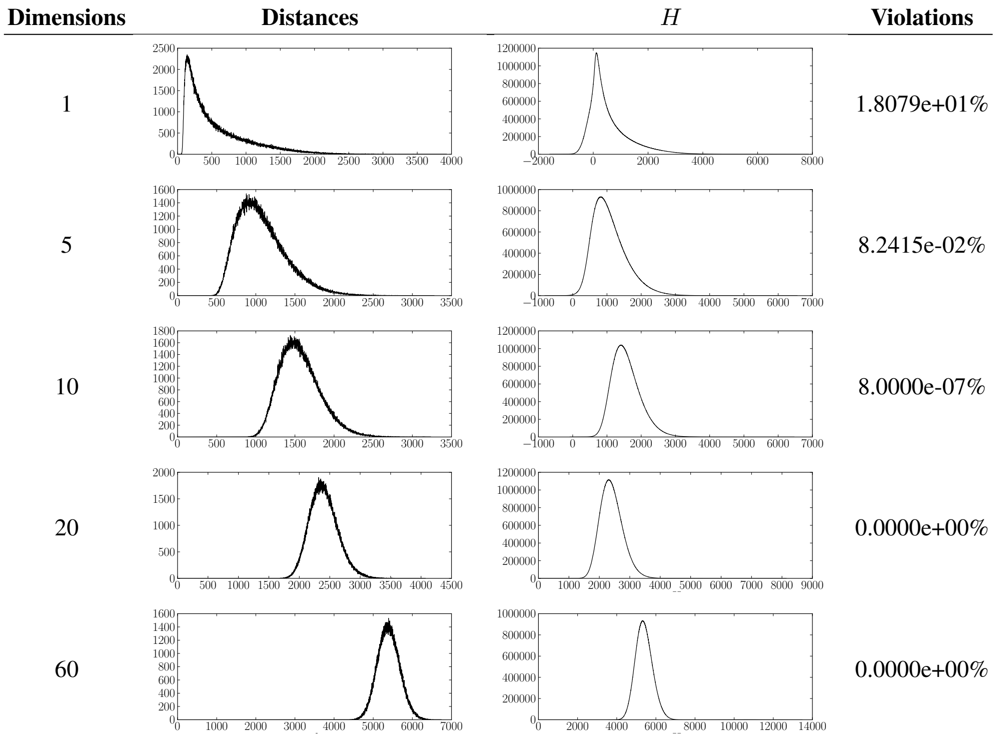

Fig. 4. Dimensions of the sequences, histograms of the distribution of distances, the distribution of H and the percentage of triplets that violate the triangular inequality, for the experiments with the random-walk synthetic corpus.

with AESA with respect to an exhaustive search. Finally, we will see how using AESA affects classification and retrieval rates.

4.1 Intrinsic Dimensionality and Triangular Inequality

In Figure 4, we can see by means of histograms the relation between the distribution of distances and the distribution of H (Section 3). We performed an experiment similar to the one in [19], with random-walk sequences, but varying the number of dimensions (in [19] this experiment was done for just one dimension). We generated 1000 sequences for each number of dimensions, then we checked 1000000000 triplets. The fact of having the distribution of distances near to 0 (as it happens with sequences of one dimension) makes more probable to find triplets that violate the triangular inequality in the distribution of H. On the other hand, if the distribution of distances is far from the value 0 (as it happens with sequences of 20 dimensions), the percentage considerably decreases.

With respect to real world data, Table 1 shows the dimensionality and the corresponding percentage of triplets that violate the triangular inequality for each shape descriptor. As we can be observe, the greater the dimensionality the lower the percentage

Table 1. Comparison of the dimensionality with the percentage of triplets that violate the triangular inequality.

		Dimensions	**Violations**
MPEG7B	Curvature	1	2.95 %
	BAS	4	$7.25{\cdot}10^{-3}$ %
	SC	60	$7.07{\cdot}10^{-5}$ %
Silhouette	Curvature	1	$3.93{\cdot}10^{-1}$ %
	BAS	4	$2.61{\cdot}10^{-5}$ %
	SC	60	$6.54{\cdot}10^{-7}$ %

of triplets. As it is commented in Section 3, a great value in the intrinsic dimensionality makes the violation of the triangular inequality less probable.

4.2 Time

We also performed experiments of shape retrieval for the k most similar shapes, with values of k: 1, 5, 10, 20 and 40. To use AESA to obtain the k nearest neighbours we can keep a sorted list of them and prune with the last one. In classification, in many cases, it would be enough $k = 1$, although we could also use greater values. In retrieval, 10 or 20 prototypes could be enough for a first answer (or even unique) for a user of a concrete application of shape retrieval.

For BAS and SC we present a graph (Figure 5) with the average time of AESA, with respect to an exhaustive search. There is a huge improvement.

4.3 Classification and Retrieval Rates Using AESA and CDTW

Finally, we need to mention the classification and retrieval rates for the k nearest neighbours (Table 2). The only results that change are the ones of the curvature, but the difference is not so great.

Table 2. Recognition rates for an exhaustive search and AESA.

		Curvature		BAS		SC	
	k	Exhaustive	AESA	Exhaustive	AESA	Exhaustive	AESA
MPEG7B	1	90.50	**90.00**	97.64	97.64	98.78	98.78
	5	83.21	**83.23**	94.91	94.91	96.67	96.67
	10	70.94	**70.74**	88.44	88.44	91.63	91.63
	20	55.39	**55.29**	74.61	74.61	79.20	79.20
	40	63.83	**63.43**	82.85	82.85	86.73	86.73
Silhouette	1	91.87	**91.77**	96.91	96.91	98.59	98.59
	5	88.25	**88.20**	94.66	94.66	97.77	97.77
	10	80.27	80.27	90.75	90.75	95.81	95.81
	20	69.47	69.47	83.13	83.13	90.60	90.60
	40	61.20	61.20	73.85	73.85	83.03	83.03

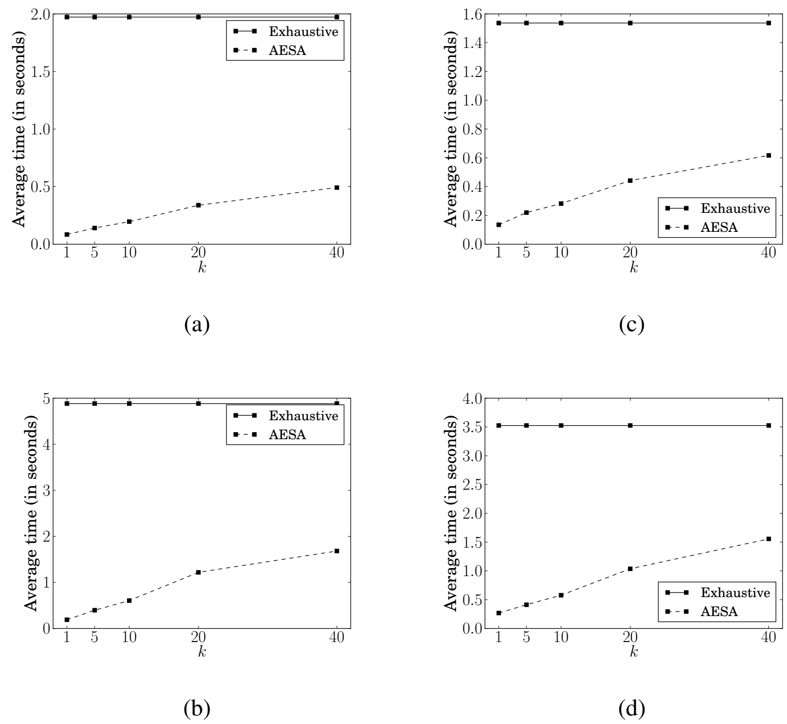

Fig. 5. Average time of an exhaustive search and AESA for (a) BAS, and (b) the shape contexts with the MPEG7B database. For (c) BAS, and (d) the shape contexts with the Silhouette database.

5 Discussion

From the experiments presented in the last section, it is clear that when the dimensionality of the cyclic sequences is sufficiently high we can obtain a very low percentage of triplets that violate the triangular inequality. In real tasks of shape classification and retrieval with AESA and CDTW, we have studied three shape descriptors with different number of dimensions. In particular the curvature, despite having a significant percentage of violations of the triangle inequality, surprisingly obtains quite acceptable rates with AESA (without being the same ones). With BAS and SC descriptors the rates are the same with respect to an exhaustive search. But if the results of the curvature are acceptable, with descriptors of higher dimensionality we can be more confident that AESA will have a good behaviour. We want to remark as well that our proposal significantly speeds up the classification and retrieval of these shape descriptors [2–8] and that our heuristic is the only alternative to an exhaustive search for them.

Of course, this proposal can be applied to other contexts based on DTW, not just the one of shape recognition and obviously it is possible to use other indexing methods based on metric spaces. In posterior work we aim to explore these contexts.

References

1. Sankoff, D., Kruskal, J., eds.: Time Warps, String Edits, and Macromolecules: the Theory and Practice of Sequence Comparison. Addison-Wesley, Reading, MA (1983)
2. Arica, N., Yarman-Vural, F.T.: BAS: a perceptual shape descriptor based on the beam angle statistics. Pattern Recognition Letters **24**(9-10) (2003) 1627–1639
3. Adamek, T., O'Connor, N.E.: A multiscale representation method for nonrigid shapes with a single closed contour. IEEE Trans. Circuits Syst. Video Techn **14**(5) (2004) 742–753
4. Alajlan, N., Rube, I.E., Kamel, M.S., Freeman, G.: Shape retrieval using triangle-area representation and dynamic space warping. Pattern Recognition **40**(7) (July 2007) 1911–1920
5. Belongie, S., Malik, J., Puzicha, J.: Shape matching and object recognition using shape contexts. IEEE Trans. Pattern Anal. Mach. Intell **24**(4) (2002) 509–522
6. Ling, H., Jacobs, D.W.: Shape classification using the inner-distance. IEEE Trans. Pattern Anal. Mach. Intell **29**(2) (2007) 286–299
7. Gopalan, R., Turaga, P., Chellappa, R.: Articulation-invariant representation of non-planar shapes. In: European Conference on Computer Vision. (2010) 286–299
8. Wang, J., Bai, X., You, X., Liu, W., Latecki, L.: Shape matching and classification using height functions. Pattern Recognition Letters (2011)
9. Keogh, E., Wei, L., Xi, X., Vlachos, M., Lee, S., Protopapas, P.: Supporting exact indexing of arbitrarily rotated shapes and periodic time series under Euclidean and warping distance measures. The VLDB Journal **18**(3) (2009) 611–630
10. Marzal, A., Palazón, V., Peris, G.: Shape Retrieval Using Normalized Fourier Descriptors Based Signatures and Cyclic Dynamic Time Warping. Structural, Syntactic, and Statistical Pattern Recognition (2006) 208–216
11. Keogh, E.J., Ratanamahatana, C.A.: Exact indexing of dynamic time warping. Knowl. Inf. Syst **7**(3) (2005) 358–386
12. Vidal, E.: An algorithm for finding nearest neighbours in (approximately) constant average time. Pattern Recognition Letters **4**(3) (1986) 145–157
13. Vidal, E.: New formulation and improvements of the nearest-neighbour approximating and eliminating search algorithm (AESA). Pattern Recognition Letters **15**(1) (1994) 1–7
14. Vidal, E., Casacuberta, F., Rulot, H.M.: Is the DTW distance really a metric? An algorithm reducing the number of DTW comparisons in isolated word recognition. Speech Communication **4**(4) (1985) 333–344
15. Casacuberta, F., Vidal, E., Rulot, H.: On the metric properties of dynamic time warping. IEEE Trans. Acoustics, Speech and Signal Processing **ASSP-35**(11) (1987) 1631
16. Vidal, E., Casacuberta, F., Benedi, J., Lloret, M.: On the verification of triangle inequality by dynamic time-warping dissimilarity measures. Speech Commun. **7**(1) (1988) 67–79
17. Vidal, E., Rulot, H.M., Casacuberta, F., Benedi, J.M.: On the use of a metric-space search algorithm (AESA) for fast DTW-based recognition of isolated words. IEEE Trans. Acoustics, Speech and Signal Processing **ASSP-36**(5) (1988) 651
18. Chávez, E., Navarro, G., Baeza-Yates, R., Marroquín, J.: Searching in metric spaces. ACM Computing Surveys (CSUR) **33**(3) (2001) 273–321
19. Lemire, D.: Faster retrieval with a two-pass dynamic-time-warping lower bound. Pattern Recognition **42**(9) (2009) 2169–2180
20. Bober, M.: MPEG-7 visual shape descriptors. IEEE Trans. Circuits Syst. Video Techn **11**(6) (2001) 716–719
21. Sharvit, D., Chan, J., Tek, H., Kimia, B.B.: Symmetry-based indexing of image databases. In: Workshop on Content-Based Access of Image and Video Libraries. (1998) 56–62