

# Balancing Deformability and Discriminability for Shape Matching

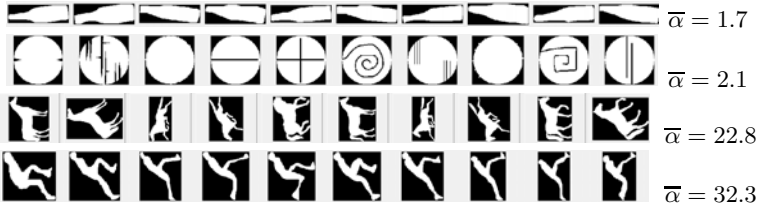
Haibin Ling, Xingwei Yang, and Longin Jan Latecki

Center for Information Science and Technology, Dept. of Computer and Information Science  
Temple University, Philadelphia, 19122, USA  
{hbling, xingwei.yang, latecki}@temple.edu

**Abstract.** We propose a novel framework, *aspect space*, to balance deformability and discriminability, which are often two competing factors in shape and image representations. In this framework, an object is embedded as a surface in a higher dimensional space with a parameter named *aspect weight*, which controls the importance of intensity in the embedding. We show that this framework naturally unifies existing important shape and image representations by adjusting the aspect weight and the embedding. More importantly, we find that the aspect weight implicitly controls the degree to which a representation handles deformation. Based on this idea, we present the *aspect shape context*, which extends shape context-based descriptors and adaptively selects the “best” aspect weight for shape comparison. Another observation we have is the proposed descriptor nicely fits context-sensitive shape retrieval. The proposed methods are evaluated on two public datasets, MPEG7-CE-Shape-1 and Tari 1000, in comparison to state-of-the-art solutions. In the standard shape retrieval experiment using the MPEG7 CE-Shape-1 database, the new descriptor with context information achieves a bull’s eye score of 95.96%, which surpassed all previous results. In the Tari 1000 dataset, our methods significantly outperform previous tested methods as well.

## 1 Introduction

To understand and analyze object deformation is one of the most important goals in pattern recognition. The study of object shapes dates back to at least a century ago in D’Arcy Thompson’s seminal work [35]. Recently, great successes have been achieved in robust image and shape representation for many computer vision tasks [24,25,18,27,9,21,14,5,36,33,13]. Many approaches gain the robustness by using representations that are invariant (or insensitive) to certain groups of deformations. For example, the popular scale invariant feature transform (SIFT) [24] selects scale invariant blob regions [20] for reliable key point matching. Some further works push this frontier to more general deformations such as affine invariant [27], projection invariant [37], or general deformation invariant [25,21]. Similar progress has been made in shape analysis as well. Moment invariants [12] are well known to be robust to similarity transformations. Bending invariants [10,6,7] for 2D and 3D shapes can be achieved by using geodesic distances. Articulation insensitivity is gained similarly by using the inner-distance [22]. Topology invariants [8,30] have been applied for both 2D and 3D shape analysis. Invariant representations are also used in texture analysis [18,39].



**Fig. 1.** Four shape classes from the MPEG7 dataset and their discriminative deformabilities  $\bar{\alpha}$  (§ 3.3)

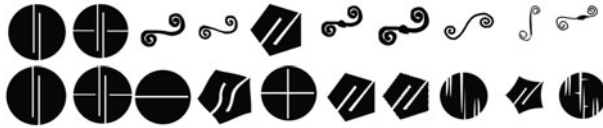
On the other hand, invariants to larger groups of deformation often come at a price of reduced discriminability [37,21] and sometimes introduce vulnerability to imaging nuisances such as noises and illumination variations. In fact, as observed in category classification tasks [42], the best local features are often a trade off between descriptive abilities and robustness to deformations.

In Fig. 1, we see that objects often undergo different degrees of deformation and a trade off between deformation robustness and discriminability is desired. For example, simple Euclidean distance based approaches (e.g., shape context [5]) can work well for rigid objects like bottles (first row in Fig. 1) while the articulation invariant representations (e.g., inner-distance shape context [22]) handles articulated shapes like human body (fourth row of Fig. 1) successfully. However, for complicated shapes like the pie-shaped devices in Fig. 1 (second row), articulation invariance becomes an over-killing that easily confuses a query shape with other shapes through articulation. This fact is illustrated in the first row of Fig. 2.

Motivated by the above work and especially by [21,20], we propose a new framework called *aspect space*<sup>1</sup> to balance the deformation and discriminability for object representation. The basic idea is to embed a 2D image or a shape (or a 3D volume) as a surface in a 3D space (or a 4D space) with a parameter named *aspect weight*, which controls the importance of intensity in the embedding. The embedded surfaces with different aspect weights then form an aspect space, such that the aspect weight naturally correlates to the degree of object deformation. By converting silhouettes to binary images, we show that the aspect space provides a unified framework for several popular shape descriptors including the shape context [5], the inner-distance shape context [22], and the bending invariant shape signature [10].

We apply the aspect space framework for shape retrieval tasks that is an important problem in computer vision. Taking advantage of the aspect weight’s ability of controlling the degree of deformation, we propose a new shape descriptor named *aspect shape context* (ASC), which extends the inner-distance shape context (IDSC) [22] by replacing the inner-distance with geodesic distances in the aspect space. In fact, both the shape context (SC) [5] and inner-distance shape context are special cases of ASC, corresponding to a zero aspect value and a sufficiently large aspect value, respectively.

<sup>1</sup> The “aspect” in our work comes from “aspect ratio”, which should not be confused with that used in the “aspect” graph.



**Fig. 2.** The top 10 nearest neighbors obtained by IDSC (the first row) and ASC (the second row) for the query of a pie-shaped device

When comparing two shapes, the distance between them is chosen as the minimum distance over different aspect weights. This way, ASC successfully balances the shape deformation and discriminability. We tested the proposed approach on a standard shape retrieval experiment with the MPEG7-CE-1 shape database. ASC demonstrates promising retrieval results and improves the previous approach. For example, in Fig. 2, one comparison between IDSC and ASC is given. The query shape is a pie-shape device. It is obvious that the retrieval results of IDSC are very different from the query shape demonstrating that articulation invariance becomes an over-killing factor. The results of the second row clearly show that ASC can balance between deformation robustness and discriminability.

In summary, we make three main contributions in this paper. First, we propose the aspect space framework, which provides a unified image and shape representation that allows controlling the degree of deformation. Second, with the framework, we designed a new shape descriptor that automatically balances the deformation and discriminability for shape matching. The descriptor demonstrates excellent performance for shape matching tasks on two benchmark shape datasets, the MPEG7 shape database [17] and the Tari 100 database [4]. Third, the novel shape descriptor can describe the relation between shapes very well, not only in the same class, but also in different classes. Thus, it provides an excellent input for diffusion based, context-sensitive shape similarity, which reaches the best ever bull’s eye score of 95.96% on the MPEG7 shape database. This increases by 20% the best bull’s eye scores on this dataset [17] published first time in CVPR 2000, which clearly demonstrates the progress in the field achieved in the last 10 years.

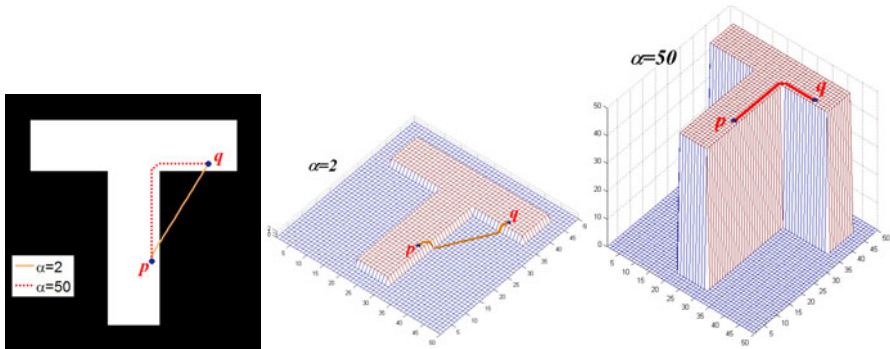
The rest of the paper is organized as follow. § 2 describes the aspect space framework and its relation to existing object and shape representations. Then, we describe the new shape descriptor, aspect shape context, in § 3. To compute a context-sensitive shape distance, we utilize the locally constrained dissuasion process described in § 4. Experimental results are reported and analyzed in § 5. Finally, § 6 concludes the paper.

## 2 Aspect Space

### 2.1 Aspect Space and Geodesics

Given an image  $I : \Lambda \rightarrow [0, 1]$ , where  $\Lambda \subset \mathbb{R}^2$  is the image domain, we define its induced *aspect space*  $\mathcal{A}(I; \alpha)$  of  $I$  as

$$\mathcal{A}(I; \alpha) = (x, y, \alpha I(x, y)), (x, y) \in \Lambda, \alpha \in [0, \infty). \quad (1)$$



**Fig. 3.** Aspect spaces and geodesics on embedding surfaces. Left: a “T” shape and two geodesics between points  $p, q$  for  $\alpha = 2$  and  $50$  respectively. Middle: the aspect space and the geodesic for  $\alpha = 2$ . Right: the aspect space and the geodesic for  $\alpha = 50$ .

Note that the embedding is equivalent to the one used in [21], i.e.,

$$(1 - \alpha)x, (1 - \alpha)y, \alpha I(x, y), \alpha \in [0, 1],$$

but the new definition is practically more convenient and numerically stabler. Also note that the embedding and the following derivations are not restricted to 2D.

A natural way to study the embedding in  $\mathcal{A}(I; \alpha)$  is to investigate the curves especially geodesics on it. Following [21], for a given parameterized curve

$$\gamma(t) = (x(t), y(t), \alpha I(x, y)), \quad t \in [0, 1]$$

on  $\mathcal{A}(I; \alpha)$  with a fixed  $\alpha$ , its curve length is  $l(\gamma) = \int_0^1 \left( \frac{\partial x}{\partial t}^2 + \frac{\partial y}{\partial t}^2 + \alpha^2 \frac{\partial I^2}{\partial t} \right)^{1/2} dt$ . Fig. 3 shows an example of embedding a “T” shape and the geodesics between two points  $p, q$  on the embedded surfaces.

It has been shown in [21] that geodesic distances become invariant to general deformation (i.e., homeomorphism) when  $\alpha$  approaches  $\infty$ , in which case the intensity term dominates  $l(\gamma)$ . On the other extreme, when  $\alpha$  equals to zero, geodesic distances degenerate to the Euclidean distances that are invariant to only rigid deformations. This brings a connection between the *geodesic intensity histogram* [21] and the *spin image* [18]. Motivated by this observation, we will use the aspect space framework for designing robust and discriminative shape descriptor (§ 3).

## 2.2 Relations to Robust Shape Descriptors

We now show how the aspect space relates to several popular shape descriptors including the shape context (SC) [5], inner-distance shape context (IDSC) [22], and the bending invariant signature [10].

The shape context [5] uses the joint distance-orientation distributions of landmarks for shape description. It has been shown to be very effective in many shape comparison

tasks. The inner-distance shape context [22], IDSC extends SC to achieve articulation invariant by using the so called inner-distance, which is defined as the length of shortest paths between landmark points when the paths are restricted within a shape boundary. To relate them in the aspect space framework, we treat a given shape  $S : A \rightarrow [0, 1]$  as a binary mask, i.e.,

$$S(x, y) = \begin{cases} 1 & \text{if } (x, y) \in \text{shape interior} \\ 0 & \text{otherwise} \end{cases} . \quad (2)$$

Consequently, the aspect space of  $S$  is defined as

$$\mathcal{A}(S; \alpha) = (x, y, \alpha S(x, y)) . \quad (3)$$

Given two points  $p, q$  on shape  $S$ , it can be shown that the geodesic distance between them on  $\mathcal{A}(S; \alpha)$  is equivalent to the inner-distance between them, for a large enough  $\alpha$ . An example is shown in Fig. 3 for  $\alpha = 50$ . The reason is, when  $\alpha$  is large enough, the geodesic path from  $p$  to  $q$  along the surface will remain on the plateau formed by the object. This means that the geodesic path, which is also a shortest path, is restricted within the shape boundary. Therefore, it reduces to the inner-distance used in IDSC.

On the other hand, when  $\alpha=0$ , the geodesic distance on the surface degenerates to the Euclidean distance, as illustrated for  $\alpha=2$  in Fig. 3. This is the distance used in SC.

The bending invariant signature uses geodesic distances along object surfaces. Let a 3D volume  $V : A' \subset \mathbb{R}^3 \rightarrow [0, 1]$  has boundary  $\partial(V)$ , we first build a 3D volume mask as following

$$V(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x} \in \partial(V) \\ 0 & \text{otherwise} \end{cases} . \quad (4)$$

Then, with a large enough aspect weight, the geodesic path on  $\mathcal{A}(V; \alpha)$  is restricted on the object boundary  $\partial(V)$ , which is used for building the bending invariant signatures [10].

It is also worth mentioning that our work is related the work on shape analysis using the diffusion distances [8,30] and the Gromov-Hausdorff framework [7,6].

### 2.3 Relations to Other Image Representations

Aside from the above representations, the aspect space also closely relates to many other image representations or statistics. For example, when  $\alpha \rightarrow \infty$ , aspect space relates to the *total variation* (TV) [29] in that only intensity values are taken into account. In this case, the geodesic distance along between points becomes, up to a constant ratio, the total variation of the geodesic path connecting them.

Another important motivation to our work is the Beltrami flow framework [34], where an image is embedded as a surface in 3D space and differential properties of the surface are used for low-level image processing. We expect future works along similar directions in the aspect space.

### 3 Balancing Deformability and Discriminability for Shape Comparison

#### 3.1 Aspect Shape Context

We propose to use the aspect space framework for shape analysis. As shown in § 2, for a given shape  $S$  as defined in equation (2), the geodesic distances between landmark points of  $S$  on  $\mathcal{A}(S; \alpha)$  varies from the Euclidean distance to the inner-distance when  $\alpha$  increases from 0 to a large positive value. This property provides a natural way to unify the shape descriptors taking advantages of both distances, namely the *shape context* (SC) [5] and the *inner-distance shape context* (IDSC) [22]. Motivated by this observation, we present the *aspect shape context* (ASC) as below:

**Definition:** Let  $\{p_i\}_{i=1}^n$  be  $n$  silhouette landmark points of  $S$ , its *aspect shape context* (ASC) is defined as a set of  $n$  histograms on its aspect space  $\mathcal{A}(S; \alpha)$  as

$$ASC(S, \alpha) = \{H_i(S, \alpha)\}_{i=1}^n, \quad (5)$$

where  $H_i(S, \alpha) \in \mathbb{R}^{n_g \times n_o}$  is a  $n_g \times n_o$  histogram measuring the joint geodesic distance-orientation distribution of landmarks w.r.t. point  $p_i$ , such that the geodesic distances and orientations are all computed w.r.t.  $\mathcal{A}(S; \alpha)$ . Specifically,  $H_i(S, \alpha)$  captures the joint distribution of geodesic distances and orientations of all other points with respect to point  $p_i$ . In ASC, for the orientation, we use the angle between boundary tangent direction and the direction of the geodesic projected to the 2D image domain, which is an extension of the *inner-angle* defined in [22]. The geodesic distance in our implementation is through the fast marching algorithm [32].

#### 3.2 Comparing Aspect Shape Context

How should we compare two shapes  $S_1, S_2$  given their ASCs? The basic idea is to compare two shapes by adaptively choosing the “best” aspect weight  $\alpha$ . We expect that a small  $\alpha$  (such that ASC works like SC) will be used for rigid shapes while a large  $\alpha$  (such that ASC works like IDSC) for non-rigid ones.

We define the *co-aspect*  $\hat{\alpha}$  as the one that produces minimum shape distances between  $S_1$  and  $S_2$ <sup>2</sup>.

$$\hat{\alpha}(S_1, S_2) = \arg \min_{\alpha} d_H(ASC(S_1, \alpha), ASC(S_2, \alpha)), \quad (6)$$

where  $d_H(\cdot, \cdot)$  computes a distance between two sets of shape context histograms at a fixed  $\alpha$  (e.g., the “shape context” distance used in [5,22]). Consequently, the distance between  $S_1, S_2$  is defined as the histogram distance that minimizes (6), i.e.,

$$d_{ASC}(S_1, S_2) = d_H(ASC(S_1, \hat{\alpha}(S_1, S_2)), ASC(S_2, \hat{\alpha}(S_1, S_2))). \quad (7)$$

Another choice is to integrate the distances over  $\alpha$ ,

$$d'_{ASC}(S_1, S_2) = \int d_H(ASC(S_1, \alpha), ASC(S_2, \alpha)) d\alpha. \quad (8)$$

<sup>2</sup> If such  $\alpha$  is not unique, the smallest one is chosen.

However, we observed that this definition is less effective than that in (7), since the “smoothing” over all  $\alpha$  may wash out the information provided by the right  $\alpha$ .

Here we clarify a potential confusion. For two topologically equivalent shapes  $S_1, S_2$ , one may guess that  $d_{ASC}(S_1, S_2) = 0$  with a very large (theoretically infinite) co-aspect, since they are the same up to a *homeomorphic transformation*. This is fortunately not true, except when  $S_1, S_2$  are related by an articulation. This observation is very important, because it shows that ASC is not biased toward arbitrary deformations while being able to handle articulation.

### 3.3 Discriminative Deformability

To further understand the relation between aspect weights to shape deformation and discriminability, we define the *discriminative deformability* for a set of shapes  $\mathbb{S} = \{S_i\}_{i=1}^{|\mathbb{S}|}$  coming from the same class. The discriminative deformability of  $\mathbb{S}$ , denoted as  $\bar{\alpha}(\mathbb{S})$ , is

$$\bar{\alpha}(\mathbb{S}) = \frac{1}{|\mathbb{S}|(|\mathbb{S}| - 1)} \sum_{1 \leq i < j \leq |\mathbb{S}|} \hat{\alpha}(S_i, S_j). \quad (9)$$

Roughly speaking, the deformability of a class of object is defined as the average aspect that minimizes the distances between image pairs from the class. Such a deformability naturally takes into account the discriminative information. For example, in Fig. 1, we have shown the discriminative deformabilities of four classes of images. For the pie-shaped devices (e.g., the second row of Fig. 1), each single shape seems to have complex structures that can easily lead to articulation. However, by checking objects in the class, we find no articulation at all and the global shapes of these devices remain fairly rigid. Consequently, it has a relatively small discriminative deformability.

## 4 Beyond Pairwise Shape Similarity

Recent work clearly demonstrated that adding context information to direct pairwise shape similarity can substantially improve shape retrieval [40,15,41]. Under context of a given shape we understand here its first  $K$  nearest neighbors. However, these methods [40,15,41] mainly focus on improving the transduction algorithms. We demonstrate that a ‘better’ original distance matrix is also very crucial for the shape retrieval with context information. The word ‘better’ does not necessarily mean a better bull’s eye score. Instead, it means the algorithm can balance the deformation and discriminability, so that the retrieval results should have really similar view to the query. In other words, even if the retrieved results are not from the same shape class, they have perceptual resemblance to the query. Otherwise, totally different objects in the top retrieval results may ruin the graph structure constructed by the shapes, which makes negative impact for the graph transduction or diffusion processes to learn the shape manifold structure. In order to show that ASC is suitable for context information based shape retrieval, we use the Locally Constrained Diffusion Process (LCDP) introduced in [41] to learn the sub-manifold structure for shapes.

In LCDP, a fully connected graph is constructed. The vertices are the data points (shapes) and each edge is labeled with the strength of the connection  $P(i, j) = k(x_i, x_j)$ ,

where  $k$  is a kernel function that is symmetric and positivity preserving. In this paper, given two shapes  $x_i$  and  $x_j$ ,  $k(x_i, x_j)$  is defined by applying a Gaussian to  $d_{ASC}(x_i, x_j)$ . We then normalize row wise the matrix  $(P(x_i, x_j))_{i,j}$  so that the sum of each row is one. We obtain a stochastic matrix  $(P(x_i, x_j))_{i,j}$ , which we denote with the same symbol. Then, the original graph  $G$  is replaced by a  $K$  nearest neighbor ( $KNNs$ ) graph  $G_K$  that has the edge weights defined as follows:  $P_K(i, j) = k(x_i, x_j)$  if  $x_j$  belongs to the  $KNNs$  of  $x_i$  and  $P_K(i, j) = 0$  otherwise. We also row wise normalize  $(P_K(x_i, x_j))_{i,j}$  to a stochastic matrix. It represents one-step transition probabilities  $P_K(x_i, x_j)$  from  $x_i$  to  $x_j$ .

LCDP only considers the paths between the  $KNNs$  of  $x_i$  and the  $KNNs$  of  $x_j$ , which can be viewed as a soft measure of their  $KNNs$ ' compatibility. The probability of transition from node  $x_i$  to  $x_j$  is high if all the the paths between points in  $KNNs(x_i)$  and in  $KNNs(x_j)$  contain high strength. This restriction make the diffusion focus on more meaningful paths, as  $KNNs$  have higher probability to be the same class. This helps to remove the negative impact from far away noises. To reach this goal, LCDP defines an iterative method to update probability of transition  $P_{KK}$

$$P_{KK}^{t+1} = P_K P_{KK}^t (P_K)^T. \quad (10)$$

where  $t$  controls the degree of diffusion,  $P_{KK}^1 = P$  and  $(P_K)^T$  is the transpose of  $P_K$ . After several iterations updating,  $P_{KK}$  can describe the manifold structure of the shapes well and it is used as the final distances for retrieval. LCDP has been proved to be able to learn the sub-manifold structure of shapes very well and obtain excellent retrieval results [41]. With the help of LCDP, bull's eye score on MPEG-7 data set can reach the highest ever 95.24%, which is discussed in details below.

## 5 Experiments

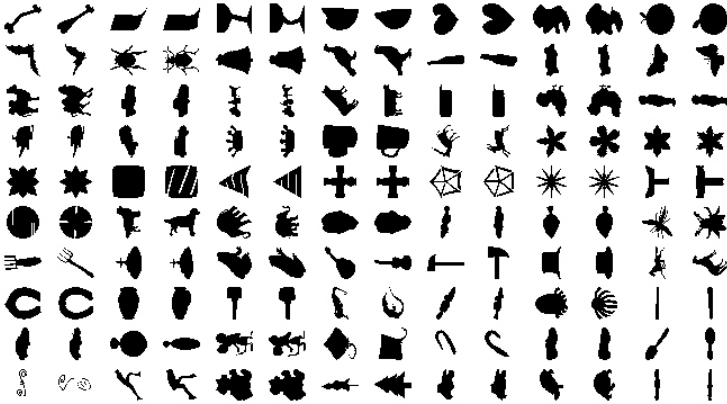
### 5.1 Shape Classification Using MPEG7 Dataset

The proposed framework and shape descriptor is tested for shape classification on a commonly used MPEG7 CE-Shape-1 part B database [17]. The database contains 1400 silhouette images from 70 classes, where each class has 20 different shapes (some examples are shown in Fig. 4). The performance of different solutions is measured by the so called bull's eye score: every shape in the database is submitted as a query and the number of shapes from the same class in the top 40 is counted. The bull's eye score is then defined as the ratio of the number of correct hits to the best possible number of hits (which is  $20 \times 1400$ ).

We tested on the MPEG7 dataset using the distance  $d_{ASC}$  defined in (7), with  $\alpha = 0, 10, \dots, 60, \infty$ , where  $\infty$  indicates the IDSC. We used 300 sample point son each shape and  $8 \times 12$  distance-orientation bins. Table 1 lists the performance along with previous reported results. It is clear that the proposed ASC descriptor achieves excellent classification rate that is comparable to the state-of-the-art. Example retrieval results by ASC and IDSC are shown in Fig. 2.

Table 2 shows a combined result of LCDP and ASC and compares it to other context-sensitive methods. The proposed method can reach the best ever result 95.96% on MPEG7 data set. For using LCDP, we set  $K = 7$  and number of iteration  $t = 19$ .





**Fig. 4.** Typical shape images from the MPEG7 CE-Shape-1, two images from each class

**Table 1.** Retrieval rates (bull’s eye) of different methods on the MPEG-7 data set

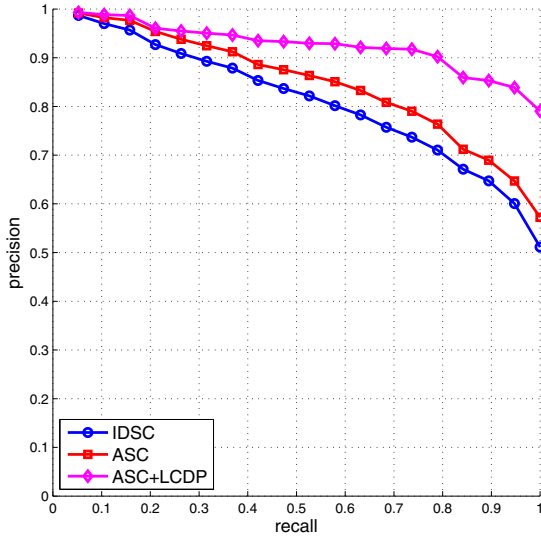
Alg.	CSS [28]	Vis. Parts [16]	Shape Context [5]	Multiscale Rep. [1]	IDSC [22]	Hier. Procrustes [26]
Score	75.44%	76.45%	76.51%	84.93%	85.40%	86.35%
Alg.	IDSC+ EMD- $L_1$ [23]	Triangle Area [2]	Shape Tree [11]	ASC (proposed)	Layered Graph [19]	Contour Flexibility [38]
Score	86.56%	87.23%	87.70%	88.30%	88.75%	89.31%

**Table 2.** Retrieval rates (bull’s eye) of different context shape retrieval methods on the MPEG-7 data set

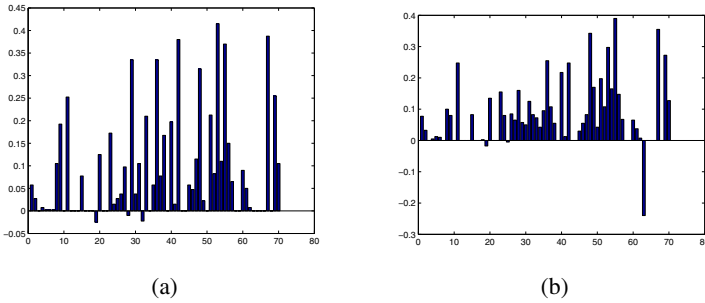
Alg.	IDSC + LP [40]	IDSC + LCDP [41]	IDSC + LCDP+gp [41]	IDSC + Mutual graph [15]	proposed approach + LCDP
Score	91.00%	92.36%	93.32%	93.40%	95.96%

We also show the precision/recall curves of IDSC, the proposed ASC, and ASC with LCDP in Fig. 5. We follow [4] for the precision/recall curves. Precision can be seen as a measure of exactness or fidelity, whereas recall is a measure of completeness, which is more informative for retrieval task. It is clear that ASC has better performance than IDSC not only in the bull’s eye score, but also in the precision/recall curve. Moreover, the LCDP improves the performance a lot, which demonstrates ASC provides a very good input to context-sensitive shape retrieval.

In order to further demonstrate the reason why the proposed approach with context information can obtain much better results, we show the bull’s eye score increment for each class after LCDP in Fig. 6(a). Compared to Fig. 6(b), which is taken from [41], none of the class bull’s eye score has obvious drop after LCDP. This demonstrates the advantage of ASC compared to IDSC. The class which has the most obvious drop in [41] is the class spoon. In Fig. 7, we compare the retrieval results on an example from



**Fig. 5.** Precision/Recall curves of IDSC, ASC and ASC+LCDP on the MPEG7 CE-Shape-1 database

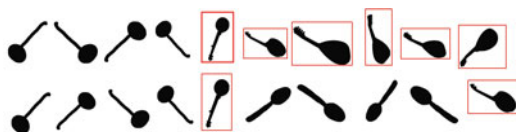


**Fig. 6.** (a) The percent gain in bull's eye retrieval rates for each class by the proposed approach. (b) The percent gain in bull's eye retrieval rates for each class in [41].

class spoon between IDSC+LCDP [41] (the first row) and ASC+LCDP (the second row). The shapes with red rectangles mean wrong retrieval results. They come from the class guitar. It is clear that the ASC+LCDP yields better results than IDSC+LCDP. Only guitars with shape very similar to the query spoon remained.

### 5.2 Shape Classification Using Tari 1000 Dataset

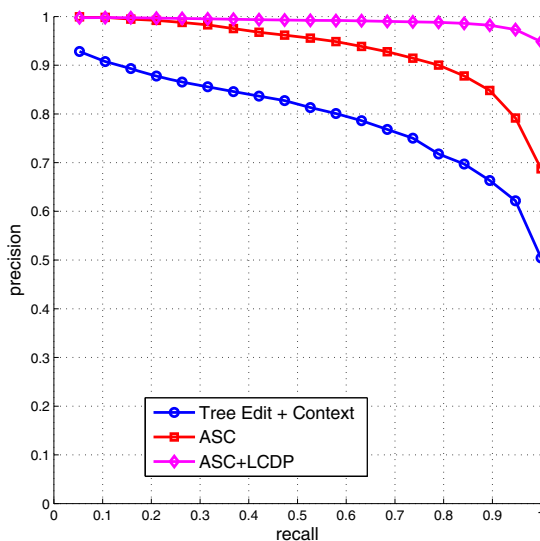
In order to demonstrate the proposed approach can handle articulated shapes, we also test it on recently constructed Tari 1000 data set [4] (see Figure 8), which is designed to have large intra-class shape deformation. It is formed by extending the 180 data set



**Fig. 7.** The retrievals of IDSC+LCDP [41] (the first row) and ASC+LCDP (the second row). The shapes with red rectangle are the wrong results.



**Fig. 8.** Typical shape images from the Tari 1000 dataset, two images from each class



**Fig. 9.** Precision/Recall curves on Tari 1000 dataset. The result of “Tree Edit+Context” is from [4].

in [3] with shapes collected from various sources including [16,31]. It consists of 1000 shapes with 50 class, each class contains 20 shapes. Many of the shapes are articulated and have large deformation. On this dataset, Baseski et al. [4] have demonstrated an effective shape matching algorithm based on using skeletal trees.

For the experiments on Tari 1000, we use exactly the same parameters as on MPEG7 dataset. For evaluation, however, the authors in [4] did not report the bull’s eye score for the whole data set. Instead, they provided the precision/recall curve. We report both

of them. ASC can reach 95.44%, and with the help of LCDP, the bull's eye can go to nearly perfect 99.79%. For using LCDP, we set  $K = 10$  and number of iteration  $t = 3$ . The precision/recall curves are shown in Fig. 9.

## 6 Conclusion and Future Work

In this paper we introduce a general framework, aspect space, to simultaneously address the deformation and discriminatingly problem. The framework naturally models the degree of deformation that is allowed in a representation by adjusting a embedding parameter named aspect weight. In this framework, we designed a new shape descriptor named aspect shape context, which improves the original shape context and inner-distance shape context in a standard shape retrieval benchmark testing, using the MPEG7 CE-Shape-1 database. It also provides a very good input for context-sensitive shape retrieval.

We also show that the framework is closely related to many existing image and shape representations. These attractive properties indicate its potential for many interesting future works. We are especially interested in the study of designing representations that address simultaneously the illumination and shape variations. Another topic we would like to pursue is the shape descriptors without knowing silhouettes. One practical issue with the proposed shape descriptor is the computational cost. In addition to seeking faster computational solution, we will also investigate supervised technology to learn properties for each object class.

## Acknowledgments

We thank anonymous reviewers for helpful comments. The work is supported in part by the NSF Grants IIS-0916624, IIS-0812118, BCS-0924164, and the AFOSR Grant FA9550-09-1-0207.

## References

1. Adamek, T., O'Connor, N.: A multiscale representation method for nonrigid shapes with a single closed contour. *IEEE Tran. Cir. & Sys. for Video Technology* 14(5), 742–753 (2004)
2. Alajlan, N., Kamel, M., Freeman, G.: Geometry-based image retrieval in binary image databases. *IEEE Trans. Pattern Anal. Mach. Intell.* 30(6), 1003–1013 (2008)
3. Aslan, C., Erdem, A., Erdem, E., Tari, S.: Disconnected skeleton: Shape at its absolute scale. *IEEE Trans. Pattern Anal. Mach. Intell.* 30(12), 2188–2203 (2008)
4. Baseski, E., Erdem, A., Tari, S.: Dissimilarity between two skeletal trees in a context. *Pattern Recognition* 42(3), 370–385 (2009)
5. Belongie, S., Malik, J., Puzicha, J.: Shape matching and object recognition using shape contexts. *IEEE Trans. Pattern Anal. Mach. Intell.* 24(4), 509–522 (2002)
6. Bronstein, A.M., Bronstein, M.M., Kimmel, R.: Generalized multidimensional scaling: a framework for isometry-invariant partial surface matching. *Proc. Nat. Acad. Sci.* 103(5), 1168–1172 (2006)
7. Bronstein, A.M., Bronstein, M.M., Bruckstein, A.M., Kimmel, R.: Analysis of two-dimensional non-rigid shapes. *Int. Journal of Computer Vision* 78(1), 67–88 (2008)

8. Bronstein, A.M., Bronstein, M.M., Kimmel, R.: A Gromov-Hausdorff framework with diffusion geometry for topologically-robust non-rigid shape matching. *Int. Journal of Computer Vision* 89(2–3), 266–286 (2010)
9. Cheng, H., Liu, Z., Zheng, N., Yang, J.: A deformable local image descriptor. In: *CVPR*, pp. 1–8 (2008)
10. Elad, A., Kimmel, R.: On bending invariant signatures for surfaces. *IEEE Trans. Pattern Anal. Mach. Intell.* 25(10), 1285–1295 (2003)
11. Felzenszwalb, P., Schwartz, J.: Hierarchical matching of deformable shapes. In: *CVPR*, pp. 1–8 (2007)
12. Hu, M.: Visual pattern recognition by moment invariants. *IRE Tran. Information Theory* 8, 179–187 (1962)
13. Johnson, A.E., Hebert, M.: Using spin images for efficient object recognition in cluttered 3D scenes. *IEEE Trans. Pattern Anal. Mach. Intell.* 21(5), 433–449 (1999)
14. Kadir, T., Zisserman, A., Brady, M.: An affine invariant salient region detector. In: Pajdla, T., Matas, J.(G.) (eds.) *ECCV 2004. LNCS*, vol. 3021, pp. 228–241. Springer, Heidelberg (2004)
15. Kotschieder, P., Donoser, M., Bischof, H.: Beyond pairwise shape similarity analysis. In: Zha, H., Taniguchi, R.-i., Maybank, S. (eds.) *Computer Vision – ACCV 2009. LNCS*, vol. 5996, pp. 655–666. Springer, Heidelberg (2009)
16. Latecki, L.J., Lakämper, R.: Shape similarity measure based on correspondence of visual parts. *IEEE Trans. Pattern Anal. Mach. Intell.* 22(10), 1185–1190 (2000)
17. Latecki, L.J., Lakamper, R., Eckhardt, U.: Shape descriptors for non-rigid shapes with a single closed contour. In: *CVPR*, vol. 1, pp. 424–429 (2000)
18. Lazebnik, S., Schmid, C., Ponce, J.: A sparse texture representation using affine-invariant regions. *IEEE Trans. Pattern Anal. Mach. Intell.* 27(8), 1265–1278 (2005)
19. Lin, L., Zeng, K., Liu, X.B., Zhu, S.C.: Layered Graph Matching by Composite Cluster Sampling with Collaborative and Competitive Interactions. In: *CVPR* (2009)
20. Lindeberg, T.: Feature detection with automatic scale selection. *Int. Journal of Computer Vision* 30(2), 79–116 (1998)
21. Ling, H., Jacobs, D.W.: Deformation invariant image matching. In: *ICCV*, pp. 1466–1473 (2005)
22. H. Ling and D. W. Jacobs. Shape classification using the inner-distance. *IEEE Trans. Pattern Anal. Mach. Intell.*, 29(2):286–299, 2007.
23. Ling, H., Okada, K.: EMD-L1: An Efficient and Robust Algorithm for Comparing Histogram-Based Descriptors. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) *ECCV 2006. LNCS*, vol. 3953, pp. 330–343. Springer, Heidelberg (2006)
24. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *Int. Journal of Computer Vision* 60(2), 91–110 (2004)
25. Matas, J., Chum, O., Urban, M., Pajdla, T.: Robust wide baseline stereo from maximally stable extremal regions. In: *BMVC* (2002)
26. McNeill, G., Vijayakumar, S.: Hierarchical procrustes matching for shape retrieval. In: *CVPR*, pp. 885–894 (2006)
27. Mikolajczyk, K., Schmid, C.: Scale & affine invariant interest point detectors. *Int. Journal of Computer Vision* 60(1), 63–86 (2004)
28. Mokhtarian, F., Abbasi, S., Kittler, J.: Efficient and robust retrieval by shape content through curvature scale space. In: *Wkshp on Image DataBases and MultiMedia Search* (1996)
29. Rudin, L., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. *Physica D* 60(1–4), 259–268 (1992)
30. Rustamov, R., Lipman, Y., Funkhouser, T.: Interior Distance Using Barycentric Coordinates. In: *Symposium on Geometry Processing* (2009)
31. Sebastian, T.B., Klein, P.N., Kimia, B.B.: Recognition of shapes by editing their shock graphs. *IEEE Trans. Pattern Anal. Mach. Intell.* 26(5), 550–571 (2004)

32. Sethian, J.A.: A fast marching level set method for monotonically advancing fronts. *Proc. Nat. Acad. Sci.* 93(4), 1591–1595 (1996)
33. Sivic, J., Zisserman, A.: Video Google: Efficient visual search of videos. In: Ponce, J., Hebert, M., Schmid, C., Zisserman, A. (eds.) *Toward Category-Level Object Recognition*, pp. 127–144 (2006)
34. Sochen, N., Kimmel, R., Malladi, R.: A general framework for low level vision. *IEEE Trans. Image Processing* 7(3), 310–318 (1998)
35. Thompson, D.W.: *On Growth and Form*. Cambridge University Press, Cambridge (1917)
36. Torralba, A., Murphy, K., Freeman, W., Rubin, M.: Context-based vision system for place and object recognition. In: *ICCV*, pp. 273–280 (2003)
37. Vedaldi, A., Soatto, S.: Features for recognition: Viewpoint invariance for non-planar scenes. In: *ICCV*, vol. 2, pp. 1474–1481 (2005)
38. Xu, C., Liu, J., Tang, X.: 2d shape matching by contour flexibility. *IEEE Trans. Pattern Anal. Mach. Intell.* 31(1), 180–186 (2009)
39. Xu, Y., Ji, H., Fermüller, C.: A projective invariant for textures. In: *CVPR*, pp. 1932–1939 (2006)
40. Yang, X., Bai, X., Latecki, L.J., Tu, Z.: Improving shape retrieval by learning graph transduction. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) *ECCV 2008, Part IV. LNCS*, vol. 5305, pp. 788–801. Springer, Heidelberg (2008)
41. Yang, X., Köknar-Tezel, S., Latecki, L.J.: Locally constrained diffusion process on locally densified distance spaces with applications to shape retrieval. In: *CVPR* (2009)
42. Zhang, J., Marszalek, M., Lazebnik, S., Schmid, C.: Local features and kernels for classification of texture and object categories: A comprehensive study. *Int. Journal of Computer Vision* 73(2), 213–238 (2007)