Locally Constrained Diffusion Process on Locally Densified Distance Spaces with Applications to Shape Retrieval

Xingwei Yang, Suzan Köknar-Tezel, and Longin Jan Latecki
Temple University
Department of Computer and Information Sciences
Philadelphia, PA 19094, USA
{xingwei, tezel, latecki}@temple.edu

Abstract

*The matching and retrieval of 2D shapes is an important challenge in computer vision. A large number of shape similarity approaches have been developed, with the main focus being the comparison or matching of pairs of shapes. In these approaches, other shapes do not influence the similarity measure of a given pair of shapes. In the proposed approach, other shapes do influence the similarity measure of each pair of shapes, and we show that this influence is beneficial even in the unsupervised setting (without any prior knowledge of shape classes). The influence of other shapes is propagated as a diffusion process on a graph formed by a given set of shapes. However, the classical diffusion process does not perform well in shape space for two reasons: it is unstable in the presence of noise and the underlying local geometry is sparse. We introduce a locally constrained diffusion process which is more stable even if noise is present, and we densify the shape space by adding synthetic points we call 'ghost points'. We present experimental results that demonstrate very significant improvements over state-of-the-art shape matching algorithms. On the MPEG-7 data set, we obtained a bull's-eye retrieval score of **93.32%**, which is the highest score ever reported in the literature.*

1. Introduction

Shape is one of the most important features of an image and it plays a key role in human perception. Human beings tend to perceive scenes as being composed of individual objects which can be best identified by their shape. Furthermore, the shape of an object is simple for a user to describe, either by giving an example or by sketching. However, using shape information to recognize objects has proven to be a difficult task for computer vision systems. Many approaches have been introduced to better describe shape features and to improve shape similarity measures. Although many of these approaches can increase the accuracy of classifying and recognizing shapes, none of them can solve the following problem. Suppose there is a space describing shapes. Since differences between shapes in the same class can be very large and differences between shapes in different classes can be very small, no pairwise shape comparison can describe shape dissimilarity correctly. Therefore, the distance between two shapes can be correctly described only if it is considered in the context of other shapes similar to them, which is the motivating idea of the proposed approach.

In our approach, the influence of other shapes is propagated as a diffusion process on a graph formed by a given set of shapes. However, as the shape space is sparse (see Sec. 2), in some cases the diffusion process can not propagate properly. It is obvious that adding more data points to the shape space would make the estimation of the data manifold more accurate. In other words, if the shape space is properly densified, a diffusion process is able to better reveal its underlying manifold structure. We propose a novel method to add synthetic data points to distance spaces that are not Euclidean spaces. We introduce synthetic points with correct distances to the existing points. To the best of our knowledge, this is the first time researchers try to solve the problem of densifying non Euclidean data manifolds, and as our experimental results illustrate, the diffusion process performs significantly better on the desified manifolds.

There have been several proposed approaches to add syn-

(a) (b) (c)

Figure 1. (c) The mean horse computed by averaging corresponding sample contour points of the aligned shapes in (a) and (b).

978-1-4244-3991-1/09/$25.00 ©2009 IEEE

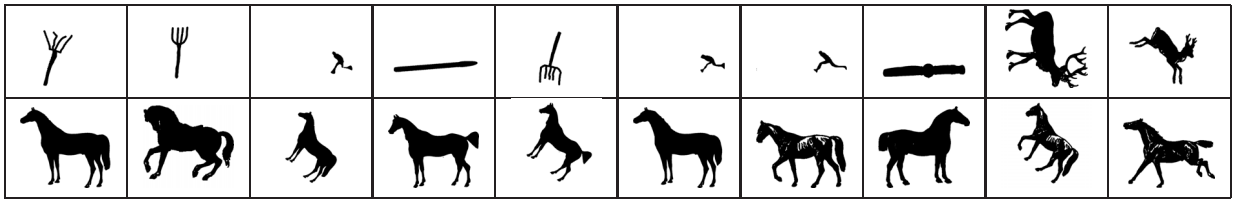

Figure 2. First row: the retrieval results of the mean horse from Fig. 1(c). Second row: the retrieval results of the ghost horse created by the averaging in distance space of the two shapes in Figs. 1(a) and (b).

thetic examples in the Euclidean space. They goal is also different, since they try to solve the problem of balancing the number of examples in different classes, specifically over-sampling minority classes. For example, the SMOTE (Synthetic Minority Over-Sampling Technique) [5] algorithm and its variations [1, 16] have been found to be successful in classification problems but their methods are not suitable for shapes, since they work only in Euclidean space. In those methods, synthetic points are added as a weighted average of the Euclidean coordinates of two existing points. However, the Euclidean distance is known to be unsuitable as a shape dissimilarity measure even if shapes are represented as vectors of their contour sample points. For example, the horse in Fig. 1(c) is computed as the average of the Euclidean coordinates of the two horses in Fig. 1(a) and Fig. 1(b). The Euclidean coordinates were obtained as sequences of 2D coordinates of 100 aligned contour sample points. Although the feature points of both horses correspond, it is difficult to recognize the shape in Fig. 1(c) as a horse. To demonstrate the problem, we submitted the mean horse as a query to the MPEG-7 CE-Shape-1 part B data set [12]. The top ten retrieval results are shown in the first row of Fig. 2, ordered from left to right. Obviously none of the retrieval results is correct, but they are similar to the mean horse. For example, the tines of the forks are similar to the 'legs' of the average horse. The second row of Fig. 2 shows the retrieval results of the 'synthetic horse' generated by the proposed approach, which are all correct. We used the Inner Distance Shape Context (IDSC) [13] as the shape distance in both cases.

The second main idea in this paper is to replace the original diffusion process with a locally constrained diffusion process. As we will demonstrate in Section 4, it is significantly more robust to noise than the original diffusion process.

The rest of the paper is organized as follows. In Section 2, we introduce metric embedding and the construction of synthetic points. In Section 3, the classical diffusion process is introduced. In Section 4 we describe the proposed locally constrained diffusion process in detail. The experimental results are shown in Section 5. In Section 6 we discuss the relation between the proposed approach and other methods.

2. Ghost points and metric embedding

In this paper, we view shape space as a set X and a distance function $\rho : X \times X \rightarrow \Re$, where $\Re$ denotes real numbers. We require only that $\rho(x, y) \geq 0$ for all $(x, y) \in X \times X$ and $\rho(x, y) = 0$ if $x = y$. Clearly, we would like ρ to be as close as possible to a metric, but this is not always possible, since there are clear arguments from human visual perception that the distance between shapes does not always satisfy the triangle inequality and the symmetry conditions. In any case, for theoretical reasons, we assume in Section 2.1 that ρ is a metric. However, as we will demonstrate in our experimental results, this assumption is not necessary for practical applications.

By embedding a metric space into a Euclidean space, we add new synthetic points to the shape space. We can do this so that the new points have correct distances to all existing points. Thus, the new points augment the shape space X but we cannot visualize them, which is the reason we call them ghost points.

2.1. Definition of ghost points

The goal of metric embedding is to embed a metric space into a Euclidean space so that the distances between points are preserved. A distance preserving mapping between two metric spaces is called an isometry.

It is known that not every four point metric space can be isometrically embedded into a Euclidean space $\Re^k$, e.g., see [9]. However, every three point metric space can be isometrically embedded into the plane $\Re^2$. Let (Δ, ρ), where $\Delta = \{x, a, b\} \subseteq X$, be a metric space with three distinct points. Then it is easy to map Δ to the vertices of a triangle on the plane. Let $h : \Delta \rightarrow \Re^2$ be the isometric embedding, which means that for any two points $y, z \in \Delta$, $\rho(y, z)^2 = ||y - z||^2$, where $|| \cdot ||$ is the standard L_2 norm that induces the Euclidean distance on the plane.

Let $\mu(a, b)$ denote the mean of two points a, b. If $a, b \in \Re^2$, then we have the usual formula $\mu(a, b) = \frac{1}{2}(a + b)$ (see Fig. 3, where $e = \mu(a, b)$).

Our first key contribution is the definition of $\mu(a, b)$ for any two points a, b in a metric space X. To define $\mu(a, b)$ in a metric space X, we need to specify $\rho(x, \mu(a, b))$ for every $x \in X$. We first isometrically embed the three point metric subspace $\Delta = \{x, a, b\} \subseteq X$ into the plane $\Re^2$ by h. We

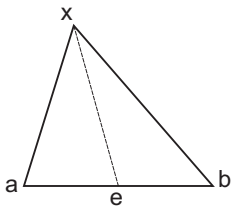

Figure 3. The construction of $\rho(x, e)$ for $e = \mu(a, b)$.

define $\mu(a, b) = h^{-1}(\frac{1}{2}(h(a) + h(b))$. Since $h(\Delta)$ defines vertices of a triangle on the plane, we can easily derive that

$$||h(x) - \frac{h(a) + h(b)}{2}||^2 = \frac{||h(x) - h(a)||^2}{2} + \frac{||h(x) - h(b)||^2}{2} - \frac{||h(a) - h(b)||^2}{4}$$

Since h is an isometry and $\mu(a, b) = h^{-1}(\frac{1}{2}(h(a) + h(b))$, we obtain (see Fig. 3)

$$\rho(x, \mu(a, b))^2 = \frac{1}{2}\rho(x, a)^2 + \frac{1}{2}\rho(x, b)^2 - \frac{1}{4}\rho(a, b)^2 \quad (1)$$

Consequently, Eq. 1 defines the distance of every point $x \in X$ to the new point $\mu(a, b)$, which we call the mean of a and b. By computing the distances of $\mu(a, b)$ to all points in X, we define a new point $\mu(a, b)$, and the augmented set $X' = X \cup \{\mu(a, b)\}$ is also a distance space. We stress that to add a new point $\mu(a, b)$ to X we do not need to compute the embedding h. We use h only to derive Eq. 1. Moreover, since the embedding h is an isometry, Eq. 1 defines correct distances from $\mu(a, b)$ to all points in X. This fact is illustrated in the second row of Fig. 2, where we see the sorted 10 closest shapes shapes to $\mu(a, b)$ with a and b being the two shapes in Figs. 1(a) and (b). As shown in the first row of Fig. 2, simple averaging in Euclidean space may not produce correct distances, since the Euclidean distance is not adequate for shape similarity. We used Inner Distance Shape Context (IDSC) [13] as our shape distance function ρ in this example.

If the space X is finite, i.e., $X = \{x_1, \ldots, x_n\}$, then the distance function $\rho : X \times X \to \Re_{\geq 0}$ is represented by a square matrix $M_\rho(X)$. Each row of the square distance matrix $M_\rho(X)$ is the distance of one shape x to all shapes in the data set, i.e., for all $y \in X$, $M_\rho(x, y) = \rho(x, y)$. The matrix for $X \cup \{\mu(a, b)\}$ is obtained by simply adding one row and one column to $M_\rho(X)$, with each entry computed using Eq. 1.

2.2. Strategies for adding ghost points

There are many possible strategies for adding ghost points. Our strategy is very simple. We add to $X = \{x_1, \ldots, x_n\}$ a point $\mu(x, NN_1(x))$ for each $x \in X$, where $NN_1(x)$ is the first nearest neighbor of x different from x, i.e., $NN_1(x) = argmin_{y \in S}(\rho(x, y))$ for $S = X \setminus \{x\}$. However, if $y = NN_1(x)$ and $x = NN_1(y)$, this strategy would insert the same ghost point twice. Therefore, we need to take care to not add duplicate ghost points. After adding the ghost points, we obtain a new shape space X'. As we will show in the experimental results, the augmented space X' densifies the original shape space X in such a way as to make the estimation of the data manifold more accurate.

This densification of space X is performed in the unsupervised setting, since we do not assume any knowledge of the class labels of points in X. To augment X in a supervised setting, we add ghost points to X as described above with the one exception that the first nearest neighbors are computed within the class of a given point, i.e., instead of $S = X \setminus \{x\}$, we define $S = \{y \in X | \ class(y) = class(x)$ and $y \neq x\}$

We use the augmented shape space X' in the diffusion process (Sec. 3) to influence the shape similarity measures between the query shape and all other shapes. After the diffusion process is run, we exclude the ghost points and calculate our retrieval and classification rates based on only the original shape data set to allow for a fair comparison to existing methods.

3. Diffusion process

Given a set of data points $X = \{x_1, \ldots, x_n\}$, we consider a fully connected graph $G = (X, E)$. The vertices of G are the data points and each edge E is labeled with the strength of the connection $E(i, j) = k(x_i, x_j)$, 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 the shape distance $\rho(x_i, x_j)$.

From the symmetric graph defined by (X, E), one can construct a reversible Markov chain on X. This is a classic technique in many fields. The degree of each node is defined as

$$D(x_i) = \sum_{j=1}^{n} k(x_i, x_j)$$

and the transition probability is defined as

$$P(x_i, x_j) = \frac{k(x_i, x_j)}{D(x_i)}.$$

It is obvious that the transition matrix P inherits the positivity-preserving property, but it is no longer symmetric. However, we have gained a conservation property:

$$\sum_{j=1}^{n} P(x_i, x_j) = 1$$

From a data analysis point of view, the reason for studying this diffusion process is that the matrix P contains geometric information about the data set X. Indeed, the transitions that it defines directly reflect the local geometry defined by the immediate neighbors of each node in the graph of the data. In other words, $P(x_i, x_j)$ represents the probability of transition in one time step from node x_i to node x_j and it is proportional to the edge-weight $k(x_i, x_j)$. For $t \geq 0$, the probability of transition from x_i to x_j in t time steps is given by $P^t(x_i, x_j)$, which is the tth power P^t of P. One of the main ideas of the diffusion framework is that the chain running forward in time, or equivalently, taking larger powers of P, allows us to integrate the local geometry and therefore reveals relevant geometric structures of X at different scales, where t plays the role of a scale parameter. In [6], the data points can be embedded into Euclidean space by diffusion maps (DM), which can then reorganize the data points according to their geometric relation as revealed by the diffusion process.

Ideally, diffusion coordinates generated by diffusion maps should reveal the intrinsic geometric structure of the underlying data manifold. However, as we illustrate by the following example, the diffusion process is still sensitive to noise. Our example illustrates that the diffusion process may fail to capture the correct topology if the actual topology of the data manifold is changed because of noise or outliers. Since noise and outliers can influence the distribution of data points, low density areas may become high density areas or vice versa, which will make the transition probability of the diffusion process incorrect. In Fig. 4, the samples are taken from a spiral as a function of arc length l with added Gaussian noise and a noise 'bridge' between inner and outer samples. Since the underlying manifold has a 1D structure, we would expect the diffusion process to be able to recover it when we use the coordinates of the second most important eigenvector, as described in [18, 23].

In Figs. 4(a) and (c), we plot the coordinates of the second most important eigenvector as a function of arc length (measured as point index). As can be clearly observed in Fig. 4(a), the function from arc length to the second diffusion coordinate is not one-to-one, which means that the intrinsic 1D structure of the spiral has not been recovered by the standard diffusion process. Correspondingly, in Fig. 4(b), the order of points according to their second diffusion coordinate is color coded. Points with similar color have similar second diffusion coordinates. The fact that the 1D structure is not recovered is shown by the yellow colored points that are present in the bottom left as well as in the top right parts of the spiral. As shown in Figs. 4(c) and (d), the proposed locally constrained diffusion process (Sec. 4) is able to recover the 1D structure of the spiral. The graph in (c) does jitter a bit since we approximate the arc length coordinates of the spiral with the point index.

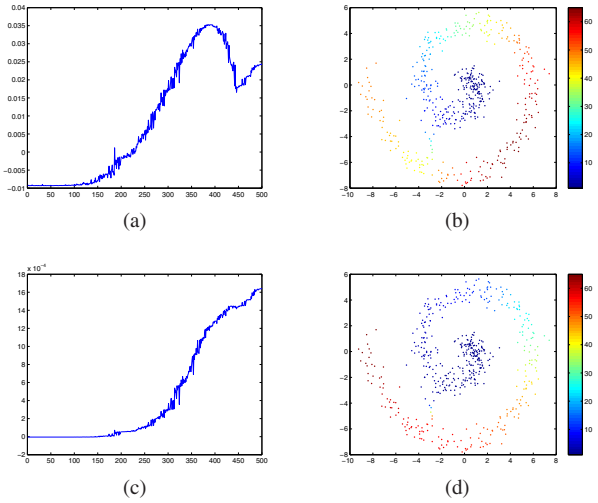

Figure 4. An example comparing the standard diffusion process (DM) to our method (LCDP). (a) is the plot of second most important eigenvector as a function of arc length. (b) shows the points color coded according to their second diffusion coordinate using DM. (c) and (d) show the same plots as (a) and (b) but using LCDP.

4. Locally Constrained Diffusion Process

As the diffusion process can be influenced even by moderate noise and outliers, in order to reduce the effect of noisy data points we introduce in this section a locally constrained diffusion process.

In the classical diffusion process setting, all paths between nodes x_i and x_j are considered when computing the probability of a walk from x_i to x_j. If there are several noisy nodes, the paths passing through these nodes will affect this probability as we demonstrated in Fig. 4.

A solution to this problem is introduced in [22], where a random walk is restricted to the K nearest neighbors of the data points by replacing the original graph G with a K nearest neighbor (KNN) graph G_K that has the edge weights defined as follows: $E_K(i,j) = k(x_i, x_j)$ if x_j belongs to the KNNs of x_i and $E_K(i,j) = 0$ otherwise. Then, the one-step transition probabilities $P_K(x_i, x_j)$ from x_i to x_j are defined

$$P_K(x_i, x_j) = \frac{E_K(i,j)}{\sum_j E_K(i,j)}.$$

Through replacing the P in Section 3 by P_K, the effect of noise is reduced, but the process is still not robust enough to noise. The reason is that the relation between the $KNN(x_i)$ and $KNN(x_j)$ is too hard and too narrow. It counts a data point x_k only if x_k is a KNN of both x_i and x_j. This causes problems if both points x_i and x_j belong to the same dense cluster, in which case they may have no common KNNs although they are very similar. In other words, although x_i and x_j are very similar to each other

and there are many short paths connecting them in graph G, they may have no common neighbor in G_K.

In order to solve this problem, we consider the paths between KNNs of x_i and 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 $KNN(x_i)$ and in $KNN(x_j)$ are short. We define

$$P_{KK}^{t+1}(x_i, x_j) = \sum_{k \in KNN(x_i), l \in KNN(x_j)} P(x_i, x_k) P_{KK}^{t}(x_k, x_l) P(x_l, x_j) \tag{2}$$

Eq. 2 can be viewed as a symmetric version of the approach in [22], and can be expressed as matrix multiplication

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

The embedding results of our proposed approach on the noisy spiral data are shown in Figs. 4(c) and (d). These figures demonstrate that the proposed locally constrained diffusion process (LCDP) is able to recover the intrinsic geometric structure of the spiral.

5. Experimental results

In this section, we demonstrate the validity of our approach for shape retrieval on two standard data sets, MPEG-7 and Swedish Leaf. We compare the Locally Constrained Diffusion Process (LCDP) to three closely related methods: diffusion process based on Locally Appropriate Metric (LAM) [22]; diffusion distances after embedding by Diffusion Maps (DM) [10]; and the Label Propagation (LP) approach in [25]. We show also the positive effect of adding ghost points in both unsupervised and supervised settings.

In all of the following experiments, the σ for the Gaussian Kernel function follows the approach in [25]. The number of K nearest neighbors is 20 for the MPEG-7 data set and 40 for the Swedish Leaf data set. The number of iterations of the diffusion process, t, is set empirically.

5.1. MPEG-7 data set

First we show the experimental results on the MPEG-7 CE-Shape-1 part B data set [12]. MPEG-7 is a standard data set and is widely used to test shape classification and retrieval methods. It contains 1400 binary images divided into 70 shape classes of 20 images each. Every shape in the data set is compared to all other shapes, and the number of shapes from the same class among the 40 most similar shapes is reported. The bull's-eye retrieval rate is the ratio of the total number of shapes from the same class to the highest number possible (which is 1400×20), thus the best possible score is 100%. To show that the proposed approach can improve shape retrieval results on existing shape distance measures, we choose the well-known shape similarity method, Inner Distance Shape Context (IDSC) [13], to compute the pairwise distances between the shapes. The bull's-eye scores of the proposed approaches and the other approaches using IDSC are shown in Table 1, and the retrieval scores (the ratio of the number of correct shapes among the first k shapes for $k = 1, \ldots, 40$) are shown in Fig. 5. The lowest overall retrieval results of Diffusion Maps (DM) illustrate the fact that embedding the shape space into Euclidean space may lead to significant loss of information. This is the only method that performs worse than the original IDSC pairwise distance measure.

Although the accuracy of LAM is higher than IDSC, it is still significantly lower than the proposed LCDP. Even without ghost points, LCDP increases the bull's-eye score to 92.36%, which is better than the highest previously reported bull's-eye score of 91.00% in [25] and demonstrates that our method does reduce the effect of noise and outlier shapes. By adding ghost points in an unsupervised setting, the bull's-eye score reaches 93.32%, the highest ever reported. It is consistent with our assumption that the ghost points densify the data space, which improves the performance of the diffusion process.

In Fig. 6, we report the percentage gain for each of the 70 shape classes in the MPEG-7 data set obtained by LCDP with unsupervised ghost points when compared to IDSC. We observe that the bull's-eye retrieval rate was improved by over 20% on 9 shape classes. This demonstrates the ability of the proposed approach to learn object appearance in the context of other shapes. But as learning involves generalization, there is always a danger of overgeneralization. Yet this graph demonstrates that this danger is very small for the proposed approach since the bull's-eye score of only one class decreases significantly. Furthermore, this decrease in accuracy can be explained by the fact that this class contains shapes of spoons which are very similar to sea-snakes, pencils, and keys in the MPEG-7 data set.

From the graph in Fig. 5, it is clear that the retrieval rate when using the unsupervised ghost points is not always better than the other approaches. For the early nearest neighbors, i.e., when k is small, it is worse than the other methods because in the unsupervised setting we assume that the very local structure of each data point is correct; that is, that the first nearest neighbor of each of the data points should be from the same class as the data point itself. However, since IDSC can not attain 100% accuracy when finding the first nearest neighbors, a few inter-class ghost points will be generated. This reduces the accuracy of the retrieval rates for small k. However, since most of the ghost points generated are intra-class (and this is what we want), the retrieval rates for later k improves significantly, and the bull's-eye score reaches 93.32% for $k = 40$. To solve this problem of

generating inter-class ghost points, we also generate ghost points in a supervised setting. With supervision, only intra-class ghost points are created and this gives us the bull's-eye score of 97.21%. Furthermore, the retrieval curve is always above the curves of the other approaches. Hence we can conclude that the performance gain in the retrieval rates is optimal when the shape space is densified in a supervised setting. We want to stress that this scenario is realistic, since we usually know the class labels of the database objects. Hence the addition of ghost points in a supervised setting can be viewed as a novel supervised learning method for relevance ranking.

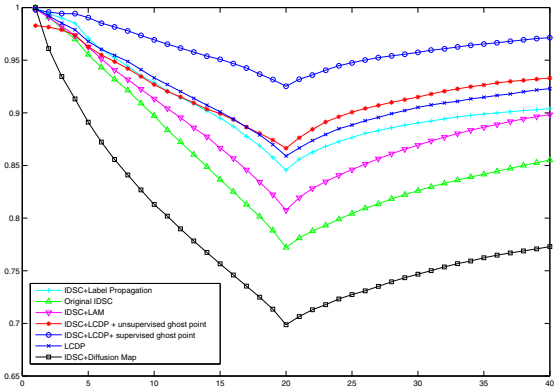

Figure 5. Comparison of our approach to other methods using IDSC.

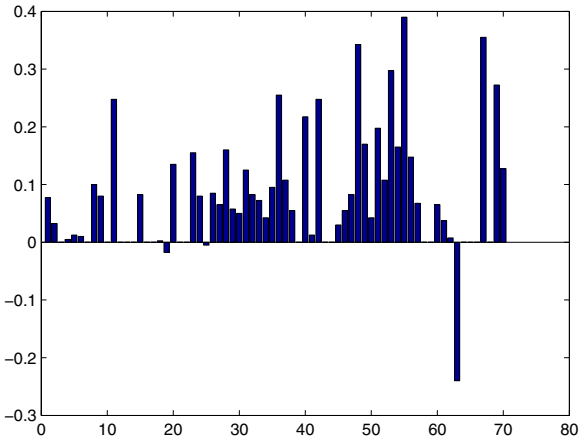

Figure 6. The percent gain in bull's-eye retrieval rates for each of the 70 shape classes of the MPEG-7 data set for IDSC [13]

5.2. Swedish Leaf data set

We also tested our approach on the Swedish leaf data set, which comes from a leaf classification project at Linköping University and the Swedish Museum of Natural History [21]. The data set contains images of leaves from 15 different Swedish tree species, with 75 leaves per species, for a total of 1125 images. Previous work focused on 1-nearest-neighbor (1NN) classification [13, 21]. In this paper, in addition to our results for 1NN classification, we also show the retrieval results as the ratio of correct shapes among the first k shapes for $k = 1, \ldots, 75$. Again, we use IDSC to find the distances between the shapes of the data set and compare our approach to two of the three methods discussed above (there are no reported results on this data set for LP [25]). The results for the different methods are shown in Fig. 7. Once again, the retrieval results for DM are significantly worse than the other approaches. Although LAM's performance is quite good, it is still worse than LCDP. LCDP without ghost points improves the retrieval results significantly with the 1NN classification rate increasing from 94.12%[13] to 98.2%, the highest score in the literature.

Consistent with the results obtained on the MPEG-7 data set, the addition of unsupervised ghost points does not improve the retrieval rates of LCDP for small k, but does improve them for larger k. Adding ghost points in the supervised setting achieves the best performance of all. The 1NN classification rates are 97.6% and 99.3% for unsupervised and supervised ghost points respectively.

The reason for the difference between the results for unsupervised and supervised ghost points is that the data set contains several classes that are very similar to each other and thus some of the ghost points added in the unsupervised setting are inter-class and we have the same problem as we discuss in Sec. 5.1. The addition of ghost points in the supervised setting solves this problem by generating only intra-class ghost points, and the retrieval rate increases significantly.

We can make some conclusions based on the experimental results. First, LCDP performs better than LP and LAM, which is consistent with the discussion in Sec. 4. Second, the effect of adding unsupervised ghost points depends greatly on the accuracy of the original shape similarity measure. If majority of the first nearest neighbors of the data points belong to the same class, adding unsupervised ghost points can achieve an enormous improvement. If this is not the case, adding unsupervised ghost points may actually cause the k nearest neighbor retrieval rate to decline for small k (though there still may be a substantial improvement for large k). Third, the ghost points generated in a supervised setting consistently and significantly improve the retrieval rates for all k. Although it is not fair to compare adding supervised ghost points to unsupervised ghost points in shape retrieval, the excellent performance indicates the application potential of ghost points in other areas of supervised and semi-supervised data mining.

Table 1. Retrieval rates (bull's-eye) of the MPEG-7 data set using Inner Distance Shape Context (IDSC).

Alg.	IDSC [13]	IDSC + LAM	IDSC +DM	IDSC +LP[25]	IDSC +LCDP	IDSC +LCDP +unsupervised GP	IDSC +LCDP +supervised GP
Score	85.40%	89.00%	78.56%	91.00%	**92.36%**	**93.32%**	**97.21%**

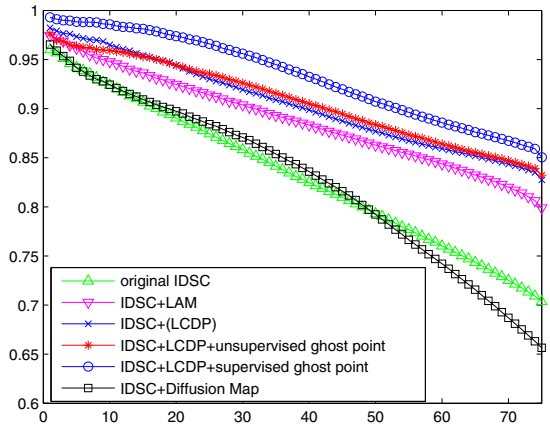

Figure 7. Retrieval curves of Swedish Leaf data set

6. Relation to other approaches

Originally, researchers focused on global features for shape matching and shape retrieval. However, the limitation of these approaches is that the descriptors are sensitive to local changes, which is a crucial property of shapes. To overcome the problems of global descriptors, local features and non-linear methods have been introduced, especially using parts of the objects. In [11], tangent distances are used to represent the difference between parts. Then the relation between sequences of parts is considered when constructing the shape correspondence. In order to capture the local features, Belongie and Malik [4] introduce the 2-dimensional non-linear histogram, Shape Context (SC), to describe the distance and angles between contour points. Since SC cannot solve the problem of matching articulated shapes, Ling and Jacobs [13] modified SC by using the geodesic distance inner shape instead of the Euclidean distance to represent shapes, which is called Inner Distance Shape Context (IDSC). Similar to IDSC, skeleton based approaches [19, 20, 3] can describe the articulated shapes well and implicitly take into account the part information. Recently, a hierarchical segment-based shape matching algorithm has been proposed by McNeill and Vijayakumar [15]. Felzenszwalb and Schwartz [7] decompose shape into segments and organize them by a novel concept, shape tree. In [24], a descriptor called Contour Flexibility is proposed to represent the deformable potential at each point along a contour. Besides, Rodriguez et. al [17] represent 2D shapes by mapping them to an analytic function on the complex plane, which is called Analytic Signature.

In [25], a graph transduction learning approach based on label propagation is introduced. It is the first approach in which the shape similarity of a pair of shapes is computed in the context of other shapes as opposed to considering only pairwise relations between two shapes as is the case in the above approaches. There are two key differences in the proposed approach. First, our diffusion process framework does not require label clamping as is the case in label propagation. Consequently, we need less than 10 iterations while label propagation requires on the order of 10,000 iterations. Second, they key step to improve the performance of a diffusion process is the introduction of ghost points, which densify the manifold structure allowing the diffusion process to better propagate.

In order to find relevant structures in complex geometries for classification and clustering, Markov chain techniques have been combined with graph-based methods. In [22], the L^1 distance between probabilities of transition is used as a metric between data points, and this metric is then employed to induce class labels. Zhou et al [26] assume a metric to arbitrarily convey a probability transition metric (proportional to weights) over the data to cluster it; [27] also assumes a metric from the beginning to arbitrarily claim a probability transition metric based on the original metric which is then used to propagate labels using accumulated correlations; In [2], Azran assumes a metric over the data to induce a probability transition metric over an M-NN graph (where labeled points are absorbing), which is used to produce a probability distribution over the labels of each unlabeled point. While our setting is similar to these approaches, we treat the Markov random walks in a unsupervised setting. In order to make it more suitable for shape space, the transition matrix is focused on k nearest neighbors and point set correspondence is constructed for robustness.

For the metric embedding, one of the most popular dimensionality reduction algorithms is Principal Component Analysis (PCA) [14]. PCA performs dimensionality reduction by projecting the original n-dimensional data onto the $d(\ll n)$-dimensional linear subspace spanned by the leading eigenvectors of the data's covariance matrix. However, in many real world problems, there is no evidence that the data is sampled from a linear subspace. Various researchers

have considered the case then the data lives on or close to a low dimensional sub-manifold from random points lying on this unknown sub-manifold. Along this direction, many subspace learning algorithms have been proposed, such as Locality Preserving Projection [8], Locally Linear Embedding [18], and ISOMAP [23]. Compared to these approaches, the proposed synthetic points directly densify the data manifold in the original distance space instead of the metric embedding. In other words, we use metric embedding to add new synthetic data points in the original distance space. A key feature of our approach is that we can do this without distorting the original distances.

7. Conclusions

In this paper, we introduce the addition of synthetic data points (ghost points) to densify sparse data spaces and the locally constrained diffusion process to reveal the intrinsic relation between shapes. We also describe the relationship between our proposed approach and other closely related methods. Instead of using the direct distance between shapes, our unsupervised approach can capture the topology of the data so that the distance measure between objects is found through the manifold enclosing the data. The experimental results demonstrate a significant increase in the retrieval rates on both toy and real data sets demonstrate the advantage of the proposed approaches on noisy and sparse data sets. The addition of ghost points can also be performed in a supervised setting, which can be viewed as a novel supervised learning method for relevance ranking. The supervised scenario is realistic, since we usually know the class labels of the database objects.

8. Acknowledgments

This work was supported in part by NSF Grant IIS-0812118 and by DOE Grant DE-FG52-06NA27508.

References

[1] R. Akbani, S. Kwek, and N. Japkowicz. Applying support vector machines to imbalanced datasets. In *ECML*, pages 39–50, 2004.

[2] A. Azran. The rendezvous algorithm: Multiclass semi-supervised learning with markov random walks. In *ICML*, 2007.

[3] X. Bai and L. J. Latecki. Path similarity skeleton graph matching. *IEEE Trans. PAMI*, 30:1282–1292, 2008.

[4] S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Trans. PAMI*, 24:705–522, 2002.

[5] N. V. Chawla, K. W. Bowyer, and W. P. Kegelmeyer. Smote: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16:321–357, 2002.

[6] R. Coifman and S. Lafon. Diffusion maps. *Applied and Computational Harmonic Analysis*, 21:5–30, 2006.

[7] P. F. Felzenszwalb and J. Schwartz. Hierarchical matching of deformable shapes. In *CVPR*, 2007.

[8] X. He, S. Yan, Y. Hu, P. Niyogi, and H. J. Zhang. Face recognition using laplacianfaces. *IEEE. PAMI*, 27:328–340, 2005.

[9] M. Jiri. *Lectures on Discrete Geometry*. Springer, 2002.

[10] S. Lafon and A. B. Lee. Diffusion maps and coarse-graining: A unified framework for dimensionality reduction graph partitioning, and data set parameterization. *IEEE PAMI*, 28:1393–1403, 2006.

[11] L. J. Latecki and R. Lakämper. Shape similarity measure based on correspondence of visual parts. *IEEE Trans. PAMI*, 22(10):1185–1190, 2000.

[12] L. J. Latecki, R. Lakämper, and U. Eckhardt. Shape descriptors for non-rigid shapes with a single closed contour. In *CVPR*, pages 424–429, 2000.

[13] H. Ling and D. Jacobs. Shape classification using the inner-distance. *IEEE Trans. PAMI*, 29:286–299, 2007.

[14] K. V. Mardia, J. T. Kent, and J. M. Bibby. *Multivariate Analysis*. Academic Press, 1980.

[15] G. McNeill and S. Vijayakumar. Hierarchical procrustes matching for shape retrieval. In *Proc. CVPR*, 2006.

[16] R. C. Prati and M. C. Monard. A study of the behavior of several methods for balancing machine learning training data. *Sigkdd Explorations*, 6:20–29, 2004.

[17] J. Rodriguez, P. Aguiar, and J. Xavier. Ansig - an analytic signature for permutation-invariant two-dimensional shape representation. In *CVPR*, 2008.

[18] S. Roweis and L. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290:2323–2326, 2000.

[19] T. B. Sebastian, P. N. Klein, and B. Kimia. Recognition of shapes by editing their shock graphs. *IEEE Trans. PAMI*, 25:116–125, 2004.

[20] K. Siddiqi, A. Shokoufandeh, S. J. Dickinson, and S. W. Zucker. Shock graphs and shape matching. *Int. J. of Computer Vision*, 35:13–32, 1999.

[21] O. Söderkvist. Computer vision classification of leaves from swedish trees. *Master's Thesis, Linköping University*, 2001.

[22] M. Szummer and T. Jaakkola. Partially labeled classification with markov random walks. In *NIPS*, 2001.

[23] J. B. Tenenbaum, V. de Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290:2319–2323, 2000.

[24] C. Xu, J. Liu, and X. Tang. 2d shape matching by contour flexibility. *IEEE Trans. Pattern Analysis and Machine Intelligence*, accepted.

[25] X. Yang, X. Bai, L. J. Latecki, and Z. Tu. Improving shape retrieval by learning graph transduction. In *ECCV*, 2008.

[26] D. Zhou and B.Scholkopf. Learning from labeled and unlabeled data using random walks. In *Pattern Recognition, Proceedings of the 26th DAGM Symposium*, 2004.

[27] X. Zhou and C. Li. Text classification by markov random walks with reward. In *DMIN05*, 2005.