Unsupervised Learning of Shape Manifolds

Nasir Rajpoot [†]   Muhammad Arif [⋆]   Abhir Bhalerao [†]

[†] Department of Computer Science, University of Warwick, CV4 7AL, UK
[⋆] Pakistan Institute of Engineering and Applied Sciences, Nilore, Pakistan

email: {nasir,arif,abhir}@dcs.warwick.ac.uk

Abstract

Classical shape analysis methods use principal component analysis to reduce the dimensionality of shape spaces. The basic assumption behind these methods is that the subspace corresponding to the major modes of variation for a particular class of shapes is linearised. This may not necessarily be the case in practice. In this paper, we present a novel method for extraction of the intrinsic parameters of multiple shape classes in an unsupervised manner. The proposed method is based on learning the global structure of shape manifolds using diffusion maps. We demonstrate that the method is effective in separating the members of different shape classes after embedding them into a low-dimensional Euclidean space.

1 Introduction

The shape of an object plays a critical role in its representation and can be used as an important cue for indexing purposes for image and video databases, or for searching of a particular phenomenon (for example, regular-shaped cells) in biomedical images, just to name a few applications. Intuitively, an object's shape can be related to the form or structure of the object. A working definition of shape adopted by most researchers in computer vision is *the quality of an object that is invariant to transformations such as translation, rotation, and scaling and to changes in brightness* [13]. An interesting variation in this definition is related to rigid transformations of parts of an object (for instance, fingers of a hand). A desirable representation of shape should *quotient out* all such transformations [9]. However, such a representation is likely to induce the so-called *curse of dimensionality* problem. To circumvent this problem, classical shape analysis methods such as [4, 12, 2] utilise principal component analysis (PCA) to reduce the dimensionality of the problem. One of the basic assumptions that these methods make is that the subspace corresponding to the major modes of variation for a particular class of shapes is linearised. The reality, however, is that such subspaces may be far from linear and hence a non-linear dimensionality reduction method may be required.

It would not be an exaggeration to say that non-linear dimensionality reduction is an active area of research [16]. A range of non-linear dimensionality reduction methods have been proposed in recent years which can be categorised broadly as *local* and *global*. *Local* embedding methods such as Laplacian eigenmaps and Locally Linear Embedding (LLE) attempt to preserve the properties of a small neighbourhood in a high-dimensional space while *global* techniques such as Isomap, Kernel PCA, diffusion maps [3] attempt

to preserve global properties of the data (see [16] for a comprehensive survey on dimensionality reduction methods). Local techniques are computationally efficient and perform well when the local geometry of the manifold is nearly Euclidean while global techniques preserve global structure of the data more faithfully [16].

A few recent studies on the problem of manifold learning with regard to shape analysis and its applications to computer vision can be found in the literature. He *et al.* [7] utilised the manifold learning of images for indexing images and using relevance feedback in the retrieval of images. Souvenir and Pless [14] proposed a manifold learning algorithm for weighted point sets, the idea being that members of different shape classes may be given different weights for efficiently separating intersecting manifolds of shape classes. Yankov and Keogh [17] presented a method for shape clustering using Isomaps for non-linear dimensionality reduction and used the classical Expectation-Maximisation (EM) algorithm for clustering the feature vectors of a reduced dimensionality. More recently, Etyngier *et al.* [6] extended the Laplacian eigenmap algorithm to enable interpolation between shape samples using local weighted means. This results in an elegant algorithm which was applied to both synthetic and real images showing its potential for image segmentation. Twining and Taylor [15] also presented a kernel PCA based extension of the active shape models.

In this paper, we propose a novel method for extracting the intrinsic parameters of multiple shape classes in an unsupervised manner, based on learning the global structure of shape manifolds with the help of diffusion maps. Assuming that similar shapes (and their transformations) belonging to a particular class lie on a sub-manifold in a high-dimensional Euclidean space $\Re^N$, the problem of extracting intrinsic shape parameters reduces to the problem of finding a low-dimensional embedding of the sub-manifold. A concept diagram of our method is shown in Figure 1. We choose the recently proposed
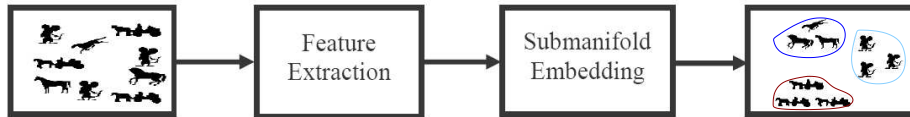

Figure 1: Concept diagram of the proposed method

diffusion maps [3] for global learning of our shape manifolds in an unsupervised manner, due to its control over the dimensionality of the sub-manifolds to embed in the lower-dimensional space. A diffusion map uses diffusion processes to understand the geometry of the high-dimensional manifold (shape manifold, in our case) by defining a Markov random walk on a graph of the data set. Like other spectral clustering methods, it uses eigenfunctions of a similarity-type matrix, termed as the Markov matrix (described in Section 3), as coordinates of the data set in a low-dimensional space. However, it differs from the other aforementioned methods in two respects: (1) it defines a robust metric on the data that reflects the connectivity of data set, and (2) it defines a time parameter in a Markov chain that determines the dimensionality of the sub-manifold [10].

The main contribution of this paper is the presentation of a framework for learning the shape manifolds for multiple classes (demonstrated by using samples from six shape classes; most shape manifold clustering methods listed above are limited to 2 or 3 classes) in an unsupervised manner. The proposed method for extraction of intrinsic shape parameters is shown to be effective using only two dimensions, even in the presence of noise. Moreover, the feature vectors used to represent shapes are invariant to scale and rotation, so no explicit alignment of shapes is required, saving valuable computational time.

The extraction of features from shapes defined as closed curves is outlined in the next section. Embedding of shape sub-manifolds in a low-dimensional space using diffusion maps is described in Section 3. Experimental results are presented and discussed in Section 4, followed by concluding remarks.

2 Shape Representation

As mentioned above, a desirable representation of shape should factor out any translation, scale, and rotation transformations. In order to learn the structure of shape manifolds of different classes in an unsupervised manner, it is crucial that such a representation employs features related to a shape that are associated with main shape characteristics. Several features for shape representation (also known as *shape descriptors*) have been proposed in the literature. It is beyond the scope of this paper to list and compare them all.

Let us regard a shape as a closed curve (*contour*) and initially represent it as a set of boundary points corresponding to the contour. We have chosen *Fourier descriptors* to represent a given shape contour. Fourier descriptors are simple yet powerful means of extracting shape features. It is worth noting that the choice of Fourier descriptors was made with simplicity in mind. The major objective of this study is to investigate the unsupervised learning of shape manifolds; no claim of optimality is made here about the representation of shapes using Fourier descriptors. Having noted that, one of the advantages of using Fourier descriptors is that they can easily be made invariant to any scale and rotation transformations (as described in the next paragraph). This implies that no explicit shape normalisation, as is the case with most well-known shape analysis methods such as [4], is required.

For a given shape, we extract its corresponding contour $\mathscr{C}$ which consists of N (fixed) boundary points (x_i, y_i), for $i = 1, 2, \ldots, N$. The number of boundary points remains fixed by employing cubic spline interpolation for re-sampling of the boundary points. A centroidal distance function r_i, for $i = 1, 2, \ldots, N$, is computed as follows [8, 11],

$$r_i = \sqrt{(x_i - \bar{x})^2 + (y_i - \bar{y})^2} \tag{1}$$

where $(\bar{x}, \bar{y})$ denotes the coordinates of the centroid of the object. The distance vector $\mathbf{r} = \{r_1, r_2, \ldots, r_N\}$ is transformed into frequency domain using FFT. The feature vector $\mathbf{f}$ can then be derived as follows,

$$\mathbf{f} = \left(\frac{|F_1|}{|F_0|}, \frac{|F_2|}{|F_0|}, \ldots, \frac{|F_{N/2}|}{|F_0|} \right)^T \tag{2}$$

where $|F_i|$ denotes the ith Fourier coefficient, with F_0 being the DC component. In the above equation, taking the magnitude of the coefficients yields rotation invariance and their division by $|F_0|$ results in scale invariance. In order to test the robustness of a shape clustering algorithm in the presence of noise, we add Gaussian noise n with zero mean and a standard deviation σ to the boundary points (x_i, y_i) as follows [8],

$$\begin{aligned} x_i' &= x_i + d_i \cdot n \cdot \sin \theta_i \\ y_i' &= y_i + d_i \cdot n \cdot \cos \theta_i \end{aligned} \tag{3}$$

for $i = 1, 2, \ldots, N$, where θ_i is the tangent angle at (x_i, y_i) and d_i denotes the distance between (x_i, y_i) and (x_{i-1}, y_{i-1}). Figure 2 shows an example contour and its corresponding centroidal distance vector without noise and with two different levels of noise.
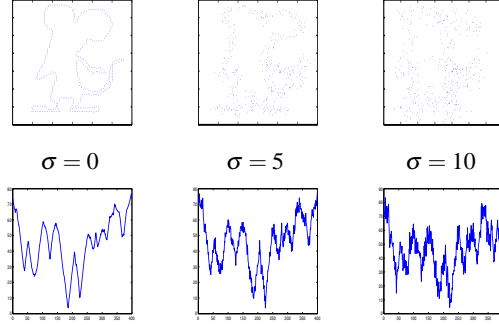
Figure 2: The contour for a Rat image with $N = 400$ boundary points (*Top row*) and the corresponding centroidal distance vector (*Bottom row*) for three levels of noise.

3 Learning the Shape Manifolds with Diffusion Maps

3.1 Computation of Shape Distances

The first step of diffusion maps based framework is to consider the feature vectors $\Omega = \{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_n\}$ corresponding to a data set of various shapes, where n denotes the total number of shapes in the data set. Like most manifold learning methods, we regard the feature vectors as nodes of an adjacency graph (node i representing the feature vector $\mathbf{f}_i$) and compute pairwise similarity matrix between $\mathbf{f}_i$ and $\mathbf{f}_j$, $i, j = 1, 2, \ldots, n$, using Gaussian kernel with width ε,

$$w(\mathbf{f}_i, \mathbf{f}_j) = \exp\left(-\frac{||\mathbf{f}_i - \mathbf{f}_j||^2}{2\varepsilon}\right) \tag{4}$$

A Markov random walk is then defined on the graph by defining the state-transition matrix $\mathbf{P}$ (also known as the Markov matrix), with its (i, j)th element p_{ij} given as follows

$$p_{ij} = \frac{w(\mathbf{f}_i, \mathbf{f}_j)}{d(\mathbf{f}_i)} \tag{5}$$

where $d(\mathbf{f}_i)$ denotes the degree of node $\mathbf{f}_i$ in the graph and is calculated as follows

$$d(\mathbf{f}_i) = \sum_{\mathbf{z} \in \Omega} w(\mathbf{f}_i, \mathbf{z}). \tag{6}$$

It can be seen from the above definitions of the Markov matrix that its ith row $\mathbf{P}_i = \{p_{ij}, \ j = 1, 2, \ldots, n\}$ corresponds to the probabilities of transition from node i to all the other nodes after one time step. It is worth noting that unlike most similarity matrices used in other manifold learning methods, this matrix is generally not symmetric.

3.2 Embedding in a Low-Dimensional Space

Once the Markov matrix for a given set of shapes belonging to different classes has been determined, sub-manifolds corresponding to the shape classes can be embedded in a lower-dimensional space as follows. If $\{\lambda_i\}$ is the sequence of eigenvalues of $\mathbf{P}$ such that $|\lambda_0| \geq |\lambda_1| \geq \cdots$, and $\{\psi_i\}$ are the corresponding eigenvectors, then a mapping from
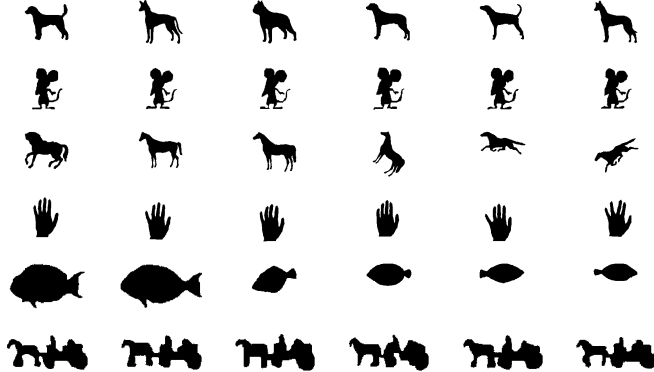
Figure 3: Some of the images used in our experiments.

the shape feature set Ω to a lower-dimensional Euclidean space $\Re^m$, where $m << N$ is the dimensionality of the lower-dimensional subspace, is given by (see [10, 3] for details),

$$\Psi : \mathbf{f} \longmapsto (\lambda_1 \psi_1(\mathbf{f}), \lambda_2 \psi_2(\mathbf{f}), \dots, \lambda_m \psi_m(\mathbf{f})) \tag{7}$$

The above relationship gives the initial low-dimensional embedding of the shape sub-manifolds at $t = 1$. In the spirit of a Markov chain, the mapping Ψ evolves as the chain progresses. The above mapping at time $t > 1$ can be computed by multiplying the diffusion coordinates $\psi_i(\mathbf{f})$, $i = 1, \dots, n$, by the $(\lambda_i)^t$ as given below [3],

$$\Psi^{(t)} : \mathbf{f} \longmapsto \left((\lambda_1)^t \psi_1(\mathbf{f}), (\lambda_2)^t \psi_2(\mathbf{f}), \dots, (\lambda_m)^t \psi_m(\mathbf{f}) \right) \tag{8}$$

Spectral fall-off and the time t of the random walk are the main factors contributing to dimensionality reduction. Consequently, for large value of t, we will be able to capture large-scale structures in the data with fewer diffusion coordinates [3]. A natural consequence of diffusion maps based embedding is that it captures the intrinsic dimensionality of the high dimensional data which could be depicted by the way the objects are organised in the new, lower-dimensional subspace.

If semantic meaning can be associated with the intrinsic shape parameters, $\Psi^{(t)}$ can be regarded as a mapping from a low-level (but high-dimensional) shape feature space to a high-level (low-dimensional) semantic space with the abstraction level of the latter going up as t increases.

4   Experimental Results and Discussion

To evaluate the clustering performance of the proposed shape manifold learning algorithm, six shape classes were selected from the Kimia database of silhouettes [5]. These classes are: *Dog*, *Rat*, *Horse*, *Hand*, *Fish*, and *Carriage* with 49, 20, 20, 16, 32 and 20 members, respectively, making a total of 157 samples. Six randomly chosen members from each of the six classes are shown in Figure 3. To study the noise sensitivity of different algorithms, the extracted boundaries of objects were corrupted by zero mean Gaussian

noise with four different noise levels following the noise addition method outlined in Section 2.

The Fourier descriptor (**FD**) feature vector for each shape contour was used to test the clustering efficiency of diffusion maps (**DM**) against raw **FD** features and **PCA**, which can be thought of as emulating the active shape model [4] (no Procrustes analysis is required here, as the **FD** features are scale and rotation invariant). The **FD** feature vectors (with a total of 199 elements for $N = 400$ boundary points) were then used as input to the diffusion maps based embedding algorithm as well as to the classical subspace projection method of **PCA**. Figure 4 shows the two-dimensional (2D) scatter plots for the three methods at $\sigma = 0$ and $\sigma = 10$; results for $\sigma = 2, 5$ are not included due to the limited space, see Table 1 and Figure 5 for quantitative results. It can be observed from Figure 4 that the first two Fourier descriptors were unable to efficiently separate all the classes except *Carriage*, *Hand*, and *Rat*. Members of the *Dog*, *Horse*, and *Fish* classes are mapped close to each other in almost all the methods. It is clear from these results that the embedding with diffusion coordinates clearly separates the former three clusters (far apart from each other) as well as better separating the latter three (which can be observed by zooming into the crowded part of the embedding using **DM**). In other words, the **DM** approach yields a higher increase in the between-class variance and reduces the within-class variance. It can also be observed from these results that noise has little effect on the clustering performance of first two Fourier descriptors as it affects the high frequency Fourier descriptors only. Interestingly, the performance of **FD** goes marginally up in the presence of noise, which can be put down to statistical variations in the feature space due to noise as well as the use of only low frequency coefficients (first and second). The **PCA** and **DM** based embeddings show signs of deterioration in their separability of clusters due to the use of high frequency coefficients which are more susceptible to noise than the low frequency coefficients.

In order to quantify the performance of the proposed shape clustering algorithm, we have used a measure known as the *class separability index* (CSI) c_i, i being the class index, defined as

$$c_i = \frac{\bar{d}_i}{\sqrt{\bar{\sigma}_i^2}} \qquad (9)$$

where $\bar{d}_i$ is the average distance from the ith cluster's centroid to the centroids of all other clusters and $\bar{\sigma}_i^2$ can be regarded as a measure of its average spread estimated by the average of variances along the cluster's principal axes. Table 1 presents results of clustering performance in terms of the CSI. Also shown in Table 1 are $\bar{c}$, which is a measure of the overall clustering performance for a particular method, and $\bar{c}_i$, which can be regarded as a measure of the complexity of the ith shape and the within-class variation for its corresponding cluster. It is evident from these results that **DM** is clearly superior to **PCA** and raw **FD** at all noise levels. As the noise level increases, the overall clustering performance goes down for both **PCA** and **DM**, although $\bar{c}$ for **DM** at $\sigma = 10$ is still higher than that for the other two methods.

Application to Detection of Nuclei in Histological Images: We also tested the proposed shape clustering algorithm on images of histological tissue samples from the prostate gland. The shapes of nuclei provide clues for both the diagnosis and prognosis of cancer using computer-assisted systems aimed at quantifying changes in the nuclei shapes. To this end, the reliable isolation of nuclei from the tissue section populated with other

	c_i	$\sigma = 0$			$\sigma = 5$			$\sigma = 10$			
		FD	**PCA**	**DM**	**FD**	**PCA**	**DM**	**FD**	**PCA**	**DM**	
Dog	c_1	3.0	2.4	**12.5**	3.2	2.5	**6.7**	3.4	2.8	**7.9**	$\bar{c}_1 = 4.9$
Rat	c_2	13.1	20.6	**110.8**	14.0	19.0	**76.6**	15.7	18.7	**80.8**	$\bar{c}_2 = 41.1$
Horse	c_3	2.8	3.7	**8.6**	2.9	3.6	**9.1**	3.0	3.5	**5.9**	$\bar{c}_3 = 4.8$
Hand	c_4	16.8	23.1	**213.1**	16.3	19.2	**72.2**	17.2	16.6	**60.2**	$\bar{c}_4 = 50.5$
Fish	c_5	2.8	3.4	**15.8**	2.8	3.4	**10.8**	3.0	4.0	**9.9**	$\bar{c}_5 = 6.5$
Carriage	c_6	11.4	12.9	**17.9**	11.5	12.8	**66.5**	11.7	12.5	**68.4**	$\bar{c}_6 = 31.0$
	$\bar{c}$	8.3	11.0	**72.0**	8.5	10.2	**40.3**	9.0	9.7	**38.9**	

Table 1: Class separability index c_i, for $i = 1, 2, \ldots, 6$, using first two Fourier descriptors (**FD**), top two PCA coordinates (**PCA**), and top two (2nd and 3rd) diffusion coordinates (**DM**), for different levels σ of noise.

undesirable objects is pivotal for the success of these systems. Figure 6(*a*) shows a typical tissue section from prostate gland. The tissue section images were binarised using a intensity-based *k*-mean clustering scheme. Diffusion maps for the **FD** feature vectors of the various object boundaries were computed. The potential nuclei identified by a simple zero-crossing applied to the top two coordinates in the embedding are shown in Figure 6(*b*). From these preliminary results, it can be observed that our algorithm has successfully detected all regular-shaped nuclei (shown in *green*), effectively separating them from the remaining objects, such as parts of nuclei, overlapping nuclei clusters, nuclear debris, and extracellular objects (shown in *red*). One of the major advantages of this detection scheme is its low computational complexity. It only needs two thresholding operations to separate regular-shaped nuclei from other objects. The spatial arrangement of these nuclei as dictated by the diffusion mapping can be semantically interpreted as well (see [1] for more details).

5   Conclusions

Extraction of the intrinsic parameters of shape is of vital importance for a wide range of applications. In this paper, we have presented a novel framework for learning the global structure of shape manifolds using diffusion maps in an unsupervised manner. The clustering results using the proposed framework show its superiority over classical shape clustering using linear dimensionality reduction. The difficulty in separating members of *Dog* and *Horse* classes stems from the fact that their contours have certain subtle differences which could not be picked up by the first two diffusion coordinates. It is our observation that the class separability indices for these two classes can be significantly increased by increasing the number of diffusion coordinates. A naïve extension of the proposed shape clustering algorithm for the shape classification problem would require re-computing the Markov matrix every time there is a new test sample. This extra computation can be avoided by employing out-of-sample extension methods (such as the Nyström formula) [16] to embed a new sample into the low-dimensional diffusion coordinate space without having to re-compute the Markov matrix. The proposed framework can also be extended for it to be applied to the general problem of shape modelling.

References

[1] M. Arif and N.M. Rajpoot. Detection of nuclei by unsupervised manifold learning. In *Proceedings Medical Image Understanding and Analysis (MIUA)*, July 2007.

[2] G. Charpiat, O. Faugeras, and R. Keriven. Approximations of shape metrics and application to shapewarping and empirical shape statistics. *Foundations of Computational Mathematics*, 2005.

[3] R. Coifman and S. Lafon. Diffusion maps. *Applied and Computational Harmonic Analysis, Special Issue on Diffusion Maps and Wavelets*, 21:5–30, July 2006.

[4] T. Cootes, D. Cooper, C. Taylor, and J. Graham. Active shape models – their training and applications. *Computer Vision and Image Understanding*, 61(1):38–59, 1995.

[5] D.Sharvit, J.Chan, H.Tek, and B.B.Kimia. Symmetry-based indexing of image databases. *Journal of Visual Communication and Image Representation*, 1998.

[6] P. Etyngier, R. Keriven, and J.-P. Pons. Towards segmentation based on a shape prior manifold. In *1st International Conference on Scale Space andVariational Methods in Computer Vision (SSVM)*, 2007.

[7] X. He, W.-Y. Ma, and H.-J. Zhang. Learning an image manifold for retrieval. In *Proceedings ACM International Conference on Multimedia*, pages 17–23, 2004.

[8] H. Kauppinen, T. Seppanen, and M. Pietikainen. An experimental comparison of autoregressive and fourier-based descriptors in 2-d shape classification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 17:201–207, 1995.

[9] D.G. Kendall. Shape manifolds, procrustean metrics, and complex projective spaces. *Bulletin of the London Mathematical Society*, 16(2):81–121, 1984.

[10] S. Lafon and A.B. Lee. Diffusion maps and coarse-graining: A unified framework fordimensionality reduction, graph partitioning, and data setparmeterization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(9), September 2006.

[11] Y. Lu, S. Schlosser, and M. Janeezko. Fourier descriptors and handwritten digit recognition. *Machine Vision and Applications*, 6:25–34, 1993.

[12] O. Faugeras M. Leventon, E. Grimson. Statistical shape influence in geodesic active contours. In *Proceedings International Conference onComputer Vision and Pattern Recognition (CVPR)*, 2000.

[13] A. Sekuler. Local and global minima in visual completion:effects of symmetry and orientation. *Perception*, 23:529–545, 1994.

[14] R. Souvenir and R. Pless. Manifold clustering. In *Proceedings International Conference on Computer Vision (ICCV)*, pages 648–653, 2005.

[15] C.J. Twining and C.J. Taylor. Kernel principal component analysis and the construction of non-linear active shape models. In *Proceedings British Machine Vision Conference (BMVC)*, 2001.

[16] L.J.P. van der Maaten, E.O. Postma, and H.J. van den Herik. Dimensionality reduction: A comparative review. Preprint, 2007.

[17] D. Yankov and E. Keogh. Manifold clustering of shapes. In *Proceedings International Conferenceon Data Mining*, 2006.
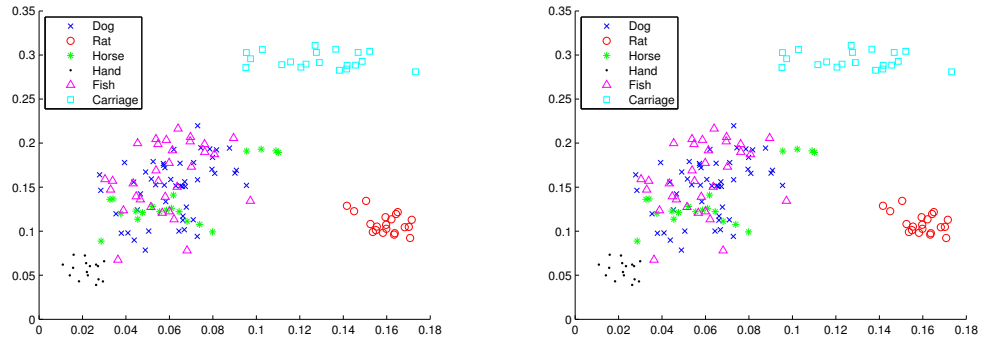
$\sigma = 0$ $\sigma = 10$

Two-dimensional feature space using top two (1st and 2nd) Fourier descriptors

$\sigma = 0$ $\sigma = 10$
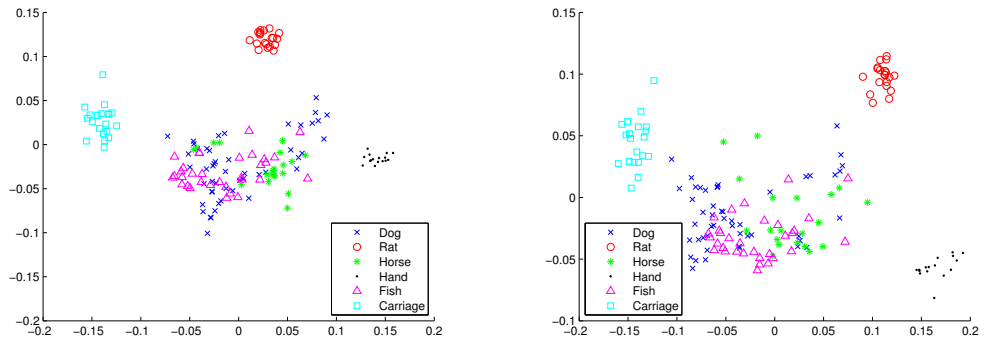
Scatter plots of top two PCA coordinates

$\sigma = 0$ $\sigma = 10$
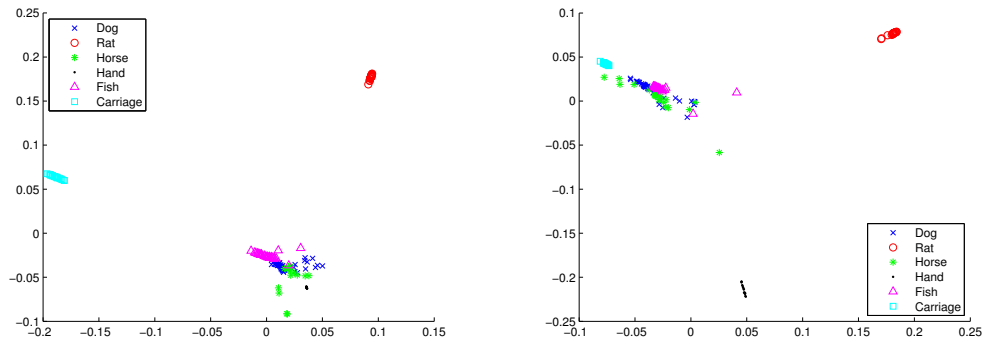
Scatter plots of top two diffusion coordinates

Figure 4: Two-dimensional feature space (*Top*) and 2D embeddings using PCA (*Middle*) and diffusion maps (*Bottom*).
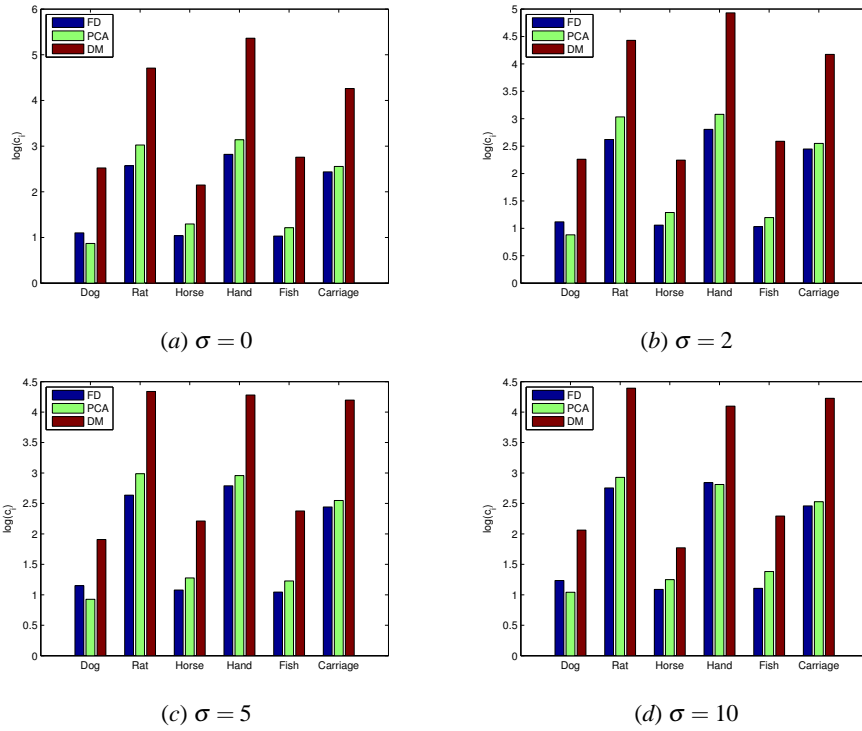
(*a*) $\sigma = 0$ (*b*) $\sigma = 2$

(*c*) $\sigma = 5$ (*d*) $\sigma = 10$

Figure 5: *Clockwise from Top-Left*: Logarithm of the class separability index c_i, for $i \in \{Dog, Rat, Horse, Hand, Fish, Carriage\}$, using top two Fourier descriptors (**FD**), PCA coordinates (**PCA**), and diffusion coordinates (**DM**) for $\sigma = 0$, 2, 5, and 10.
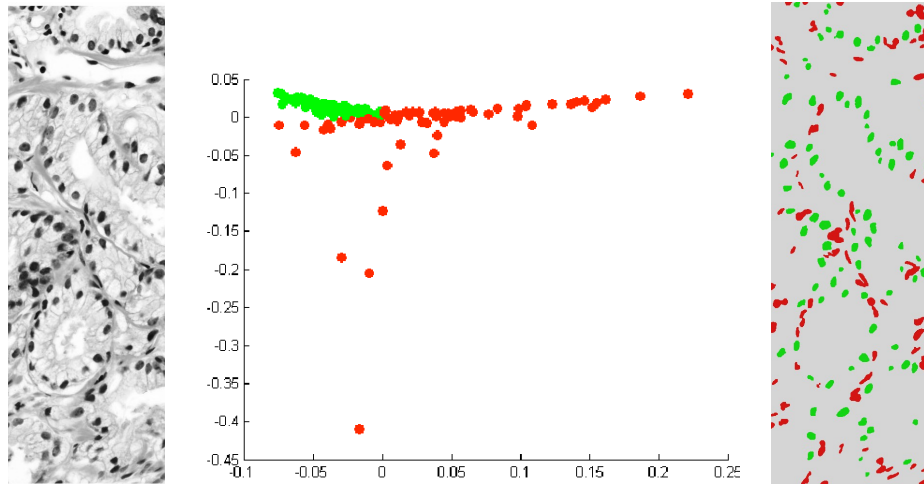

Figure 6: Detection of nuclei: A prostate histology image (*Left*), 2D embedding of the feature vectors for nuclei contours (*Middle*), and Regular-shaped nuclei (shown in green colour) detected by zero-crossings of the diffusion coordinates (*Right*).