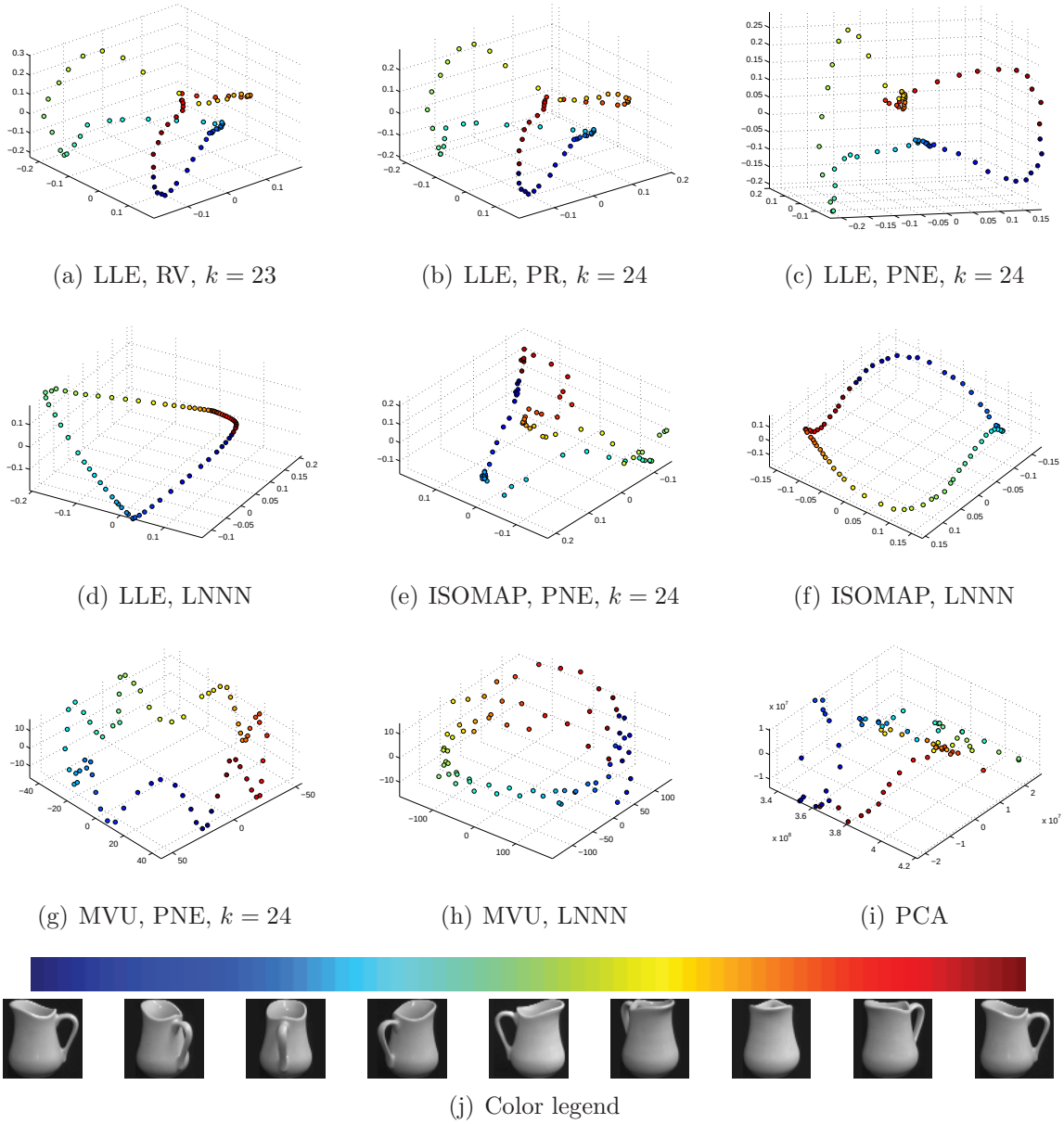


(a) LLE, RV, $k = 23$

(b) LLE, PR, $k = 24$

(c) LLE, PNE, $k = 24$

(d) LLE, LNNN

(e) ISOMAP, PNE, $k = 24$

(f) ISOMAP, LNNN

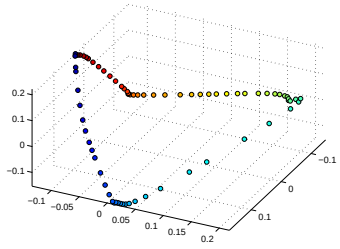
(g) MVU, PNE, $k = 24$

(h) MVU, LNNN

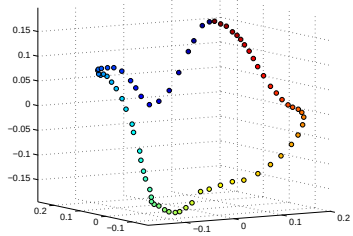
(i) PCA

(j) Color legend

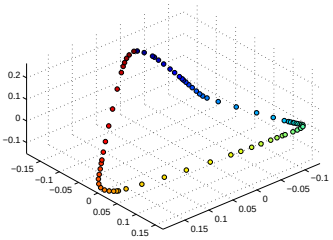
Figure 7.8: Dimensionality reduction on the pitcher database



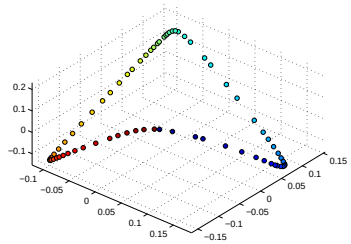
(a) LLE, RV, $k = 6$



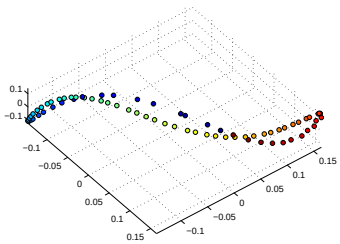
(b) LLE, PR, $k = 25$



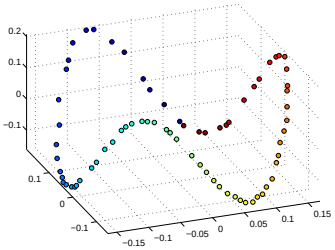
(c) LLE, PNE, $k = 24$



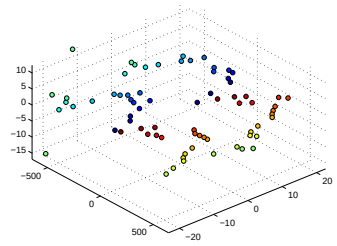
(d) LLE, LNNN



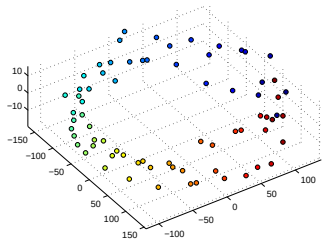
(e) ISOMAP, PNE, $k = 4$



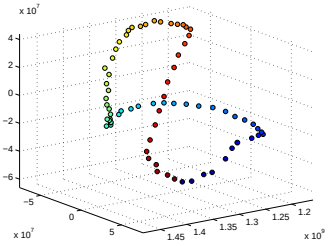
(f) ISOMAP, LNNN



(g) MVU, PNE, $k = 4$



(h) MVU, LNNN



(i) PCA

(j) Color legend

Figure 7.9: Dimensionality reduction on the Maneki Neko database

7.3 Discussion

It is clear, from tables 7.3 and 7.5, that the Empirical Regularization (ER) method proposed in [3] yields better embedding results than automatic regularization methods (Discarded Eigenvalues (DE) and Stability and Acurracy Tradeoff (SAT)), when using the artificial data sets. Nonetheless, ER method needs to manually set up Δ variable in equation (4.1), which can be difficult for data with dimension greater than 3, because there is not a reference or control about embedding results, that is, a blind embedding. That shortage is evidenced in the results obtained for the meteorology database (tables 7.7 and 7.10); in these cases, the empirical regularization can not overcome the performance of the automatic methods.

Discarded Eigenvalues method, proposed in [2], shows errors in embedding results, specially in cases when input manifolds have irregularities, for example, in the the Swiss roll with hole and the fishbowl data sets. In these cases, the variance calculated in equation (4.2) can not be proper as regularization value for $\mathbf{G}$, particularly when neighborhoods are unconnected or present discontinuities.

On the other hand, our proposed automatic regularization method (Stability and Acurracy Tradeoff (SAT)) was effective for finding out the value of regularization parameter. Embedding results are suitable and correspond with the expected unfolding surfaces. Moreover, for the Meteorology database ($p = 6$), the proposed method presents the best performance. In general, the SAT method computes suitable embedding for both artificial and real-world data sets.

In regard to the automatic selection of the number of nearest neighbor, the RV cost function (4.25), as a global trend, when the size of the neighborhood is augmented, the value of σ_R^2 is diminished, because a transformation for NLDR employing a large number of neighbors results in a linear transformation, and the residual variance does not quantify the local geometric structure, then this measure can not identify a suitable number of neighbors k. Figures 7.1(a) to 7.1(c), 7.3(a) to 7.3(c), 7.5(a) to 7.5(c), and 7.8(a) show some examples of the unsuitable embedding obtained by means of the Residual Variance method. In this case, data in low-dimensional space are overlapped and the RV method does not detect it. Besides, it is possible to notice that the number of nearest neighbors is too large. Nevertheless, Figures7.7(a) to 7.7(c), and 7.9(a) do not display the situation above pointed out and allows to calculate an appropriate embedding. The inconsistent results obtained by using the residual variance make of it an unreliable measure. In [24] the ambiguous results are attributed to the fact that Euclidean distance becomes an unreliable indicator for proximity.

The Procrustes Rotation (PR) criterion proposed in [25] takes into account the local geometric structure but does not consider the global behavior of the manifold. Then, far neighborhoods can be overlapped in the low-dimensional space and this measure shall not detect it, which can be seen in Figures 7.1(d) to 7.1(f), 7.3(d) to 7.3(f), 7.5(d) to 7.5(e), 7.8(b), and 7.9(b). Tables 7.7 and 7.10 corroborate the wrong embedding results for the meteorology database, which are obtained when the number of neighbors is chosen by means of the PR method.

The proposed PNE measure (4.32) computes a suitable number of neighbors for both artificial and real-world manifolds. Besides, the embedding results calculated using this

criterion are in accordance to the expected visual unfolding, (specially in Figures 7.1(i), 7.3(i), 7.9(c)). This criterion preserves the local geometry of data and the global behavior of the manifold.

The Figures 7.1(l), 7.3(l), 7.8(d), and 7.9(d) shown suitable embedding results for computing a local number of nearest neighbors (LNNN) for each input. The proposed methodology calculates each k_i preserving the linearity of every patch, taking into account the density of each region and the neighborhood size of the other regions in the manifold, in order to retain both local geometry and global structure. Moreover, the global optimization techniques require to compute an embedding for each possible value of k, solving the eigenvalue problem in LLE many times, resulting in a large computational cost process. The local optimization methodology solves just once the eigenvalue problem, reducing greatly the computational cost compared with global optimization techniques. Then, the local selection of the number of neighbors outperforms almost all embedding results obtained using techniques for global selection, specially when intrinsic properties of the neighborhoods change throughout the manifold. Particularly, in these cases, the size of neighborhoods lying in linear patches is not regular. Then, the local algorithm proposed can adapt to this kind of fluctuations.

From Figures 7.2, and 7.4 is possible to notice that the proposed methodologies for the automatic selection of the number of neighbors can also be set in ISOMAP and MVU. The embedding error and the number of neighbors preserved (tables 7.4, 7.6, 7.9, and 7.12) show that these NLDR techniques produce appropriate low dimensional outputs, when the numbers of neighbors are chosen according to LNNN, particularly for the artificial data sets. Nevertheless, in the fishbowl data set 7.6 and for some visualizations results on the pitcher and the Maneki Neko databases (Figures 7.8(e), 7.9(e), and 7.9(g)) the low dimensional mappings of the high dimensional data are wrong.

In summary, the proposed techniques for both regularization and selection of the number of neighbors outperform the approaches found in the state of art. Our methods can produce suitable embeddings on both artificial and real-world databases, whereas the other techniques just show good performance in a few particular examples. The LNNN method shows appropriate mappings in the most of the analyzed databases, and the computational cost of this method is less than for the PNE criterion.

Chapter 8

Results about Correntropy-LLE

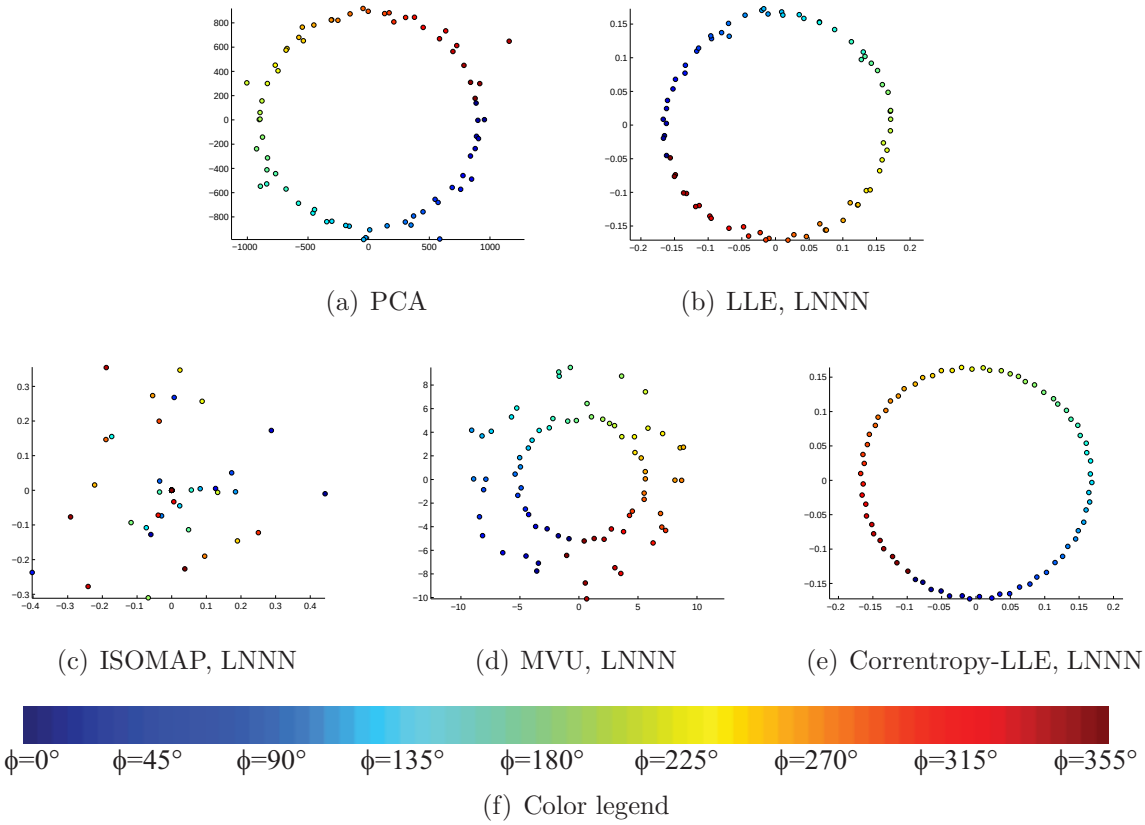
In order to verify the improvements of the embedding results for corrupted functional data, using correntropy measure in locally linear embedding instead of Euclidean distance, we employ three different data bases, which allow us to visually determine whether the embedding was correctly calculated and to confirm the capabilities of the proposed approach. These databases are: 1) sine with non-Gaussian noise, 2) pitcher with noise, and 3) Maneki Neko with noise.

We compare the Correntropy-LLE technique for visualization against LLE with Euclidean distance, ISOMAP, MVU, and PCA. According to the computation of the intrinsic dimensionality by means of the correlation dimension and self-similarity measure (section 4.3), the dimension of the output space is set to $m = 2$ for the artificial data set and $m = 3$ for the real-world data sets. Actually, for the corrupted real-world databases, the intrinsic dimension is greater than three because of the noise. Nevertheless, we are facing a visualization problem, then, we need to plot the embedding results in a 3-dimensional space at the most. Moreover, based on the results of the previous chapter, the number of nearest neighbors is chosen by means of the LNNN method, which computes an specific number of neighbors for each input object.

The σ parameter is initially chosen by means of the Silverman's rule (5.4), it is called σ_0. Nevertheless, in the most of the cases, this value (σ_0) is too small, and it is necessary to refine the search of σ inside the interval $[\sigma_0,\ 100\sigma_0]$ according to the maximization of the NPN criteria (6.1).

The embedding results for the sine with non-Gaussian noise, pitcher with noise and Maneki Neko with noise are presented in Figures 8.1, 8.2, and 8.3, respectively. The subfigures 8.1(f), 8.2(f), and 8.3(f) are indicating the match between the colors of the low dimensional points and the rotation of the real objects.

The quality of the embedding results are assessed by means of the number of neighbors preserved (see Table 8.1).



(a) PCA (b) LLE, LNNN

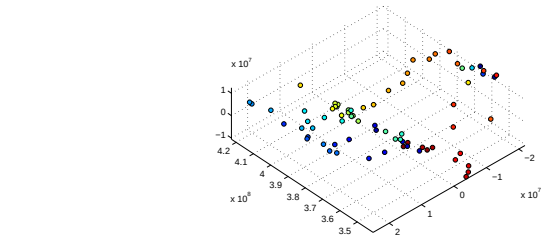
(c) ISOMAP, LNNN (d) MVU, LNNN (e) Correntropy-LLE, LNNN

(f) Color legend

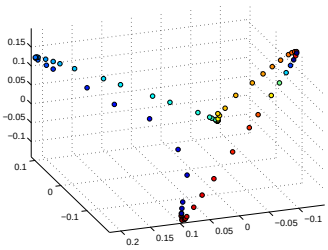
Figure 8.1: Dimensionality reduction on the sine with non-Gaussian noise database

Data Sets	**Number of Preserved Neighbors (target = 1)**			
	LLE	ISOMAP	MVU	Correntropy LLE
Noisy Sine	0.5111	0.1167	0.6368	0.9954
Noisy Pitcher	0.7062	0.7949	0.7265	0.8550
Noisy Maneki Neko	0.8843	0.8002	0.8764	0.8885

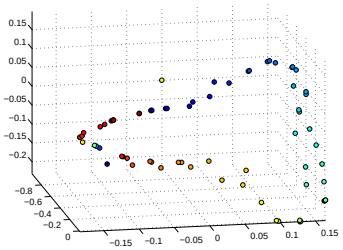
Table 8.1: Number of Preserved Neighbors computed by (6.1). Noisy data sets.



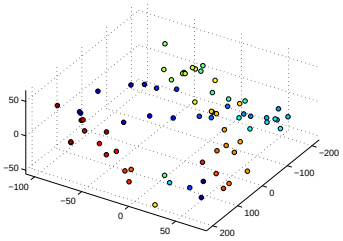
(a) PCA



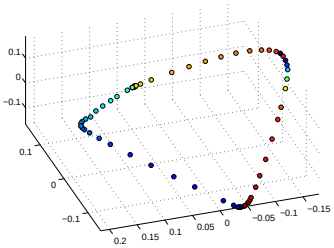
(b) LLE, LNNN



(c) ISOMAP, LNNN



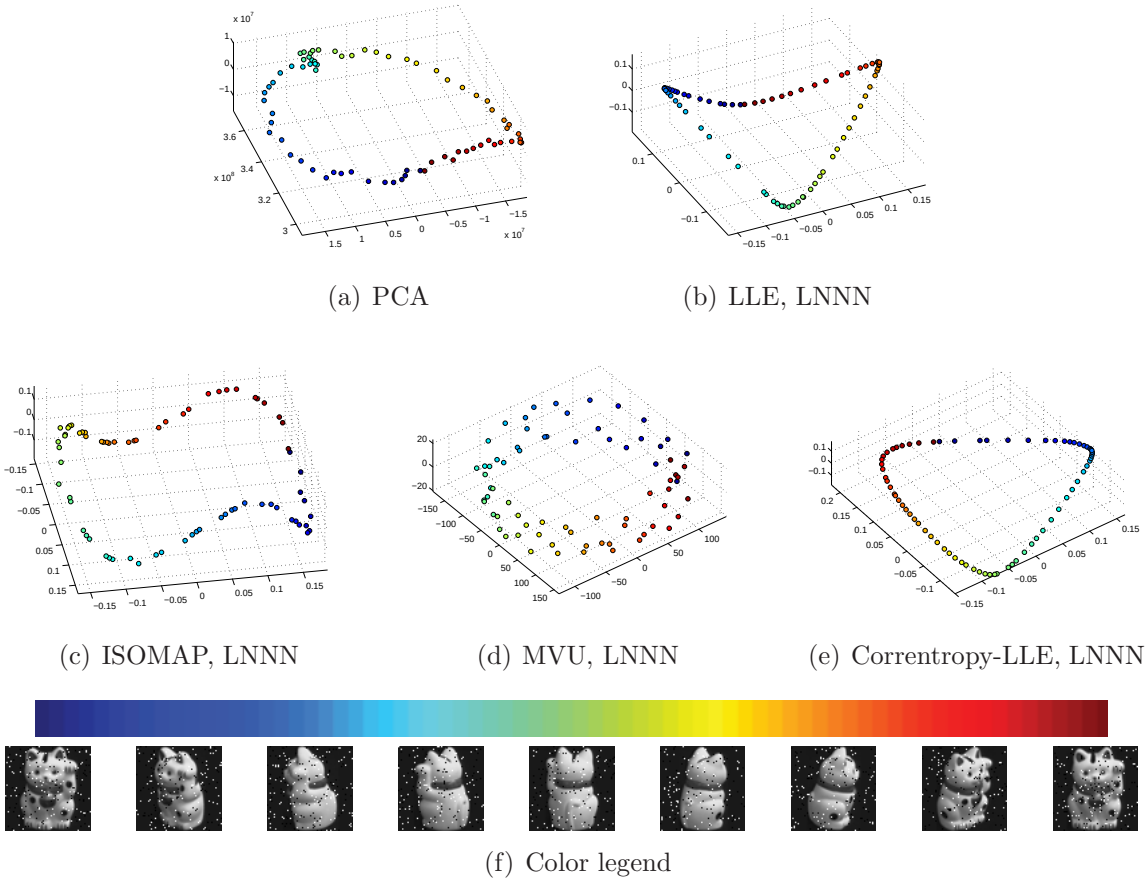
(d) MVU, LNNN



(e) Correntropy-LLE, LNNN

(f) Color legend

Figure 8.2: Dimensionality reduction on the pitcher with noise database



(a) PCA

(b) LLE, LNNN

(c) ISOMAP, LNNN

(d) MVU, LNNN

(e) Correntropy-LLE, LNNN

(f) Color legend

Figure 8.3: Dimensionality reduction on the Maneki Neko with noise database

8.1 Discussion

In all experiments only our proposed approach, Correntropy-LLE, could control the distortion caused by artifacts on the low dimensional output. Although the linear transformation (PCA) found out the low dimensional manifold for the artificial data set (Fig. 8.1(a)), results were strongly affected for both Gaussian and non-Gaussian noises. Additionally, PCA could not capture the underlying structure for real-world data sets, and neither avoid distortions produced by the artifacts. Its embedding results exhibit overlapped trajectories and neighborhoods conformed by dissimilar objects. The results obtained using LLE with Euclidean distance are not more encouraging than results obtained using PCA, at least for the artificial data set. For the pitcher database the manifold was not suitably unfolded. LLE could avoid overlapped trajectories but it could not represent the soft rotation what images show, see Fig. 8.2(b). Indeed, Euclidean distance could not capture the existent relations among pixels. It is not appropriated for measuring distances between functional data. Besides, non-gaussian noise once again deforms embedding results for artificial and real-world data sets, putting objects far away from the unfolded manifold. The ISOMAP technique can not discover the underlying structure behind the artificial data set (Figure 8.1(c)), and for the real-world examples, the low dimensional representations (Figures 8.2(c) and 8.3(c)) do not reflect the behavior of the high dimensional objects (rotation and scaling). For the artificial data set, MVU shows a double low dimensional manifold, one for the corrupted signal and other for the signals without noise, in Figure 8.1(d) a double ring is presented. Similarly, the embedding results on the real-world data sets, using MVU, show distorted manifolds (Figures 8.2(d) and 8.3(d)), the points are scattered in the low-dimensional space and they are not following a clear trajectory.

In contrast, the results achieved with Correntropy-LLE are consistent in all cases. Low dimensional structures were found and the effect of the non-gaussian noise was removed. Very small oscillations (Figure 8.1(e)) are due to gaussian noise added to signals. Underlying structures for both pitcher and Maneki Neko data sets (Figures 8.2(e) and 8.3(e)) are smooth and accurately represent the hidden process behind the pictures. We choose three dimensions for represent the real-world data sets because original pictures have angle changes, which need two dimensions, and besides they have scale changes, requiring one more dimension. It is possible to notice from embedding results for LLE and Correntropy-LLE that there are some neighborhoods containing dissimilar objects, this occurs when differences between objects are not captured by pixel differences.

These observations are corroborated quantitatively by computing the NPN criterion (Table 8.1), pitifully the results for PCA can not be quantified by means of NPN because it does not work based on neighborhood preservation, which is the base of the NPN cost function.

On the other hand, the kernel bandwidth controls how similar two objects are in a neighborhood, this property of the correntropy provides an appropriate way to remove the effects of the non-gaussian noise on the unfolded representation. Nonetheless, to adjust this parameter it is possible to employ density estimation techniques such as Silverman's rule. Next, for a suitable tuning of this parameter, it is possible to refine the search of σ inside a large interval, i.e. $[\sigma_0,\ 100\sigma_0]$, according to the maximization of the NPN criteria (6.1), which can be computationally expensive.

Chapter 9

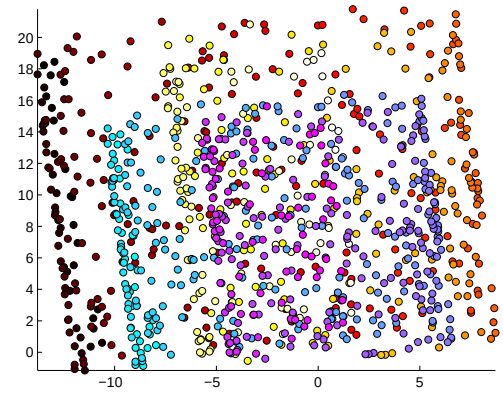
Results about NLDR Employing Class Label Information

Unlike the previously presented experiments, in this chapter, high dimensional points from multiple manifolds (classes) are mapped into a low dimensional space, by means of an improved version of the LLE algorithm, which takes into account class label information. This approach is employed in order to outperform both visualization and classification results obtained with the conventional LLE algorithm or other NLDR techniques when data correspond to several manifolds.

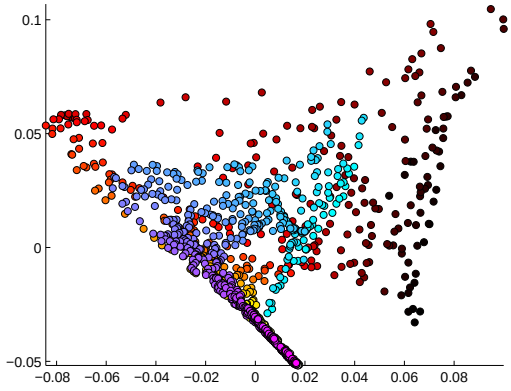
The double roll database is tested in visualization and classification labors. Moreover, a new data set conformed by the first five classes (0 to 4) extracted from the MNIST database is also employed with the same purposes. On the other hand, the MNIST with 10 classes and the PCG databases are just tested for classification.

We compare the proposed approach, which we call class label LLE (section 5.2), against PCA, LLE, δ-LLE (section 3.3), and ISOMAP. In all cases $\delta = 1$, and the number of nearest neighbors is chosen by means of the LNNN method. For the classification labors, the intrinsic dimension m is selected by examining the eigenvalue spectra of local covariance matrices $\mathbf{G}$ (refers to (4.33)), where $\nu = 0.95$ (amount of retained variance by projecting the data set).

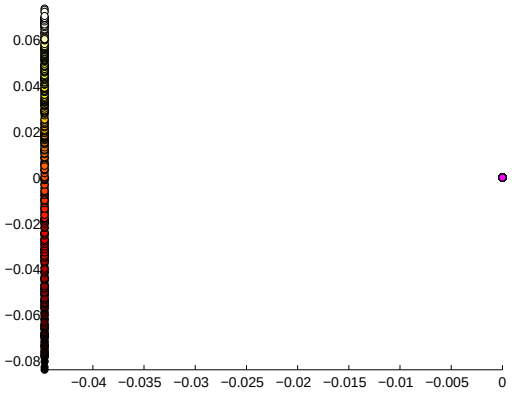
The two dimensional embedding for the double roll and the MNIST (with 5 classes) databases are presented in Figure 9.1 and 9.2, respectively. The tables 9.1, 9.2, and 9.3 present the classification accuracy, the confidence interval, and the dimension of the output space, each table is associated with a respective kind of classifier (LBNC, QBNC, and KNNC).



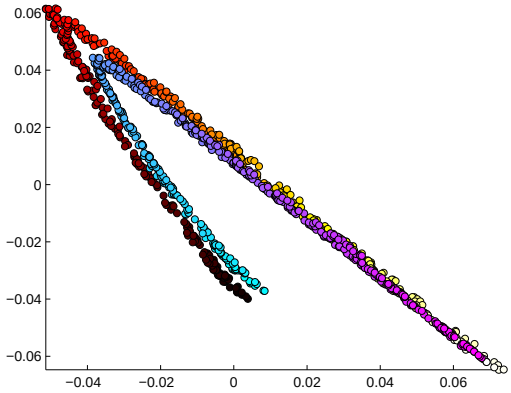
(a) PCA



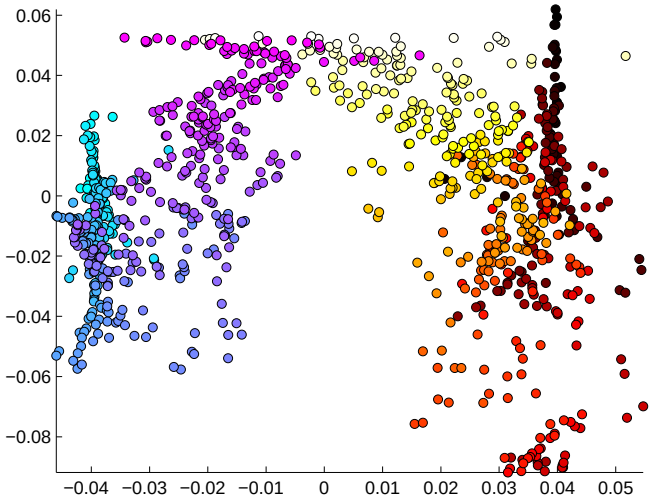
(b) LLE, LNNN



(c) δ-LLE, LNNN

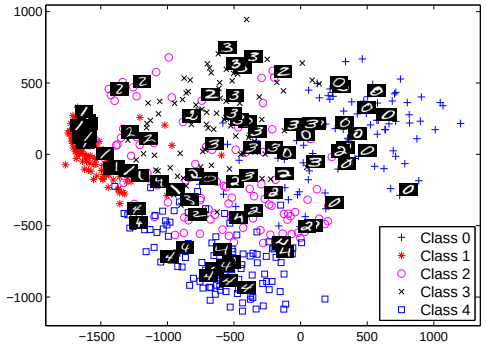


(d) ISOMAP, LNNN

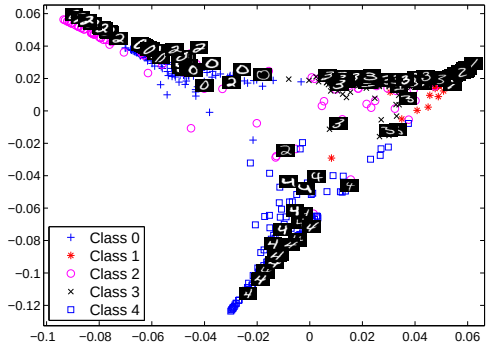


(e) Class Label LLE, LNNN

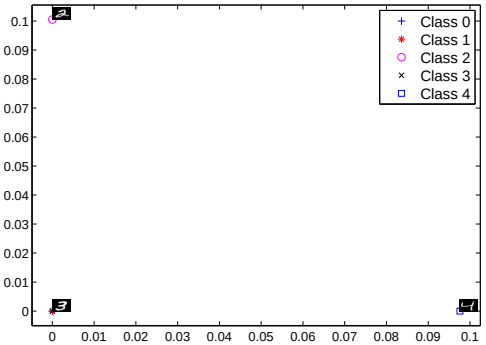
Figure 9.1: Dimensionality reduction for visualization on the double roll (2 classes) database



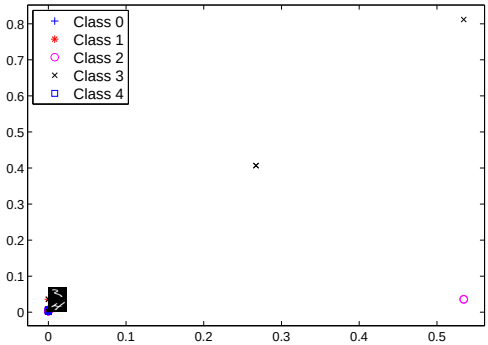
(a) PCA



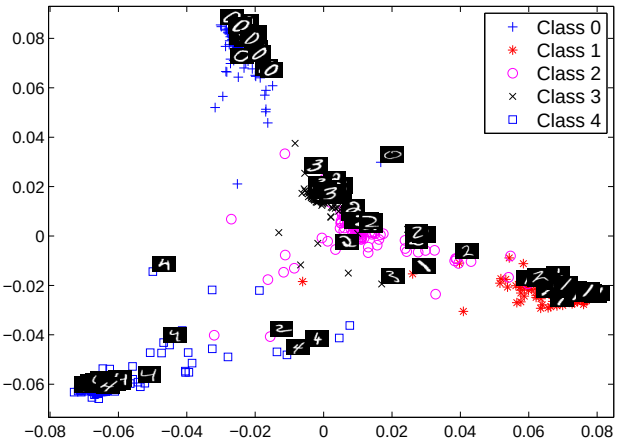
(b) LLE, LNNN



(c) δ-LLE, LNNN



(d) ISOMAP, LNNN



(e) Class Label LLE, LNNN

Figure 9.2: Dimensionality reduction for visualization on the MNIST (5 classes) database

Data Sets	**Classification Accuracy (Acc) – LBNC**				
	PCA	ISOMAP	LLE	δ-LLE	Class label LLE
Double Roll	Acc = 0.5777	Acc = 0.5153	Acc = 0.5153	Acc = 0.4697	**Acc = 0.9667**
$m = 2$	CI = [0.5655, 0.5899]	CI = [0.4966, 0.5341]	CI = [0.4678, 0.5629]	CI = [0.4194, 0.5200]	CI = [0.9604, 0.9729]
MNIST 0-4	Acc = 0.8373	Acc = 0.4233	Acc = 0.7020	Acc = 0.1933	**Acc = 0.8253**
$m = 4$	CI = [0.8215, 0.8532]	CI = [0.3574, 0.4893]	CI = [0.6770, 0.7270]	CI = [0.1933, 0.1933]	CI = [0.8003, 0.8504]
MNIST 0-9	Acc = 0.7236	Acc = 0.1270	Acc = 0.6047	Acc = 0.0980	**Acc = 0.6872**
$m = 7$	CI = [0.7076, 0.7397]	CI = [0.1045, 0.1495]	CI = [0.5875, 0.6219]	CI = [0.0980, 0.0980]	CI = [0.6701, 0.7042]
PCG	Acc = 0.5750	Acc = 0.5500	Acc = 0.8640	Acc = 0.3780	**Acc = 0.9159**
$m = 15$	CI = [0.5475, 0.6025]	CI = [0.4790, 0.6210]	CI = [0.8547, 0.8733]	CI = [0.3258, 0.4303]	CI = [0.9032, 0.9285]

Table 9.1: Classification Accuracy and Confidence Interval – Linear Bayes Normal Classifier

Data Sets	**Classification Accuracy (Acc) – QBNC**				
	PCA	ISOMAP	LLE	δ-LLE	Class label LLE
Double Roll	Acc = 0.6527	Acc = 0.5427	Acc = 0.5757	Acc = 0.5	**Acc = 0.9673**
$m = 2$	CI = [0.6410, 0.6644]	CI = [0.5044, 0.5810]	CI = [0.5327, 0.6186]	CI = [0.5, 0.5]	CI = [0.9618, 0.9728]
MNIST 0-4	**Acc = 0.8807**	Acc = 0.3780	Acc = 0.5693	Acc = 0.2267	Acc = 0.7493
$m = 4$	CI = [0.8643, 0.8970]	CI = [0.3265, 0.4295]	CI = [0.4655, 0.6731]	CI = [0.2267, 0.2267]	CI = [0.7063, 0.7924]
MNIST 0-9	**Acc = 0.8772**	Acc = 0.1017	Acc = 0.5703	Acc = 0.1172	Acc = 0.6470
$m = 7$	CI = [0.8285, 0.8459]	CI = [0.0875, 0.1158]	CI = [0.5335, 0.6070]	CI = [0.1162, 0.1182]	CI = [0.6151, 0.6788]
PCG	Acc = 0.6171	Acc = 0.6567	Acc = 0.8220	Acc = 0.5006	**Acc = 0.9024**
$m = 15$	CI = [0.5863, 0.6478]	CI = [0.6110, 0.7024]	CI = [0.8018, 0.8421]	CI = [0.4994, 0.5018]	CI = [0.8822, 0.9227]

Table 9.2: Classification Accuracy and Confidence Interval – Quadratic Bayes Normal Classifier

Data Sets	**Classification Accuracy (Acc) – k-NNC**				
	PCA	ISOMAP	LLE	δ-LLE	Class label LLE
Double Roll	Acc = 0.7293	Acc = 0.6520	Acc = 0.6413	**Acc = 0.9997**	Acc = 0.9733
$m = 2$	CI = [0.7213, 0.7374]	CI = [0.6032, 0.7008]	CI = [0.6066, 0.6761]	CI = [0.9990, 1]	CI = [0.9683, 0.9783]
MNIST 0-4	Acc = 0.8840	Acc = 0.3347	Acc = 0.7553	Acc = 0.7073	**Acc = 0.8887**
$m = 4$	CI = [0.8672, 0.9008]	CI = [0.2862, 0.3831]	CI = [0.7181, 0.7926]	CI = [0.6947, 0.7200]	CI = [0.8698, 0.9075]
MNIST 0-9	**Acc = 0.8233**	Acc = 0.0868	Acc = 0.6429	Acc = 0.4625	Acc = 0.7986
$m = 7$	CI = [0.8107, 0.8359]	CI = [0.0675, 0.1062]	CI = [0.6210, 0.6648]	CI = [0.4317, 0.4933]	CI = [0.7746, 0.8226]
PCG	Acc = 0.7720	Acc = 0.6939	Acc = 0.9287	**Acc = 0.9567**	Acc = 0.9390
$m = 15$	CI = [0.7482, 0.7957]	CI = [0.6653, 0.7225]	CI = [0.9167, 0.9406]	CI = [0.9426, 0.9703]	CI = [0.9294, 0.9486]

Table 9.3: Classification Accuracy and Confidence Interval – k-Nearest Neighbor Classifier

9.1 Discussion

For the artificial data set, the low dimensional space computed by means of PCA (Figure 9.1(a)) presents overlapped classes, and it is possible to see that the individual manifolds were not correctly unfolded. The LLE algorithm unfold the manifolds but it does not have the ability for separating the classes, then one can notice a curved plane superimposed on another plane (Figure 9.1(b)). The δ-LLE technique maximizes the separation between classes, constructing clusters for them (Figure 9.1(c)). For this reason, the local structure of the high dimensional data is not preserved in the low dimensional space. A three dimensional representation of the double roll data set after employing the δ-LLE algorithm produces two orthogonal and separated lines (one per roll), then the local geometry of the data is not recovered, that is, the rolls can not be suitable unfolded by means of the δ-LLE. Although ISOMAP avoids the class overlapping (Figure 9.1(d)) the low dimensional output does not present an appropriate representation for the high dimensional data.

On the other hand, the proposed technique (class label LLE) unfolds each one of the rolls, while the local geometry of the high dimensional data is preserved and the classes are separated. It should be noted that the overlapped area between classes in Figure 9.1(e) corresponds with the closet points from different classes in the high dimensional space (Figure 6.8).

For the MNIST (5 classes) database, the visual representation obtained with PCA reflects the behavior of the data (Figure 9.2(a)). It possibly suggests that the manifolds are linear. The δ-LLE and ISOMAP techniques represent all the points of a class on the same low dimensional point (Figures 9.2(c) and 9.2(d)). In these cases, the local geometry of the high dimensional data is not preserved. Moreover, several classes are placed at the same position; particularly in δ-LLE, it becomes a sign that the class label information is not appropriately took into account. The conventional LLE algorithm tries to preserve the local structure of the data, but because of this technique is not supervised, then the class overlapping is not avoided (Figure 9.2(b)).

The class label LLE technique captures the local structure of the data and it ensures the separability between the classes, in Figure 9.2(e) is possible to see five different clouds of points (one per class).

In regard to classification, our approach presents, in the most of the cases, the best classification accuracy (Tables 9.1, 9.2, and 9.3). Class label LLE allows to use a classifier with very simple decision boundary with high classification performance, because the technique can unfold the non-linear input data structures. The results produce, as such as possible, simpler low dimensional manifolds whose topology is preserved, whereas at the same time the class label information is considered, ensuring the class separability. The δ-LLE algorithm seems to obtain a high classification performance, when the technique is used in conjunction with a non-linear boundary classifier on the databases with two classes. Nonetheless, as the number of classes grows, the classification accuracy strongly diminishes, because of the overtraining induced by this kind of NLDR.

In summary, the proposed supervised NLDR technique shows the ability of representing several manifolds at the same time for visualization tasks. And for classification, our approach is efficient and competitive, regarding to other similar methods. Additionally, class label LLE has the an advantage, it does not require complex classifier. On the contrary,

the simplest boundary decision classifier (lineal boundary) achieves a high classification performance, which is specially desired when hardware implementations are needed.

Part IV

Conclusions and Future Work

Chapter 10

Conclusions

This work presents three main contributions. Firstly, it was developed a methodology for an objective estimation of each one of the free parameters of the locally linear embedding algorithm, which allows to reduce the variability and uncertainty about the embedding results. Particularly, for the computation of the regularization parameter a new method was formulated, which finds a balance between the Stability and the Accuracy of the required solution, because the increase of these parameters leads to augment the error in the solution of the cost function (4.4). In other words, the method calculates the regularization term based on constraining the size of the weighted vector norm. As a result, the proposed method extends utilization of LLE to a broader range of applications. The proposed method calculates suitable embeddings for artificial (with and without discontinuities) databases as well as real-world ones. When this automatical regularization method is employed there is not necessary an expert user for tuning the algorithm. Besides, manifolds and/or embedding spaces with more than 3 dimensions can not be visually assessed, then there is not possible to set up the regularization parameter in a manual way. The proposed automatic method called Stability and Accuracy Tradeoff (SAT) is better than the state of art for automatic selection of LLE regularization parameter (i.e. Discarded Eigenvalues), and presents better performance than the empirical regularization technique on real-world data sets. Besides, two new criteria for choosing automatically the number of nearest neighbors needed in LLE algorithm are presented. One of them called Preservation Neighborhood Error (PNE) computes a global number of neighbors for the whole manifold, and it can be used as a measure for quantifying the embedding quality for NLDR techniques. The second proposed approach computes an specific number of neighbors for each point in the manifold, taking into account the linearity and density of each patch. Moreover, this local criterion solves just once the eigenvalue problem of LLE, reducing the computational cost in contrast to global optimization techniques presented in this work. Always, the best embedding results were obtained using one of these proposed approaches instead of the methods found in the state of art. Both of our approaches preserve local geometry of data and global behavior of the manifold. Nevertheless, local criterion requires special attention when it is used on manifolds with a large amount of outliers. In this sense, the local cost function can overfit the number of neighbors, producing an unconnected manifold, thus the embedding results can be inconsistent. In this case it is preferable to use the proposed algorithm for global selection.

Secondly, the use of a new cost function for comparing functional data, which can be employed as similarity measure in the core of the LLE algorithm, was proposed. Indeed, this cost function called Correntropy defines suitable neighborhoods and improve the quality of the embedding. With this in mind, we formulate a new technique for NLDR called Correntropy-LLE, which is designed to work with functional data unlike most approaches for analyzing data on nonlinear manifolds. When the correntropy similarity measure is employed instead of Euclidean distance in the LLE algorithm, the unfolded structure describes more clearly the underlying process behind the input data. Besides, the low dimensional representations are smooth, without distortions, and generally, the neighborhoods are conformed for very similar objects. The algorithm takes advantage of the metric induced by the correntropy similarity measure, which reduces the effects of artifacts, distortions and non-gaussian noise present in the objects. Our proposal outperforms the embedding results obtained with PCA, the conventional LLE algorithm, ISOMAP, and MVU, which was corroborated visually and quantitatively.

And as a third contribution of this work, a non-linear dimensionality reduction algorithm based on LLE, which preserves the local geometry of the high dimensional data and provides a discriminative strategy during the embedding procedure, was constructed. Specifically, the conventional LLE algorithm was reformulated in search of relevant underlying variables behind the observed high dimensional data, which must be related to another reference signal, that is, the class label information. This underlying variables span a low dimensional space, where the objects can be reconstructed as locally linear combinations of its neighbor. The low dimensional data preserves the local geometry relations, the global structure of the manifolds, and it describes the behavior defined by the class label information. Our approach exhibits, in the most of the developed experiments, the best classification accuracy compared to similar PCA, ISOMAP, conventional LLE, and supervised δ-LLE. The proposed supervised NLDR algorithm, called Class Label LLE, shows the ability of representing several manifolds at the same time for visualization tasks, and it is efficient and competitive, regarding to other similar methods in classification labors. Additionally, a fortitude of the Class Label LLE algorithm is that it allows to work with simple boundary decision classifier achieving a high classification accuracy, which is specially desired for hardware implementations.

Chapter 11

Future Work

Following the researching line described in this thesis, three main projects could be taken up, which involve the proposed improvements for the LLE algorithm.

1. In order to obtain an information-theoretic point of view for the proposed class label LLE, it is necessary to look for a mathematical relation between our approach and the information bottleneck method. Then, a new version of the algorithm based on information measures could be developed.

2. To make the hardware implementation of the proposed algorithms for developing projects in film animation, entertainment, advertising and marketing.

3. It would be interesting to derive an explicit mapping between the high and low dimensional spaces, specially, to predict new samples, e.g. for visualizing objects from unseen angles.

Part V

Appendix

Appendix A

Abbreviations

Acc Classification accuracy.

CI Confidence interval.

DE Discarded eigenvalues.

ER Empiric regularization.

ISOMAP Isometric feature mapping.

ITL Information-theoretic learning.

KNNC k-Nearest neighbor classifier.

LBNC Linear Bayes normal classifier.

LDA Linear discriminant analysis.

LLE Locally linear embedding.

LNNN Local number of nearest neighbors.

LS Least square.

MDS Multidimensional scaling.

MVU Maximum variance unfolding

NLDR Non-Linear dimensionality reduction.

NPN Number of preserved neighbors.

PCA Principal component analysis.

PCG Phonocardiography.

PNE Preservation neighborhood error.

PR Procrustes rotation.

QBNC Quadratic Bayes normal classifier.

RV Residual variance.

SAT Stability and accuracy tradeoff.

SDP Semidefinite programming.

Bibliography

[1] A. K. Jain, R. P. W. Duin, and J. Mao, "Statistical pattern recognition: A review," *IEEE Transactions on Pattern Recognition and Machine Intelligence*, vol. 22, no. 1, pp. 4–37, 2000.

[2] D. de Ridder and R. P. W. Duin, "Locally linear embedding for classification," Pattern Recongnition Group, Delft University of Technology, Delft, The Netherlands, Tech. Rep., 2002.

[3] L. K. Saul and S. T. Roweis, "Think globally, fit locally: Unsupervised learning of low dimensional manifolds," *Machine Learning Research*, vol. 4, pp. 119–155, 2003.

[4] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, pp. 2323–2326, 2000.

[5] O. Kouropteva, O. Okun, and M. Pietikäinen, "Supervised locally linear embedding algorithm for pattern recognition," in *IbPRIA, LNCS 2652*, 2003, pp. 386–394.

[6] G. Daza-Santacoloma, J. D. Arias-Londoño, J. I. Godino-Llorente, N. Sáenz-Lechón, V. Osma-Ruíz, and G. Castellanos-Domínguez, "Dynamic feature extraction: An application to voice pathology detection," *Intelligent Automation and Soft Computing*, 2009.

[7] A. Jog, A. Joshi, S. Chandran, and A. Madabhushi, "Classifying ayurvedic pulse signals via consensus locally linear embedding," in *International Conference on Bio-inspired Systems and Signal Processing*, Porto, Portugal, January 2009.

[8] M. Polito and P. Perona, "Grouping and dimensionality reduction by locally linear embedding," in *NIPS*, 2001.

[9] J. C. Principe, D. Xu, and J. W. F. III, "Information theoretic learning," in *Unsupervised Adaptive Filtering*, S. Haykin, Ed. New York: John Wiley & Sons, 2000, ch. 7.

[10] J. Tenenbaum, V. de Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, pp. 2319–2323, 2000.

[11] J. Tenenbaum, "Mapping a manifold of perceptual observations," in *NIPS*, 1998, pp. 682–688.

[12] A. R. Webb, *Statistical Pattern Recognition*, 2nd ed. Indianapolis, IN, USA: John Wiley & Sons, Ltd, 2002.

[13] L. K. Saul and S. T. Roweis, "An introduction to locally linear embedding," AT&T Labs and Gatsby Computational Neuroscience Unit, Tech. Rep., 2000.

[14] Z. Zhang and J. Wang, "MLLE: Modified locally linear embedding usign multiple weights," in *NIPS*, 2006.

[15] B. Li, C.-H. Zheng, and D.-S. Huang, "Locally linear discriminant embedding: An efficient method for face recognition," *Pattern Recognition*, vol. 41, pp. 3813–3821, 2008.

[16] D. de Ridder, O. Kouropteva, O. Okun, M. Pietikäinen, and R. P. W. Duin, "Supervised locally linear embedding," in *International Conference on Artificial Neural Networks*, 2003.

[17] K. Q. Weinberger and L. K. Saul, "An introduction to nonlinear dimensionality reduction by maximum variance unfolding," in *21st National Conference on Artificial Intelligence*, 2006.

[18] D. Zhao, "Formulating lle using alignment technique," *Pattern Recognition*, vol. 39, pp. 2233–2235, 2006.

[19] R. Karbauskaite, G. Dzemyda, and V. Marcinkevicius, "Selecting a regularisation parameter in the locally linear embedding algorithm," in *Int. Conference on Continuous Optimization and Knowledge-Based Technologies*, 2008.

[20] J. Vanderplas and A. Connolly, "The dimensionality of data: Locally linear embedding of sloan galaxy spectra," *The Astronomical Journal*, vol. 138, pp. 1365–1379, 2009.

[21] G. Daza-Santacoloma, C. D. Acosta-Medina, and G. Castellanos-Domínguez, "Regularization parameter choice in locally linear embedding," *Neurocomputing*, vol. 73, pp. 1595–1605, 2010.

[22] C. D. Meyer, *Matrix Analysis and Applied Linear Algebra*. SIAM, 2000.

[23] P. C. Hansen, *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion*. SIAM, 2000.

[24] O. Kouropteva, O. Okun, and M. Pietikäinen, "Selection of the optimal parameter value for the locally linear embedding algorithm," in *The 1st International Conference on Fuzzy Systems and Knowledge Discovery*, 2002.

[25] Y. Goldberg and Y. Ritov, "Local procrustes for manifold embedding: a measure of embedding quality and embedding algorithms," *Machine learning*, 2009.

[26] J. Valencia-Aguirre, A. Álvarez Mesa, G. Daza-Santacoloma, and G. Castellanos-Domínguez, "Automatic choice of the number of nearest neighbors in locally linear embedding," in *14th Iberoamerican Congress on Pattern Recognition – CIARP*, 2009.

[27] A. C. Rencher, *Methods of multivariate analysis*, 2nd ed. Hoboken, NJ, USA: Wiley-Interscience, 2002.

[28] H. Kantz and T. Schreiber, *Nonlinear Time Series Analysis*, 2nd ed. Cambridge University Press, 2004.

[29] J. O. Ramsay and B. W. Silverman, *Functional Data Analysis*, ser. Statistics. New York, NY, USA: Springer, 2005.

[30] R. Silipo, G. Deco, R. Vergassola, and H. Bartsch, "Dynamics extraction in multivariate biomedical time series," *Biological Cybernetics*, vol. 79, pp. 15–27, 1998.

[31] P. Hall, D. S. Poskitt, and B. Presnell, "A functional data-analytic approach to signal discrimination," *Technometrics*, vol. 43, no. 1, pp. 1–9, 2001.

[32] F. Ferraty and P. Vieu, "Curves discrimination: A nonparametric functional approach," *Computational Statistics & Data Analysis*, vol. 44, pp. 161–173, 2003.

[33] X. Leng and H.-G. Müller, "Classification using functional data analysis for temporal gene expression data," *Bioinformatics*, vol. 22, no. 1, pp. 68–76, 2006.

[34] I. Santamaría, P. P. Pokharel, and J. C. Principe, "Generalized correlation function: Definition, properties, and applications to blind equalization," *IEEE Transactions on Signal Processing*, vol. 54, no. 6, pp. 2187–2197, June 2006.

[35] W. Liu, P. P. Pokharel, and J. C. Principe, "Correntropy: Properties and applications in non-gaussian signal processing," *IEEE Transactions on Signal Processing*, vol. 55, no. 11, pp. 5286–5298, November 2007.

[36] K. Madsen, H. B. Nielsen, and O. Tingleff, *Optimization with Constraints*, 2nd ed. Informatics and Mathematical Modelling – Technical Universiy of Denmark, 2004.

[37] M. Pillati and C. Viroli, "Supervised locally linear embedding for classification: An application to gene expression data analysis," in *Book of Short Papers, CLADAG 2005*, Parma, Italy, 2005, pp. 147–150.

[38] M. Loog and D. de Ridder, "Local discriminant analysis," in *The 18th International Conference on Pattern Recognition*, 2006.

[39] X. Geng, D.-C. Zhan, and Z.-H. Zhou, "Supervised nonlinear dimensionality reduction for visualization and classification," *IEEE Transactions on Systems, Man, and Cybernetics - Part B: Cybernetics*, vol. 35, pp. 1098–1107, 2005.

[40] N. Tishby, F. C. Pereira, and W. Bialek, "The information bottleneck method," in *The 37th annual Allerton Conference on Communication, Control, and Computing*, 1999.