ViSOM for Dimensionality Reduction in Face Recognition

Weilin Huang and Hujun Yin

School of Electrical and Electronic Engineering
The University of Manchester
Manchester, M60 1QD, UK
weilin.huang@postgrad.manchester.ac.uk, h.yin@manchester.ac.uk

Abstract. The self-organizing map (SOM) is a classical neural network method for dimensionality reduction and data visualization. Visualization induced SOM (ViSOM) and growing ViSOM (gViSOM) are two recently proposed variants for a more faithful, metric-based and direct data representation. They learn local quantitative distances of data by regularizing the inter-neuron contraction force while capturing the topology and minimizing the quantization error. In this paper we first review related dimension reduction methods, and then examine their capabilities for face recognition. The experiments were conducted on the ORL face database and the results show that both ViSOM and gViSOM significantly outperform SOM, PCA and related methods in terms of recognition error rate. In the training with five faces, the error rate of gViSOM dimension reduction followed by a soft k-NN classifier reaches as low as 2.1%, making ViSOM an efficient approach for data representation and dimensionality reduction.

1 Introduction

Dimensionality reduction techniques have been widely used for data preprocessing, which provide basis for further analysis, management and storage of the data. It extracts meaningful information from high-dimensional data and represents the data by fewer dimensions, and thus can greatly facilitate data analysis, clustering and classification. For instance, in face recognition, each face is represented by a large number of pixel values. It is difficult or inefficient to directly operate on such high-dimensional data. Thus reducing dimensionality has become an important issue in data intensive pattern recognition.

Principal component analysis (PCA) is a primary technique and is regarded as the foundation for many dimensionality reduction techniques. PCA seeks a linear projection that best represents the data in the least-squares sense. It has been widely used in data analysis due to its computational simplicity and analytical tractability. *Eigenface* [1] is a famous application of PCA in face recognition. However, the linearity of PCA limits its power for complex and increasingly large data sets, as it is not capable of revealing nonlinear structure of the data defined by beyond second order statistics. Several PCA-based nonlinear techniques for

J.C. Príncipe and R. Miikkulainen (Eds.): WSOM 2009, LNCS 5629, pp. 107–115, 2009.

dimensionality reduction have been proposed recently. For example, kernel-PCA [2] extends PCA to nonlinear by using a kernel function in the input space. Local linear embedding (LLE)[3], a local PCA method, learns the underlying manifold of the data by minimizing an embedding function; while Isomap [4] captures the topology structure by computing the geodesic manifold distances between data points. Meanwhile, there has been previous work on applying these nonlinear techniques for face recognition [5,6,7,8].

Neural networks provide alternative approaches to nonlinear data projection and dimension reduction. The SOM [9] is one of the classical methods for clustering, dimension reduction and data visualization. Dimension reduction is achieved by establishing a topological order of the projection between input data and their corresponding neurons on the map. The applications of SOM in face recognition and comparisons with PCA-based methods can be found in [10,11]. For a more natural and direct display of data structure, ViSOM [12] has been proposed and improved recently by a growing variant, gViSOM [13]. The inter-point distances are locally preserved on the map along with the topology. It has been shown that ViSOM provides a better visual exhibition of data points and their distribution on the map than SOM [12,13]. ViSOM (or gViSOM) represents a metric scaling of the input space and has comparable capability for highly nonlinear manifold learning with other nonlinear PCA methods, such as LLE and Isomap [13]. A review on nonlinear dimensionality reduction is given in [14]. In this paper, we apply the ViSOM and gViSOM for dimensionality reduction in face recognition. We examine their performances and compare them with SOM and several PCA-based methods.

The rest of the paper is organized as follows. PCA-based algorithms, both linear and nonlinear, are briefly reviewed in Section 2, followed by the introduction of SOM-based methods in Section 3. Section 4 presents the classifiers used in the experiment, and the experimental details and results are shown in Section 5. Finally, Section 6 concludes the paper.

2 PCA-Based Methods

PCA [1] is a classical linear dimension reduction method aiming at finding principal orthogonal directions from a data set by solving an eigenvalue problem. While discarding a large number of minor components, a small number of principal components are retained to form a linear, low-dimensional subspace, known as *eigenface* in face recognition. Raw face images are projected onto the *eigenface* subspace first, and the classification is carried out in the reduced space.

Kernel-PCA [2] projects input data onto a high-dimensional feature space by using a hypothetic nonlinear function. Then the standard PCA is performed on the high-dimensional data set via a kernel function. There are two commonly used kernel functions: *polynomial* and *Gaussian radial basis*.

LLE [3] is a local PCA method and is capable of mapping high-dimensional nonlinear data onto a single global coordinate system of lower dimensionality. The neighborhood or topology is preserved in the embedding space by minimizing the cost functions in the input space and output space respectively. The optimal weights of input data can be found by solving a least squares problem of the cost function in the input space, while the embedding vectors in the output space are computed as an eigenvalue problem.

Isomap [4] seeks to learn the underlying manifold structure of a data set by computing the geodesic manifold distances between all pairs of data points. It first defines a neighborhood graph, over which each point is connected to all its neighbors in the input space. Then the geodesic distances of all pairs of points are computed via the shortest path on the neighborhood graph (using Floyd's algorithm). Finally multidimensional scaling is applied to the distance matrix to construct the embedding of the data to preserve intrinsic geometry structure of the data.

Curvilinear component analysis (CCA) [15] is another method to represent nonlinear data structure in a lower-dimensional space. The intrinsic geometric property of the data is revealed by preserving local distance relationships via an error function. A neighborhood function is used for local topology preservation and emphases on maintaining shorter distances than longer ones.

3 SOM-Based Methods

3.1 SOM

SOM is an unsupervised learning loosely based on the retinatopic mapping: an ordered projection of visual retina to visual cortex [16]. It uses a set of neurons ranged often in a 2-D lattice to form a topological mapping of the input space. The SOM learns the topological structure of the input by updating the weight vectors of a neighborhood of the winning neurons when being presented with an input. In the case of dimensionality reduction and data visualization, high-dimensional data are projected onto a low-dimensional SOM, represented by the order, location or index of the neuron on the map. The data structure learned by SOM reveals the relative or ordinal relationships among input data. However, it is unable to reproduce the quantitative distances between the input points on the reduced space. In many applications, a more faithful and metric scaling of the input space is more desirable in data visualization and dimensionality reduction [12,13].

3.2 ViSOM

The visualization induced SOM (ViSOM) [12] has been proposed to extend the SOM for faithful (local) distance preservation on the map. The ViSOM preserves the distance quantities along with the topology of data set. The updating force of SOM, $[x(t)-w_k(t)]$, can be decomposed into two components: $[x(t)-w_v(t)]+[w_v(t)-w_k(t)]$. The first term is the updating force from the winner v to the input $x(t)$, which is the same to the updating force of the winner. The second term is a lateral force that brings the neighboring neurons to the winner. This lateral contraction force is regulated in order to maintain a uniform inter-neuron distance locally on the map in ViSOM [12]. The ViSOM algorithm is briefly described below.

1. Select the winner v when an input $x(t)$ is presented, and update its weight

$$\Delta w_v(t) = \alpha(t)[x(t) - w_v(t)] \tag{1}$$

2. Update the weights of the nodes in the neighborhood according to

$$\Delta w_l(t) = \alpha(t)\eta(\varphi, l, t)[x(t) - w_v(t)] + \beta[w_v(t) - w_l(t)] \tag{2}$$

   where $\beta = d_{vl}/(\delta D_{vl}) - 1$, d_{vl} is the distance of neuron weights in the input space, D_{vl} is the distance of neuron indexes on the map, and δ is the resolution parameter. The neighborhood function η is similar to that in the SOM, the width of the neighborhood decreases from an initially large value to a final small value but not just to one as in the SOM.
3. Refresh the map by using the weights of randomly chosen neurons as the input at a small percentage of updating times.

The ViSOM learns the local quantitative distances of the input data by regularizing the inter-neuron contraction force, while it captures the ordering and minimizes the quantization error. The distance of two (local) projected points on the map is proportional to the distance of these two points in the input space, making feature representation and data visualization more faithful and quantitatively measurable. The resolution of the map can be enhanced by incorporating the local linear projection (LLP) method [17], which projects a data point onto the sub plane spanned by the two closest edges instead of to the winner.

3.3 gViSOM

It has been shown that SOMs of prefixed size are difficult to converge to highly nonlinear manifolds [13]. For improving the local distance-preserving capability of ViSOM, an incremental or growing ViSOM (gViSOM) has been proposed [13] for embedding and metric-scaling nonlinear manifolds. Details of the gViSOM algorithm are as follows,

1. Start with a small initial map (e.g. 5×5) of either rectangular or hexagonal. Place the initial map onto a linear subspace of either the entire or a local region of the data space. Set the desired resolution and the neighborhood size.
2. Randomly draw a data sample from the data space and find the winning neuron with the shortest distance.
3. If the sample falls within the neighborhood, update the weights of the neurons of the neighborhood using the ViSOM algorithm; otherwise go back to Step 2.
4. At regular iteration intervals (e.g. 2000 iterations), if the growing condition is met (that is, the data is underrepresented by the existing map), grow the map by adding a column or row to the side with the highest activities (measured by the winning frequencies). The added column or row is a linear extrapolation of the existing map. Other growing structures can be used, such as incrementing polygons instead of entire column or row for a free structure of the map and efficient use of neurons.

5. As in the ViSOM, at regular intervals (every certain number of iterations), refresh the map (neurons) probabilistically.
6. Check if the map has converged. If not go back to Step 2.
7. Project all data samples onto the map, either to the neurons or by the LLP resolution enhancement.

4 Classifiers

For classification, three common classifiers were used in our system: the Nearest-Neighbor (NN), soft k-Nearest Neighbor (soft k-NN) and the Linear Discriminant Analysis (LDA). NN is the simplest classifier assigning a test sample to the class of the most similar example in the training set. In soft k-NN classifier [11], each principal component outputs a confidence value, which gives the degree of support for each component in every face representation, and then the final decision is given by considering all of these confidence values. LDA [18] is an efficient and widely used linear classifier. It tries to find the linear projection of the data set that minimizes within-class scatter while maximizes between-class separation. The ratio of the determinant of the between-class scatter matrix and the within-class scatter matrix in the projected space is maximized by solving an eigenvalue problem.

5 Experiments and Results

In the experiment, the described methods were used for reducing data dimensions in the preprocessing of raw face images, and then one of the classifiers was used for classification. The performances of the dimension reduction methods were evaluated based on the same classifier. The experiment was conducted on a publicly available database, the ORL database (of Olivetti Research Laboratory), which consists of 40 subjects with 10 different face images for each subject. All images in the database were taken against a dark homogeneous background with an up-right, frontal position and have the same size of 92×112. Face images vary slightly in term of lighting conditions, facial expressions or facial details. Examples of two subjects are shown in Fig.1.

Fig. 1. Examples of ORL face images

PCA-Based Methods. In PCA-based methods, the number of dimensions (92×112) of face image is reduced to 60. Two types of kernel-PCA, *polynomial* (KPCA1) and *Gaussian Radial Basis* (KPCA2), were used with degree of 2 and radius of 30, respectively. The sizes of neighborhood used by LLE and Isomap were set to 30 and 120 respectively. The 60-dimensional face representations are then used for training and testing by a NN, soft k-NN or LDA classifier.

SOM-Based Dimensionality Reduction. The face images are first locally sampled by moving a window of size 5×5 over the entire image by 4 pixels each time. The sampled images are reconstructed to the sizes of $25 \times 23 \times 28$ after sampling. That is, each sampled face image contains 23×28 25-dimensional subsamples. These 25-dimensional samples are used as the inputs for SOM-based maps, which are trained by implementing the SOM-based algorithms with 50000 updates. For each method, its size and parameters have been optimized to its best performance. For example, the sizes of SOM, ViSOM and gViSOM varied from 5×5 to 30×30, and the chosen sizes represent the cases with the best performances (i.e. 10×10 for SOM, 30×30 for ViSOM with δ of 0.5 and 16×19 for gViSOM starting at 5×5 with δ of 0.6).

Then all 25-dimensional samples in each face image are passed through the trained SOM, ViSOM and gViSOM, and represented by the 2-D index values of the corresponding winners on the maps. Thereby, on the trained map, each face image has a corresponding 2-D face projection (as shown in Fig. 2), which is used for further classification. Each dimension of the face projection can be reconstructed as a feature face (with size of 23×28, examples of two subjects are shown in Fig. 3), which resemble features of the original face images. As can be seen, ViSOM methods resemble better the original features due to its the metric preserving property in feature extraction. For a full and objective evaluation, the performances of SOM-based and PCA-based methods were investigated on the same classifier for each experiment on all subjects of the ORL database. The number of training images was varied from 3, 4, 5 to 6 per subject and the remaining 7, 6, 5, and 4 were used as test images respectively. The results reported are the average results of 10 independent implementations with different

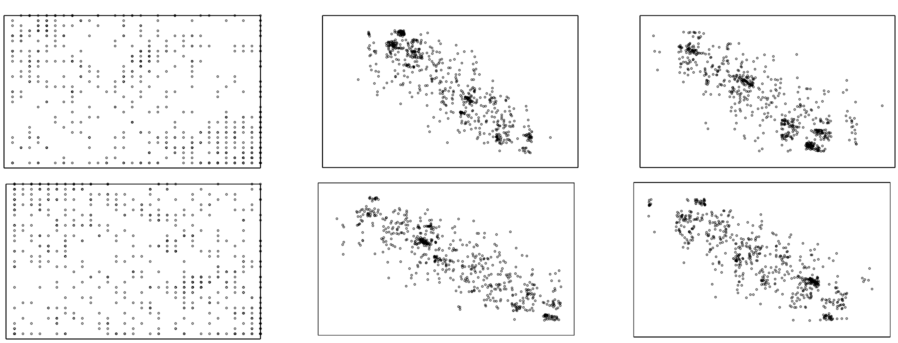

Fig. 2. Face projections of SOM (left), ViSOM (center) and gViSOM (right)

Fig. 3. Feature faces of SOM (left), ViSOM (center) and gViSOM (right)

Table 1. Error rates of PCA-based methods followed by a NN, soft k-NN or LDA classifier

No. of training faces	*Error rates*(%)					
	PCA	KPCA1	KPCA2	LLE	ISOMAP	CCA
	NN Classifier					
3	13.25	13.54	12.25	**11.25**	14.21	12.25
4	8.08	8.64	**7.17**	7.29	8.54	8.17
5	**5.65**	5.75	5.80	5.70	6.85	5.80
6	**3.56**	**3.56**	4.06	4.13	4.13	4.06
	soft k-NN Classifier					
3	13.43	15.50	11.75	**11.29**	14.14	12.68
4	8.69	9.42	9.08	**7.25**	8.79	8.46
5	6.15	6.45	7.00	**5.50**	6.90	5.85
6	4.26	4.50	5.87	3.94	4.75	**3.81**
	LDA Classifier					
3	**9.36**	11.07	10.07	9.71	13.46	12.15
4	**5.08**	6.00	5.96	6.75	8.37	7.21
5	3.80	4.15	4.95	**3.75**	6.90	5.40
6	3.12	3.31	3.19	**2.69**	4.31	4.31

randomly chosen training images. Meanwhile, the same choices of training (and test) images were used by all the methods to ensure an unbiased comparison. The results of PCA-based methods followed by the NN, soft k-NN or LDA classifier are shown in Tables 1. The performances of SOM-based methods with the NN or soft k-NN classifier are listed in Table 2, $ViSOM^*$ and $gViSOM^*$ denote the projections with the LLP resolution enhancement.

The tables show that with more training samples, error rates decrease in all methods as expected. The SOM have the similar performances to PCA-based methods with the NN classifier; ViSOM and gViSOM yield markedly improved performances than SOM and PCA-based methods with about 2% lower error rates in every implementation. With soft k-NN classifier, LLE has slightly lower error rates than other PCA-based methods; while SOM-based methods have about 2-3% improvements over the LLE, and gViSOM with LLP has even

Table 2. Error rates of SOM-based methods followed by a NN or soft k-NN classifier

No. of training faces	*Error rates*(%)				
	SOM	ViSOM	ViSOM*	gViSOM	gViSOM*
	NN Classifier				
3	11.57	10.61	**10.50**	10.86	10.79
4	7.50	6.42	**6.37**	6.46	6.54
5	5.85	**4.30**	4.35	4.40	4.50
6	3.81	2.69	**2.63**	2.75	2.88
	soft k-NN Classifier				
3	8.04	7.75	7.32	7.21	**6.71**
4	4.46	3.88	3.79	**3.67**	**3.67**
5	3.20	2.80	2.40	2.55	**2.10**
6	1.88	1.25	1.19	0.81	**0.75**

better performances with more than 1% further improvement. The error rates of gViSOM with LLP in training five and six faces are as low as 2.1% and 0.75% respectively. These results are also better than the results of the PCA-based methods followed by a LDA classifier, with performance improvements of 2-6% in error rate.

6 Conclusions

In this paper, we have applied the recently proposed ViSOM and gViSOM for face recognition. The capabilities of them for dimensionality reduction and feature extraction are compared with the standard SOM and several nonlinear PCA methods. The experimental results on a real-world face database show that SOM-based methods achieve a comparable or better performance than widely used PCA-based methods for feature extraction and dimensionality reduction; while metric preserving ViSOM and gViSOM outperform the SOM and these PCA-based methods with significant 2-6% improvement in the error rate. The gViSOM followed by a soft k-NN classifier gives the lowest error rate. This demonstrates that faithful representation of high dimensional data is important in pattern recognition and ViSOM offers an effective nonlinear data projection for face recognition.

References

1. Turk, M., Pentland, A.: Eigenfaces for recognition. Journal of Cognitive Neuroscience 3, 71–86 (1991)
2. Scholköpf, B., Smola, A., Müller, K.-R.: Nonlinear component analysis as a kernel eigenvalue problem. Neural Computation 10, 1299–1319 (1998)
3. Roweis, S.T., Saul, L.K.: Nonlinear dimensionality reduction by locally linear embedding. Science 290, 2323–2326 (2000)

4. Tenenbaum, J.B., de Silva, V., Langford, J.C.: A global geometric framework for nonlinear dimensionality reduction. Science 290, 2319–2323 (2000)
5. Pang, Y.H., Teoh, A.B.J., Wong, E.K., Abas, F.S.: Supervised locally linear embedding in face recognition. In: Int. Symp. on Biometrics and Security Technologies, pp. 1–6 (2008)
6. Yang, M.H.: Kernel eigenfaces vs. kernel Fisherfaces: face recognition using kernel methods. In: Proc. 5th IEEE Int. Conf. on Automatic Face and Gesture Recognition (RGR 2002), Washington DC, pp. 215–220 (2002)
7. Yang, M.H., Ahuja, N., Kriegman, D.: Face recognition using kernel eigenfaces. IEEE Int. Conf. on Image Processing 1, 37–40 (2000)
8. Yang, M.H.: Extended Isomap for pattern classification. In: Proc. National Conference on Artificial Intelligence, pp. 224–229 (2002)
9. Kohonen, T.: Self-Organizing Maps, 2nd edn. Springer, Heidelberg (1997)
10. Lawrence, S., Giles, C.L., Tsoi, A., Back, A.: Face recognition: Aconvolutional neural-network approach. IEEE Trans. Neural Networks 8, 98–113 (1997)
11. Tan, X., Chen, S., Zhou, Z., Zhang, F.: Recognizing partially occluded, expression variant faces from single training image per person with SOM and soft k-NN ensemble. IEEE Trans. on Neural Networks 16, 875–886 (2005)
12. Yin, H.: ViSOM-a novel method for multivariate data projection and structure visualization. IEEE Trans. on Neural Networks 13, 237–243 (2002)
13. Yin, H.: On multidimensional scaling and the embedding of self-organizing maps. Neural Networks 21, 160–169 (2008)
14. Yin, H.: Nonlinear dimensionality reduction and data visualization: A review. Int. Journal of Automation and Computing 3, 294–303 (2007)
15. Demartines, P., Hérault, J.: Curvilinear component analysis: a self-organizing neural network for nonlinear mapping of data sets. IEEE Trans. on Neural Networks 8, 148–154 (1997)
16. Von der Malsburg, C., Willshaw, D.J.: Self-organization of orientation sensitive cells in the striate cortex. Biological Cybernetic 14, 85–100 (1973)
17. Yin, H.: Resolution enhancement for the ViSOM. In: Proc. Workshop on Self-Organizing Maps (WSOM 2003), pp. 208–212 (2003)
18. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. IEEE Trans. on Pattern Analysis and Machine Intelligence 19, 711–720 (1997)