2015 Fifth International Conference on Instrumentation and Measurement, Computer, Communication and Control

Embedding new samples via locality-constrained sparse representation for non-linear manifold learning

Liu Yang[1,*], Yunyan Wei[3],
[1]School of Engineering,
Mudanjiang Normal University,
Mudanjiang, China
yangliu3560@gmail.com
[2]State Grid Dalian Electric Power Supply Company,
Dalian, China
63693389@qq.com

Feng Pan[2], Xiaohui Li[3]
[3]College of Computer Science and Information Technology,
Key Laboratory of Intelligent Information Processing of
Jilin Universities,
Northeast Normal University
Changchun, China
13944809411@139.com
lixh195@nenu.edu.cn

***Abstract*—How to embed the new observations (or samples) into the low-dimensional space is a crucial problem in non-linear manifold learning techniques. This issue can be converted into the problem of finding an accurate mapping that transfers the unseen data samples into an existing manifold. In this paper, a locality-constrained sparse representation algorithm is proposed to deal with the out-of-sample embedding problem for manifold learning. Through taking the data locality information into consideration, the local data structure can be well preserved in our proposed algorithm. To justify the superiority of the proposed method, our approach has been tested on several challenging face datasets and compared with other out-of-sample embedding techniques. The experimental results show that the proposed method can not only achieve the competitive recognition rate than the existing methods, but also save more time than the traditional nonlinear dimensionality reduction methods for the out-of-sample problem.**

Keywords—Non-linear manifold learning; Out-of-sample embedding; Sparse representation; Locality constraint

I. INTRODUCTION

The manifold learning has been proved to be a promising nonlinear dimensionality reduction (NDR) method for high dimensional data analysis [1-3]. Different from the classical linear dimensionality reduction methods, such as principal component analysis (PCA) [4] and linear discriminate analysis (LDA) [4] that only focus on the global Euclidean structure of data, the manifold learning techniques aim to find a projection space in which the intrinsic local geometry of the high-dimensional data can be well preserved. The representative manifold embedding algorithms include ISOMAP [2], LLE [3], LE [5], etc. ISOMAP aims to find a subspace that preserves the geodesic distance among observations so as to unfolding the manifold. LLE seeks to preserve the geometry of the local neighborhoods. It uses linear weighting, which reconstructs a given sample by its neighbors, to represent the geometry of the local neighborhoods and preserves these weighting coefficients for reconstruction of the given sample's projection in low dimensional space. LE preserves the locality of the local neighborhoods by manipulations on an undirected graph, which indicates neighbor relations of pairwise data points.

Unfortunately, a main drawback of the manifold learning methods is that they do not provide an explicit mapping function which embeds the high-dimensional data into the low-dimensional subspace [6]. Therefore, in order to obtain the low-dimensional representations of new coming samples, the learning procedure, containing all previous samples and new samples as inputs, has to be repeatedly implemented. It is obvious that this consumes much time for subsequently arrived data. Many existing manifold learning techniques do not contain an out-of-sample extension naturally, so it is necessary to find ways for extending manifold learning techniques to handle new samples.

Recently, Raducanu et al. [7] adopt the sparse representation approach (SRC [8]) to solve the out-of-sample problem for manifold learning. They demonstrate that the sparse representation theory is an alternative for out-of-sample embedding. However, the traditional sparse representation approach ignores the local information of the data. Thus it may select the samples not close to the new observations for representation. In [2, 3, 5], some researchers have pointed out that the data locality plays a very important role in manifold learning techniques. Furthermore, some recent researches also proved that the locality is more essential than sparsity in recognition and classification tasks [9]. Therefore, in order to capture the locality information of the data and improve the performance of the manifold learning method for recognition and classification tasks, a locality-constrained sparse representation algorithm is proposed in this paper to handle the out-of-sample problem in non-linear manifold learning. In our algorithm, a locality adaptor is introduced into the sparse representation procedure to obtain a more accurate embedding result.

The remaining of this paper is organized as follows. In Section 2, we introduce our proposed approach for the out of-sample problem based on data locality. Experimental results on four face datasets are presented in Section 3. Finally, in Section 4, we conclude the paper.

* Corresponding author

978-1-4673-7723-2/15 $31.00

DOI 10.1109/IMCCC.2015.8

II. LOCALITY-CONSTRAINED SPARSE REPRESENTATION FOR OUT-OF-SAMPLE EMBEDDING

As highlighted in the introduction, data locality has been widely utilized in the nonlinear dimension reduction [2, 3] and sparse representation. Thus, in order to achieve better embedding results of the new observations, the local information must be taken into account. In this section, we will detail the proposed locality-constrained sparse representation algorithm.

A. Projection of new samples

Firstly, we assume that $Y_s = (y_1, \cdots, y_N)$ is the low-dimensional embedding results of the seen samples $X_s = (x_1, \cdots, x_N)$. Then, the embedding result (denoted by y_{N+1}) of an unseen sample x_{N+1} in observed space can be computed by solving the following problem:

$$\begin{aligned} &\min \sum_{i=1}^{N} || y_{N+1} - y_i ||^2 \, W_{(N+1)i} \\ &= \sum_{i=1}^{N} (y_{N+1} - y_i)^T (y_{N+1} - y_i) W_{(N+1)i} \end{aligned} \tag{1}$$

where $W_{(N+1)\,i}$ is a similarity coefficient between x_{N+1} and x_i .

The derivative with respect to y_{N+1} of (1) is formulated as follow:

$$2\sum_{i=1}^{N} (y_{N+1} - y_i) W_{(N+1)i} = 0 \tag{2}$$

At last, the embedding y_{N+1} is given by:

$$y_{N+1} = \frac{\sum_{i=1}^{N} W_{(N+1)i} y_i}{\sum_{i=1}^{N} W_{(N+1)i}} \tag{3}$$

From (3), we can see that the embedding result of an unseen sample can be obtained by linear combination of all fixed embedded samples where the linear coefficients are set to the similarities between the unseen sample and the existing samples. When $W_{(N+1)\,i}$ is set to a kernel function, (3) is equivalent to the Laplacian Eigenmaps Latent Variable Model (LVM) [10]. When $W_{(N+1)\,i}$ is set to a sparse representation vector, (3) is equivalent to the method [7].

B. Computation of the similarity coefficients via locality-constrained sparse representation

From Section 2.1, we can see that the problem of out-of-sample embedding can be solved by linear combining the existing embedding results. Thus, let the vector $a = (W_{(N+1)1}, \cdots, W_{(N+1)N})^T$ be the similarities between the unseen sample x_{N+1} and all seen samples, our objective reduces to compute the vector a .

The proposed locality-constrained sparse representation is formulated as follow:

$$\begin{aligned} &\min || x_{N+1} - Xa ||_F^2 + \lambda || p \otimes a ||_2^2 \\ &s.t. \ \mathbf{1}^T a = 1 \end{aligned} \tag{4}$$

where λ is the regularization parameter balancing the two terms. The symbol $\otimes$ denotes the element-wise multiplication. The shift-invariant constraint $\mathbf{1}^T a = 1$ enforces the coding results a to remain the same even if the origin of the data coordinate system is shifted as proved in [11].

In this study, the locality adaptor p which measures the similarity between the unseen sample and seen samples is defined as:

$$p_{(N+1)i} = \exp\left(\frac{|| x_{N+1} - x_i ||^2}{\beta}\right) \tag{5}$$

where β is a suitable positive scalar. It is usually set to the average of squared distances between all pairs. From (5), we can find that if the seen sample is close to the unseen sample, the locality adaptor calculated by them is small.

It is clear that the coefficients with respect to the seen samples far away from the unseen sample are penalized heavily and those coefficients with respect to the seen samples close to the unseen sample will be well encouraged.

To determine the solution a for (4), we consider the Lagrange function $L(a,\eta)$, which is defined as:

$$L(a,\eta) = || x_{N+1} - Xa ||_2^2 + \lambda || p \otimes a ||_2^2 + \eta(\mathbf{1}^T a - 1) \tag{6}$$

which can be re-formulated as:

$$L(a,\eta) = a^T C a + \lambda a^T diag(p)^2 a + \eta(\mathbf{1}^T a - 1) \tag{7}$$

where $C = (x_{N+1}\mathbf{1}^T - X)^T (x_{N+1}\mathbf{1}^T - X)$ and $diag(p)$ is a diagonal matrix whose nonzero elements are the entries of p .

Let $\partial L(a,\eta) / \partial a = 0$, we have

$$\Phi a + \eta \mathbf{1} = 0 \tag{8}$$

where $\Phi = 2(C + \lambda diag(p)^2)$. Once we pre-multiply (8) by $\mathbf{1}^T \Phi^{-1}$, we obtain $\eta = -(\mathbf{1}^T \Phi^{-1} \mathbf{1})^{-1}$. Substituting η into (8) gives the analytical solution of (4) as:

$$\begin{aligned} \widetilde{a} &= (C + \lambda diag(p)^2)^{-1} \mathbf{1} \\ a &= \widetilde{a} / (\mathbf{1}^T \widetilde{a}) \end{aligned} \tag{9}$$

Then, the locality-constrained sparse representation coefficient vector a has been computed. Further, the similarity coefficients $W_{(N+1)i}$ are set to $W_{(N+1)i} = |\alpha_i|$, $(i = 1, \ldots, N)$, where α_i is the *i*-th element in the vector α .

Finally, through substituting the $W_{(N+1)i}$ into (3), we can get the embedding result of the unseen sample.

III. EXPERIMENTS AND RESULTS

In this section, we validate the efficacy of the proposed algorithm on four face datasets including Yale [12], AR [13], CMU PIE [14] and Extended YaleB [15]. We compare the performance of our proposed method with five state-of-the-art algorithms. In the first method LE_all, the training samples and the testing samples are all regarded as seen samples. Then, we use LE method to obtain the low-dimensional representations of all the samples. In the second method LE_part, the training samples are regarded as seen samples and the testing samples are regarded as unseen samples. For each unseen sample, we repeatedly use LE method to obtain the low-dimensional representation. In the third method LVM [10], a kernel function is firstly utilized to compute the similarities between the unseen sample and the existing samples. Then, the embedding of unseen sample is obtained by (3). In the fourth method [16], a simple linear regression is employed to infer a matrix transform A that best approximates the existing mapping through the linear equation $Y_s = A^T X_s$. In the fifth method, the sparse representation approach is adopted to solve the out-of-sample problem [7]. The proposed method and other algorithms used for comparison are all developed in Matlab and executed on a computer with Intel Core i7-2600 CPU at 3.4 GHz and 8 GB physical memory.

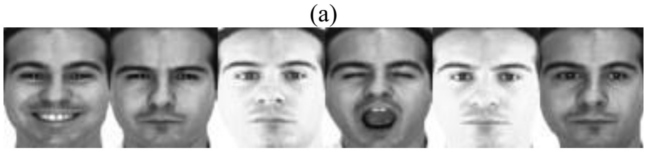
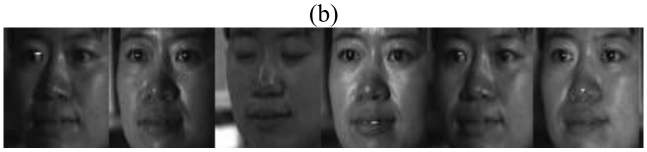
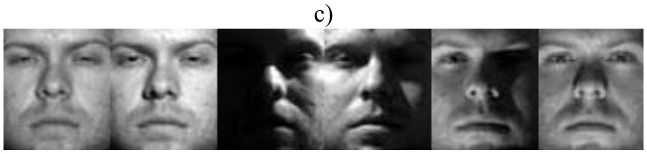

(a)

(b)

c)

(d)

Fig. 1. Some images of the databases used in the experiment. (a) Yale, (b) AR, (c) CMU PIE and (d) Extended YaleB

A. Datasets description

The AR database [13] contains over 4000 frontal images for 126 individuals. There are 26 face images available for each person, and the images are taken under different variations, including illumination, expression, and facial occlusion/disguise in two separate sessions. We choose a subset of the dataset consisting of 50 male and 50 female subjects. For each subject, 14 images with only illumination and expression variations were selected.

CMU PIE face database [14] includes 68 subjects with 41368 face images as a whole. The face images were captured by 13 synchronized cameras and 21 flashes, under varying pose, illumination, and expression. In our experiment, we use a reduced dataset containing 1632 face images of 68 individuals.

The Extended YaleB database [15] consists of 2432 frontal-face images of 38 subjects. The face images were captured under various laboratory-controlled lighting conditions.

In our experiment, all images are resized to the resolution of 32×32 pixels. Some images from the four databases described above are shown in Fig. 1.

B. Experimental results and analysis

To make the computation of the embedding process more efficient, the dimensionality of the original face samples was reduced by applying PCA.

For each face dataset, we randomly choose 30% images for training, and the remaining images are used for testing. The random training set selection is repeated 10 times. Here, the testing implies: (1) the out-of-sample embedding of the unseen data, (2) assigning it a class-label through the Nearest Neighbor classifier in the embedded space (recognition).

The average recognition rates versus feature dimensions of different algorithms can be seen in Figs. 2-5. From these figures, it can be found that with the increase in feature dimension at the beginning, the recognition rates of all algorithms are also improved. However, the trend is not maintained for LE_all and LE_part. After they achieve their top performances, the recognition results of LE_all and LE_part begin to decrease with the increase in feature dimension. We can also see that the performances of the proposed method are better than other algorithms on all four databases. This means that the local information is crucial for embedding the unseen samples in non-linear manifold learning.

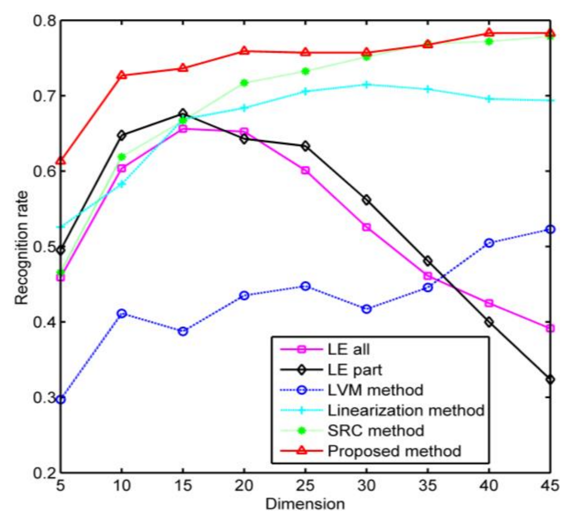


Fig. 2. Experimental results on the Yale dataset

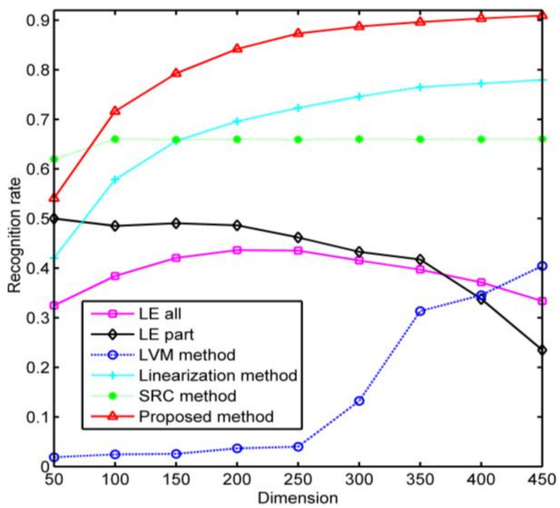


Fig. 3. Experimental results on the AR database.

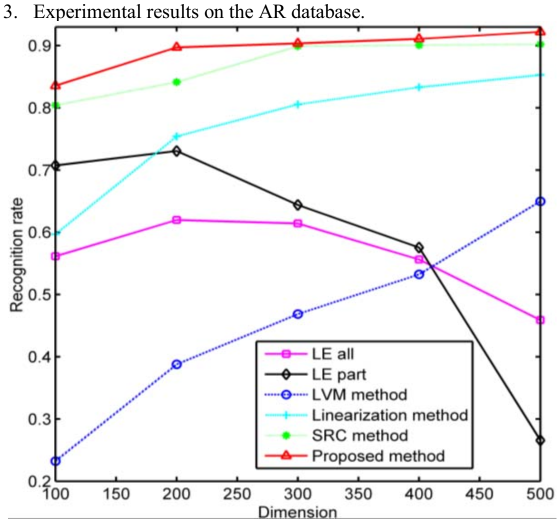


Fig. 4. Experimental results on the CMU PIE database.

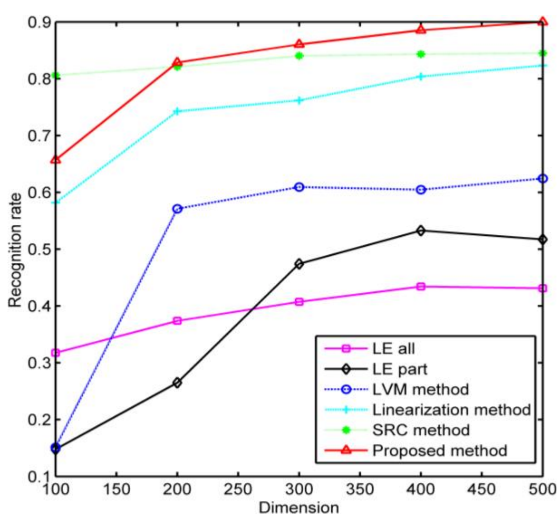


Fig. 5. Experimental results on the Extended YaleB database.

The best average recognition results obtained by different algorithms on four face databases are shown in Table I. From this table, the following points can be observed. Firstly, we can find that the recognition performances of SRC and the proposed method are better than other four methods. This is due to that both of these two methods are regression based algorithm and can obtain more accurate similarities between the unseen and existing samples. Secondly, it can be seen that the recognition rates obtained by our proposed method are superior to SRC. This is because the local information of the data is taken into consideration in our method.

TABLE I. THE BEST AVERAGE RECOGNITION RESULTS OBTAINED BY DIFFERENT ALGORITHMS ON FOUR FACE DATABASES

Methods	Yale	AR	CMU PIE	Extend YaleB
LE_all	0.6562 (15)	0.4364 (200)	0.6198 (200)	0.4342 (400)
LE_part	0.6762 (15)	0.5000 (50)	0.7306 (200)	0.5329 (400)
LVM method	0.5229 (45)	0.4044 (450)	0.6498 (500)	0.6244 (500)
Linearization method	0.7148 (30)	0.7794 (450)	0.8528 (500)	0.8233 (500)
SRC method	0.7783 (45)	0.6604 (450)	0.9020 (500)	0.8448 (500)
Proposed method	0.7829 (45)	0.9092 (450)	0.9219 (500)	0.8998 (500)

Note: The numbers in parentheses are the corresponding subspace dimension with the best result.

The average running times of three methods on four datasets is listed in Table II. In this study, we use publicly available packages such as l_1-*ls* [17] to solve the sparse representation coefficient vector in SRC. From Table II, the results clearly show that the proposed method saves much time than the traditional manifold learning and SRC methods for the out-of-sample problem.

TABLE II. THE AVERAGE RUNNING TIMES (S) OF THREE METHODS ON FOUR DATASETS

Databases	Proposed method	SRC method	LE-part
Yale	2.73	8.32	2.95
AR	111.40	235.75	287.67
CMU PIE	171.48	347.56	378.48
Extended YaleB	567.98	1099.96	1291.58

IV. CONCLUSIONS

Inspired by the pioneer work of sparse representation and the importance of data locality, a locality-constrained sparse representation method for solving the out-of-sample embedding problem in non-linear manifold learning is proposed. The experimental results show that not only the proposed method is able to achieve the competitive recognition rate than the existing methods, but also it can save more time than the traditional nonlinear dimensionality reduction methods for the out-of-sample problem. The results demonstrate the effectiveness of the proposed algorithm. In the future, in order to improve the recognition performance, we will try to find a new relationship between the new samples and the existing manifold, which can be a more accurate representation for out-of-sample embedding.

ACKNOWLEDGMENT

This work is supported by Heilongjiang Province Education Department's Scientific Research Project (No. 12541835).

REFERENCES

[1] H.S. Seung and D.D. Lee, “The manifold ways of perception”, Science, vol. 290, no. 5500, pp. 2268-2269, Dec. 2000.

[2] J.B. Tenenbaum, V.de Silva and J.C. Langford, “A global geometric framework for nonlinear dimensionality reduction”, Science, vol. 290,no. 5500, pp. 2319-2323, Dec. 2000.

[3] S.T. Roweis and L.K. Saul, “Nonlinear dimensionality reduction by locally linear embedding”, Science, vol. 290, no. 5500, pp. 2323-2326, Dec. 2000.

[4] A.M. Martinez and A.C. Kak, “PCA versus LDA”, IEEE Trans. Pattern Anal. Mach. Intell., vol. 23, no. 2, pp. 228-233, Feb. 2001.

[5] M. Belkin and P. Niyogi, “Laplacian eigenmaps for dimensionality reduction and data representation”, Neural Comput., vol. 15, no. 6, pp. 1373-1396, Jun. 2003.

[6] M.W. Trosset, C.E. Priebe, “The out-of-sample problem for classical multidimensional scaling”, Computational Statistics & Data Analysis., vol. 52, no. 10, pp. 4635-4642, Jun. 2008.

[7] B. Raducanu and F. Dornaika, “Embedding new observations via sparse-coding for non-linear manifold learning”, Pattern Recognition., vol. 47, no. 1, pp. 480-492, Jan. 2014.

[8] J. Wright, A. Yang, A. Ganesh, S. Sastry and Y. Ma, “Robust face recognition via sparse representation”, IEEE Transactions on Pattern Analysis and Machine Intelligence., vol. 31, no. 2, pp. 210-227, Feb. 2009.

[9] J.J. Wang, J.C. Yang, K. Yu, F.J. Lv, T. Huang and Y.H. Gong, “Locality-constrained Linear Coding for Image Classification”, IEEE International Conference on Computer Vision and Pattern Recognition (CVPR) , pp. 3360-3367, Jun. 2010.

[10] M.A. Carreira-Perpinán, Z. Lu, “The Laplacian Eigenmaps latent variable model”, Journal of Machine Learning Research, pp. 59-66, 2007.

[11] K. Yu, T. Zhang and Y. Gong, “Nonlinear learning using local coordinate coding”, Advances in Neural Information Processing Systems, vol. 9, pp. 2223-2231, Dec. 2009.

[12] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman, “Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection”, IEEE Trans. Pattern Anal. Mach. Intell., vol. 19, no. 7, pp. 711-720, Jul. 1997.

[13] A.M. Martinez and R. Benavente, “The AR face database”, CVC, Barcelona, Spain, Tech. Rep. 24, 1998.

[14] T. Sim, S. Baker, and M. Bsat, “The CMU pose, illumination, and expression database”, IEEE Trans. Pattern Anal. Mach. Intell., vol. 25, no. 12, pp. 1615-1618, Dec. 2003.

[15] A.S. Georghiades, P.N. Belhumeur, and D.J. Kriegman, “From few to many: Illumination cone models for face recognition under variable lighting and pose”, IEEE Trans. Pattern Anal. Mach. Intell., vol. 23, no. 6, pp. 643-660, Jun. 2001.

[16] D. Cai, X. He, J. Han, “Spectral regression for efficient regularized subspace learning”, IEEE 11th International Conference on Computer Vision, pp. 1-8, Oct. 2007.

[17] K. Koh, S.Kim, S.Boyd, “l1_ls: A matlab solver for large-scale l1-regularized least squares problems”, Stanford University, 2008