ISBN 978-981-07-8391-4
Proceedings of 2013 3rd International Conference on Computer Engineering and Bioinformatics (ICCEB 2013)
Bangkok, Thailand, November 23-24, 2013, pp. 8 -12

Locality Regularization Graph Embedding in Face Verification

Pang Ying Han[1], Teo Chuan Chin[1+], Andrew Teoh[2], Hiew Fu San[3], Ooi Shih Yin[1]

[1] Multimedia University, 75450 Melaka, Malaysia
[2] Yonsei University, Seoul, South Korea
[3] Infineon (Malaysia) Sdn Bhd, Melaka, Malaysia

Abstract. Graph embedding techniques attempt to construct a high locality projection in such a way that projected same class samples should be close to each other. However, estimation of population data locality could be severely biased due to limited number of training samples. This biased estimation could trigger overfitting problem, leading to poor generalization. In this paper, we propose three new dimensionality reduction techniques. Projection features are regularized by utilizing a local Laplacian matrix to better approach true data locality for higher locality preserving. These techniques are developed based on the manipulation of locality regularization. The difference between them would be the adoption of different local Laplacian matrices. In view of this, a common name is given, which is Locality Regularization Graph Embedding (denoted as LRGE). The robustness of these techniques is tested on CMU PIE and FERET databases. Experimental results validate the effectiveness of these techniques in face verification.

Keywords: Locality, Regularization, Graph embedding, Local Laplacian matrix, Face verification

1. Introduction

Graph embedding techniques are relatively new emerging techniques in face verification. These techniques are based on manifold preserving criterion that seeks underlying data structures by modelling local manifold based on data similarities on an affinity graph [1]. Representative examples of these techniques are Neighbourhood Preserving Embedding (NPE) [2], Locality Preserving Projection (LPP) [3], Marginal Fisher Analysis (MFA) [4] and etc.

In graph embedding framework, it is assumed that same class samples should be as close as possible to each other. Hence, high data locality preserving property is desired on dimensionality reduction techniques. With this, the projected samples from the same class are close to each other in the lower dimensional feature space. However, practically, the statistics and the nature of data population are estimated based on a finite number of available training samples. Due to limited number of training samples, the estimation could be biased and overfitted. Consequently, the discriminant function could be affected, leading to poor generalization. Therefore, another promising direction for better class discrimination is to regulate the projection directions. With this, locality preserving capability of the projection features is enhanced.

In this paper, a graph embedding regularization is studied to regularize projection directions for solving the overfitting problem and improving the locality preserving of projection features. Along with this research, three dimensionality reduction techniques are proposed. The robustness of these techniques is tested on two popular publicly available face databases: CMU Pose, Illumination, and Expression (CMU PIE) [5] and Facial Recognition Technology (FERET) [6]. Experimental results validate the effectiveness of these techniques in face verification.

[+] Corresponding author: Teo Chuan Chin. Tel.: +6 06 252 3214; fax: +6 06 231 8840.
E-mail address: ccteo@mmu.edu.my

2. Overview of Graph Embedding

Let $\mathbf{X} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n]$ with $\{\boldsymbol{x}_i \in \boldsymbol{R}^d | i = 1,2,\dots,n\}$ be a set of n numbers of d-dimensional data, $G = \{\mathbf{X}, \mathbf{W}\}$ be a weighted graph with vertex $\mathbf{X}$, $|\mathbf{X}| = n$ and weight matrix $\mathbf{W} \in \boldsymbol{R}^{n \times n}$. Each element in W signifies the similarity of vertex pairs [4]. Different definitions of graph G correspond to different graph embedding algorithms. For simplicity, one-dimensional case is explained and the low dimensional representations of vertices are represented as a vector $\boldsymbol{y} = [y_1, y_2, \dots, y_n]$. y_i is the low dimensional representation of vertex x_i. The target of the mapping is to make the vertices stay as close as possible via manifold criterion as defined below,

$$\boldsymbol{y}^* = \underset{\boldsymbol{y}\mathbf{B}\boldsymbol{y}^{\mathrm{T}}=b}{\operatorname{argmin}} \sum_{i \neq j} \left\| y_i - y_j \right\|^2 W_{ij} \tag{1}$$

where b is a constant and $\mathbf{B}$ is a constraint matrix for avoiding a trivial solution. With some simple algebraic manipulation, we have

$$\sum_{i,j} \left\| y_i - y_j \right\|^2 W_{ij} = 2\boldsymbol{y}\mathbf{L}\boldsymbol{y}^{\mathrm{T}} \tag{2}$$

where $\mathbf{L}$ is Laplacian matrix and $\mathbf{D}$ is a diagonal matrix defined as

$$\mathbf{L} = \mathbf{D} - \mathbf{W}, \ D_{ii} = \sum_j W_{ij}, \ \ \forall i \neq j \tag{3}$$

3. Locality Regularization Graph Embedding

Locality Regularization Graph Embedding (LRGE) is a linear graph embedding approach that based upon a locality preserving regulation model. In this approach, projection features are regulated to approach the true data locality for better locality preserving, leading to better discriminating capability.

3.1. Locality Preserving Capability in Graph Embedding

In graph embedding, data are distributed on an underlying manifold, M. Suppose that there is a map h: $M \to \mathbf{R}$. The gradient of h from the manifold space M to $\mathbf{R}$ is denoted as $\nabla h(\mathrm{x})$. For small $\delta(\mathrm{x})$ [7],

$$\left| h\big(\mathrm{x} + \delta(\mathrm{x})\big) - h(\mathrm{x}) \right| \approx |\langle \nabla h(\mathrm{x}), \delta(\mathrm{x}) \rangle| \leq \|\nabla h(\mathrm{x})\| \|\delta(\mathrm{x})\| \tag{4}$$

From Equation (4), it is noticed that data points near x will be mapped to data points near $h(\mathrm{x})$ if $\|\nabla h(\mathrm{x})\|$ is small. Belkin and Niyogi [8] defined a function as a measure metric of locality preserving capability on average of the map h,

$$\frac{\int_M \|\nabla h(\mathrm{x})\|^2 \, d\mathrm{x}}{\int_M |h(\mathrm{x})|^2 \, d\mathrm{x}} \tag{5}$$

The above equation has later been discretized by He and Niyogi [3] as $f(\upsilon) = \frac{\upsilon^{\mathrm{T}}\mathbf{XLX}^{\mathrm{T}}\upsilon}{\upsilon^{\mathrm{T}}\mathbf{XDX}^{\mathrm{T}}\upsilon}$ where $\mathbf{X}$ is a finite number of data samples and υ is a linear projective map. The smaller value of $f(\upsilon)$ is, the better the locality preserving capability of the vector υ.

3.2. Formulation of Locality Regularization Graph Embedding

Laplacian matrix, $\mathbf{L}$ in Equation (3) can be specifically divided into: global Laplacian $\mathbf{L}_{glo}$ and local Laplacian $\mathbf{L}_{loc}$. $\mathbf{L}_{glo}$ is the Laplacian of a graph in which all vertices are adjacent regardless their class membership and the weight of each edge is inversely proportional to the number of vertices ($W_{ij} \propto \frac{1}{n}$, n is the total number of vertices). $\mathbf{L}_{loc}$ is the Laplacian of a graph in which all vertices from the same class are adjacent. $\mathbf{L}_{loc}$ could be further specified into binary local Laplacian $\mathbf{L}_{bin\,loc}$, intra-class local Laplacian $\mathbf{L}_{class\,loc}$ and adjustable local Laplacian $\mathbf{L}_{adj\,loc}$. $\mathbf{L}_{bin\,loc}$ is the simple-minded Laplacian of a graph in which all vertices are grouped into c disjoint clusters and the vertices per cluster (with same class membership) are adjacent with 1-valued weight of each edge; $\mathbf{L}_{class\,loc}$ is the Laplacian of a graph in which all vertices are grouped into c disjoint clusters and the vertices per cluster (with same class membership) are adjacent with the weight of each edge that inversely proportional to the number of vertices of the corresponding cluster ($W_{ij} \propto \frac{1}{n_i}$, n_i is the total number of vertices in cluster i); $\mathbf{L}_{adj\,loc}$ is the Laplacian of a graph where $\mathbf{L}_{adj\,loc} = \mathbf{D} - \mathbf{W}$, $D_{ii} = \sum_j W_{ij}$, $\forall i \neq j$ and $\mathbf{W}$ is a weight matrix that signifies the similarity of vertex pairs that with same class membership. In this work, the edge weights $\mathbf{W}$ in $\mathbf{L}_{adj\,loc}$ are computed based on heat kernel [9].

As mentioned previously, graph embedding techniques assume samples from the same classes to be close to each other. Hence, this local neighbourhood geometry could be used to model the intrinsic data manifold and $\mathbf{L}_{loc}$ that comprising $\mathbf{L}_{bin\ loc}$, $\mathbf{L}_{class\ loc}$ and $\mathbf{L}_{adj\ loc}$ well reflects the locality. Thus, $\mathbf{L}_{loc}$ is substituted to $\mathbf{L}$ in Equation (3) to model the data manifold by solving the singular value decomposition problem,

$$\boldsymbol{\varphi} = \mathbf{V}^{\mathrm{T}}\mathbf{X}\mathbf{L}_{loc}\mathbf{X}^{\mathrm{T}}\mathbf{V} \tag{6}$$

where $\mathbf{V} = [\upsilon_{1,}\ \upsilon_{2,}\ \dots, \upsilon_d\ ]$ is the eigenvector matrix of $\mathbf{X}\mathbf{L}_{loc}\mathbf{X}^{\mathrm{T}}$ and $\boldsymbol{\varphi}$ is a diagonal matrix whose diagonal elements are the eigenvalue φ_i of the corresponding eigenvector υ_i , ${\upsilon_i}^{\mathrm{T}}\upsilon_i = 1$. Suppose that the eigenvalues are in descending order and Fig. 1 illustrates a plot of φ_i's corresponding to eigenvectors υ_i 's.

Based on the Rayleigh quotient format of Equation (5), there is a relationship between the locality preserving function $f(\upsilon_i)$ and the eigenvalue φ_i corresponding to υ_i in Equation (6). The smaller value of φ_i, the higher the locality preserving capability of the corresponding projective vector υ_i .

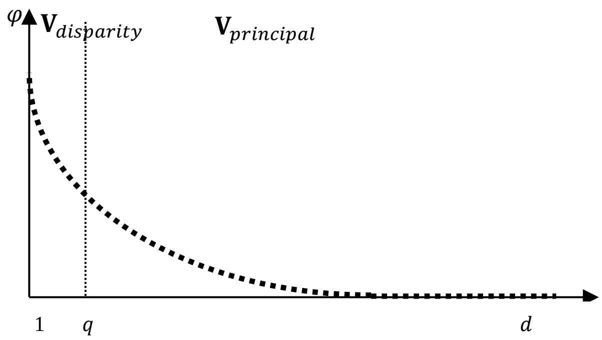

Fig. 1: A Plot of φ vs d: decomposition of the whole space V into disparity and principal subspaces.

In Fig. 1, it is observed that lower locality preserving capability, that corresponds to larger value of φ, is found on the first few eigenvectors. Hence, in devising a locality preserving regulation model, the entire space $\mathbf{V}$ is decomposed into two subspaces and different weights are imposed to each subspace. Specifically, larger weights are imposed to the subspace with high locality preserving; whereas smaller weighting factors are assigned to the subspace with low locality preserving. In this work, $\mathbf{V}$ is decomposed into two subspaces as follows: (I) *disparity subspace* that corresponds to low locality preserving $\mathbf{V}_{disparity} = [\upsilon_1, \upsilon_2, \dots, \upsilon_q]$, (II) *principal subspace* for high locality preserving $\mathbf{V}_{principal} = [\upsilon_{q+1}, \upsilon_2, \dots, \upsilon_d]$, as shown in Fig. 1.

The end point of the disparity subspace, denoted as q, is estimated by,

$$\varphi_q \approx \varphi_{disparity\ fence} \tag{7}$$

where $\varphi_{disparity\ fence} = \tau(Q3 + 1.5IQR)$ and τ is a scaling parameter for generalization.

Locality regularization model is developed on $\mathbf{V}_{disparity}$ and $\mathbf{V}_{principal}$ based on the principal that highlighted previously. Based on the principle, a piecewise weighting function is proposed as follow,

$$\gamma_i = \begin{cases} \frac{1}{\sqrt{\varphi_i}}, & 1 \le i \le q \\ \frac{1}{\sqrt{\varphi_q}}, & q+1 \le i \le d \end{cases} \tag{8}$$

Using the weighting function and the eigenvectorsυ_i 's, the data locality of the training data is improved by regularizing the data via,

$$\widetilde{\mathbf{X}} = \widetilde{\mathbf{V}}^{\mathrm{T}}\mathbf{X} \tag{9}$$

where $\widetilde{\mathbf{V}} = [\gamma_i \upsilon_i\ ]_{i=1}^{d}$. This regularization process prepares the data with more localized features, which is beneficial to discrimination, for the subsequent dimensionality reduction process.

LRGE is a linear graph embedding with the edge weights defined as

$$W_{jk}^{LRGE} = \begin{cases} \frac{1}{n_i} - \frac{1}{n}, & \text{if } jth \text{ point and } kth \text{ point are belonged to the ith class } C_i \\ -\frac{1}{n}, & \text{otherwise} \end{cases} \quad (10)$$

Based on the objective function of graph embedding in Equation (6), the objective function of LRGE is defined as,

$$\phi^* = \underset{\phi^{\mathrm{T}}\phi=1}{\operatorname{argmax}} \ \phi^{\mathrm{T}}\widetilde{\mathbf{X}}\mathbf{W}^{LRGE}\widetilde{\mathbf{X}}^{\mathrm{T}}\phi \quad (11)$$

where $\mathbf{W}^{LRGE}$ is a weight matrix and $\widetilde{\mathbf{X}} = [\widetilde{x}_1, \widetilde{x}_2, \ldots, \widetilde{x}_n]$ is the regularized training data computed in Equation (9). Since $\mathbf{L}_{loc}$ comprises $\mathbf{L}_{bin\ loc}$, $\mathbf{L}_{class\ loc}$ and $\mathbf{L}_{adj\ loc}$, LRGE could be specifically named as LRGE_$\mathbf{L}_{bin\ loc}$, LRGE_ $\mathbf{L}_{class\ loc}$ and LRGE_$\mathbf{L}_{adj\ loc}$ for the adoption of different Laplacian matrices.

4. Experimental Results and Discussions

The robustness of the three proposed techniques, LRGE_$\mathbf{L}_{bin\ loc}$, LRGE_ $\mathbf{L}_{class\ loc}$ and LRGE_$\mathbf{L}_{adj\ loc}$, is tested on two face databases: CMU PIE and FERET. The average error rates (AERs) (that is the average value of false accept rate (FAR) and false reject rate (FRR)) measured in this experiment serve as a performance measurement metric for the quality of the dimensionality reduction techniques.

The verification performance of the proposed techniques is evaluated and compared with other well-known dimensionality reduction techniques, such as PCA [10], LPP, NPE, LDA[11], supervised LPP - SLPP, supervised NPE - SNPE, MFA, LSDA [12], and ERE [13]. The best results obtained by the techniques on each database are recorded in Table 1.

The proposed LRGE_$\mathbf{L}_{bin\ loc}$, LRGE_ $\mathbf{L}_{class\ loc}$ and LRGE_$\mathbf{L}_{adj\ loc}$ exhibit their superiority in small number of features. The results suggest that LRGE is able to retrieve more discriminating features in the lower dimensional subspace.

Furthermore, it is observed that both verification performances of ERE and LRGE are comparable and superior to other techniques. This is due to the regularizations of class covariance matrix Σ_i in ERE and locality preserving function in LRGE enable better population statistics estimation.

ERE and LRGE can be distinguished in terms of algorithm designs. In ERE, based on noise fraction, eigenspace of class covariance matrix Σ_i is decomposed into three subspaces: face (very small noise variance to face variance ratio), noise (noise component is dominant in the variance) and null (zero variance based on training samples) subspaces. On the other hand, in LRGE, eigenspace of locality matrix $\mathbf{X}\mathbf{L}_{loc}\mathbf{X}^{\mathrm{T}}$ is decomposed based on locality preserving capabilityφ. Hence, only two subspaces are formed, which are disparity (with low locality preserving) and principal (with high locality preserving) subspaces.

Besides that, we examine a setting where training samples of testing subjects are not available. So, the training is learned from samples of other group of subjects. This testing is known as subject-independent strategy. There is no overlapping in subject between the training and testing sets. This strategy can measure the robustness and generalization of dimensionality reduction techniques effectively.

In the experiments, FERET 2 database, comprising 200 subjects with 10 samples per subject, is adopted. Samples from 100 subjects are selected for training; whereas, samples from the remaining 100 subjects are used for testing. The best results obtained by various techniques based on subject-independent strategy are also recorded in Table 1. It is observed that the verification performances of the supervised techniques are affected by the test protocols, i.e. subject-dependent test and subject-independent test. These techniques exhibit inferior verification performance in subject-independent test. The main reason is that the training samples and testing samples are from different faces / subjects. So, when new faces are inputted for verification, the techniques may not be well generalized to the new samples, leading to poorer verification performance. However, the proposed LRGE_$\mathbf{L}_{bin\ loc}$, LRGE_ $\mathbf{L}_{class\ loc}$ and LRGE_$\mathbf{L}_{adj\ loc}$ demonstrate the lowest verification error rate with 29%. This validates that the regularization of projection features in LRGE assures their data discriminating capability.

Table. 1: Performance Comparisons of LRGE on CMU PIE, FERET and FERET 2 Databases.

Method	CMU PIE	FERET	FERET 2
	Error rate (%)	Error rate (%)	Error rate (%)
Unsupervised methods			
PCA	66	40.04	43.42
LPP	45.23	40.37	42.11
NPE	50.57	40.1	42.27
Supervised methods			
LDA	28.67	39.06	43.81
SLPP	33.73	30.3	32.18
SNPE	31.33	31	38.78
MFA	27.64	35.8	42.78
LSDA	34.8	41.25	41.64
ERE	19.85	27.28	33.4
LRGE_$L_{bin\ loc}$	**20.57**	**25.54**	**29.44**
LRGE_$L_{class\ loc}$	**20.4**	**25.6**	**29.4**
LRGE_$L_{adj\ loc}$	**20.58**	**25.25**	**29.46**

5. Acknowledgements

The authors acknowledge the financial support of Telekom Research and Development Sdn. Bhd. of Malaysia

6. References

[1] Y. Fu, S. Yan, and T. Huang. Classification and feature extraction by simplexization. *IEEE Transactions on Information Forensics and Security*. 2008, **3**(1): 91-100.

[2] X. He, C. Deng, S. Yan, and H. Zhang. Neighborhood Preserving Embedding. In: *Proc. of the Tenth IEEE International Conference on Computer Vision*. 2005, pp. 1208-1213.

[3] X. He and P. Niyogi. Locality preserving projections. *Advances in Neural Information Processing Systems* ***16*** , 2003.

[4] S. Yan, D. Xu, B.Y. Zhang, H.J. Zhang, Q. Yang, and S. Lin. Graph embedding and extensions: a general framework for dimensionality reduction. *IEEE Trans. On Pattern Analysis and Machine Intelligence*. 2007, **29**(1): 40-51.

[5] T. Sim, S. Baker, and M. Bsat. The CMU Pose, Illumination, and Expression Database. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2003, **25**(12):1615-1618.

[6] P.J. Phillips. Overview of the face recognition grand challenge. In: *Proc. of the IEEE International Conference on Computer Vision and Pattern Recognition*. 2003, pp. 947-954.

[7] C. Deng, X. He, Y. Hu, J. Han, and H. Thomas. Learning a spatially smooth subspace for face recognition. In: *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*. 2007, pp. 1-7.

[8] M. Belkin, and P. Niyogi. Laplacian eigenmaps and spectral techniques for embedding and clustering. In: *Proc. of Conference on Advances in Neural Information Processing System 15*, 2001.

[9] X. He, S. Yan, Y. Hu, P. Niyogi, P. and H. Zhang. Face recognition using laplacianfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 2005, **27**(3): 328-340.

[10] M. Turk and A. Pentland. Eigenfaces for recognition. *J. Cognitive Neuroscience*. 1991, **3**(1): 71-86.

[11] P. Belhumeur, J. Hespanha, and D. Kriegman. Eigenfaces vs. Fisherfaces: recognition using class specific linear. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 1997, **19**: 711-720.

[12] C. Deng, X. He, K. Zhou, J. Han, and H. Bao. Locality sensitive discriminant analysis. In: *Proc. of International Joint Conference Artificial Intelligence*. 2007, pp. 708-713.

[13] X. Jiang, B. Mandal, and A.C. Kot. Eigenfeature regularization and extraction in face recognition. *IEEE Trans. Pattern Anal. Mach. Intell*, 2008, **30**(3): 383-394.