Expert Systems with Applications 37 (2010) 6718–6721

Contents lists available at ScienceDirect

Expert Systems with Applications

journal homepage: www.elsevier.com/locate/eswa

Short Communication

A novel local preserving projection scheme for use with face recognition

Yong Xu [a,*], Fengxi Song [b], Ge Feng [a], Yingnan Zhao [c]

[a] Harbin Institute of Technology, Shenzhen Graduate School, Shenzhen, Guang Dong 518055, China
[b] New Star Research Institute of Applied Technology, Hefei, Anhui 230031, China
[c] Nanjing University of Information Science and Technology, Nanjing, Jiangsu 210044, China

ARTICLE INFO

Keywords:
Locality preserving projection (LPP)
Objective function
Face recognition
Feature extraction

ABSTRACT

When locality preserving projection (LPP) is applied to face recognition, it usually suffers from the small sample size (SSS) problem, which means that the eigen-equation of LPP cannot be solved directly. In order to address this issue, we propose a novel LPP scheme. This scheme transforms the objective function of LPP into a new function, which allows the resultant eigen-equation to be directly solved no matter whether the SSS problem occurs or not. Moreover, the fact that the proposed scheme has an adjustable parameter enables us to be able to obtain the best classification accuracy by adjusting the parameter. Our analysis comprehensively reveals the theoretical properties of the proposed scheme and its relationship with other LPP methods. Our analysis also shows that the conventional LPP can be regarded as a special form of the proposed scheme, which also implies that the classification accuracy of the conventional LPP will be lower than the best classification accuracy of the proposed scheme.

1. Introduction

LPP can preserve the local structure of the data set when reducing the dimension of samples (Belkin & Niyogi, 2001; He, Cai, & Min, 2005; He & Niyogi, 2003; He, Yan, Hu, & Zhang, 2003; Min, Lu, & He, 2004). This means that after the LPP transformation, samples that were in close proximity in the original space remain in close proximity in the new space. Since LPP uses the same linear mapping to transform all the samples into a new space, it can also be viewed as a linear feature extraction method (Yang, Zhang, Yang, & Niu, 2007; Yan et al., 2007).

In the past several years, some improvements to the conventional LPP have been proposed. For example, in order to exploit the class membership relation, supervised locality preserving projection (SLPP) methods (Yu, Teng, & Liu, 2006; Zheng, Yang, Tan, Jia, & Yang, 2007) have been proposed.

LPP obtains its solution by solving an eigen-equation. It should be 8pointed out that for image recognition applications, the conventional LPP usually suffers from the small size (SSS) problem and the corresponding eigen-equation cannot be solved directly (He, Yan, Hu, Niyogi, & Zhang, 2005; Pan & Ruan, 2008). There are the following two reasons: The first reason is that the conventional LPP should first transform the 2D image matrices into one-dimensional vectors (hereafter they are called sample vectors) and consequently the sample vectors have a very high dimensionality. The second reason is that for the image recognition application, the number of the training samples is usually smaller than the dimensionality of the sample vectors and the matrices in the eigen-equation. As a result, the matrices in the generalized eigenvalue equation of the conventional LPP are singular and the SSS problem occurs (Feng, Hu, & Zhou, 2008; Hu, Feng, & Zhou, 2007; Xu & Zhang, 2008).

In this paper we propose a novel scheme, namely the LPP subtraction scheme, to implement LPP. The proposed scheme considers the locality projection preserving issue as a multi-objective programming issue. This allows us to obtain a nice LPP solution that can be easily obtained in all the cases including the SSS case. This paper has the following main contributions: First, it proposes, for the first time, the LPP subtraction scheme, which allows us to directly solve the eigen-equation of LPP no matter whether the SSS problem occurs or not. Second, it comprehensively reveals the theoretical properties of the LPP subtraction scheme. Third, it clearly shows that the conventional LPP can be regarded as a special form of the proposed scheme, which indeed implies the best classification accuracy of the proposed scheme will be better than the classification accuracy of the conventional LPP. Additionally, our experiments provide adequate information about the variation with the value of the parameter of the classification accuracy of the LPP subtraction scheme.

The remainder of this paper is organized as follows. In Section 2, we first briefly present the conventional LPP. In Section 3, we propose the LPP subtraction scheme, present its theoretical properties, reveal its relationship with other LPP methods and describe its feature extraction procedure. In Section 4, we test the proposed LPP scheme using a large face database and compare the classification

* Corresponding author. Tel.: +86 752 26032458; fax: +86 752 26032461.
E-mail address: laterfall2@yahoo.com.cn (Y. Xu).

0957-4174/$ - see front matter © 2010 Elsevier Ltd. All rights reserved.
doi:10.1016/j.eswa.2010.02.107

results of the proposed LPP scheme and other methods. Section 5 offers our conclusion.

2. LPP subtraction scheme

In this section we first present the eigen-equation of LPP and then propose the LPP subtraction scheme that can easily work out the LPP solution no matter whether the SSS problem occurs or not. The goal of the conventional LPP is to work out the maximum eigenvalue solution of

$$XWX^Tz = \lambda XDX^Tz, \tag{1}$$

where $X = [x_1 \quad x_2 \quad \dots \quad x_N]$. x_i represents the one-dimensional vector corresponding to the ith sample. In practice, one should convert (1) into $(XDX^T)^{-1}XWX^Tz = \lambda z$ and then solve this equation to obtain its maximum eigenvalue solution. In the case where the SSS problem occurs, since XDX^T is singular, this equation cannot be solved directly.

We present the LPP subtraction scheme as follows. Let $\overline{W} = XWX^T$, $\overline{D} = XDX^T$. We can regard that the maximum eigenvalue solution of (1) is equivalent to the solution that maximizes the following objective functions:

$$\max\left(\frac{z^T\overline{W}z}{z^Tz}\right), \tag{2}$$

$$\min\frac{z^T\overline{D}z}{z^Tz}. \tag{3}$$

Then we face a problem of solving a multi-objective programming problem with the objective functions (2) and (3). This multi-objective programming problem cannot be solved directly, so we convert it into a single-objective programming problem. We use the additive rule to combine (2) and (3) to produce the following single objective optimization problem:

$$\max\left(\frac{z^T\overline{W}z}{z^Tz} - \frac{z^T\overline{D}z}{z^Tz}\right) = \max\frac{z^T(\overline{W} - \mu\overline{D})z}{z^Tz}, \tag{4}$$

where μ represents a real number.

If z^Tz satisfies the constraint $z^Tz = 1$, then (4) can be converted into the objective function $\max z^T(\overline{W} - \mu\overline{D})z$. Thus, in order to obtain the vector z that satisfies this objective function under the constraint $z^Tz = 1$, we define the following Lagrangian function:

$$L(z, \mu) = z^T(\overline{W} - \mu\overline{D})z - \lambda(z^Tz - 1). \tag{5}$$

If (5) reaches its extreme maximum value, $\frac{z^T(\overline{W}-\mu\overline{D})z}{z^Tz}$ will also obtain its extreme maximum value.

Let $\frac{\partial L(z,\mu)}{\partial z} = 0$, then we have

$$(\overline{W} - \mu\overline{D})z = \lambda z. \tag{6}$$

We note that (6) is an eigen-equation. As a result, we can conclude that the optimal projecting axis should be the eigenvector corresponding to the largest eigenvalue of the matrix $\overline{W} - \mu\overline{D}$. If k projecting axes are needed, we should take the k eigenvectors corresponding to the first k largest eigenvalues of $\overline{W} - \mu\overline{D}$ as the k projecting axes. This scheme is referred to as LPP subtraction scheme. By substituting Eq. (6) into $z^T(\overline{W} - \mu\overline{D})z$ we obtain

$$z^T(\overline{W} - \mu\overline{D})z = \lambda. \tag{7}$$

The major advantages of the LPP subtraction scheme are as follows: First, differing from the conventional LPP scheme which is inapplicable to the case where $\overline{D}$ is singular, the eigen-equation of the conventional LPP scheme (Eq. (6)) can be directly solved in all the cases no matter whether the matrix $\overline{D}$ is singular or not. However, when the SSS problem occurs, we cannot directly solve the eigen-equation of the conventional LPP. Second, when we directly solve (6) no matrix inverse operation should be performed, whereas when solving the eigen-equation of the conventional LPP scheme we should calculate the inverse matrix of XDX^T in advance. As we know, the matrix inverse operation needs a high computational cost. Thus, solving the eigen-equation (6) with fixed μ of the LPP subtraction scheme needs a lower computational cost than solving the eigen-equation of the conventional LPP scheme.

3. Analysis on the LPP subtraction scheme

3.1. The LPP subtraction scheme with nonsingular $\overline{D}$

In this subsection we will analyze properties of the LPP subtraction scheme that has nonsingular $\overline{D}$ and show that under the condition of nonsingular $\overline{D}$ the conventional LPP is a special form of the proposed LPP subtraction scheme. Let z_0 represent the optimal projection direction determined by the LPP subtraction scheme. Since the LPP subtraction scheme has the parameter μ, we define the following function

$$f(\mu) = \frac{z_0^T(\overline{W} - \mu\overline{D})z_0}{z_0^Tz_0}. \tag{8}$$

Eq. (8) can show how the capability of locality preserving projection of z_0 varies with the parameter μ. Indeed we can also use Eq. (8) to assess the variation with μ of the capability of locality preserving projection of arbitrary projection direction z, if we replace z_0 in (8) with z.

We note that $f(\mu) = \max_{z\neq 0}\frac{z^T(\overline{W}-\mu\overline{D})z}{z^Tz}$. Based on (7), we know that $f(\mu)$ is indeed the largest eigenvalue of the matrix $\overline{W} - \mu\overline{D}$. This can be easily demonstrated by using the extremum property of generalized Rayleigh quotient.

In the context below, we explore more properties of the LPP subtraction scheme by proving two theorems.

Theorem 1. *$f(\mu)$ is a monotone decreasing function. Especially, if the matrix $\overline{D}$ is nonsingular, $f(\mu)$ will be a strictly monotone decreasing function.*

Proof. Let $\mu_1 < \mu_2$. Suppose that z_1 and z_2 are the unit eigenvectors corresponding to the largest eigenvalues of the matrices $\overline{W} - \mu_1\overline{D}$ and $\overline{W} - \mu_2\overline{D}$, respectively. Then we have $f(\mu_1) = z_1^T(\overline{W} - \mu_1\overline{D})z_1 \geqslant z_2^T(\overline{W} - \mu_1\overline{D})z_2$. One the other hand, $z_2^T(\overline{W} - \mu_1\overline{D})z_2 = z_2^T(\overline{W} - \mu_2\overline{D})z_2 + (\mu_2 - \mu_1)z_2^T\overline{D}z_2 = f(\mu_2) + (\mu_2 - \mu_1)z_2^T\overline{D}z_2$. Since $\overline{D}$ is semi-positive definite, we have $z_2^T\overline{D}z_2 \geqslant 0$. Thus, $f(\mu_1) \geqslant f(\mu_2)$ is certain. That is, $f(\mu)$ is a monotone decreasing function. □

If $\overline{D}$ is non-singular, then $\overline{D}$ is positive definite and we have $z_2^T\overline{D}z_2 > 0$. As a consequence, we can conclude that if $\overline{D}$ is non-singular, $f(\mu_1) > f(\mu_2)$ will be always satisfied. That is, if $\overline{D}$ is non-singular, $f(\mu)$ is a strictly monotone decreasing function.

Theorem 2. *If $\overline{D}$ is nonsingular, it can be demonstrated that the conventional LPP is a special form of the proposed LPP subtraction scheme.*

Proof. First of all, we know that $f(0) = \max_{z\neq 0}\frac{z^T\overline{W}z}{z^Tz} = \lambda_{\overline{W}} > 0$, where $\lambda_{\overline{W}}$ stands for the largest eigenvalue of $\overline{W}$. If z is an arbitrary unit vector, we obtain

$$z^T\overline{W}z \leqslant \lambda_{\overline{W}}. \tag{9}$$

Let $\lambda_{\overline{D}}$ denote the smallest eigenvalue of $\overline{D}$. For an arbitrary unit vector z, we have

$$z^T\overline{D}z \geqslant \lambda_{\overline{D}} > 0, \tag{10}$$

The combination of (9) and (10) allows us to obtain

$$f(\mu) \leqslant \lambda_{\overline{W}} - \mu\lambda_{\overline{D}}. \tag{11}$$

According to (11), we know that $f(\mu_1) < 0$ when $\mu_1 > \lambda_{\overline{W}}/\lambda_{\overline{D}}$.

According to Cao (1980), for a continuous function $f(\mu)$, there must exist such a point μ' in the interval $(0, \mu_1)$ that $f(\mu') = 0$. Since $f(\mu)$ defined in this paper is a strictly monotone function, there will be an unique μ' that satisfies $f(\mu') = 0$.

$f(\mu') = 0$ means that $(\overline{W} - \mu'\overline{D})z(\mu') = 0$, i.e. $\overline{W}z(\mu') = \mu'\overline{D}z(\mu')$, which is indeed the eigen-equation of the conventional LPP. This shows that if the parameter μ of the LPP subtraction scheme is set to such a value that allows $f(\mu) = 0$ to be satisfied, then the LPP subtraction scheme is identical to the conventional LPP. Thus, we can conclude that the conventional LPP is a special form of the proposed LPP subtraction scheme. □

3.2. Asymptotic characteristic of the LPP subtraction scheme

With this subsection, we show the asymptotic characteristic of the LPP subtraction scheme. We start with the following theorem.

Theorem 3. *Under the condition that $\overline{D}$ is nonsingular, the $z(\mu)$ produced by the proposed subtraction scheme approximates the unit eigenvector corresponding to the smallest eigenvalue $\lambda_{\overline{D}}$ of $\overline{D}$ when the parameter μ approximates positive infinity.*

Proof. Let z' represent the unit eigenvector corresponding to the smallest eigenvalue $\lambda_{\overline{D}}$ of matrix $\overline{D}$. Let z_μ be the unit eigenvector corresponding to the largest eigenvalue of matrix $\overline{W} - \mu\overline{D}$. Then, for any unit vector z, we have:

$$z_\mu^T(\overline{W} - \mu\overline{D})z_\mu \geqslant z^T(\overline{W} - \mu\overline{D})z. \quad (12)$$

Let $v_\mu = z_\mu - z'$. Then we have the following inequality:

$$(v_\mu + z')^T(\overline{W} - \mu\overline{D})(v_\mu + z') \geqslant z'^T(\overline{W} - \mu\overline{D})z'. \quad (13)$$

This can be rewritten as:

$$\mu(v_\mu^T\overline{D}v_\mu + 2v_\mu^T\overline{D}z') \leqslant z_\mu^T\overline{W}z_\mu - (z')^T\overline{W}z' \leqslant z_\mu^T\overline{W}z_\mu + (z')^T\overline{W}z' \leqslant 2\lambda_{\overline{W}}, \quad (14)$$

where $\lambda_{\overline{W}}$ still represents the maximum eigenvalue of $\overline{W}$.

Since $\overline{D}z' = \lambda_{\overline{D}}z'$, the inequality (14) can be rewritten as:

$$v_\mu^T\overline{D}v_\mu + 2\lambda_{\overline{D}}v_\mu^Tz' \leqslant 2\lambda_{\overline{W}}/\mu. \quad (15)$$

Since $z_\mu = v_\mu + z'$ and $z_\mu^Tz_\mu = (z')^Tz' = 1$ we have:

$$v_\mu^Tv_\mu = -2v_\mu^Tz'. \quad (16)$$

By substituting Eq. (16) into (14), we obtain:

$$v_\mu^T\overline{D}v_\mu - \lambda_{\overline{D}}v_\mu^Tv_\mu \leqslant 2\lambda_{\overline{W}}/\mu. \quad (17)$$

From (17), we have $\lim_{\mu\to\infty}(2\lambda_{\overline{W}}/\mu) = 0$. In other words, when $\mu \to \infty$, we obtain $v_\mu^T\overline{D}v_\mu - \lambda_{\overline{D}}v_\mu^Tv_\mu \leqslant 0$. On the other hand, since $\lambda_{\overline{D}}$ is the smallest eigenvalue of the matrix, we should have $v_\mu^T\overline{D}v_\mu \geqslant \lambda_{\overline{D}}v_\mu^Tv_\mu$. As a result, we know that:

$$\lim_{\mu\to\infty} v_\mu^T\overline{D}v_\mu = \lambda_{\overline{D}}v_\mu^Tv_\mu. \quad (18)$$

According to the definition of v_μ we have $\lim_{\mu\to\infty} v_\mu = 0$, i.e.

$$\lim_{\mu\to\infty} z_\mu = z'. \quad (19)$$

This completes the proof of Theorem 3. □

3.3. Feature extraction and classification procedure

For a multi-class problem, there are more than one effective discriminant vector. Thus, we can first take the eigenvector corresponding to the maximum eigenvalue of Eq. (6) as the first projecting axis. If k projecting axes are required, we should take the k eigenvectors corresponding to the first k largest eigenvalues of Eq. (6) as the projecting axes. These projecting axes are orthogonal to each other and they form a non-redundant coordinate system to represent the samples.

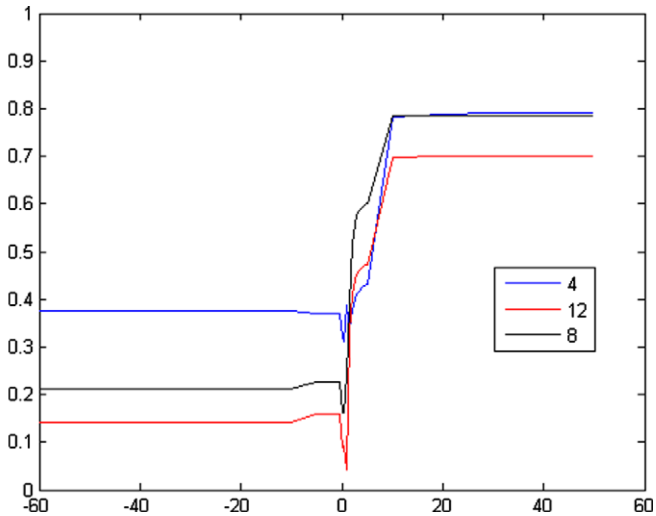

Fig. 1. Variation with the value of the parameter μ of the classification error rate of the supervised LPP subtraction scheme. This figure shows the error rates when 4, 8 and 12 images of each subject are used as training samples, respectively. The x-axis represents the value of the parameter μ while the y-axis stands for the mean of the error rate.

The feature extraction and classification procedure of the multi-class problem is as follows:

Step 1. Solve the eigenvectors corresponding to the first k largest eigenvalues of Eq. (6). Take them as projecting axes.

Step 2. Project each training and each test sample onto each of the k projecting axes to produce its respective feature vector.

Step 3. Use the nearest neighbor to classify the test samples.

4. Experiments

We test the LPP subtraction scheme using the AR face database. In this database, each subject has 26 images. We transformed each face image of the AR face database to a 50×40 matrix using the down-sampling scheme shown in Xu and Jin (2008). We tested the methods in the cases where different numbers of images per subject were used as training samples.

We first generated a series of random numbers and then selected training samples using these random numbers. For example, when we performed the experiment on eight training samples per subject, we first generated a series of random numbers including eight numbers with the range of from 1 to 26. Then we checked the filename of each face image to determine whether it should be taken as a training sample or not. For a face image of each subject, if the number included in its filename was same as one of the random number, then the image was selected as one training sample; otherwise it was used as one test sample. In order to obtain representative experimental result under the condition of a fixed number of training samples per subject, we tested each of the methods 5 times each using different series of random numbers.

We should point out that the LPP subtraction scheme proposed in this paper is an unsupervised LPP scheme. As we know, the difference between unsupervised LPP and supervised LPP is only from the definitions of the matrices W and D. If we define the matrices W and D in the same way as work (Zheng et al., 2007), the proposed LPP subtraction scheme becomes a supervised LPP scheme and can be referred to as supervised LPP subtraction scheme. We use Fig. 1 to show the variation with the value of the parameter μ of the mean of the classification error rates of the supervised LPP subtraction scheme. Fig. 1 shows that when μ is set to a negative value, the supervised LPP subtraction scheme usually obtains a low classification error rate. On the other hand, when μ is set to a large enough positive value, the supervised LPP subtraction scheme produces a high classification error rate. Fig. 1 also shows that when μ is set to a small enough positive value the supervised LPP subtraction scheme obtains its lowest error rate.

5. Conclusion

The LPP subtraction scheme proposed in this paper allows us to obtain a LPP solution in a computationally tractable way. The analysis on the theoretical properties of the proposed scheme clearly shows its asymptotic characteristics and reveals the relationship between the LPP subtraction scheme and other LPP methods. Compared to the conventional LPP, the LPP subtraction scheme has the following main advantages: The first is that the corresponding eigen-equation can be directly solved no matter whether the SSS problem occurs or not. The second is that solving the eigen-equation with fixed μ of the LPP subtraction scheme needs a lower computational cost than solving the eigen-equation of the conventional LPP scheme. The third is that by adjusting the parameter the proposed scheme is able to produce a higher recognition accuracy than the conventional LPP. Experimental results provide adequate information about the variation with the value of the parameter of the classification error of the LPP subtraction scheme.

Acknowledgments

This article is partly supported by National Nature Science Committee of China under Grants Nos. 60975006, 60803090 and 60702076.

References

Belkin, M., & Niyogi, P. (2001). Laplacian eigenmaps and spectral techniques for embedding and clustering. In *Proceedings of the conference on advances in neural information processing systems*.

Cao, Z. (1980). *Eigenvalue problems*. Shanghai: Shanghai Science and Technology Press.

Feng, G., Hu, D., & Zhou, Z. (2008). A direct locality preserving projections (DLPP) algorithm for image recognition. *Neural Processing Letters, 27*(3), 247–255.

He, X., & Niyogi, P. (2003). Locality preserving projections. In *Proceedings of the conference on advances in nerual information processing systems*.

He, X., Yan, S., Hu, Y., & Zhang, H. (2003). Learning a locality preserving subspace for visual recognition. In *9th IEEE international conference on computer vision (ICCV 2003)* (pp. 385–393).

He, X., Cai, D., & Min, W. (2005). Statistical and computational analysis of locality preserving projection. In *Proceedings of the 22nd international conference on machine learning* (pp. 281–288).

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

Hu, D., Feng, G., & Zhou, Z. (2007). Two-dimensional locality preserving projections (2DLPP) with its application to palmprint recognition. *Pattern Recognition, 40*(1), 339–342.

Min, W., Lu, K., & He, X. (2004). Locality preserving projection. *Pattern Recognition, 37*(4), 781–788.

Pan, X., & Ruan, Q.-Q. (2008). Palmprint recognition with improved two-dimensional locality preserving projections. *Image and Vision Computing, 26*(9), 1261–1268.

Xu, Y., & Jin, Z. (2008). Down-sampling face images and low-resolution face recognition. In *The third international conference on innovative computing, information and control, June 18–20, 2008, Dalian, China* (pp. 392–395).

Xu, Y., & Zhang, D. (2008). A new solution scheme of unsupervised locality preserving projection method for the SSS problem. In N. da Vitora Lobo et al. (Eds.), *SSPR&SPR 2008. LNCS* (Vol. 5342, pp. 785–791).

Yang, J., Zhang, D., Yang, J.-Y., & Niu, B. (2007). Globally maximizing, locally minimizing: Unsupervised discriminant projection with applications to face and palm biometrics. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 29*(4), 650–664.

Yan, S., Xu, D., Zhang, B., Zhang, H.-J., Yang, Q., & Lin, S. (2007). Graph embedding and extension: A general framework for dimensionality reduction. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 29*(1).

Yu, W., Teng, X., & Liu, C. (2006). Face recongnition using discriminant locality preserving projections. *Image and Vision Computing, 24*, 239–248.

Zheng, Z., Yang, F., Tan, W., Jia, J., & Yang, J. (2007). Gabor feature-based face recognition using supervised locality preserving projections. *Signal processing, 87*, 2473–2483.