Improved Principal Component Regression for Face Recognition Under Illumination Variations

Shih-Ming Huang and Jar-Ferr Yang, *Fellow, IEEE*

***Abstract*—The uncontrollable illumination problem is a great challenge for face recognition. In this paper, we propose a novel face recognition framework, the improved principal component regression classification (IPCRC) algorithm, which could overcome the problem of multicollinearity in linear regression. The IPCRC approach first performs principal component analysis (PCA) process to project the face images onto the face space. The first n principal components are intentionally dropped to boost the robustness against illumination changes. Then, the linear regression classification (LRC) is executed on the projected data and the identity is determined by the minimum reconstruction error. Experiments carried out on Yale B and FERET facial databases reveal that the proposed framework outperforms the state-of-the-art methods and demonstrates promising abilities against severe illumination variation.**

***Index Terms*—Face recognition, improved principal component regression classification.**

I. Introduction

FACE recognition approach [1] is designed to distinguish a specific identity from the unknown objects characterized by face images and is a very active research field because of its wide range of applications, such as video surveillance, access control etc. In our daily life, illumination variation is a very common problem and is a great challenge to face recognition. Varying illumination conditions deteriorate face recognition performance dramatically because the same person may appear different faces under varying lighting conditions.

In the literature, numerous researches have been proposed to achieve successful face recognition. These approaches could be categorized into two categories namely *reconstructive* and *discriminative* methods. The reconstructive approaches such as principal component analysis (PCA) [2]–[4] and independent component analysis (ICA) [5], [6] have been reported to be robust for the problem related to noisy conditions. The discriminative approach, such as linear discriminant analysis (LDA) [3], [4], has been known to yield better results in clean conditions. Subsequently, a variety of LDA-based approaches have been presented to achieve higher performance. Although the LDA-based face recognition algorithm reaches better performance under variable lighting conditions, the performance is still not satisfactory under severe illumination variation. To resolve such problem, the idea by applying the kernel trick is to nonlinearly map the input image from the input space to a higher dimensional feature space and then perform LDA in the mapped feature space. By performing this mapping nonlinearly, the nonlinear distribution becomes linearly separable in the higher dimensional feature space. Therefore, many kernel-based PCA and LDA approaches [7] such as the kernel PCA (KPCA) and the kernel Fisher discriminant (KFD) have been developed in face recognition. Moreover, Baudat and Anouar [8] proposed the generalized discriminant analysis (GDA) method by extending the KFD method. Their experiments on Iris and Seed data show that the GDA outperforms linear techniques, and the kernel-based methods achieve higher performance. Lu *et al.* proposed the kernel direct LDA (KDDA) method [9] by combining the GDA and direct LDA (D-LDA). In 2007, Huang *et al.* proposed the kernel subspace LDA (KSLDA) method [10], which could automatically tune the parameters of the Gaussian kernel. Experimental results show that the existing LDA-based methods are good and feasible to tackle illumination variations. However, the performance is still not satisfactory for severe illumination variations.

Recently, in 2010, a linear regression classification (LRC) algorithm [11] has been introduced for face recognition, which is based on that face images from a specific class are known to lie on a linear subspace. The regression coefficients are estimated by using the least square method, and then the decision is made in favor of the class with the minimum reconstruction error. Experiments have shown that the downsampled image could be used for classification directly. Although the LRC outperforms the state-of-the-art methods in some cases, it still could not withstand severe illumination variations.

The novelty of this paper is to propose an improved principal component regression (IPCR), which could overcome the problem of multicollinearity in linear regression. The main difference between PCR and IPCR approaches is to drop different principal components for LRC. The PCR drops the last n principal components whereas the IPCR drops the first n ones. Finally, the drop of the first n principal components gains the improvement significantly in face recognition under illumination variations.

II. Principal Component Regression Classification

In this section, a novel face recognition algorithm based on the principal component regression (PCR) is proposed. The PCR is a linear regression method which could overcome the

Manuscript received November 25, 2011; revised January 09, 2012; accepted January 12, 2012. Date of publication January 23, 2012; date of current version February 02, 2012. This work was supported in part by the National Science Council of Taiwan under Grant NSC 100-2218-E-006-017. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Lisimachos Paul Kondi.

The authors are with Institute of Computer and Communication Engineering, Department of Electrical Engineering, National Cheng Kung University, Tainan 70101, Taiwan (e-mail: smhuang@video5.ee.ncku.edu.tw; jf.yang@video5.ee.ncku.edu.tw).

Digital Object Identifier 10.1109/LSP.2012.2185492

problem of multicollinearity. Since face images have similar structure, the image vectors are highly correlated such that it is very impossible to reliably estimate regression coefficients. Although the PCR could counteract the problem of multicollinearity, the mean-square-error criterion, underlying the PCA embedded in the PCR, prefers the low frequencies including the illumination variation to the high frequencies containing the discriminative information. Thus, we suggest further dropping the first n principal components to boost the robustness against illumination changes.

The proposed improved principal component regression classification (IPCRC) algorithm is a two-step classification method. In the first step, the PCA of the measured data transforms the observed variables into the new decorrelated components. Next, the first n principal components are further dropped. In the second step, the projected coefficients are then used in regression such that we can estimate reliable regression coefficients for each subject for face recognition.

The PCA is widely used for dimensionality reduction in computer vision fields, especially for face recognition technology. The *eigenfaces* based on the PCA was presented in face recognition [2]. In the PCA, the data is represented as a linear combination of an orthonormal vectors that maximize the data scatter (covariance) across all images. As to the PCA transform, many variants [12] have been presented to face recognition. In our research, the PCA (i.e., the traditional PCA) and the PCA with zero average (PCAZ) will be applied in linear regression.

A. Traditional PCA

Mathematically, given a set of M training images and each gray scale training image is in size of $a \times b$ and is represented as $\boldsymbol{v}_m \in \Re^{a\times b}$, $m = 1, 2, \ldots, M$. Each image is transformed to a column vector, $\boldsymbol{x}_m \in \Re^{L\times 1}$, where $L = ab$. By stacking all $\boldsymbol{x}_m$, we have the collected data given as

$$\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m, \ldots, \boldsymbol{x}_M] \in \Re^{L\times M}. \tag{1}$$

By subtracting the mean vector $\boldsymbol{\mu}_x = (1/M)\sum_{m=1}^{M}\boldsymbol{x}_m$, we have the zero-mean image column data matrix, i.e., $\widetilde{\boldsymbol{X}} = \boldsymbol{X} - \boldsymbol{\mu}_x$. To perform the PCA, the sample covariance matrix is given by

$$\boldsymbol{\Sigma} = \frac{1}{M}\widetilde{\boldsymbol{X}}\widetilde{\boldsymbol{X}}^T. \tag{2}$$

After performing the eigen-decomposition of the covariance matrix, $\boldsymbol{\Sigma}$, we obtain

$$\boldsymbol{\Lambda} = \boldsymbol{P}^T\boldsymbol{\Sigma}\boldsymbol{P} \tag{3}$$

where $\boldsymbol{P} = [\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_k]$ and $\boldsymbol{\Lambda} = diag(\lambda_1, \lambda_2, \ldots, \lambda_k)$ with the sorted eigenvalues as $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_k$ and their associated eigenvectors $\boldsymbol{p}_1, \boldsymbol{p}_2, \ldots, \boldsymbol{p}_k$ of $\boldsymbol{\Sigma}$. Thus, the ordinary PCA projection is given by

$$\boldsymbol{Y} = \boldsymbol{P}^T\widetilde{\boldsymbol{X}}. \tag{4}$$

Therefore, the correlation of input data can be removed. Due to the orthogonal property among eigenvectors, one principal component cannot be linearly predicted from the others.

B. PCA With Zero Average (PCAZ)

Given a set of M training samples, the image data matrix, $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m, \ldots, \boldsymbol{x}_M]$, instead of the subtraction of the mean of column vectors, the PCAZ would subtract the data vector by the mean of row vectors. In other words, each column vector $\boldsymbol{x}_m$ will remove its mean, $\overline{\boldsymbol{x}}_m = (1/L)\sum_{l=1}^{L}\boldsymbol{x}_{m,l}$, where $\boldsymbol{x}_{m,l}$ is the l^{th} component of $\boldsymbol{x}_m$. Then, the zero-average data matrix becomes $\hat{\boldsymbol{X}} = [\boldsymbol{x}_1 - \overline{x}_1, \ldots, \boldsymbol{x}_M - \overline{x}_M]$. Finally, similar to the PCA, the sample covariance matrix for the PCAZ is given by

$$\hat{\boldsymbol{\Sigma}} = \frac{1}{M}\hat{\boldsymbol{X}}\hat{\boldsymbol{X}}^T. \tag{5}$$

After eigen-decomposition of the covariance matrix $\hat{\boldsymbol{\Sigma}}$, we have

$$\hat{\boldsymbol{\Lambda}} = \hat{\boldsymbol{P}}^T\hat{\boldsymbol{\Sigma}}\hat{\boldsymbol{P}} \tag{6}$$

where $\hat{\boldsymbol{P}} = [\hat{\boldsymbol{p}}_1, \hat{\boldsymbol{p}}_2, \ldots, \hat{\boldsymbol{p}}_k]$ and $\hat{\boldsymbol{\Lambda}} = diag(\hat{\lambda}_1, \hat{\lambda}_2, \ldots, \hat{\lambda}_k)$ with eigenvalues as $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \cdots \geq \hat{\lambda}_k$ and their corresponding eigenvectors $\hat{\boldsymbol{p}}_1, \hat{\boldsymbol{p}}_2, \ldots, \hat{\boldsymbol{p}}_k$ of $\hat{\boldsymbol{\Sigma}}$. Thus, the PCAZ projection becomes

$$\hat{\boldsymbol{Y}} = \hat{\boldsymbol{P}}^T\hat{\boldsymbol{X}}. \tag{7}$$

C. Linear Regression Classification (LRC)

Assume we have N subjects with p_i training images from the i^{th} class, $i = 1, 2, \ldots, N$. Through the projection of the image space onto the face space, the matrix $\boldsymbol{T}$ contains all feature vectors (the projected coefficients) from N subjects and thus $\boldsymbol{T} = [\boldsymbol{W}_1, \ldots, \boldsymbol{W}_i, \ldots, \boldsymbol{W}_N]$. In order to apply linear regression classification algorithm to estimate class specific model, we have to group the feature vectors regarding the class-membership. Hence, for the i^{th} class, we have

$$\boldsymbol{W}_i = [\boldsymbol{w}_{i,1}, \ldots, \boldsymbol{w}_{i,j}, \ldots, \boldsymbol{w}_{i,p_i}] \in \Re^{L\times p_i} \tag{8}$$

where each vector $\boldsymbol{w}_{i,j}$ is a column vector of $\boldsymbol{W}_i$. Thus, in the training phase, the i^{th} class is represented by a vector space $\boldsymbol{W}_i$, which is called the predictor for each subject.

If $\boldsymbol{y}$ belongs to the i^{th} class, it can be represented as a linear combination of the training images from the i^{th} class and can be defined as

$$\boldsymbol{y} = \boldsymbol{W}_i\boldsymbol{\beta}_i + \boldsymbol{e}, \quad i = 1, 2, \ldots, N \tag{9}$$

where $\boldsymbol{\beta}_i \in \Re^{p_i\times 1}$ is the vector of regression parameters and $\boldsymbol{e}$ is an error vector whose components are independent random variables with mean 0 and variance σ^2. The goal of regression is to find $\widetilde{\boldsymbol{\beta}}_i$, which minimizes the residual errors as

$$\widetilde{\boldsymbol{\beta}}_i = \arg\min_{\boldsymbol{\beta}_i} \|\boldsymbol{W}_i\boldsymbol{\beta}_i - \boldsymbol{y}\|_2^2, \quad i = 1, 2, \ldots, N. \tag{10}$$

The regression coefficients can be solved through the least-square estimation and can be written as a matrix form as

$$\widetilde{\boldsymbol{\beta}}_i = \left(\boldsymbol{W}_i^T \boldsymbol{W}_i\right)^{-1} \boldsymbol{W}_i^T \boldsymbol{y}, \quad i = 1, 2, \ldots, N \cdot \tag{11}$$

The estimated vector of parameters $\widetilde{\boldsymbol{\beta}}_i$ and predictors $\boldsymbol{W}_i$ are used to predict the response vector $\widetilde{\boldsymbol{y}}_i$ for the i^{th} class as

$$\widetilde{\boldsymbol{y}}_i = \boldsymbol{W}_i \widetilde{\boldsymbol{\beta}}_i, \quad i = 1, 2, \ldots, N \,. \tag{12}$$

By substituting (11) into (12), we have

$$\widetilde{\boldsymbol{y}}_i = \boldsymbol{W}_i \left(\boldsymbol{W}_i^T \boldsymbol{W}_i\right)^{-1} \boldsymbol{W}_i^T \boldsymbol{y}, \quad i = 1, 2, \ldots, N \cdot \tag{13}$$

Therefore, we can get a class specific hat matrix (projection matrix) as

$$\widetilde{\boldsymbol{y}}_i = \boldsymbol{H}_i \boldsymbol{y}, \quad i = 1, 2, \ldots, N \tag{14}$$

where $\widetilde{\boldsymbol{y}}_i$ is the projection of $\boldsymbol{y}$ onto the subspace of the i^{th} class by the projection matrix, $\boldsymbol{H}_i = \boldsymbol{W}_i(\boldsymbol{W}_i^T \boldsymbol{W}_i)^{-1} \boldsymbol{W}_i^T$. It is noted that the projection matrix $\boldsymbol{H}_i$ is symmetric and also idempotent.

The LRC is developed based on the minimum reconstruction error. In other words, if the original vector belongs to the subspace of class i, the predicted response vector $\widetilde{\boldsymbol{y}}_i$ will be the closest vector to the original vector. Therefore, the identity i^* could be determined by calculating the Euclidean distance measure between the predicted response vectors and the original vector as

$$\begin{aligned} i^* &= \arg\min_i \|\widetilde{\boldsymbol{y}}_i - \boldsymbol{y}\|, \quad i = 1, 2, \ldots, N \\ &= \arg\min_i \|\boldsymbol{H}_i \boldsymbol{y} - \boldsymbol{y}\|. \end{aligned} \tag{15}$$

III. Comparison With Related Works

When data are strongly correlated among the variables, the problem of multicollinearity in linear regression (LR) may destabilize the regression model. To alleviate this problem, a variety of unbiased methods, such as ridge regression (RR), and principal component regression (PCR), have been introduced.

A. Linear Regression (LR)

The goal of the linear regression is to find $\widetilde{\boldsymbol{\beta}}_i$, which minimizes the residual errors. The variance of the regression parameter vector $\widetilde{\boldsymbol{\beta}}_i$ in the linear regression model is expressed as

$$Var(\widetilde{\boldsymbol{\beta}}_i)_{LR} = \sigma^2 \sum_{j=1}^{J} \frac{1}{d_{ij}} \boldsymbol{v}_{ij} \boldsymbol{v}_{ij}^T \tag{16}$$

where $\boldsymbol{W}_i = \boldsymbol{U}_i \boldsymbol{D}_i \boldsymbol{V}_i{}^T$ by the singular value decomposition (SVD), $\boldsymbol{v}_{ij}$ is the j^{th} column eigenvector of $\boldsymbol{V}_i$, and d_{ij} is the j^{th} eigenvalue corresponding to the $\boldsymbol{v}_{ij}$.

B. Ridge Regression (RR)

The goal of the ridge regression is to find $\widetilde{\boldsymbol{\beta}}_i$ to minimize the residual errors and their penalty as

$$\widetilde{\boldsymbol{\beta}}_i = \arg\min_{\boldsymbol{\beta}_i} \{\|\boldsymbol{W}_i \boldsymbol{\beta}_i - \boldsymbol{y}\|_2^2 + \lambda \|\boldsymbol{\beta}_i\|_2^2\}, \quad i = 1, 2, \ldots, N \cdot \tag{17}$$

Comparing with (10), the ridge regression adds a penalty, $\lambda\|\boldsymbol{\beta}_i\|_2^2$ to the regression model to reduce the variance of the model. The estimate of the regression parameter vectors can be computed by

$$\widetilde{\boldsymbol{\beta}}_i = (\boldsymbol{W}_i^T \boldsymbol{W}_i + \lambda \boldsymbol{I})^{-1} \boldsymbol{W}_i^T \boldsymbol{y}. \tag{18}$$

The variance of the regression parameter vector $\widetilde{\boldsymbol{\beta}}_i$ in the ridge regression model is expressed as

$$Var(\widetilde{\boldsymbol{\beta}}_i)_{RR} = \sigma^2 \sum_{j=1}^{J} \frac{d_{ij}}{\left(d_{ij} + \lambda\right)^2} \boldsymbol{v}_{ij} \boldsymbol{v}_{ij}^T. \tag{19}$$

Thus, the ridge regression adds a constant λ to the diagonal term of $\boldsymbol{W}_i^T \boldsymbol{W}$ in (18), which makes the variance in (19) be smaller than that in (16). Empirically, for some cases, λ is set to be 14.

C. Principal Component Regression (PCR)

The goal of the principal component regression is to find $\widetilde{\boldsymbol{\beta}}_i$, which minimizes the residual errors in PCA space. In this work, we propose dropping the first $(Q-1)$ principal components as

$$Var(\widetilde{\boldsymbol{\beta}}_i)_{IPCR} = \sigma^2 \sum_{j=Q}^{J} \frac{1}{d_{ij}} \boldsymbol{v}_{ij} \boldsymbol{v}_{ij}^T. \tag{20}$$

Empirically, we set Q to be 3. In other words, our system drops the first two principal components. This contribution is novel to the PCR because the existing suggestion is to drop the last n principal components as

$$Var(\widetilde{\boldsymbol{\beta}}_i)_{PCR} = \sigma^2 \sum_{j=1}^{P} \frac{1}{d_{ij}} \boldsymbol{v}_{ij} \boldsymbol{v}_{ij}^T \tag{21}$$

where $P < J$. The variances of these two types are both smaller than that of the LR. Moreover, our suggestion could be supported that the first three principal components correspond to the variation in lighting in the PCA-based methods [3]. As to the PCA methods, the traditional PCA removes the mean image from all training samples, while the PCAZ removes the mean of each image. So PCAZ is insensitive to illumination variations and can achieve robust recognition for different illuminations [12]. Furthermore, IPCR drops the first n principal components. Therefore, the proposed solution is useful for recognizing face images with illumination variation. Note that when all principal components are used, the multicollinearity would be reproduced.

IV. Experimental Results

We have examined our algorithm on two publicly available face databases: Yale B [13] and FERET [14] databases. All images are cropped and resized to 30×25 pixels. We followed the experimental protocol as proposed in the literature [10] and

TABLE I
ACCURACY (%) COMPARISONS ON YALE B AND FERET

Methods	Yale B				FERET
	Subset 2	Subset 3	Subset 4	Subset 5	
Eigenface	89.81	47.04	21.90	--	60.66
Direct-LDA	98.15	70.09	30.79	--	79.60
Fisherface	95.14	75.14	34.76	--	78.04
Subspace-LDA	99.84	94.44	49.80	--	80.28
KPCA	92.96	49.44	22.38	--	63.72
KDDA	99.63	89.26	38.89	--	78.86
GDA	99.81	96.94	49.92	--	81.64
KSLDA	99.81	96.99	52.22	--	82.40
LRC	100.00	100.00	91.86	52.11	76.20
RR	100.00	100.00	92.57	53.68	78.66
PCA+LRC	100.00	99.17	85.00	50.53	82.00
PCAZ+LRC	**100.00**	**100.00**	**95.00**	**64.21**	**82.80**

compared the proposed method with the KSLDA, KDDA, etc., where the recognition rates are directly obtained from [10].

The Yale B contains images of ten subjects with nine poses and 64 illuminations per pose. The frontal face images with 64 different illuminations of all subjects are used for evaluation. The Yale B is divided into five subsets based on the angle of the light source directions. Training is conducted using Subset 1 and the remaining subsets (Subset 2 to 5) are used for testing. In the FERET database, we selected 250 people, with four frontal-view images from each subject. These 1000 face images are with illumination and expression variations. Two images of each person are randomly selected for training, and the other two images are for testing.

Table I reveals that the PCAZ+LRC outperforms the LRC, RR, and PCA+LRC. Moreover, the proposed method performs better than the LDA-based approaches. On the other hand, for Yale B database, the LRC and RR perform better than the PCA+LRC. The reason is that the PCA is sensitive to severe illumination variations. For the FERET, the PCA+LRC performs better than the LRC and RR. This is because the FERET contains illumination and expression variations. Thus the proposed method possesses higher robustness to illuminations than the other methods.

V. CONCLUSION

An improved principal component regression classification approach has been proposed for variable lighting face recognition. Furthermore, our method drops the first two principal components, which reduces the variance of the IPCR and improves the robustness against severe illumination variation. Interestingly, from our study, we found that the regression-based approaches outperform the LDA-based approaches for face recognition under illumination changes. Particularly, the proposed PCAZ+LRC could achieve higher accuracy. Additionally, because of the PCA process, the PCRC may have advantages for computation, compared with the LRC without PCA process.

REFERENCES

[1] W. Zhao, R. Chellappa, A. Rosenfeld, and P. J. Phillips, "Face recognition: A literature survey," *ACM Comput. Surv.*, vol. 35, no. 4, pp. 399–458, 2003.

[2] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cogn. Sci.*, pp. 71–86, 1991.

[3] P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 19, no. 7, 1997.

[4] A. M. Martinez and A. C. Kak, "PCA versus LDA," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 23, pp. 228–233, 2001.

[5] P. C. Yuen and J. H. Lai, "Face representation using independent component analysis," *Pattern Recognit.*, vol. 35, no. 6, pp. 1247–1257, 2002.

[6] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski, "Face recognition by independent component analysis," *IEEE Trans. Neural Netw.*, vol. 13, no. 6, pp. 1450–1464, 2002.

[7] M. H. Yang, "Kernel eignefaces vs. Kernel fisherfaces: Face recognition using kernel methods," in *Proc. Fifth IEEE Int. Conf. Automatic Face and Gesture Recognition*, May 2002, pp. 215–220.

[8] G. Baudat and F. Anouar, "Generalized discriminant analysis using a kernel approach," *Neural Comput.*, vol. 12, pp. 2385–2404, 2000.

[9] J. W. Lu, K. Plataniotis, and A. N. Venetsanopoulos, "Face recognition using kernel direct discriminant analysis algorithms," *IEEE Trans. Neural Netw.*, vol. 14, no. 1, pp. 117–126, Jan. 2003.

[10] J. Huang, P. C. Yuen, W. Chen, and J. Lai, "Choosing parameters of kernel subspace LDA for recognition of face images under pose and illumination variations," *IEEE Trans. Syst., Man, Cybern. B*, pp. 847–862, 2007.

[11] I. Naseem, R. Togneri, and M. Bennamoun, "Linear regression for face recognition," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 32, no. 11, pp. 2106–2112, Jul. 2010.

[12] J. Yang, D. Zhang, and J. Yang, "Constructing PCA baseline algorithms to reevaluate ICA-based face-recognition performance," *IEEE Trans. Syst., Man, Cybern. B*, pp. 1015–1021, 2007.

[13] A. S. Georghiades, P. N. Belhumeur, and D. W. Jacobs, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 23, no. 6, pp. 630–660, Jun. 2001.

[14] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET valuation methodology for face-recognition algorithms," *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 22, no. 10, pp. 1090–1104, Oct. 2000.