We are IntechOpen,
the world's leading publisher of
Open Access books
Built by scientists, for scientists

4,800
Open access books available

122,000
International authors and editors

135M
Downloads

Our authors are among the

154
Countries delivered to

TOP 1%
most cited scientists

12.2%
Contributors from top 500 universities

WEB OF SCIENCE™

Selection of our books indexed in the Book Citation Index
in Web of Science™ Core Collection (BKCI)

Interested in publishing with us?
Contact book.department@intechopen.com

Numbers displayed above are based on latest data collected.
For more information visit www.intechopen.com

Advances of Robust Subspace Face Recognition

Yang-Ting Chou, Jar-Ferr Yang and
Shih-Ming Huang

Additional information is available at the end of the chapter

Abstract

Face recognition has been widely applied in fast video surveillance and security systems and smart home services in our daily lives. Over past years, subspace projection methods, such as principal component analysis (PCA), linear discriminant analysis (LDA), are the well-known algorithms for face recognition. Recently, linear regression classification (LRC) is one of the most popular approaches through subspace projection optimizations. However, there are still many problems unsolved in severe conditions with different environments and various applications. In this chapter, the practical problems including partial occlusion, illumination variation, different expression, pose variation, and low resolution are addressed and solved by several improved subspace projection methods including robust linear regression classification (RLRC), ridge regression (RR), improved principal component regression (IPCR), unitary regression classification (URC), linear discriminant regression classification (LDRC), generalized linear regression classification (GLRC) and trimmed linear regression (TLR). Experimental results show that these methods can perform well and possess high robustness against problems of partial occlusion, illumination variation, different expression, pose variation and low resolution.

Keywords: subspace projection, principal component analysis, linear discriminant analysis, linear regression classification, robust linear regression classification, ridge regression, improved principal component regression, unitary regression classification, linear discriminant regression classification, generalized linear regression classification, trimmed linear regression

1. Introduction

From the tragic 911 incident in 2011, more and more researches focus on the security issues with computational intelligence. How to avoid the tragic event from happening again and how to quickly identify the terrorists and suspects before or after the tragic event happens are very important. Therefore, the effectiveness of security is being examined almost everywhere. The discoveries of current security vulnerabilities along with the exploration of new methods should be constantly investigated to improve the security systems. Security measures with computational intelligence are used to improve the safety of our everyday lives.

The security issue is that we should recognize the wanted criminals and invaders from their biometric characteristics such as face, fingerprint, iris, palm and so on. Among these biometrics signals, the face image is easier and more direct to be captured by distanced cameras than others. For instance, the face images of any suspect who walks through the hotel lobby will be recorded by cameras. Hence, computer vision technologies with cameras can be applied to realize intelligent video surveillance systems. Since many face images of criminals and terrorists are available in the police department, they can be used to identify if the unknown face images are them from the distributed cameras. Thus, an efficient face recognition system could help to improve security. The face recognition systems would not only be helpful in identifying the criminals and terrorists, but also be used to search missing persons or identify the incident of weak person. Thus, face recognition systems with surveillance cameras have been already installed in many locations such as department store, airports and supermarkets. Besides, if the face recognition systems installed at home can timely detect the user's facial expression, the smart service for the user can be properly introduced accordingly.

The goal of face recognition is to distinguish a specific identity and its outlook from face images. However, in realistic situations, such as video surveillance and access control, face recognition task might encounter great challenges such as different facial expressions, illumination variations, partial occlusions and even low resolution problems, which will degrade the face recognition performance and result in severe security complications. For example, the image captured by a CCTV camera at a distance would have a very low resolution which degrades the recognition performance significantly. Besides, in the testing phase, the face image is a factor which is out of control. In other words, the person may not be on a frontal pose and may not be a pure image, that is the person may be wearing glasses, hat, or mask, or even with some lighting influence and expressions. Over past years, subspace projection optimizations have been widely proposed to solve this problem with linear [1] and non-linear [2] approaches. The principal component analysis (PCA) [3, 4] and linear discriminant analysis (LDA) [5] are the two typical examples of linear transform approaches which attempt to seek a low-dimensional subspace for dimensionality reduction. The nonlinear projection approaches also have been used in many literatures like the kernel PCA (KPCA) [6] and kernel LDA (KLDA) [7] which can uncover the underlying structure when the samples lie on a nonlinear manifold structure in the image space.

Recently, the linear regression classification (LRC) proposed in 2010 by Naseem *et al*. [8] has been treated as an effective subspace projection method, which performs well on face recog-

nition. Moreover, the robust linear regression classification (RLRC) [9] estimating regression parameters by using the robust Huber estimation was introduced to achieve robust face recognition under illumination variation and random pixel corruption. Ridge regression (RR) [10] estimated the regression parameters by using a regularized least square method to model the linear dependency in the spatial domain. Huang *et al.* and Chou *et al.* presented several improved approaches of LRC, including improved-PCA-LRC [11], LDA-LRC [12], unitary-LRC [13], and generalized-LRC [14, 15] for dealing with different situations like facial expressions, lighting changes, and pose variations. Lai *et al.* [16] utilized the least trimmed square (LTS) as a robust estimator to detect the contaminated pixels from query for boosting the performance under the partial occlusion situation.

The rest of this chapter is organized as follows. With the overview of fundamentals and facial representation, several famous face recognition algorithms are first presented in Section 2. Section 3 is dedicated to present several advances of subspace projection optimizations for robust face recognition technologies including RLRC, RR, improved principal component regression (IPCR), unitary regression classification (URC), linear discriminant regression classification (LDRC), generalized linear regression classification (GLRC) and trimmed linear regression (TLR). The performances of the aforementioned projection methods will be shown in Section 4. Finally, conclusions are drawn in Section 5.

2. Fundamentals of face recognition and representation

As shown in **Figure 1**, the typical face recognition system contains two major parts: face detection and face recognition. In this section, the face detection methods are first briefly introduced. Then, the well-known subspace project methods are reviewed. Finally, the similarity measures of image feature vectors are overviewed. Generally, the unknown data vector will be projected into a certain subspace, a similarity measure will be used to classify it. To narrow down the computation and increase the recognition accuracy, the first step of the recognition system, called face detection, is to detect and crop the face region from the image or video.
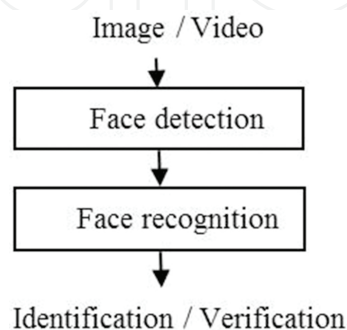
Figure 1. The simplified flow chart of face recognition system.

2.1. Face detection

The methods of face detection [17–20], can be separated into neural network, feature-based, and color-based approaches. Neural network approach [21] trains the facial class and non-facial class while a new image or video can be detected based on the prior training data. The well-known method is AdaBoost learning algorithm [22, 23]. Feature-based approach is to utilize the facial feature for detecting facial region. For example, the corresponding positions of the eyes, nose, and mouth are useful features; moreover, the shape of face, which is almost like an ellipse, can be included. Rule-based algorithm [24] and elliptical edge [25] are two popular feature-based methods. Color-based approach as [26] adopts the variance of skin color to detect if the region is face or not. For example, the face region in grayscale should not change immense while the eyes, mouth and hair should be darker than the other part of face.

Once the face areas are detected by a selected face detection method, their face images in size of $a \times b$ pixels could be projected into another subspace such as principal space, kernel space, frequency space and so on, in order to find a proper set of features for boosting the recognition performance. Assume there are C subjects. Each class is with N training color images. For the ith class, $i = 1, 2, \ldots, C$, the jth training color image in size of $a \times b$ pixels with K components is formed a data matrix as $\boldsymbol{v}_{i,j,k} \in \boldsymbol{R}^{a \times b \times K}$ for $j = 1, 2, \ldots, N$ and $k = 1, 2, \ldots, K$. For example, $K = 3$ color components, $k = 1, 2,$ and 3 denote the red, green, blue channels, respectively. For some recognition algorithms, $\boldsymbol{v}_{i,j,k} \in \boldsymbol{R}^{a \times b \times K}$ is transformed to grayscale as $\boldsymbol{g}_{i,j} = c_1 \boldsymbol{v}_{i,j,1} + c_2 \boldsymbol{v}_{i,j,2} + c_3 \boldsymbol{v}_{i,j,3}$, where are $c_1, c_2,$ and c_3 are fixed in visualization. The gray image $\boldsymbol{g}_{i,j}$ is reshaped into one column vector as $\boldsymbol{x}_{i,j} \in \boldsymbol{R}^{M \times 1}$ where $M = a \times b$. In the testing phase, an unknown color image, $\boldsymbol{z} \in \boldsymbol{R}^{M \times K}$, is given. In order to predict unknown $\boldsymbol{z}$ by training data, it should be transformed to grayscale, be normalized and be reshaped into a column vector as $\boldsymbol{y} \in \boldsymbol{R}^{M \times 1}$.

2.2. Subspace projection methods

The famous subspace projection methods, such as PCA and LDA are reviewed in the following sections.

2.2.1. Principal component analysis (PCA)

The PCA method is widely used for dimensionality reduction in the computer vision field, especially for face recognition technology. In the PCA, the data is represented as a linear combination of an orthonormal set of vectors that maximize the data scatter across all images. The first principal component represents the most variability of the image as possible while the second one represents the second most, and so on. The flow chart to find PCA transformation bases is shown in **Figure 2**. The main objective of the PCA is to reduce the dimension of the feature image $\boldsymbol{x}_{i,j}$ to retain a few principal components. This means that most of the useless information would be reduced, and the remaining data could be well represented in a lower dimension space by the PCA.

As shown in **Figure 2**, the derivations of the PCA transformation bases are stated in the following equations. First, the feature face image should remove the global mean to become:

$$\widetilde{x}_{i,j} = x_{i,j} - \overline{x}_{\text{global}} \tag{1}$$

where $\overline{x}_{global} = \frac{1}{C \cdot N} \sum_C \sum_N x_{i,j}$ is a global mean vector of all facial image vectors.
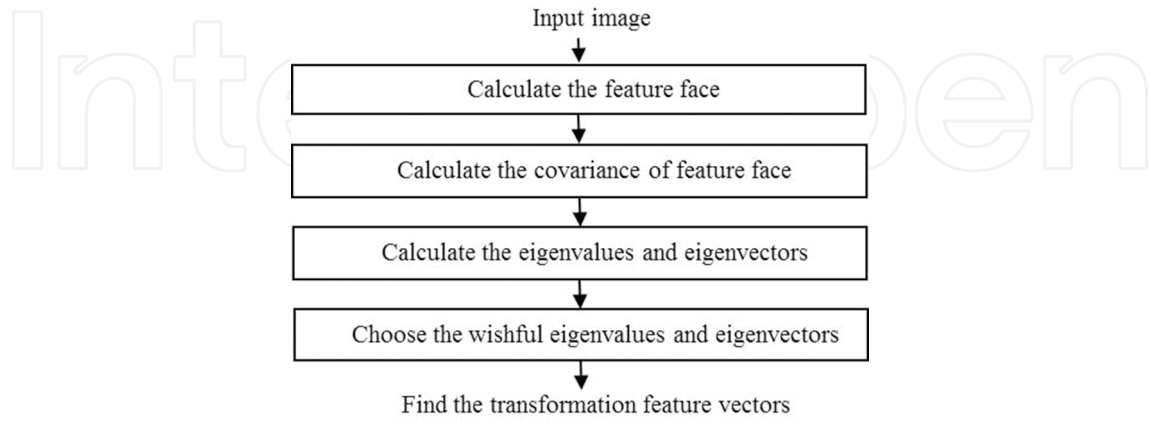
Figure 2. The flow chart for finding PCA transformation.

After the computation of the feature face images, we can obtain $M \times M$ covariance matrix of all feature face images as:

$$Qu = ru \tag{2}$$

Based on the covariance matrix, the eigenvectors and eigenvalues can be retrieved by singular value decomposition (SVD) or eigen-decomposition as:

$$P_{PCA} = \{u_1, u_2, \dots, u_P\}, P \leq M \tag{3}$$

where $r = \{r_1, r_2, \dots, r_M\}$ is a set of total M descending-ordered eigenvalues and their corresponding eigenvectors $u = \{u_1, u_2, \dots, u_M\}$ According to the expected dimension, we can choose P principal components. Thus, the PCA transformation with the P largest eigenvectors, the PCA transformation P_{PCA} with $P \times M$ size can be formed by the corresponding P eigenvalues as:

$$w_{PCA,i,j} = P_{PCA}^T \widetilde{x}_{i,j} \tag{4}$$

Finally, we can achieve the PCA features, $w_{PCA,i,j} \in R^{P \times 1}$, by multiplying PCA transformation and the feature image vector as:

$$P_{LDA} = \underset{P}{\text{argmax}} \frac{|P^T S_B P|}{|P^T S_W P|} \tag{5}$$

On the other hand, the testing image vector y can be projected onto PCA subspace by P_{PCA}. The PCA subspace $\hat{y}_{PCA}$ can be written as:

$$\hat{y}_{LDA} = P_{LDA}^T y \tag{6}$$

And the similarity measure based on this feature data vector is calculated to determine the final result.

2.2.2. Linear discriminant analysis

Fisher proposed the LDA for recognition which is a kind of statistical analysis method like the PCA. But the difference is that the LDA can discriminate the different subjects even though the maximum variance subspaces among them are overlapped as shown in **Figure 3**. The goal of the LDA is that these projections onto a line will be well separated by disparate classes and be well concentrated by the same class.

Image / Video

Face detection

Face recognition

Identification / Verification
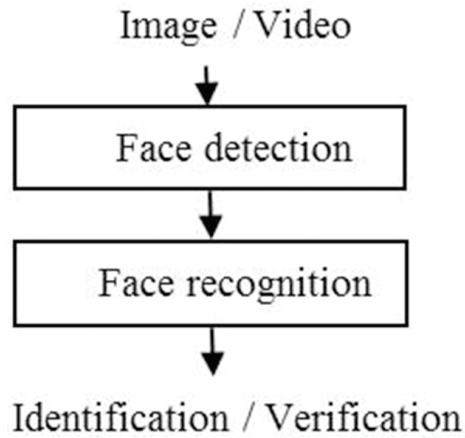
Figure 3. Comparison of LDA and PCA in projection space.

Thus, the concept of LDA is to seek the optimal projection by maximizing the ratio of between-class and within-class scatter. Fisher utilizes a criterion to optimize this problem as:

$$L_{1,i,j} = \left\| w_{i,j} - \hat{y} \right\|_1 = \sum_{m=1}^{M} \left| w_{i,j}^{(m)} - \hat{y}^{(m)} \right| \tag{7}$$

where $S_B = \sum_{i=1}^{C} \sum_{q=1,q \neq i}^{C} \left(\bar{x}_{local,i} - \bar{x}_{local,q} \right)\left(\bar{x}_{local,i} - \bar{x}_{local,q} \right)^T$ is the between-class matrix where $\bar{x}_{local,i} = \frac{1}{N} \sum_{j=1}^{N} x_{i,j}$ is a local mean vector of the ith class. And $S_W = \frac{1}{C} \sum_{i=1}^{C} \left(X_i - \bar{x}_{local,i} \right)\left(X_i - \bar{x}_{local,i} \right)^T$ is the within-class matrix where X_i is concatenated by the ith data set of N training gray images. Then, the optimal projection matrix, W_{LDA}, can be solved by computing generalized SVD or eigen-decomposition as:

$$\cos\theta_{i,j} = \frac{\boldsymbol{w}_{i,j}^T \widehat{\boldsymbol{y}}}{\|\boldsymbol{w}_{i,j}\|_2 \|\widehat{\boldsymbol{y}}\|_2} = \frac{\left|\sum_{h=1}^{M} \boldsymbol{w}_{i,j}^{(h)} \widehat{\boldsymbol{y}}^{(h)}\right|^2}{\sum_{h=1}^{M}\left|\boldsymbol{w}_{i,j}^{(h)}\right|^2 \sum_{h=1}^{M}|\widehat{\boldsymbol{y}}^{(h)}|^2} \tag{8}$$

where Λ is the diagonal eigenvalue matrix. We apply the optimal projection matrix to convert the face feature vector $\boldsymbol{x}_{i,j}$ into a new discriminant vector, $\boldsymbol{w}_{Fisher,i,j}$ as:

$$\boldsymbol{w}_{Fisher,i,j} = \boldsymbol{P}_{LDA}^T \boldsymbol{x}_{i,j} \tag{9}$$

In the same way, the testing image vector is projected onto LDA subspace by $\boldsymbol{P}_{LDA}$ and can be represented as:

$$\boldsymbol{y} = \boldsymbol{X}_i \boldsymbol{\beta}_i, i = 1,2, \dots, C \tag{10}$$

And the final result can be determined by using similarity measure based on this feature vector.

2.3. Similarity measures

There exist three distance measures [27–29] such as the city block distance (Taxicab geometry, L_1), Euclidean distance (L_2) and L_∞ norm distance. These distance measures are defined from two column vectors $\boldsymbol{w}_{i,j}$ and $\widehat{\boldsymbol{y}}$ which can be obtained from the subspace projection like PCA subspace $\{\boldsymbol{w}_{PCA,i,j}, \widehat{\boldsymbol{y}}_{PCA}\}$, LDA subspace $\{\boldsymbol{w}_{LDA,i,j}, \widehat{\boldsymbol{y}}_{LDA}\}$, and the other projections with dimensionality of M or P. The distance measures, L_1, L_2, and L_∞ can be respectively written as:

$$\widehat{\boldsymbol{y}}_{LRC,i} = \boldsymbol{X}_i \left(\boldsymbol{X}_i^T \boldsymbol{X}_i\right)^{-1} \boldsymbol{X}_i^T \boldsymbol{y}, i = 1,2, \dots, C \tag{11}$$

$$\widehat{\boldsymbol{y}}_{LRC,i} = \boldsymbol{H}_i \boldsymbol{y}, i = 1,2, \dots, C \tag{12}$$

and

$$i^* = \underset{i}{\operatorname{argmin}} \left\|\widehat{\boldsymbol{y}}_{LRC,i} - \boldsymbol{y}\right\|_2 = \underset{i}{\operatorname{argmin}} \left\|\boldsymbol{H}_i \boldsymbol{y} - \boldsymbol{y}\right\|_2, i = 1,2, \dots, C \tag{13}$$

where $\boldsymbol{x}_{i,j}^{(m)}$ and $\boldsymbol{y}^{(m)}$ are the mth component of $\boldsymbol{x}_{i,j}$ and $\boldsymbol{y}$ column vectors, respectively.

However, these vectors satisfy the Cauchy-Schwarz inequality as:

$$\widehat{\boldsymbol{\beta}}_{RR,i} = \underset{\boldsymbol{\beta}_i}{\operatorname{argmin}} \{\|\boldsymbol{y} - \boldsymbol{X}_i \boldsymbol{\beta}_i\|_2^2 + \lambda\|\boldsymbol{\beta}_i\|_2^2\}, i = 1,2, \dots, C \tag{14}$$

To ignore the amplitudes of two feature data vectors, the similarity measure can be also defined by a cosine criterion as:

$$\cos\theta_{i,j} = \frac{w_{i,j}^T \hat{y}}{\|w_{i,j}\|_2 \|\hat{y}\|_2} = \frac{\left|\sum_{h=1}^M w_{i,j}^{(h)} \hat{y}^{(h)}\right|^2}{\sum_{h=1}^M \left|w_{i,j}^{(h)}\right|^2 \sum_{h=1}^M \left|\hat{y}^{(h)}\right|^2} \tag{15}$$

3. Advances of subspace projection optimization

In this section, the advances of subspace projection optimization are presented for robust face recognition system. Then, the well-known subspace projection methods including LRC, RLRC, RR, IPCR, URC, LDRC, GLRC and TLR are introduced.

3.1. Linear regression classification (LRC)

For applying the linear regression to estimate the class specific model, all N training gray images from the same class are concatenated as:

$$P_{PCAZ} = \{u_{n+1}, u_{n+2}, \ldots, u_P\}, P \leq M \tag{16}$$

where X_i is in the size of $M \times N$ and is called class-specific model. In other words, the ith class is represented by a vector space X_i, which is called the regressor for each subject, in the training phase.

In the testing phase, if an unknown column vector y belongs to the ith class, its linear combination can be rewritten in terms of the training data from the ith class and can be formulated as:

$$w_{PCAZ,i,j} = P_{PCAZ}^T \tilde{x}_{i,j} \tag{17}$$

where $\beta_i \in R^{N \times 1}$ is the vector of regression parameters. The goal of the linear regression is to find the regression parameters by minimizing the residual errors as:

$$\hat{\beta}_{PCAZ,i} = \underset{\beta_i}{\mathrm{argmin}} \left\{ \left\| y_{PCAZ} - w_{PCAZ,i}\beta_i \right\|_2^2 \right\}, i = 1,2,\ldots,C \tag{18}$$

The regression coefficients, β_i, can be solved through the least-square estimation method and can be represented as:

$$\hat{\beta}_{PCAZ,i} = \left(X_{PCAZ,i}^T X_{PCAZ,i}\right)^{-1} X_{PCAZ,i}^T y_{PCAZ}, i = 1,2,\ldots,C \tag{19}$$

For each class i, the regressed vector $\hat{y}_{LRC,i}$ can be predicted through the regression parameters $\hat{\beta}_{LRC,i}$ and predictors X_i as

$$\underset{\boldsymbol{P}_{URC}}{\operatorname{argmin}} \sum_{i=1}^{C} \sum_{j=1}^{N} tr\left[\boldsymbol{P}_{URC}^{T}(\boldsymbol{x}_{i,j} - \tilde{\boldsymbol{x}}_{i})(\boldsymbol{x}_{i,j} - \tilde{\boldsymbol{x}}_{i})^{T} \boldsymbol{P}_{URC}\right] = \underset{\boldsymbol{P}_{URC}}{\operatorname{argmin}} tr\left[\boldsymbol{P}_{URC}^{T} \boldsymbol{E}_{URC} \boldsymbol{P}_{URC}\right] \tag{20}$$

By substituting Equation (19) into Equation (20), the predicted response vector $\overset{\wedge}{\boldsymbol{y}}_{i}$ can be rewritten as:

$$\underset{\boldsymbol{P}_{LDRC}}{\operatorname{argmin}} \frac{E_{BC}}{E_{WC}} \tag{21}$$

Theoretically, we can treat Equation (21) as a class-specific projection as:

$$\frac{E_{BC}}{E_{WC}} == \frac{\frac{1}{NC(C-1)}\sum_{i=1}^{C}\sum_{j=1}^{N}\sum_{q=1,q\neq i}^{C}\left\|\tilde{x}_{i,j} - \hat{x}_{i,j,q}^{inter}\right\|^{2}}{\frac{1}{NC}\sum_{i=1}^{C}\sum_{j=1}^{N}\left\|\tilde{x}_{i,j} - \hat{x}_{i,j}^{intra}\right\|^{2}} \tag{22}$$

where $\overset{\wedge}{\boldsymbol{y}}_{LRC,i}$ is the projection of $\boldsymbol{y}$ onto the subspace of the ith class by the projection matrix $\boldsymbol{H}_{i} = \boldsymbol{X}_{i}\left(\boldsymbol{X}_{i}^{T}\boldsymbol{X}_{i}\right)^{-1}\boldsymbol{X}_{i}^{T}$.

In the LRC approach, the minimum reconstruction error is adopted for determining the final result. In other words, the distance between predicted response vector $\overset{\wedge}{\boldsymbol{y}}_{LRC,i}$ and unknown column vector $\boldsymbol{y}$ will be smallest when the unknown column vector belongs to the training vector space of class i. Therefore, the identity i^* can be determined by minimizing the Euclidean distance between the predicted response vector and unknown vector as:

$$\boldsymbol{E}_{w} = \frac{1}{NC}\sum_{i=1}^{C}\sum_{j=1}^{N}(\boldsymbol{x}_{i,j} - \boldsymbol{x}_{i,j}^{intra})(\boldsymbol{x}_{i,j} - \boldsymbol{x}_{i,j}^{intra})^{T} \tag{23}$$

3.2. Robust linear regression classification (RLRC)

The LRC has been claimed that classical statistical methods are robust, but they are only robust in the fact of true cases. Once the data distribution is in fact of false cases, the regression parameter under original least square estimation could be inaccurate. In other words, the original least square estimation is inefficient and can be biased in the presence of outliers. There exist several approaches for robust estimation like *R*-estimator [30, 31] and *L*-estimator [30, 32]. However, *M*-estimator is now shown superiority due to their generality, efficiency and high breakdown point [30, 33]. Based on the *M*-estimator, the optimal function becomes:

$$\underset{\boldsymbol{P}_{LDRC}}{\operatorname{argmin}} \frac{E_{BC}}{E_{WC}} = \underset{\boldsymbol{P}_{LDRC}}{\operatorname{argmin}} \frac{\boldsymbol{P}_{LDRC}^{T}\boldsymbol{E}_{b}\boldsymbol{P}_{LDRC}}{\boldsymbol{P}_{LDRC}^{T}\boldsymbol{E}_{w}\boldsymbol{P}_{LDRC}} \tag{24}$$

where

$$\rho(\|\mathbf{y} - \boldsymbol{X}_i\boldsymbol{\beta}_i\|) = \begin{cases} \frac{1}{2\gamma}\|\mathbf{y} - \boldsymbol{X}_i\boldsymbol{\beta}_i\|^2 & , \text{ for } \|\mathbf{y} - \boldsymbol{X}_i\boldsymbol{\beta}_i\| \leq \gamma \\ \|\mathbf{y} - \boldsymbol{X}_i\boldsymbol{\beta}_i\| - \frac{1}{2}\gamma, & \text{ for } \|\mathbf{y} - \boldsymbol{X}_i\boldsymbol{\beta}_i\| > \gamma \end{cases} \tag{25}$$

and $\rho(\bullet)$ is a symmetric function and γ being a tuning constant, also called the Huber threshold.

3.3. Ridge regression (RR)

The goal of the RR is to find and minimize the residual errors and their penalty as:

$$\boldsymbol{E}_b\boldsymbol{u}_l = \lambda_l\boldsymbol{E}_w\boldsymbol{u}_l, l = 1,2,\dots,\varphi \tag{26}$$

where λ is the regularization parameter. Comparing with linear regression, the RR adds a penalty, $\lambda\|\boldsymbol{\beta}_i\|_2^2$, to the regression model to reduce the variance of the model. The regression parameter vectors can be computed by:

$$\boldsymbol{y}_k = \boldsymbol{X}_{i,k}\boldsymbol{\beta}_{GLRC,i}, i = 1,2,\dots,C; \ k = 1,2,\dots,K \tag{27}$$

3.4. Improved principal component regression (IPCR)

Multicollinearity denotes the interrelations among the independent variables. In the linear regression, the regression estimation could be imprecise because the multicollinearity phenomenon would inflate the variance and covariance. To overcome the problem of multicollinearity, various approaches have been proposed. IPCR is one of the powerful approaches.

The IPCR is a two-step classification method. In the first step, the PCAZ is adopted to transform the observed variables into the new decorrelated components. Then, the first n components are dropped because these components are very sensitive to the lighting changes. Mathematically, the PCA process is used in all training samples including covariance matrix evaluation as Equation (2), and eigen-decomposition estimation as Equation (3). Then, we can obtain a set of eigenvectors, $\boldsymbol{u}=\{u_1, u_2,\dots, u_M\}$, and a set of eigenvalues, $r=\{r_1, r_2,\dots, r_M\}$ with $r_1 \geq r_2 \geq \dots \geq r_M$. As above mentioned, we drop first n components and the projection matrix can be express as:

$$\widehat{\boldsymbol{\beta}}_{GLRC,i} = \underset{\boldsymbol{\beta}_{GLRC,i}}{\text{argmin}} \left\{ \sum_{k=1}^{K}(\boldsymbol{y}_k - \boldsymbol{X}_{i,k}\boldsymbol{\beta}_{GLRC,i})^T(\boldsymbol{y}_k - \boldsymbol{X}_{i,k}\boldsymbol{\beta}_{GLRC,i}) \right\} \tag{28}$$

The PCAZ features, $w_{PCAZ,i,j} \in \boldsymbol{R}^{P \times 1}$, can be obtained by multiplying the projection matrix and the average image vector as:

$$\boldsymbol{r}_k = \sum_{i=1}^{C}|\widehat{\boldsymbol{y}}_{i,k} - \boldsymbol{y}_k| \tag{29}$$

In order to apply LRC to estimate class specific model, feature vectors should be grouped according to the class-membership. Hence, for the ith class, we have

$w_{PCAZ,i}=[w_{PCAZ,i,1}, w_{PCAZ,i,2}, \ldots, w_{PCAZ,i,N}]$. In the testing phase, an unknown column vector, y, is transformed to PCAZ subspace as $y^{(PCAZ)}$. In the second step, the new subspace of PCAZ projection is used in LRC such that we can seek more reliable regression coefficients for each subject for face recognition. The goal of regression becomes to minimize the residual errors as:

$$y_k = X_{i,k}\widetilde{\beta}_{GLRC,i} \tag{30}$$

The regression parameter vectors can be rewritten as a matrix form as:

$$\widetilde{\beta}_{GLRC,i} = \underset{\beta_{GLRC,i}}{\mathrm{argmin}} \sum_{k=1}^{K} \alpha_k \left(y_k - X_{i,k}\widetilde{\beta}_{GLRC,i}\right)^2 \tag{31}$$

3.5. Unitary regression classification (URC)

The total within-class projection error from all classes cannot be taken in previous mentioned methods for classification that would degrade the recognition accuracy. The URC is proposed to minimize the total within-class projection error from all classes for LRC to improve the robustness for pattern recognition.

Instead of original space, we hope to find a global unitary rotation $P_{URC}=[s_1,\ldots,s_\Psi]$ with $\Psi \le M$, which can rotate the original data space to a new compact w_{URC} data space as:

$$w_{URC,i,j} = P_{URC}^T x_{i,j} \tag{32}$$

to achieve the total minimum projection error of all training data stated as:

$$\underset{P_{URC}}{\mathrm{argmin}} \sum_{i=1}^{C} \sum_{j=1}^{N} \left\| w_{URC,i,j} - \widetilde{w}_i \right\|^2 \tag{33}$$

where $\widetilde{w}_i = \widetilde{H}_{URC,i} w_{URC,i,j}$ is the within-class projection to make the objective function be well-posed. In w_{URC} data space, the ith class projection matrix can be obtained by following $\widetilde{H}_{URC,i}=W_{URC,i}\left(W_{URC,i}^T W_{URC,i}\right)^{-1}W_{URC,i}^T$ where $W_{URC,i}=[W_{URC,i,1}, W_{URC,i,2}, \ldots, W_{URC,i,N}]$. The unitary rotation matrix, P_{URC}, is used to achieve the total minimum within-class projection error for LRC. From minimum reconstruction error, the objective function in T data space can be represented as:

$$\begin{aligned}
&\underset{P_{URC}}{\mathrm{argmin}} \sum_{i=1}^{C} \sum_{j=1}^{N} \left\| w_{URC,i,j} - \widetilde{H}_{URC,i} w_{URC,i,j} \right\|^2 \\
&= \underset{P_{URC}}{\mathrm{argmin}} \sum_{i=1}^{C} \sum_{j=1}^{N} \left\| P_{URC}^T x_{i,j} - \widetilde{H}_{URC,i} P_{URC}^T x_{i,j} \right\|^2
\end{aligned} \tag{34}$$

By substituting $W_{URC,i}=P_{URC}^T X_i$ into $\widehat{H}_{URC,i}=W_{URC,i}\left(W_{URC,i}^T W_{URC,i}\right)^{-1}W_{URC,i}^T$, the objective function becomes:

$$\underset{P_{URC}}{\text{argmin}} \sum_{i=1}^{C} \sum_{j=1}^{N} tr\left[P_{URC}^T(x_{i,j} - \tilde{x}_i)(x_{i,j} - \tilde{x}_i)^T P_{URC}\right] = \underset{P_{URC}}{\text{argmin}} \, tr\left[P_{URC}^T E_{URC} P_{URC}\right] \tag{35}$$

where $E_{URC} = \sum_{i=1}^{C} \sum_{j=1}^{N} (x_{i,j} - \tilde{x}_i)(x_{i,j} - \tilde{x}_i)^T$, also called within-class projection error matrix. The projection matrix, $P_{URC} = [s_1,\ldots,s_\Psi]$, can be solved by evaluating eigen-decomposition as:

$$E_{URC} s_l = \lambda_l s_l, l = 1,2,\ldots,\Psi \tag{36}$$

where $\lambda_\Psi \geqq \ldots \geqq \lambda_l \geqq \ldots \geqq \lambda_1 \geqq 0$.

3.6. Linear discriminant regression classification (LDRC)

Although the previous methods including LRC, RLRC, and IPCRC can perform well on face recognition, we cannot guarantee that the projection subspace in LRC or IPCRC is most discriminatory. When the projection subspaces among the different subjects overlap, the recognition result would be incorrect. To obtain an effective discriminant subspace for LRC, the LRC with discriminant analysis is presented by maximizing the ratio of the between-class reconstruction error (BCRE) to the within-class reconstruction error (WCRE) by the LRC.

Mathematically, all images are collected from C classes as $X = [X_1, X_2, \ldots, X_C] = [x_{1,1}, \ldots, x_{i,j}, \ldots, x_{C,N}]$. LDRC is to find an optimal projection by maximizing the BCRE over the WCRE for the LRC such that the LRC on the optimal subspace has better discrimination for classification. The goal of LDRC is to maximize the objective function as:

$$\underset{P_{LDRC}}{\text{argmin}} \frac{E_{BC}}{E_{WC}} \tag{37}$$

where $P_{LDRC} = [u_1, u_2, \ldots, u_\varphi]$ is the optimal projection matrix, and E_{BC} and E_{WC} denote the BCRE and WCRE, respectively. The original space, $x_{i,j}$, can be mapped into the subspace, $\tilde{x}_{i,j} = P_{LDRC}^T x_{i,j}$. Hence, the objective function can be rewritten as:

$$\frac{E_{BC}}{E_{WC}} == \frac{\frac{1}{NC(C-1)} \sum_{i=1}^{C} \sum_{j=1}^{N} \sum_{q=1,q\neq i}^{C} \left\|\tilde{x}_{i,j} - \hat{x}_{i,j,q}^{inter}\right\|^2}{\frac{1}{NC} \sum_{i=1}^{C} \sum_{j=1}^{N} \left\|\tilde{x}_{i,j} - \hat{x}_{i,j}^{intra}\right\|^2} \tag{38}$$

where $\tilde{x}_{i,j,q}^{inter} = H_q^{\tilde{x}} \tilde{x}_{i,j}$ denotes the inter-class projection of $\tilde{x}_{i,j}$ by the LRC from the different q^{th} class and $x_{i,j}^{intra} = H_{i,j}^{\tilde{x}} \tilde{x}_{i,j}$ denotes the intra-class projection of $\tilde{x}_{i,j}$ by the LRC in the same class. The $x_{i,j}$ is used to instead of $\tilde{x}_{i,j}$ as:

$$\frac{E_{BC}}{E_{WC}} = \frac{\frac{1}{NC(C-1)} \sum_{i=1}^{C} \sum_{j=1}^{N} \sum_{q=1,q\neq i}^{C} \left\|\tilde{x}_{i,j} - \hat{x}_{i,j,q}^{inter}\right\|^2}{\frac{1}{NC} \sum_{i=1}^{C} \sum_{j=1}^{N} \left\|\tilde{x}_{i,j} - \hat{x}_{i,j}^{intra}\right\|^2} = \frac{\frac{1}{NC(C-1)} \sum_{i=1}^{C} \sum_{j=1}^{N} \sum_{q=1,q\neq i}^{C} \left\|P_{LDRC}^T x_{i,j} - H_q^{\tilde{x}} P_{LDRC}^T x_{i,j}\right\|^2}{\frac{1}{NC} \sum_{i=1}^{C} \sum_{j=1}^{N} \left\|P_{LDRC}^T x_{i,j} - H_{i,j}^{\tilde{x}} P_{LDRC}^T x_{i,j}\right\|^2} \tag{39}$$

With some algebraic deduction, the form becomes:

$$\frac{E_{BC}}{E_{WC}} = \frac{\frac{1}{NC(C-1)}\sum_{i=1}^{C}\sum_{j=1}^{N}\sum_{q=1,q\neq i}^{C} tr\left[P_{LDRC}^{T}(x_{i,j}-x_{i,j,q}^{inter})(x_{i,j}-x_{i,j,q}^{inter})^{T}P_{LDRC}\right]}{\frac{1}{NC}\sum_{i=1}^{C}\sum_{j=1}^{N} tr\left[P_{LDRC}^{T}(x_{i,j}-x_{i,j}^{intra})(x_{i,j}-x_{i,j}^{intra})^{T}P_{LDRC}\right]} = \frac{tr(P_{LDRC}^{T}E_{b}P_{LDRC})}{tr(P_{LDRC}^{T}E_{w}P_{LDRC})} \tag{40}$$

where

$$E_{b} = \frac{1}{NC(C-1)}\sum_{i=1}^{C}\sum_{j=1}^{N}\sum_{q=1,q\neq i}^{C}(x_{i,j}-x_{i,j,q}^{inter})(x_{i,j}-x_{i,j,q}^{inter})^{T} \tag{41}$$

and

$$E_{w} = \frac{1}{NC}\sum_{i=1}^{C}\sum_{j=1}^{N}(x_{i,j}-x_{i,j}^{intra})(x_{i,j}-x_{i,j}^{intra})^{T} \tag{42}$$

is inter-class and intra-class reconstruction error, respectively. In other words, the objective function can be represented as:

$$\underset{P_{LDRC}}{argmin}\ \frac{E_{BC}}{E_{WC}} = \underset{P_{LDRC}}{argmin}\ \frac{P_{LDRC}^{T}E_{b}P_{LDRC}}{P_{LDRC}^{T}E_{w}P_{LDRC}} \tag{43}$$

For solving the optimization problem, Equation (43) can be reformulated as the following:

$$\underset{P_{LDRC}}{argmin}\ P_{LDRC}^{T}E_{b}P_{LDRC}, s.t.\ P_{LDRC}^{T}E_{w}P_{LDRC} = \vartheta \tag{44}$$

where ϑ is a constant. The projection matrix, $P_{LDRC}=[u_1, u_2,\dots, u_\varphi]$, can be solved by evaluating eigen-decomposition as:

$$E_{b}u_{l} = \lambda_{l}E_{w}u_{l}, l = 1,2,\dots,\varphi \tag{45}$$

where $\lambda_1 \geqq \dots \geqq \lambda_l \geqq \dots \geqq \lambda_\varphi$.

3.7. Generalized linear regression classification (GLRC)

In real-world recognition applications, the input images generally have multiple components which can overcome the unexpected effects such as pose variations, limited image information and so on. For color face recognition, the GLRC with membership grade (MG) criteria is proposed to defend the unexpected effects.

Mathematically, each channel component is separately normalized and transformed to one column vector such that $v_{i,j,k}\in R_{p\times q\times K} \to x_{i,j,k}\in R_{d\times K}$, where $d = p\cdot q$. In the ith class, the kth component of N training images is collected as:

$$X_{i,k} = \left[x_{i,1,k}, x_{i,2,k}, \dots, x_{i,N,k} \right] \in R^{d \times N} \tag{46}$$

for $i = 1, 2, \dots, C$ and $k = 1, 2, \dots, K$, where $X_{i,k}$ is treated as the kth-channel collected training data of the ith class in the training phase.

For the test image, the kth-channel testing image, z_k, is normalized and reshaped into a column vector as $y_k \in R^{d \times 1}$. For the kth component, the linear combination of $X_{i,k}$ from the ith class for the test vector y_k becomes:

$$y_k = X_{i,k} \beta_{GLRC,i}, i = 1,2,\dots,C; \ k = 1,2,\dots,K \tag{47}$$

where $\beta_{GLRC,i} \in R^{N \times 1}$ is an ideal projection vector of the ith-class regression parameter for all channels. In order to estimate the projection vector, the objective function becomes:

$$\widehat{\beta}_{GLRC,i} = \underset{\beta_{GLRC,i}}{\mathrm{argmin}} \left\{ \sum_{k=1}^{K} \left(y_k - X_{i,k} \beta_{GLRC,i} \right)^T \left(y_k - X_{i,k} \beta_{GLRC,i} \right) \right\} \tag{48}$$

After solving the optimization problem, the regression vector can be expressed as:

$$\widehat{\beta}_{GLRC,i} = \left(\sum_{k=1}^{K} X_{i,k}^T X_{i,k} \right)^{-} \left(\sum_{k=1}^{K} X_{i,k}^T y_k \right) \tag{49}$$

In order to achieve optimal performance, the different components should be treated as unequally important. Thus, the absolute sum of prediction residual of the kth component after the direct least square optimization is given as:

$$r_k = \sum_{i=1}^{C} \left| \widehat{y}_{i,k} - y_k \right| \tag{50}$$

where $\widehat{y}_{i,k} = X_{i,k} \beta_{GLRC,i}$, $i = 1, 2, \dots, C$. Based on the statistical opinion, we define the importance of the kth component to be inverse of the normalized absolute sum of prediction residual, which is expressed by:

$$\alpha_k = \frac{1}{r_k + \varepsilon} \sum_{k=1}^{K} r_k \tag{51}$$

where ε is a tiny value which is used to avoid $r_k = 0$. The larger the residual, r_k is, the less important the kth component will be. For the GRLC optimization, we propose the linear combination of $X_{i,k}$ of the kth component in the ith class for the test vector y_k becomes:

$$y_k = X_{i,k} \widetilde{\beta}_{GLRC,i} \tag{52}$$

where $\tilde{\beta}_{GLRC,i} \in R^{N \times 1}$ is the vector of the ith-class total regression parameters to achieve the GRLC optimization as:

$$\tilde{\beta}_{GLRC,i} = \underset{\beta_{GLRC,i}}{\text{argmin}} \sum_{k=1}^{K} \alpha_k \left(y_k - X_{i,k} \tilde{\beta}_{GLRC,i} \right)^2 \tag{53}$$

The optimal total regression parameter vector, $\tilde{\beta}_{GLRC,i}$ can be given by:

$$\tilde{\beta}_{GLRC,i} = \left(\sum_{k=1}^{K} \alpha_k X_{i,k}^T X_{i,k} \right)^{-} \left(\sum_{k=1}^{K} \alpha_k X_{i,k}^T y_k \right) \tag{54}$$

The prediction, $\hat{y}'_{i,k}$ is then expressed as $\hat{y}'_{i,k} = X_{i,k} \tilde{\beta}_{GLRC,i}$.

For identity recognition, the minimum prediction error of the GRLC should be further designed to compute the similarity between the prediction vector $\hat{y}'_{i,k}$ and the query vector y. The similarity in terms of minimization of prediction errors of total K components can be designed by the following MG criteria as:

$$i^* = \underset{i}{\text{argmin}} \left\{ \sum_{k=1}^{K} \left(1 + \left(\frac{d_{i,k}}{\bar{d}_k + \varepsilon} \right)^t \right)^{-1} \right\} \tag{55}$$

where $d_{i,k} = \alpha_k \| \hat{y}'_{i,k} - y_k \|$, $\bar{d}_k = \frac{1}{N} \sum_{i=1}^{N} \alpha_k \| \hat{y}'_{i,k} - y_k \|$ and t is the pre-selected fuzzy factor.

3.8. Trimmed linear regression (TLR)

For the occlusion situations, the previous methods including LRC, RLRC, IPCR, URC, LDRC, and GLRC are not suitable because the existing methods treat all pixels as equally import. Conversely, if the outliers can be detected and trimmed from the testing image and the corresponding training samples, the mechanism still can work. Hampel identifier [34, 35] for outlier detection is highly thought of by the researchers because it can make out the extreme values easily. An advantage of Hampel identifier is that it adopts median absolution deviation (MAD), which is a powerful measure in statistics, for removing the masking data. Mathematically, the Hampel identifier can be expressed as:

$$\frac{|\Delta - media(\Delta)|}{MAD/0.6745} > 2.24 \tag{56}$$

where Δ is a data set, $media(\Delta)$ denotes the media value of Δ data set. The number of 0.6745 is a probable error of standard deviation. When the ratio is larger than 2.24, the data will be abandoned. For example, there is a data set, [2, 3, 3, 4, 4, 250]. The sample mean is 44.33, sample variance is 100.76, sample median is 3.5, MAD equals to 0.5, and the detection rule by mean and median is:

$$\frac{|250-44.33|}{100.76} = 2.04 \tag{57}$$

and

$$\frac{|250-3.5|}{0.5 \cdot 0.6745} = 332.52 \tag{58}$$

respectively. We can observe that the Hampel identifier excludes the outlier easier than the other one.

For the face recognition, the error of estimation can be presented as:

$$\boldsymbol{e}_i = \boldsymbol{y} - \boldsymbol{X}_i \boldsymbol{\beta}_i \tag{59}$$

where error is a zero mean distribution. In order to detect the occlusion part, each pixel should suffice the Hampel identifier estimation as:

$$\varepsilon' = \left\{ \zeta \mid \frac{|e_i[\zeta]-0|}{\frac{median(|e_i[\zeta]-0|)}{0.6745}} < 2.24 \right\} \tag{60}$$

where ε' is the indices of all pixels, that is $\varepsilon' = \{1, 2, \ldots, M\}$. The real median of noise is zero. From the Equation (60), the pure pixels, ε', are found out. In other words, the pure pixels are taken for regression estimation. The training data can be rewritten as $X_{TLR,i} = [x_{TLR,i,1}, x_{TLR,i,2}, \ldots, x_{TLR,i,N}] \in R^{\tau \times N}$ and testing sample becomes $y_{TLR} \in R^{\tau \times 1}$ where τ is the number of elements in ε' and $\tau < M$. The objective function becomes:

$$\widetilde{\boldsymbol{\beta}}_{TLR,i} = \mathop{\text{argmin}}_{\boldsymbol{\beta}_{TLR,i}} \left\| \boldsymbol{y}_{TLR} - \boldsymbol{X}_{TLR} \widetilde{\boldsymbol{\beta}}_{TLR,i} \right\|_2^2 \tag{61}$$

The regression parameter vectors can be represented as:

$$\widehat{\boldsymbol{\beta}}_{TLR,i} = \left(\boldsymbol{X}_{TLR,i}^T \boldsymbol{X}_{TLR,i} \right)^{-1} \boldsymbol{X}_{TLR,i}^T \boldsymbol{y}_{TLR} \tag{62}$$

4. Experimental results

In order to verify the recognition accuracy, the well-known databases including Yale B, AR, FERET, and FEI are utilized. In the experiments, we evaluate the mentioned method against low resolution problem coupled with facial expressions, illumination changes, pose variations, and partial occlusions.

4.1. Yale B database

The Yale B database contains 10 subjects [36, 37]. Each subject has 64 illumination images with 9 different poses. The Yale B can be divided into five subset based on angle of the light source directions as shown in **Figure 4**. In the experiments, the first subset with normal pose is used for training and the remaining subsets (Subset 2 to 5) with normal pose are utilized for testing. All images are cropped and resized to 30×25 pixels. **Table 1** reveals that IPCRC performs better than the traditional subspace projection like PCA and LDA. Moreover, the IPCRC can also outperform the LRC, RLRC and RR. The reason is that the original subspace cannot represent the data distribution very well. Besides, PCA subspace is very sensitive to illuminant variations. However, IPCRC not only can transform to PCA subspace, but also can defend the illumination variations by removing the top n components. Thus, IPCRC possesses higher robustness to illuminations than the other methods.

Methods	Subset 2	Subset 3	Subset 4	Subset 5
PCA	89.81	47.04	21.90	15.26
LDA	95.14	75.14	34.76	10.00
LRC	100.00	100.00	91.86	52.11
RLRC	100.00	100.00	92.86	60.00
RR	100.00	100.00	92.57	53.68
IPCRC	100.00	100.00	95.00	64.21
LDRC	100.00	100.00	97.14	56.84
URC	100.00	100.00	90.71	53.68

Table 1. Accuracy (%) comparisons on Yale B.
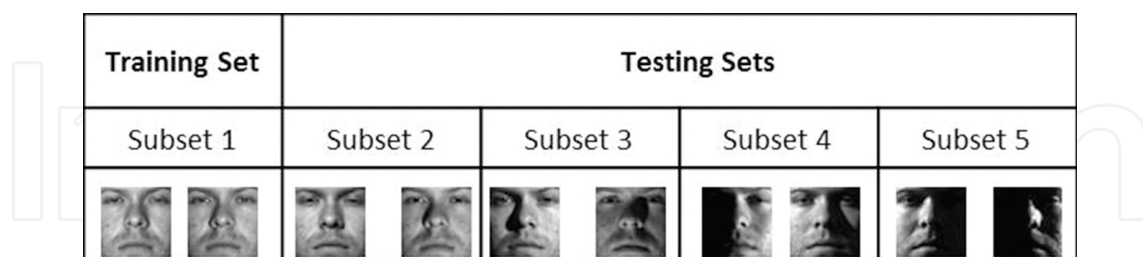
Figure 4. The experimental design and some samples of cropped and aligned illustration from Yale B face database.

### 4.2. FERET database

Furthermore, we experiment on the FERET face database [38, 39] for the purpose of verifying the performance among the different subspace projections. In the experiments, we select four facial images including fa, fb, ql, and qr from 300 subjects as **Figure 5**. All images are converted, cropped, and downsampled to 30×25 pixels with grayscale. As the **Figure 5** shown, the fa and

fb samples are small pose and rotation changes; conversely, the ql and qr samples are major pose variations. In order to obtain a reliable result, cross-validation experimental procedure is adopted. In other words, three images per person are used for training while the fourth image is used for testing. **Table 2** shows that the average recognition accuracy (ARA) in URC performs outstandingly. We can observe that the RLRC and IPCRC are highly sensitive to pose variations but in spite of these, methods perform well in noisy and illuminated face images, respectively.

	PCA	LDA	LRC	RLRC	RR	IPCRC	URC
fa	80.67	87.33	94.00	91.67	85.67	92.33	96.00
fb	81.00	84.33	92.33	90.00	83.67	91.00	95.33
ql	65.67	63.00	71.00	69.33	63.33	66.00	73.00
qr	68.33	72.00	75.00	74.33	70.00	68.67	84.33
ARA	73.92	76.67	83.08	81.33	75.67	79.50	87.17

Table 2. Accuracy (%) comparisons on FERET.
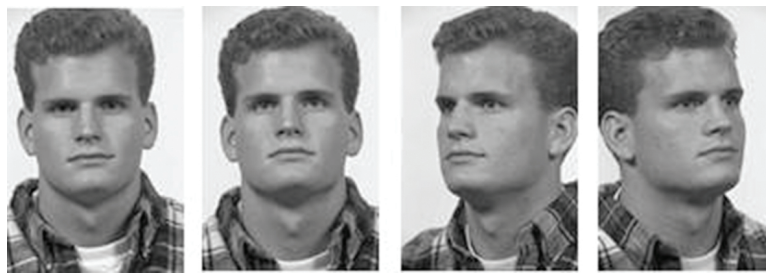
Figure 5. Samples (fa, fb, ql, qr) of one subject from FERET face database.

4.3. AR database

AR face database [40, 41] was conducted by Martinez and Benavente in 1998. This database contains 4000 mug shots of 126 subjects (70 males and 56 females) with different variations such as facial expressions, lighting changes and partial occlusions. For normal case, each subject contains 26 images in two sessions. The first session (AR1 ~ AR13), containing 13 photos, includes facial expression, different lighting changes, and partial occlusions (sunglasses and scarf) with lighting changes. The second session (AR14 ~ AR26) duplicates the same way of first session two weeks later as shown in **Figure 6**. In the experiments, 100 subjects are selected and all images are cropped and resized into 30×25 pixels with grayscale. We classify the images into four different expressions including neutral (AR4, AR14), happy (AR2, AR3), angry (AR1, AR17), and screaming (AR15, AR16) expressions. The single-one-expression training strategy is adopted to present the performance. For example, if neutral expression images are used for training, the happy, angry, and screaming expressions are used as query images. **Table 3** reveals that the LDRC achieves the best performance in all cases. Moreover,

we can observe that the happy expression images for training obtain higher performance than the others; conversely, the screaming expression images for testing can obtain lowest performance. On the other hand, the partial occlusion situations are used to discussion. In this experiments, the expression variation images (AR1~AR4, AR14~AR17) are utilized as training set, and testing sets are separated in two cases including sunglasses (AR8, AR21) and scarf (AR11, AR24). All images are cropped and resized into 42×30 pixels with grayscale. In the **Table 4**, we can observe two points. First, the TLRC can perform better than the other methods under sunglasses occlusion or scarf occlusion. Second, the upper bound occlusion seems to obtain higher performance than the lower bound occlusion. In other words, the mouth features are more useful than the eye features.

Training Set	Testing Set	PCA	LDA	LRC	RLRC	RR	IPCRC	LDRC	GLRC
N	H,A,S	88.75	86.53	89.03	87.50	84.17	88.19	91.39	89.00
H	N,A,S	90.97	89.17	92.08	91.67	82.33	92.08	93.61	54.33
A	N,H,S	90.00	85.69	90.00	89.17	83.33	89.44	90.42	89.83
S	N,H,A	88.19	84.58	87.78	86.53	77.88	87.58	90.42	55.50

Table 3. Accuracy (%) comparisons on AR.

Figure 6. Samples of one subject from AR face database.

Training Set	Testing Set	PCA	LDA	SRC	LRC	RLRC	RR	IPCRC	GLRC	TLRC
AR1~ AR4; AR14~ AR17	Sunglasses (AR8, AR21)	42.5	20.5	87.0	65.5	47.5	59.0	44.5	90.5	100.0
	Scarf (AR11, AR24)	7.0	33.5	59.5	12.5	9.5	10.5	6.0	35.5	94.5

Table 4. Accuracy (%) comparisons under partial occlusion problem on AR.

4.4. FEI database

The FEI face database [42, 43] contains 200 subjects (100 males and 100 females). Each subject has 14 images with different pose variations (image1~image10), facial expressions (image11~image12), and illumination variations (image13~image14) as shown in **Figure 7**. In the experiments, all images are resized to 24×20 pixels with grayscale and the "leave-one-out strategy" is adopted. From **Table 5**, it can be seen that the IPCRC is more robust to severe lighting variation (image 14) and URC is good at facial profiles (image 1, image 10). All in all, the ARA of URC performs the best.

		PCA	LDA	LRC	RLRC	RR	IPCRC	LDRC	URC	GLRC
Test Image	1	91.0	91.5	92.0	89.0	89.0	86.0	95.0	95.0	97.0
	2	99.5	99.0	100.0	100.0	100.0	99.0	99.5	100.0	100.0
	3	97.0	99.0	99.5	99.5	99.5	99.5	99.0	100.0	100.0
	4	98.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	5	97.5	99.5	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	6	96.5	99.5	99.0	99.0	99.0	99.0	99.5	100.0	99.5
	7	99.5	100.0	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	8	99.0	99.5	100.0	100.0	100.0	100.0	100.0	100.0	100.0
	9	97.0	99.5	99.5	99.5	99.5	99.5	100.0	100.0	100.0
	10	79.5	83.0	83.5	78.0	78.0	72.0	86.0	91.5	91.5
	11	98.0	100.0	99.5	99.5	99.5	99.5	100.0	99.5	100.0
	12	97.0	99.5	99.0	99.0	99.0	97.5	99.0	98.5	99.5
	13	47.0	87.5	97.0	97.5	97.5	99.0	98.5	99.5	94.5
	14	23.5	39.5	79.5	91.0	91.0	92.5	83.0	88.5	77.5
ARA		92.04	96.73	97.62	97.00	97.00	96.23	97.11	98.77	97.11

Table 5. Accuracy (%) comparisons on FEI.

Figure 7. Samples of one subject from FEI database.

4.5. Discussions

From the experimental results, we can observe that IPCRC has a good performance under illumination situation. The reason is that the first n components in IPCRC are removed. The first n components are very sensitive to the lighting changes. However, although IPCRC has better performance under the lighting changes, it cannot handle the pose variations and occlusion problems very well. For the pose variations, the URC performs better than the other subspace methods because URC attempts to minimize the total intra-class reconstruction error to find an optimal projection which can decrease the pose influence. LDRC embeds discriminant analysis into the LRC for seeking an optimal projection matrix such that the LRC on that subspace has high discriminatory ability for classification. In other words, LDRC can perform better than LRC and IPCRC in most cases. In the occlusion situation, the TLRC can effectively remove the masking data and project onto a more reliable subspace.

5. Conclusions

In this chapter, we presented several subspace projection methods for robust face recognition to deal with different practical situations such as pose variations, lighting changes, facial expressions, and partial occlusions.

For illumination variation task in face recognition, an improved principal component classification can be used to solve the multicollinearity problem and can perform better recognition accuracy than the original linear regression and RR. For the pose variations, a URC has been presented to minimize the total within-class projection error from all classes for LRC to improve the robustness for pattern recognition. Moreover, a LDRC has been proposed to overcome facial expressions by maximizing the ratio of the BCRE to the WCRE by the LRC. For the partial occlusions, a trimmed regression classification is used to remove unreliable pixels by the Hampel identifier. Finally, experimental results have revealed the comparisons with different subspace projection optimizations.

Author details

Yang-Ting Chou, Jar-Ferr Yang[*] and Shih-Ming Huang

*Address all correspondence to: jarferryang@gmail.com

Institute of Computer and Communication Engineering, Department of Electrical Engineering, National Cheng Kung University, Tainan, Taiwan

References

[1] Yang, J.; Zhang, D.; Frangi, A.F.; Yang, J-Y. Two-dimensional PCA: a new approach to appearance-based face representation and recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2004;26(1):131–137.

[2] Shawe-Taylor, J.; Cristianini, N., editors. Kernel methods for pattern analysis. Cambridge University Press Inc., 2004. ISBN: 0521813972.

[3] Turk, M.; Pentland, A. Eigenfaces for recognition. Journal of Cognitive Neuroscience. 1991;3(1):71–86.

[4] Belhumeur, P.N.; Hespanha, J.P.; Kriegman, D. Eigenfaces vs. fisherfaces: recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1997;19(7):711–720.

[5] Martínez, A.M.; Kak, A.C. Pca versus lda. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2001;23(2):228–233.

[6] Schölkopf, B., Smola, A.; Müller, K.R. Nonlinear component analysis as a kernel eigenvalue problem. Neural Computation. 1998;10(5):1299–1319.

[7] Yang, M.H. Kernel eigenfaces vs. kernel fisherfaces: face recognition using kernel methods. In: 2013 10th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition; 2002. p. 0215.

[8] Naseem, I.; Togneri, R.; Bennamoun, M. Linear regression for face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2010;32(11):2106–2112.

[9] Naseem, I.; Togneri, R.; Bennamoun, M. Robust regression for face recognition. Pattern Recognition. 2012;45(1):104–118.

[10] Xue, H.; Zhu, Y.; Chen, S. Local ridge regression for face recognition. Neurocomputing. 2009;72(4):1342–1346.

[11] Huang, S.M.; Yang, J.F. Improved principal component regression for face recognition under illumination variations. IEEE Signal Processing Letters. 2012;19(4):179–182.

[12] Huang, S.M.; Yang, J.F. Linear discriminant regression classification for face recognition. IEEE Signal Processing Letters. 2013;20(1):91–94.

[13] Huang, S.M.; Yang, J.F. Unitary regression classification with total minimum projection error for face recognition. IEEE Signal Processing Letters. 2013;20(5):443–446.

[14] Chou, Y.T.; Yang, J.F. Identity recognition based on generalised linear regression classification for multi-component images. IET Computer Vision. 2016;10(1):18–27.

[15] Chou, Y.T.; Yang, J.F. Object recognition based on generalized linear regression classification in use of color information. In: 2014 IEEE Asia Pacific Conference on In-Circuits and Systems; 2014. 272–275.

[16] Lai, J.; Jiang, X. Robust face recognition using trimmed linear regression. In: ICASSP; 2013. 2979–2983.

[17] Yang, G.; Huang, T.S. Human face detection in a complex back-ground. Pattern Recognition. 1994;27(1):53–63.

[18] Hotta, K.; Kurita, T.; Mishima, T. Scale invariant face detection method using higher-order local autocorrelation features extracted from log-polar image. In: Third IEEE International Conference on Automatic Face and Gesture Recognition; 1998. 70–75.

[19] Han, C.C.; Liao, H.Y.M.; Yu, G.J.; Chen, L.H. Fast face detection via morphology-based pre-processing. Pattern Recognition. 2000;33(10):1701–1712.

[20] Fasel, B. Fast multi-scale face detection. IDIAP-Com-04-1998. 1998.

[21] Rowley, H.A.; Baluja, S.; Kanade, T. Neural network-based face detection. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1998;20(1):23–38.

[22] Viola, P.; Jones, M.J. Robust real-time face detection. International Journal of Computer Vision. 2004;57(2):137–154.

[23] Lienhart, R.; Maydt, J. An extended set of haar-like features for rapid object detection. In: 2002 International Conference on Image Processing; 2002. 1–900.

[24] Kotropoulos, C.; Pitas, I. Rule-based face detection in frontal views. In: 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing; 1997. 2537–2540.

[25] Birchfield, S. An elliptical head tracker. In: IEEE Conference Record of the Thirty-First Asilomar Conference on Signals, Systems & Amp; 1997. 1710–1714.

[26] Wu, H.; Chen, Q.; Yachida, M. Face detection from color images using a fuzzy pattern matching method. IEEE Transactions on Pattern Analysis and Machine Intelligence. 1999;21(6):557–563.

[27] Orfanidis, S. SVD, PCA, KLT, CCA, and All That. Rutgers University Electrical and Computer Engineering Department, Optimum Signal Pro-cessing. 2007;1–77.

[28] Golub, G.H.; Van Loan, C.F., editors. Matrix Computations. JHU Press. 2012.

[29] Watkins, D.S., editors. Fundamentals of Matrix Computations. John Wiley & Sons Inc.. 2004. ISBN: 978-0-470-52833-4.

[30] Huber, P.J., editors. Robust Statistics. Berlin Heidelberg: Springer. 2011.

[31] Heyde, C.C., editors. Quasi-Likelihood and Its Application: A General Approach to Optimal Parameter Estimation. Springer-Verlag New York Berlin Heidelerg Inc.. 1997. ISBN: 0-387-98225-6.

[32] Fraiman, R.; Meloche, J.; García-Escudero, L.A.; Gordaliza, A.; He, X.; Maronna, R.; Yohai, V.J.; Sheather, S.J.; McKean, J.W.; Small, C.G.; Wood, A.; Fraiman, R.; Meloche, J., editors. Multivariate L-Estimation. Test. 1999; 8(2): 255-317.

[33] Hampel, F.R.; Ronchetti, E.M.; Rousseeuw, P.J.; Stahel, W.A., editors. Robust Statistics: The Approach Based on Influence Functions. John Wiley & Sons Inc.. 2011. ISBN: 9781118186435.

[34] Wilcox, R.R., editors. Applying Contemporary Statistical Techniques. Elsevier Inc.. Gulf Professional Publishing. 2003. ISBN: 978-0-12-751541-0.

[35] Wilcox, R.R., editors. Introduction to Robust Estimation and Hypothesis Testing. Academic Press. 2012.

[36] Georghiades, A.S.; Belhumeur, P.N.; Kriegman, D. From few to many: illumination cone models for face recognition under variable lighting and pose. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2001;23(6):643–660.

[37] Georghiades, A.S.; Belhumeur, P.N.; Kriegman, D. Extended Yale Face Database B. Available from: http://vision.ucsd.edu/~leekc/ExtYaleDatabase/

[38] Phillips, P.J.; Moon, H.; Rizvi, S.A.; Rauss, P.J. The FERET evaluation methodology for face-recognition algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence. 2000;22(10):1090–1104.

[39] Phillips, P.J.; Moon, H.; Rizvi, S.A.; Rauss, P.J. The FERET Database. Available from: http://www.itl.nist.gov/iad/humanid/feret/feret_master.html

[40] Martinez, A.M. The AR face database. In: CVC Technical Report #24; 1998.

[41] Martinez, A.M. The AR Face Database. Available from: http://www2.ece.ohio-state.edu/~aleix/ARdatabase.html

[42] OLIVEIRA; JR, L. L.; Thomaz, C. E. Captura e alinhamento de imagens: Um banco de faces brasileiro. Relatório de iniciação científica, Depto. Eng. Elétrica da FEI, São Bernardo do Campo, SP, 2006. 10: 1-10.

[43] OLIVEIRA; JR, L. L.; Thomaz, C. E. FEI Face Database. Available from: http://www.fei.edu.br/~cet/facedatabase.html