

RESEARCH

Open Access



Class-specific kernel linear regression classification for face recognition under low-resolution and illumination variation conditions

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

Abstract

In this paper, a novel class-specific kernel linear regression classification is proposed for face recognition under very low-resolution and severe illumination variation conditions. Since the low-resolution problem coupled with illumination variations makes ill-posed data distribution, the nonlinear projection rendered by a kernel function would enhance the modeling capability of linear regression for the ill-posed data distribution. The explicit knowledge of the nonlinear mapping function can be avoided by using the kernel trick. To reduce nonlinear redundancy, the low-rank- r approximation is suggested to make the kernel projection be feasible for classification. With the proposed class-specific kernel projection combined with linear regression classification, the class label can be determined by calculating the minimum projection error. Experiments on 8×8 and 8×6 images down-sampled from extended Yale B, FERET, and AR facial databases revealed that the proposed algorithm outperforms the state-of-the-art methods under severe illumination variation and very low-resolution conditions.

Keywords: Face recognition, Kernel linear regression classification

1 Introduction

Numerous studies [1] have been greatly proposed for face recognition recently. In realistic situations such as video surveillance, face recognition may encounter many great challenges, especially low-resolution problems, caused by the cameras at a distance. Additionally, the low-resolution (LR) problems might be coupled with other effects such as illumination variations. Therefore, it is desired to devise a face recognition method for both very low-resolution and illumination variation problems.

In the literature, numerous researches based on the subspace projection methods have been proposed to achieve successful face recognition. The principle component analysis (PCA) [2–6] and independent component analysis (ICA) [7–9] have been reported to be robust in noisy conditions. The linear discriminant analysis (LDA) [3–6] yields better results in clean conditions and lighting

changes than the PCA. Moreover, the kernel variants such as kernel PCA (KPCA) [10–12] and kernel LDA (KLDA) [12–14] have been presented to achieve better performance by nonlinearly mapping the data from the original space to a very high-dimensional feature space, which is called the reproducing kernel Hilbert space (RKHS). Therefore, KPCA and KLDA by nonlinearly mapping could be improved from high-order statistics, whereas the PCA and LDA only utilize the first- and second-order statistics. Thus, for highly nonlinear data distribution, these kernel methods are more suitable for low-resolution and illumination variation conditions. Moreover, some geometrically motivated approaches such as locality preserving projection (LPP) [15] and neighboring preserving embedding (NPE) [16] have been shown effectiveness for face recognition.

Recently, the sparse representation classification (SRC) [17, 18] and a linear regression classification (LRC) algorithms [19] have been proposed for face recognition. Although SRC-based approaches perform very well in many situations, the execution time in SRC-based approaches is more than that in LRC-based approaches. For pursuing the

* Correspondence: jarferyyang@gmail.com

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

accuracy and speed, the LRC would be a good choice for further investigation. The LRC is based on that face images from a specific class are known to lie on a class-specific linear subspace [3, 20]. The regression coefficients can be estimated by using the least-square method, and the decision can be determined by the minimum reconstruction error. Experiments reported have shown that the down-sampled low-resolution face image could be used for face classification directly. However, as the results reported, the LRC could not withstand severe illumination variations. In addition, a robust linear regression classification algorithm (RLRC) [21] has been introduced to address the problem of robust face recognition. However, low-resolution problems did not be addressed.

The performances achieved by the existing methods, such as PCA, LDA, LPP, and NPE methods, decrease under low-resolution condition because of the loss of high-frequency information [22]. Boom et al. in [23] showed that low-resolution face images below 32×32 pixels degraded the performances of the PCA and LDA seriously. In [24], the face images with 20×20 and 10×10 resolution dramatically deteriorated the recognition performance compared with those with 40×40 pixels in video-based face recognition systems. In [25], face resolution below 36×48 reduced the expression recognition performance.

To overcome the problem of low-resolution face recognition, several works have been presented to resolve this problem by using the super-resolution (SR) method [26–30]. One is to train relationship between the low-resolution (LR) and its corresponding high-resolution (HR) face images [29]. The other uses the canonical correlation analysis to compute the coherent features between the LR and its HR face images [30]. Here, we can observe that the LR and HR image pairs are needed for SR methods.

1.1 Problem statements

In realistic situations, it may be a case that only the LR face images are available in training set for identifying criminals, so it is imperative to overcome the problem that some training individuals may not be the same as gallery. In other words, the HR face images for specific persons are not available for modeling the relationship and computing the similarity in SR approaches [26–30]. Thus, how to perform LR face recognition directly without the HR information is a critical and practical topic.

To conquer the illumination variation problems, several approaches [3, 31–35] have been proposed. For instance, numerous preprocessing methods, such as the histogram equalization, gamma correction, and logarithm transform, are widely used for the illumination normalization. Other methods including the gradient operation, Gabor filters, and LDA-based approaches are well-known illumination invariant methods. However,

these methods would fail because the important features in high-frequency details for face recognition are lost under the LR problems.

1.2 Contributions

We propose a novel face recognition algorithm to improve the limitation of the LRC [15] by embedding the kernel method into the linear regression. The key of the proposed method is to apply a nonlinear mapping function to twist the original space into a higher dimensional feature space for better linear regression. Moreover, in order to make the proposed kernel projection feasible, a constrained low-rank approximation [36–38] is proposed to obtain low-rank- r singular value approximation. The low-rank approximation is a rank reduction method which minimizes the difference between a given matrix and an approximation matrix. Simulations carried on the extended Yale B, FERET, and AR facial databases reveal that the proposed kernel linear regression classification (KLRC) can achieve good performance for LR face recognition under variable lighting changes without any preprocessing. At the same time, the proposed algorithm can reconstruct the very low-resolution face image under illumination variations with high quality measured by quality assessments.

1.3 Paper outline

The rest of this paper is organized as follows. Section 2 reviews the LRC approach and presents the motivations. Section 3 formulates the proposed KLRC method with a constrained low-rank approximation algorithm. Section 4 shows the comparisons with the related work. Section 5 gives experimental results. Finally, we draw conclusions in Section 6.

2 Background and motivations

2.1 Linear regression classification (LRC)

Assume we have N subjects with p_i training images from the i th class, $i = 1, 2, \dots, N$. Each gray scale training image is in size of $a \times b$ pixels and is represented as $v_{i,j} \in \mathcal{R}^{a \times b}$, $i = 1, 2, \dots, N$ and $j = 1, 2, \dots, p_i$. Then, each training image is transformed to a column vector as $w_{i,j} \in \mathcal{R}^{q \times 1}$, where $q = a \times b$. For applying the linear regression to estimate the class-specific model, we stack all column vectors $w_{i,j}$ regarding the class-membership. Hence, for the i th class, we have

$$W_i = [w_{i,1}, \dots, w_{i,j}, \dots, w_{i,p_i}] \in \mathcal{R}^{q \times p_i}, \quad (1)$$

where each vector $w_{i,j}$ is a column vector of W_i . Thus, in the training phase, the i th class is represented by a vector space W_i , which is called the regressor for each subject.

If \mathbf{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

$$\mathbf{y} = \mathbf{W}_i \boldsymbol{\beta}_i + \mathbf{e}, i = 1, 2, \dots, N, \quad (2)$$

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

$$\tilde{\boldsymbol{\beta}}_i = \arg \min_{\boldsymbol{\beta}_i} \|\mathbf{W}_i \boldsymbol{\beta}_i - \mathbf{y}\|_2^2, i = 1, 2, \dots, N \quad (3)$$

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

$$\tilde{\boldsymbol{\beta}}_i = (\mathbf{W}_i^T \mathbf{W}_i)^{-1} \mathbf{W}_i^T \mathbf{y}, i = 1, 2, \dots, N \quad (4)$$

The vector of estimated parameters, $\tilde{\boldsymbol{\beta}}_i$, and the predictor, \mathbf{W}_i , are used to predict the response vector $\tilde{\mathbf{y}}_i$ for the i th class as

$$\tilde{\mathbf{y}}_i = \mathbf{W}_i \tilde{\boldsymbol{\beta}}_i, i = 1, 2, \dots, N \quad (5)$$

By substituting (4) for $\tilde{\boldsymbol{\beta}}_i$ in (5), the optimal prediction in the least-square sense becomes

$$\tilde{\mathbf{y}}_i = \mathbf{W}_i (\mathbf{W}_i^T \mathbf{W}_i)^{-1} \mathbf{W}_i^T \mathbf{y}, i = 1, 2, \dots, N \quad (6)$$

Theoretically, we can treat the above equation as a class-specific projection as [39],

$$\tilde{\mathbf{y}}_i = \mathbf{H}_i \mathbf{y}, i = 1, 2, \dots, N, \quad (7)$$

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

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 $\tilde{\mathbf{y}}_i$ will be the closest vector to the original vector. The identity i^* could be determined by calculating the Euclidean distance measure between the predicted response vectors and the original vector as

$$\begin{aligned} i^* &= \arg \min_i \|\tilde{\mathbf{y}}_i - \mathbf{y}\|, i = 1, 2, \dots, N \\ &= \arg \min_i \|\mathbf{H}_i \mathbf{y} - \mathbf{y}\| \end{aligned} \quad (8)$$

2.2 Motivations

The LRC has been developed based on the concept that samples from a specific person are known to lie on a class-specific linear subspace and demonstrated that it

could achieve good performance for the low-resolution face images, but not good for severe illumination variations. This is because illumination variations make the data distribution more complicated. So the face images captured under variable lighting conditions may cause the linear subspace approaches inappropriate. In other words, the linear subspace methods would fail when they violate the Lambertian assumption regarding the illumination problem. Especially, when the low-resolution problem is coupled with illumination variations, the linear subspace methods, such as the PCA and LDA, and the linear regression classification (LRC) could not counteract the problem. In this paper, a kernel linear regression classification (KLRC) with a constrained low-rank approximation is proposed for low-resolution face recognition under illumination variations. The KLRC with the nonlinear mapping function can evaluate the LRC in the higher dimensional feature space and can achieve good results.

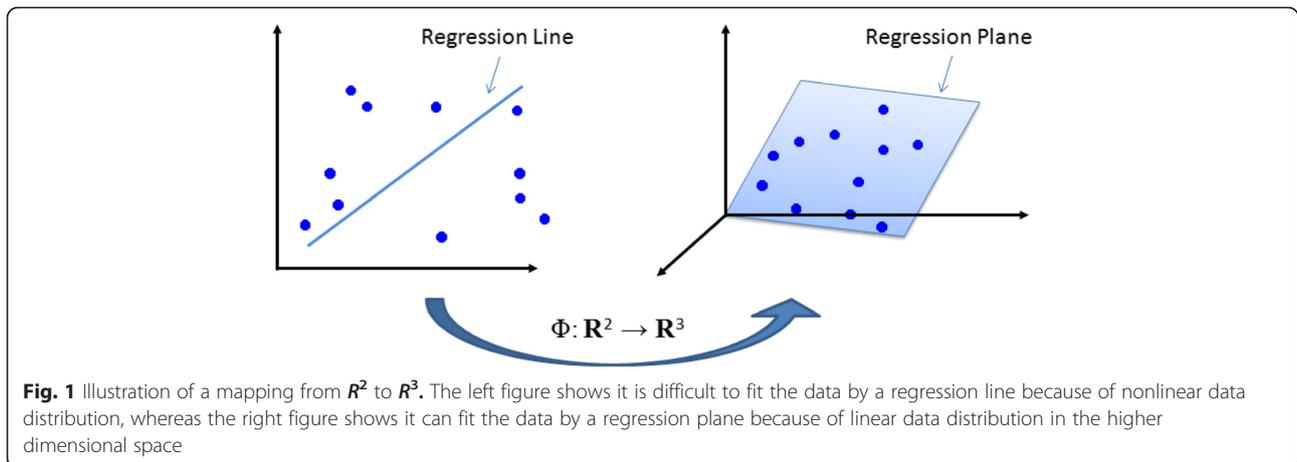
3 KLRC

Assume the original input space can be always mapped to some higher dimensional feature space where the data set is distributed linearly. As shown in Fig. 1, the left figure shows that it is difficult to fit the data by a regression line because of nonlinear data distribution, whereas the right figure shows that it is easy to fit the data by a regression plane because of data distribution linearly by a mapping function from \mathbf{R}^2 to \mathbf{R}^3 . Thus, it can be expected that a nonlinear mapping prior to linear regression could improve the limitation of the LRC under severe illumination variations. In order to formulate a general equation and solve the problem systematically, we will discuss it in details later. Here, we first introduce the kernel linear regression classification (KLRC) method. The KLRC is also developed based on the theory that samples from a specific class are known to lie on a linear subspace by a nonlinear mapping. The key is to apply a nonlinear mapping function to the input space and then evaluate it by the LRC in the higher dimensional feature space. The dimension of the resulting feature space could be very large. Fortunately, this explicit knowledge of the nonlinear mapping function can be computationally avoided by using the kernel trick [40].

For the following theoretical derivation, a vector space should be defined as

$$\mathbf{Z}_i = \begin{bmatrix} \mathbf{z}_{i,1} \\ \vdots \\ \mathbf{z}_{i,j} \\ \vdots \\ \mathbf{z}_{i,q} \end{bmatrix} = \mathbf{W}_i, \quad (9)$$

where each vector $\mathbf{z}_{i,j}$ is a row vector of \mathbf{Z}_i . Specifically, each row vector in \mathbf{Z}_i is projected from the original



space \mathcal{H}^{p_i} to a high-dimensional space \mathcal{H}^f by a nonlinear mapping function $\Phi(\mathbf{z}_{i,j}) : \mathcal{H}^{p_i} \rightarrow \mathcal{H}^f, f > p_i$. Therefore, now, \mathcal{H}^f is the space spanned by $\Phi(\mathbf{z}_{i,j})$. The projected row vectors can be used for linear regression as

$$\mathbf{y} = \Phi(\mathbf{Z}_i)\boldsymbol{\beta}_i, \quad i = 1, 2, \dots, N \tag{10}$$

Because of the increase in dimensionality, the mapping function $\Phi(\mathbf{z}_{i,j})$ is made implicitly by using the kernel function satisfying Mercer’s theorem. Furthermore, by using the dual representation $\boldsymbol{\beta}_i = \Phi(\mathbf{Z}_i)^T \boldsymbol{\alpha}_i$, the linear regression stated in (10) becomes

$$\mathbf{y} = \Phi(\mathbf{Z}_i)\Phi(\mathbf{Z}_i)^T \boldsymbol{\alpha}_i = \mathbf{K}_i \boldsymbol{\alpha}_i, \quad i = 1, 2, \dots, N, \tag{11}$$

where the kernel matrix \mathbf{K}_i is positive semi-definite. Typically, kernel functions include the polynomial kernel and Gaussian kernel, which satisfy Mercer’s theorem.

We first perform singular value decomposition (SVD) on the kernel matrix \mathbf{K}_i as

$$\mathbf{K}_i = \mathbf{U}\mathbf{S}\mathbf{V}^T, \tag{12}$$

where \mathbf{U} and \mathbf{V}^T are left and right SVD orthonormal matrices and $\mathbf{S} = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_g\}$ is an rectangular diagonal matrix with the descend-sorted singular values on the diagonal with $\lambda_1 \geq \dots \geq \lambda_k \geq \dots \geq \lambda_g \geq 0$. To achieve a robust estimation, we propose a constrained rank- r approximation of \mathbf{K}_i defined as

$$\mathbf{K}_i^r = \mathbf{U}\mathbf{S}^r\mathbf{V}^T, \tag{13}$$

where

$$\mathbf{S}^r = \text{diag}\{\lambda_1, \lambda_2, \dots, \lambda_r, 0, 0, \dots, 0\}, \tag{14}$$

by discarding $(g - r)$ least SVD components. The number of principle SVD components, r , is determined by

$$r = \arg \max_k \{ \forall \lambda_k | \lambda_k > \lambda_{\text{median}} + \mu(\lambda_{\text{median}} - \lambda_g) \}, \tag{15}$$

where μ is a selected control factor, λ_g is the smallest

singular value, and λ_{median} , which is the median of all singular values, is expressed as

$$\lambda_{\text{median}} = \text{median}\{\forall \lambda_k | k < g\} \tag{16}$$

After the construction of the constrained low-rank approximation stated in (13), we could obtain a kernel linear regression model for the i th class as

$$\mathbf{y} \approx \mathbf{K}_i^r \boldsymbol{\alpha}_i + \mathbf{e} \tag{17}$$

Then, the kernel linear regression aims to minimize the residual errors as

$$\tilde{\boldsymbol{\alpha}}_i = \arg \min_{\boldsymbol{\alpha}_i} \|\mathbf{K}_i^r \boldsymbol{\alpha}_i - \mathbf{y}\|_2^2. \tag{18}$$

The above solution can be also solved by the least-square estimation since it has the same form as stated in (2). After the low-rank approximation, we can use the pseudo-inverse of \mathbf{K}_i^r to obtain the least-square solution as

$$\tilde{\boldsymbol{\alpha}}_i = (\mathbf{K}_i^r)^- \mathbf{y}, \tag{19}$$

where the pseudo-inverse of \mathbf{K}_i^r is expressed by

$$(\mathbf{K}_i^r)^- = \mathbf{U}(\mathbf{S}^r)^- \mathbf{V}^T, \tag{20}$$

with

$$(\mathbf{S}^r)^- = \text{diag}\{\lambda_1^{-1}, \lambda_2^{-1}, \dots, \lambda_r^{-1}, 0, 0, \dots, 0\} \tag{21}$$

Since $\mathbf{K}_i^r (\mathbf{K}_i^r)^- \neq \mathbf{I}$, it will be feasible to compute the minimum reconstruction error between the original vector and projected vector for determining the classification results.

In the classification phase, the response vector \hat{y}_i for the i th class can be predicted by

$$\hat{\mathbf{y}}_i = \mathbf{K}_i^r \tilde{\boldsymbol{\alpha}}_i. \tag{22}$$

By substituting (19) in (22), we can obtain

$$\tilde{\mathbf{y}}_i = \mathbf{P}_i \mathbf{y}, \quad (23)$$

and obtain a class-specific kernel projection matrix as

$$\mathbf{P}_i = \mathbf{K}_i^r (\mathbf{K}_i^r)^{-}, \quad (24)$$

where $\tilde{\mathbf{y}}_i$ is the projection of \mathbf{y} onto the kernel subspace of the i th class by the class-specific kernel projection matrix, \mathbf{P}_i . It is noted that $\mathbf{K}_i^r (\mathbf{K}_i^r)^{-} \neq \mathbf{I}$ is necessary for the KLRC computation.

The KLRC is also developed based on the minimum reconstruction error. So in the recognition phase, the identity i^* could be determined by calculating the Euclidean distance measure between the predicted response vectors and the original vector as

$$\begin{aligned} i^* &= \arg \min_i \|\tilde{\mathbf{y}}_i - \mathbf{y}\|, \quad i = 1, 2, \dots, N \\ &= \arg \min_i \|\mathbf{P}_i \mathbf{y} - \mathbf{y}\| \end{aligned} \quad (25)$$

4 Comparison with the related works

4.1 Analysis of the regression parameter

To simplify the analysis, we assume that \mathbf{W}_i is a square matrix. We have $\mathbf{W}_i = \mathbf{U}_i \mathbf{D}_i \mathbf{V}_i^T$ by SVD with $\mathbf{U}_i^T \mathbf{U}_i = \mathbf{I}$, because \mathbf{W}_i is a square matrix, $\mathbf{U}_i^T = \mathbf{U}_i^{-1}$, and $\mathbf{U}_i \mathbf{U}_i^T = \mathbf{I}$. Similarly, $\mathbf{V}_i \mathbf{V}_i^T = \mathbf{V}_i^T \mathbf{V}_i = \mathbf{I}$. In addition, the linear kernel $k(\mathbf{z}_{ij}, \mathbf{z}_{ij}) = \langle \mathbf{z}_{ij}, \mathbf{z}_{ij} \rangle = \mathbf{z}_{ij} \mathbf{z}_{ij}^T$ is used in the KLRC for the theoretical analysis below.

4.1.1 LRC

The goal of the LRC is to find $\tilde{\boldsymbol{\beta}}_i$, which minimizes the residual errors. Statistically, the linear regression model is an unbiased estimate. Also, the variance of the regression parameter vector $\tilde{\boldsymbol{\beta}}_i$ in the linear regression model is expressed as

$$\begin{aligned} \text{Var}(\tilde{\boldsymbol{\beta}}_i)_{\text{LRC}} &= E \left\{ (\tilde{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i) (\tilde{\boldsymbol{\beta}}_i - \boldsymbol{\beta}_i)^T \right\} \\ &= E \left\{ (\mathbf{W}_i^T \mathbf{W}_i)^{-1} \mathbf{W}_i^T \mathbf{e} \mathbf{e}^T \mathbf{W}_i (\mathbf{W}_i^T \mathbf{W}_i)^{-1} \right\} \\ &= \sigma^2 (\mathbf{W}_i^T \mathbf{W}_i)^{-1} = \sigma^2 (\mathbf{V}_i \mathbf{D}_i \mathbf{U}_i^T \mathbf{U}_i \mathbf{D}_i \mathbf{V}_i^T)^{-1} \\ &= \sigma^2 \sum_{j=1}^r \frac{1}{d_{ij}} \mathbf{v}_{ij} \mathbf{v}_{ij}^T, \end{aligned} \quad (26)$$

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

4.1.2 KLRC

The goal of the KLRC is to find $\tilde{\boldsymbol{\alpha}}_i$, which minimizes the residual errors. Statistically, the kernel linear regression model is also an unbiased estimator since the kernel linear regression model in (17) has the same form as in (2). On the other side, the variance of the regression

parameter vector in the kernel linear regression model is expressed as

$$\begin{aligned} \text{Var}(\tilde{\boldsymbol{\alpha}}_i)_{\text{KLRC}} &= E \left\{ (\tilde{\boldsymbol{\alpha}}_i - \boldsymbol{\alpha}_i) (\tilde{\boldsymbol{\alpha}}_i - \boldsymbol{\alpha}_i)^T \right\} \\ &= E \left\{ (\mathbf{K}_i^r)^{-1} \mathbf{e} \mathbf{e}^T (\mathbf{K}_i^r)^{-1} \right\} \\ &= \sigma^2 (\mathbf{K}_i^r (\mathbf{K}_i^r)^T)^{-1} \\ &= \sigma^2 \left((\mathbf{Z}_i^r (\mathbf{Z}_i^r)^T) (\mathbf{Z}_i^r (\mathbf{Z}_i^r)^T)^T \right)^{-1} \\ &= \sigma^2 \left((\mathbf{W}_i^r (\mathbf{W}_i^r)^T) (\mathbf{W}_i^r (\mathbf{W}_i^r)^T)^T \right)^{-1} \\ &= \sigma^2 (\mathbf{U}_i^r \mathbf{D}_i^r \mathbf{D}_i^r \mathbf{U}_i^r (\mathbf{U}_i^r)^T)^{-1} \\ &= \sigma^2 \sum_{j=1}^r \frac{1}{d_{ij}^2} \mathbf{u}_{ij} \mathbf{u}_{ij}^T, \end{aligned} \quad (27)$$

where $r < J$, $\mathbf{W}_i^r = \mathbf{U}_i^r \mathbf{D}_i^r \mathbf{V}_i^r T$ by the SVD, and d_{ij} is the j th eigenvalue corresponding to the j th eigenvector. Compared (26) with (27), the variance of the regression parameter vector in the KLRC is smaller than that in the LRC. Therefore, it can be expected that the KLRC can provide more reliable regression parameter to the regression model for classification.

5 Experimental results

For verifications, we examine the proposed algorithms on the facial images, which are down-sampled from the extended Yale B (EYB) [41], AR [42], and FERET [43] face databases. In the experiments, we evaluate the proposed method against low-resolution problems coupled with illumination variations. In this section, all experimental results report the top 1 recognition accuracy (%).

The experiments are designed to evaluate the effectiveness of the proposed method in coping with unseen lighting changes under the LR condition. In the experiments, we compare the proposed methods, KLRC-p and KLRC-g, with the PCA+Euclidean, PCA+Mahalanobis [44], KPCA-p, KPCA-g, LDA+Euclidean, LDA+Mahalanobis, KLDA-p, KLDA-g, LRC, RLRC, SRC, LPP, NPE, improved principal component regression (IPRC) [45], unitary regression classification (URC) [46], linear discriminant regression classification (LDRC) [47], and local binary pattern (LBP) [48], where p and g denote the polynomial kernel and Gaussian kernel, respectively. The PCA-based and LDA-based approaches utilize 85 % dimensionality for experiments. It should be noted that this paper assumes the corresponding HR face images for the LR face images are not available as stated in Section 1.1. Hence, the existing face

recognition systems with the SR method are not suitable for this problem.

5.1 Experiments on EYB

The EYB contains images of 38 subjects with 9 poses and 64 illuminations per pose. The frontal face images of all subjects with 64 different illuminations are used for evaluation. The EYB is divided into five subsets based on the angle of the light source directions. As a result, there are total 2432 images: 266 (7 images per person), 456 (12 images per person), 456 (12 images per person), 532 (14 images per person), and 722 (19 images per person) images in subsets 1 to 5, respectively. All images are cropped, lowpassed, and resized to low-resolution images in size of 8×8 pixels, as shown in Fig. 2. Subset 1 is conducted for training, and the remaining subsets (subsets 2 to 5) are used for testing.

First of all, we investigate the performances under different image resolutions. Figure 3 shows the average recognition rate over the remaining subsets. Results reveal that the proposed KLRC-p and KLRC-g outperform the other methods consistently. Low-resolution face images with 8×8 pixels degrade the performances significantly. Nonetheless, the proposed KLRC algorithms still could perform well under the very low-resolution condition. Also, the results show that the performances achieved by the PCA-based, LDA-based, and the other subspace projection methods are drastically reduced for resolution below 16×16 pixels. Moreover, it is interesting to point out that the KLRC-p and KLRC-q under 8×8 pixels outperform the PCA-based and LDA-based approaches under 32×32 pixels significantly and achieve comparable recognition rate by the LRC-based approaches under 32×32 pixels. LBP [48] is a local feature method which can effectively defense the illumination variations under 32×32 pixels. However, LBP cannot perform very well under very low-resolution situation

since the facial information is not enough for local feature extraction. This success should be attributed to perform linear regression in the higher dimensional feature space.

Then, we further focus on the low-resolution images. As shown in Table 1, the KLRC-p and KLRC-g outperform the LRC, RLRC, IPCR, URC, LDRC, PCA+Euclidean, PCA+Mahalanobis, kernel-based PCA, LDA+Euclidean, LDA+Mahalanobis, kernel-based LDA, SRC, LPP, NPE, and LBP for the low-resolution face recognition under illumination variations. The PCA-based methods are the worst ones. Note that although it is widely accepted that the discriminant-based approaches offer higher robustness to lighting variations than the PCA-based approaches [3], the discriminant-based approaches still cannot withstand the low-resolution problem coupled with illumination variations in our work. It is because the low-resolution image contains insufficient high-frequency components containing the discriminative information for discriminant analysis. In [45], Huang et al. had shown that the regression-based method can perform better than the discriminant-based approach for face recognition under illumination variations. The proposed KLRC-p and KLRC-q have gained improvement significantly. In addition, although the RLRC performs better than the LRC, the RLRC could not obtain satisfactory performance because of performing regression in the original linear space.

5.2 Experiments on AR

For further verifications, we conducted experiments on AR face database. The AR database, built by Martinez and Benavente, totally contains 3510 mug shots of 135 subjects (76 males and 59 females) with different facial expressions, lighting changes, and partially occlusions. Each subject contains 26 images in two sessions. The first session, containing 13 images, includes the neutral expression, smile, anger, screaming, different lighting

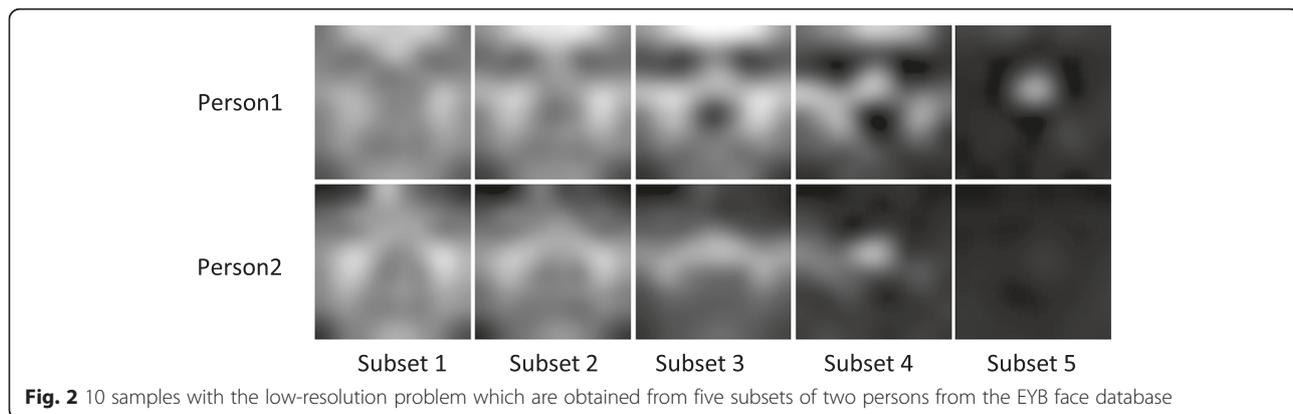


Fig. 2 10 samples with the low-resolution problem which are obtained from five subsets of two persons from the EYB face database

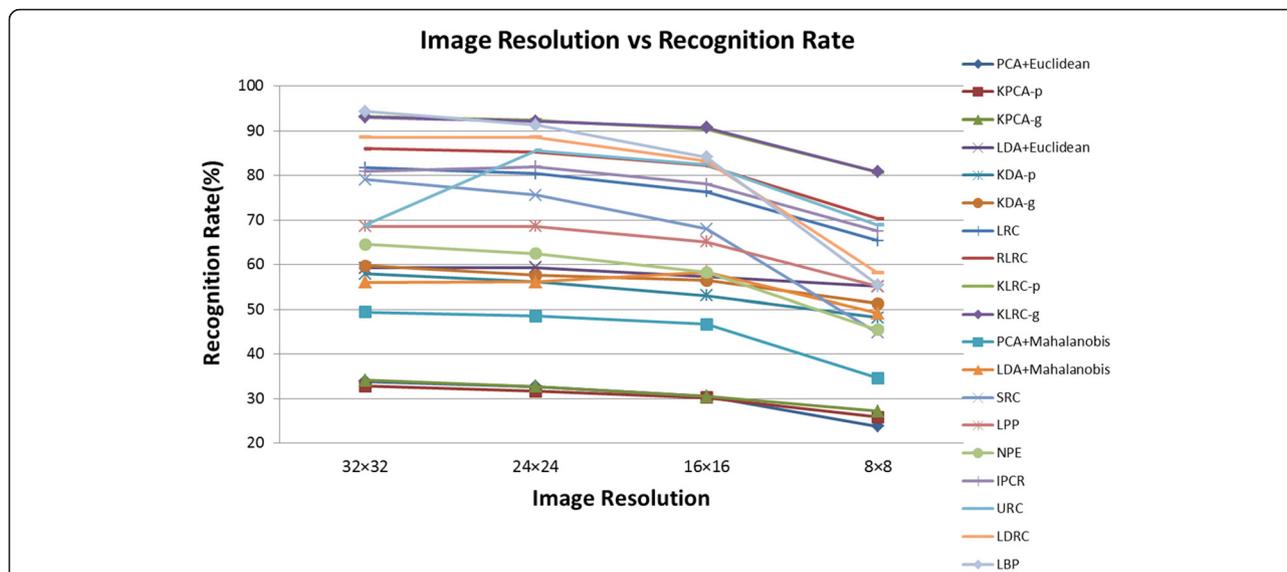


Fig. 3 The recognition rate under different image resolutions on the EYB face database

changes, and two realistic partial occlusions with lighting changes. The second session duplicates the first session in the same way 2 weeks later.

To evaluate the effectiveness of the proposed approach in coping with variable illuminations, only the face

images with illumination variations were considered in the experiments. All color face images in the AR are converted to gray levels, cropped, and down-sampled to the size of 8×6 pixels. Note that no face alignment is done in the copped face images. As shown in Fig. 4, 120 subjects with 8 face images under illumination variations, including no lighting, left lighting, right lighting, and full lighting, are chosen for evaluation. Training is conducted on the images with no lighting, and the remaining lighting conditions (left lighting, right lighting, and full lighting) are used for testing.

Table 1 Performance (%) comparisons of different methods on EYB in size of 8×8 pixels

Methods\testing faces	Subset 2	Subset 3	Subset 4	Subset 5	Average
PCA+Euclidean	70.18	19.52	2.63	2.63	23.74
PCA+Mahalanobis	66.67	34.17	19.29	18.42	34.64
KPCA-p	75.57	22.85	2.13	2.77	25.83
KPCA-g	76.67	25.48	3.57	2.91	27.16
LDA+Euclidean	100.00	89.91	26.50	4.02	55.11
LDA+Mahalanobis	88.33	72.50	27.86	7.89	49.15
KLDA-p	98.68	64.69	24.66	4.57	48.15
KLDA-g	99.56	74.34	26.54	4.71	51.29
LRC	100.00	98.03	51.69	11.77	65.37
RLRC	100.00	98.68	60.34	22.16	70.30
SRC	95.83	41.67	17.86	23.68	44.76
LPP	100.00	80.00	30.71	13.68	50.10
NPE	97.50	49.17	19.29	15.79	45.44
IPCR	100.00	96.67	56.43	16.84	67.49
URC	100.00	97.50	57.14	21.05	68.92
LDRC	100.00	75.83	39.29	17.37	58.12
LBP	96.67	70.00	31.43	23.68	55.40
KLRC-p	<i>100.00</i>	<i>98.68</i>	<i>77.82</i>	<i>46.26</i>	<i>80.69</i>
KLRC-g	<i>100.00</i>	<i>98.68</i>	<i>79.51</i>	<i>45.01</i>	<i>80.80</i>

Italicized data means the proposed methods' results

The experimental results are tabulated in Table 2, which reflects that the proposed KLRC-p and KLRC-g can attain higher recognition rate than the LRC, RLRC, IPCR, URC, LDRC, PCA+Euclidean, PCA+Mahalanobis, kernel-based PCA, LDA+Euclidean, LDA+Mahalanobis, kernel-based LDA, NPP, NPE, and LBP for low-resolution face recognition under illumination variations.

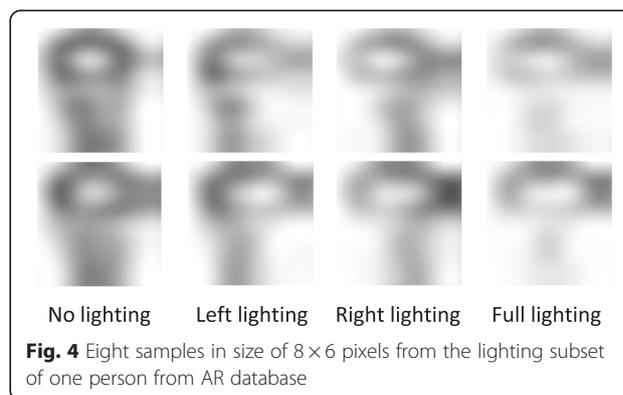


Fig. 4 Eight samples in size of 8×6 pixels from the lighting subset of one person from AR database

Table 2 Performance (%) comparisons of different methods on AR in size of 8×6 pixels

Methods\training data	Left lighting	Right lighting	Full lighting	Average
PCA+Euclidean	41.67	34.17	8.75	28.20
PCA+Mahalanobis	40.00	32.08	22.50	31.53
KPCA-p	41.25	31.25	12.50	28.33
KPCA-g	42.42	34.58	16.67	31.22
LDA+Euclidean	74.42	44.00	50.25	56.22
LDA+Mahalanobis	52.92	34.17	54.17	46.09
KLDA-p	75.25	47.33	41.50	54.69
KLDA-g	73.50	51.75	47.17	57.47
LRC	75.33	54.08	42.00	57.14
RLRC	74.42	57.75	41.90	58.03
SRC	80.00	72.50	34.17	62.22
LPP	60.83	42.92	30.83	44.86
NPE	82.08	52.50	30.42	55.00
IPCR	41.25	31.25	8.33	26.94
URC	43.75	32.92	17.50	31.39
LDRC	75.00	48.75	55.42	59.72
LBP	61.25	55.42	46.67	54.45
KLRC-p	75.42	57.92	51.25	61.53
KLRC-g	73.75	59.17	53.75	62.22

Italicized data means the proposed methods' results

From the experimental results, we can observe that SRC performs well as the proposed KLRC, especially under low ill-posed situation. However, the execution time in SRC-based approaches generally is more than that in LRC-based approaches [49]. The discriminant-based approaches, LRC, RLRC, IPCR, URC, and LDRC, could not work well for low-resolution face recognition under illumination variations.

5.3 Experiments on FERET

We further conduct experiments on the FERET face database to evaluate the performance on the illumination and expression variations since the different facial expression is inevitable to face recognition.

The FERET includes 250 people with four frontal view images from each subject. These 1000 face images with illumination and expression variations are resized to 8×6 pixels. Two images of each person are randomly selected for training, and the other two images are for testing. The experimental results are tabulated in Table 3, which has reported that the KLRC-p and KLRC-g can achieve higher recognition rate than the LRC, RLRC, IPCR, URC, LDRC, PCA+Euclidean, PCA+Mahalanobis, kernel-based PCA, LDA+Euclidean, LDA+Mahalanobis, kernel-based LDA, SRC, NPP, NPE, and LBP. As the results shown, we can observe that the KLRC-p and KLRC-g can work well for low-resolution problems with illumination and expression variations. It is reasonable because the LR face image will lose facial expression information [25]. On the other hand, in FERET, the lighting variations are slight, so the improvement is limited.

6 Conclusions

In this paper, the statistical analyses and experiment results verified that the proposed class-specific kernel linear regression classification performs the best for the low-resolution face recognition under illumination variations. With the kernel trick, the non-linear and increased-dimension mapping function enhances the modeling capability for low-resolution and illumination variations. Furthermore, the constrained low-rank approximation has been proposed to perform low-rank approximation automatically to make the kernel projection feasible for classification. The comparisons with the state-of-the-art methods indicate a comparable performance for the proposed KLRC-p and KLRC-g. We have demonstrated that the proposed KLRC-p and KLRC-g perform better than the PCA+Euclidean, PCA+Mahalanobis, KPCA-p, KPCA-g, LDA+Euclidean, LDA+Mahalanobis, KLDA-p, KLDA-g, LRC, RLRC, SRC, LPP, NPE, IPCR, URC, LDRC, and LBP for low-resolution face recognition under variable lighting. In summary, the KLRC-p and KLRC-g

Table 3 Performance (%) comparisons of different methods on FERET in size of 8×6 pixels

Methods	PCA+Euclidean	PCA+Mahalanobis	KPCA-p	KPCA-g	LDA+Euclidean
Recognition rate	70.0	36.2	71.2	71.4	72.0
Methods	LDA+Mahalanobis	KLDA-p	KLDA-g	LRC	RLRC
Recognition rate	58.3	72.4	72.4	74.0	74.2
Methods	SRC	LPP	NPE	IPCR	URC
Recognition rate	73.8	61.2	28.0	63.5	71.6
Methods	LDRC	LBP	KLRC-p	KLRC-g	
Recognition rate	76.0	53.8	76.8	76.4	

Italicized data means the proposed methods' results

dramatically improve the LRC to possess good robustness for very low-resolution face recognition under severe illumination variations.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

This work was supported in part by National Science Council of Taiwan, under Grant NSC 103-2221-E-006-109 -MY3.

Received: 14 July 2015 Accepted: 19 February 2016

Published online: 27 February 2016

References

- W Zhao, R Chellappa, A Rosenfeld, PJ Phillips, Face recognition: a literature survey. *ACM Comput Surv* **35**(4), 399–458 (2003)
- M. Turk and A. Pentland. Eigenfaces for recognition, *Journal of Cognitive Science* **3**(1), pp.71-86, 1991.
- P Belhumeur, J Hespanha, D Kriegman, Eigenfaces vs. fisherfaces: recognition using class specific linear projection. *IEEE Trans on Pattern Analysis and Machine Intelligence* **19**, 7 (1997)
- AM Martinez, AC Kak, PCA versus LDA. *IEEE Trans on Pattern Analysis and Machine Intelligence* **23**, 228–233 (2001)
- X Jiang, Linear subspace learning-based dimensionality reduction. *IEEE Signal Process Mag* **28**(2), 16–26 (2011)
- X Jiang, Asymmetric principal component and discriminant analyses for pattern classification. *IEEE Trans on Pattern Analysis and Machine Intelligence* **31**(5), 931–937 (2009)
- PC Yuen, JH Lai, Face representation using independent component analysis. *Pattern Recogn* **35**(6), 1247–1257 (2002)
- MS Bartlett, JR Movellan, TJ Sejnowski, Face recognition by independent component analysis. *IEEE Trans on Neural Networks* **13**(6), 1450–1464 (2002)
- J. Yang, D. Zhang, and J. Yang, Constructing PCA baseline algorithms to reevaluate ICA-based face-recognition performance, *IEEE Trans. on Systems, Man, and Cybernetics* **37**(4), Part B, pp.1015-1021, 2007.
- B Moghaddam, Principal manifolds and probabilistic subspaces for visual recognition. *IEEE Trans on Pattern Analysis and Machine Intelligence* **24**(6), 780–788 (2002)
- B Scholkopf, A Smola, K-R Muller, Nonlinear component analysis as a kernel eigenvalue problem. *Neural Comput* **10**, 1299–1319 (1998)
- M. H. Yang, Kernel eignefaces vs. kernel fisherfaces: face recognition using kernel methods, *Proc. Fifth IEEE Int'l Conf. Automatic Face and Gesture Recognition*, pp. 215-220, May 2002.
- G Baudat, F Anouar, Generalized discriminant analysis using a kernel approach. *Neural Comput* **12**, 2385–2404 (2000)
- S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K. R. Muller, Fisher discriminant analysis with kernels, *Neural Networks for Signal Proc. IX* **1**(1), pp. 41–48, 1999.
- X He, S Yan, Y Hu, P Niyogi, HJ Zhang, Face recognition using Laplacianfaces. *IEEE Trans on Pattern Analysis and Machine Intelligence* **27**, 3 (2005)
- X He, D Cai, S Yan, H Zhang, Neighborhood preserving embedding. *Proc ICCV'05* **2**, 1208–1213 (2005)
- J Wright, AY Yang, A Ganesh, SS Sastry, Y Ma, Robust face recognition via sparse representation. *IEEE Trans on Pattern Analysis and Machine Intelligence* **31**(2), 210–227 (2009)
- X Jiang, J Lai, Sparse and dense hybrid representation via dictionary decomposition for face recognition. *IEEE Trans on Pattern Analysis and Machine Intelligence* **37**(5), 1067–1079 (2015)
- I Naseem, R Togneri, M Bennamoun, Linear regression for face recognition. *IEEE Trans on Pattern Analysis and Machine Intelligence* **32**(11), 2106–2112 (2010)
- R Barsi, D Jacobs, Lambertian reflection and linear subspaces. *IEEE Trans on Pattern Analysis and Machine Intelligence* **25**(3), 218–233 (2003)
- I Naseem, R Togneri, M Bennamoun, Robust regression for face recognition. *Pattern Recogn* **45**(1), 104–118 (2012)
- Z. Wang and Z. Miao, Scale-robust feature extraction for face recognition, in 17th European Signal Processing Conference (EUSIPCO), pp. 1082-1086, 2009.
- B. J. Boom, G. M. Beumer, L. J. Spreeuwiers, and R. N. J. Veldhuis, The effect of image resolution on the performance of a face recognition system, in *Proc. IEEE Int. Conf. CARV*, pp. 1–6, 2006.
- A. Hadid and M. Pietikainen, From still image to video-based face recognition: an experimental analysis, in *Proc. IEEE Int. Conf. AFGR*, pp. 813–818, 2004.
- L. Tian, Evaluation of face resolution for expression analysis, in *Proc. IEEE Int. Conf. CVPR*, 82–82, 2004.
- B Gunturk, A Batur, Y Altunbasak, MH Hayes III, R Mersereau, Eigenface-domain super-resolution for face recognition. *IEEE Trans Image Process* **12**(5), 597–606 (2003)
- SW Lee, J Park, SW Lee, Low-resolution face recognition based on support vector data description. *Pattern Recogn* **39**(9), 1809–1812 (2006)
- P. H. Hennings-Yeomans, S. Baker, and B. Vijaya Kumar, Simultaneous super-resolution and feature extraction for recognition of low-resolution faces, in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, 2008.
- W.W.W. Zou and P.C. Yuen, Very low-resolution face recognition problem, *IEEE Trans. on Image Processing* **21**(1), pp. 327–340, 2011.
- H Huang, H He, Super-resolution method for face recognition using nonlinear mappings on coherent features. *IEEE Trans on Neural Networks* **22**(1), 121–130 (2011)
- S. Shan, W. Gao, B. Cao, and D. Zhao, Illumination normalization for robust face recognition against varying lighting conditions, in *Proc. IEEE Workshop on AMFG*, pp. 157–164, 2003.
- W Chen, JE Meng, S Wu, Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain. *IEEE Trans on Systems, Man, and Cybernetics-Part B: Cybernetics* **36**, 2 (2006)
- TP Zhang, YY Tang, B Fang, ZW Shang, XY Liu, Face recognition under varying illumination using gradientfaces. *IEEE Trans Image Process* **18**(11), 2599–2606 (2009)
- X Tan, B Triggs, Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE Trans Image Process* **19**(6), 1635–1650 (2010)
- B Wang, W Li, W Yang, Q Liao, Illumination normalization based on Weber's law with application to face recognition. *IEEE Signal Processing Letters* **18**, 462–465 (2011)
- J Ye, Generalized low rank approximations of matrices. *Mach Learn* **61**(1–3), 167–191 (2005)
- K. L. Clarkson and D. P. Woodruff, Low rank approximation and regression in input sparsity time, *Proceedings of the forty-fifth annual ACM symposium on Theory of computing*, pp. 81–90, 2013.
- D. Woodruff, Low rank approximation lower bounds in row-update streams, *Advances in Neural Information Processing Systems*, pp. 1781–1789, 2014.
- DC Hoaglin, RE Welsch, The hat matrix in regression and ANOVA. *Am Stat* **32**(1), 17–22 (1978)
- KR Muller, S Mika, G Ratsch, K Tsuda, B Scholkopf, An introduction to kernel-based learning algorithms. *IEEE Trans on Neural Networks* **12**(2), 181–201 (2001)
- AS Georghiadis, PN Belhumeur, DW Jacobs, From few to many: illumination cone models for face recognition under variable lighting and pose. *IEEE Trans on Pattern Analysis and Machine Intelligence* **23**(6), 630–660 (2001)
- A. Martinez and R. Benavente, The AR face database, *CVC Technical Report* 24, 1998.
- PJ Phillips, H Moon, SA Rizvi, PJ Rauss, The FERET valuation methodology for face-recognition algorithms. *IEEE Trans on Pattern Analysis and Machine Intelligence* **22**(10), 1090–1104 (2000)
- V Perlibakas, Distance measures for PCA-based face recognition. *Pattern Recogn Lett* **25**(6), 711–724 (2004)
- SM Huang, JF Yang, Improved principal component regression for face recognition under illumination variations. *IEEE Signal Processing Letters* **19**(4), 179–182 (2012)
- SM Huang, JF Yang, Linear discriminant regression classification for face recognition. *IEEE Signal Processing Letters* **20**(1), 91–94 (2013)
- SM Huang, JF Yang, Unitary regression classification with total minimum projection error for face recognition. *IEEE Signal Processing Letters* **20**(5), 443–446 (2013)
- T Ahonen, A Hadid, M Pietikainen, Face recognition with local binary patterns, in *Computer vision-ecv* (Springer, Berlin Heidelberg, 2004), pp. 469–481
- J Qian, J Yang, X Yong, General regression and representation model for classification. *PLoS One* **9**, 12 (2014)