Accepted Manuscript

Improved Principal Component Analysis and Linear regression classification for face recognition

Yani Zhu , Chaoyang Zhu , Xiaoxin Li

PII: S0165-1684(17)30413-9
DOI: 10.1016/j.sigpro.2017.11.018
Reference: SIGPRO 6669

To appear in: *Signal Processing*

Received date: 15 April 2017
Revised date: 24 November 2017
Accepted date: 25 November 2017

Please cite this article as: Yani Zhu , Chaoyang Zhu , Xiaoxin Li , Improved Principal Component Analysis and Linear regression classification for face recognition, *Signal Processing* (2017), doi: 10.1016/j.sigpro.2017.11.018

Highlights

- IPCA is used to extract the useful information from the original face images through reducing the dimension of feature vectors to better recognize images.
- Linear regression classification (LRC) is employed to deal with the problem of face recognition as a linear regression problem.
- LRC uses the least-square method to decide the class with the minimum reconstruction error.

Improved Principal Component Analysis and Linear regression classification for face recognition

Yani Zhu[a,b]*, Chaoyang Zhu[b], Xiaoxin Li[a]

[a] Zhejiang University of Technology , Hangzhou, 310014, China

[b] Hangzhou Dianzi University, Hangzhou 310018,China

*Corresponding author.

E-mail address: 10764341@qq.com

Abstract: In this paper, an improved principal component analysis (IPCA) is presented for face feature representation. IPCA is mainly designed to extract the useful information from original face images through reducing the dimension of feature vectors. Linear regression classification (LRC) algorithm is employed to treat the face recognition as a linear regression issue. LRC uses the least-square method to decide the class label with the minimum reconstruction error. Experiments are conducted on the Yale B, CMU_PIE and JAFFE databases. The proposed IPCA algorithm and LRC algorithm achieve better recognition results than that of state-of-the-art algorithms.

Keywords: Face Recognition, IPCA, LRC, Classification

1. Introduction

Face recognition problem has received extensive attention recently. As one of the most important applications of image understanding, it has been widely applied in biometric systems, security control, human identification, human-computer communication and other fields. With the latest development of multimedia signal transmission and processing, face recognition is widely used in video surveillance system, which is one of the most recent applications.

Typically, in pattern recognition problems, the high dimensionality of the data vector is considered as a potential source of redundancy measurement. Object manifold learning therefore opens up the high dimension of this “underlying source” through the appropriate transformation to the low dimensional data vector measurement [1]. Therefore, in the feature extraction stage, the image is transformed into the low dimensional vector in the face space. The main objective is to find a basis function for this transformation, which can be represented in the face space. In encoding theory, iterative measurement is more easily carried out in the presence of noise and security recovery information [2]. Therefore, maintaining robustness in low dimensional feature space is a hot topic in object recognition. Many methods of dimensionality reduction have been reported in a large number of documents. In terms of robustness, these methods are broadly divided into two major categories, namely reconstruction and discriminative approaches [3]. Reconstruction methods (such as principal components analysis (PCA) [4], independent component analysis (ICA) [5] and non-negative matrix factorization (NMF) [6-7]) mainly utilize the redundancy of visual data to produce sufficient reconstruction performance characterization. The purpose of PCA is to find a set of projection vectors that can maximize the variance of data. Because PCA is an unsupervised method, it may not be a suitable feature extraction algorithm. Maximizing the variance of data can not guarantee the separability of different classes. Therefore, the discriminant method (such as linear discriminant analysis (LDA) [8]) is sensitive to outliers and usually yields better results.In contrast, LDA is a supervised algorithm that attempts to find a set of projection vectors for maximizing the ratio of the between-class scatter to the within-class scatter. LDA projection enables data from a class as far as possible from different categories of data at the same time.On the other hand, due to the flexible decision boundaries, the optimal decision boundaries are usually determined directly from the data. In addition to these traditional approaches, it has recently been shown that the non orthodox features of the images and the random projections of the sampled images can provide the same good service as well. In fact, the dimension of the feature space and the design of the classifier may be more important than the choice of feature space. In recent years, the concept of sparse representation has been gradually applied to face recognition, and good recognition results have been achieved [9-11].

In literature study, the problem of noise pollution is investigated using a new robust linear Huber estimator and the class labels are determined based on the most accurate estimate of the subspace [12,13]. Although the usual polluted / missing pixels compensation can be resolved in the learning and classification stage, robust regression classification is rarely mentioned in the current literature. The regression classification algorithm mainly concentrates on the strong learning method, which leads to the development of the high breakdown point through the detection of the potential independent learning data in learning stage [14]. Most of these regression learning methods are adapted to replace the classical position discrimination model and robust counter scattering matrix based on the estimation, such as the minimum volume ellipsoid (MVE) and the minimum covariance determinant (MCD) estimates and projection pursuit methods [15], etc. And some estimation methods no longer use standard least squares error. An important implementation method is introduced in the target recognition problem, which improves the learning process of the traditional subspace method by introducing a hypothesis and test paradigm. The basic sampling method consists of a set of linear equations, each of which has a powerful solution consisting of subsets of image points within projection space. Through the principle of minimum description length (MDL) and the assumption of mutual competition, the coefficients of the feature space are measured. In previous studies, a complex method based on discrimination and reconstruction model [13] has been proposed to construct the subspace of above two methods which have the power of discrimination and the reconstruction. This method is applied in the generation of a probabilistic imaging model of the liver. However, randomized pixel loss damage problem is firstly described as a linear robust regression task. Recently, regression analysis based on classification methods is popular for robust face recognition. Naseem et al. presented a linear regression classifier (LRC) for face classification [9]. Because a single object class model exists in a linear combination of subspace, a linear model is developed to represent the linear combination of images. The inverse problem is solved by least square method, and the decision rule is beneficial to the minimum reconstruction error class. The LRC algorithm can be categorized as the nearest subspace (NS) problem. Compared with the existing structured sparse error coding models, which perform error detection and error support separately, Qian. et al. [16] proposed a novel method which integrated error detection and error support into one regression model. In [17], an adaptive linear discriminant regression classification (ALDRC) algorithm was proposed. ALDRC used different weights to describe the different contributions of training samples. The algorithm used weighted information to compute intra-class and within-class reconstruction errors, and then attempted to find an optimal projection matrix to maximize the ratio of intra-class reconstruction error over the within-class reconstruction error. A conclusion can be drawn from [18-23], existing classification methods can be grouped into four categories: Bayesian estimates approaches, subspace learning approaches, line regression approached, and sparse representation approaches.

In this paper, an improved PCA algorithm is proposed to extract features of the face images. IPCA can not only reduce the dimensionality of the image, but also maintain image features as well. Next, SVM and LRC methods are chosen for face classification and identification. Experimental results on three face databases show that the two classification methods have different advantages.

2. Improved Principal Component Analysis

PCA is an effective feature extraction technique and data representation method. It received an immense amount of attention in pattern recognition, image processing and computer vision. PCA produces linear combinations of the original data and aims to find the best vector space which represents the distribution of face images and reduces a large of data set to a lower dimension for getting effective results. The feature space defined by eigenvector greatly reduces the dimension of the original space, which reduces the computation time of face detection and recognition [24]. The main goal in the PCA algorithm lies in reducing the large dimensions of face data to the dimensions of the smallest spaces. PCA is considered as a multivariate analysis method based on eigenvector. PCA algorithm can be implemented by two main methods. The first one is achieved by decomposition of the eigenvalue of data covariance matrix [25]; on the other hand, the second one is performed by decomposition of a singular value of the data matrix. PCA results are expressed as a component or factor scores and standardized component score weight. Therefore, the resulted image can be expressed as the value of each eigenfaces it was related to. After reducing the dimensionality of the data set, the resulting images are called the eigenfaces (or eigenvectors). Through PCA eigenfaces method, each pixel is considered as an image as a separate dimension [26-27] .

Compared with the traditional PCA, the improved algorithm uses the mean of each class instead of the specific image within the class. Since the average of each class is a linear combination of within-class images, the average of each class retains a large number of variations of the specific image [28]. In other words, the compression process of each image is more conducive to image recognition. In addition, another obvious advantage of the improved PCA is that the training time is greatly reduced.

The gray image of the k-th input face is represented by, where m denotes image pixels. Firstly, the average value of N images is computed by $\bar{x}=\frac{1}{N}\sum_{k=1}^{N}x_k$; Secondly, covariance matrix of vectors is computed：

$$W=\frac{1}{N}\sum_{k=1}^{N}(x_k-\bar{x})(x_k-\bar{x})^T \quad (1)$$

This covariance matrix is also called the overall dispersion matrix. It consists of two parts: intra-class of discrete matrices and between-class of scatter matrices. IPCA does not consider the intra-class scatter matrix, and only calculates the between-class scatter matrix [29]. There are N pieces of images in image collections, they are detached to c classes, N_i is the quantities of the i-th subjects, where x_i denotes the row vector of the i-th facial image. The average value in every class is：$\bar{x}_i=\frac{1}{N_i}\sum_{k=1}^{N_i}x_k$, the general average value is：$\bar{x}=\frac{1}{N}\sum_{k=1}^{N}x_k$, between-class dispersion matrix is:

$$S=\frac{1}{N_i}\sum_{i=1}^{c}(\bar{x}_i-\bar{x})(\bar{x}_i-\bar{x})^T \quad (2)$$

$$x_k=[x_{k1},x_{k2},...,x_{km}]^T$$

S is a symmetric matrix, it can be diagonalized as:

$$S=\sum_{r=1}^{N}\lambda_r V_r V_r^T=V\Lambda V^T \quad . \quad (3)$$

where λ_r is the eigenvalue of W, V_r is the corresponding eigenvector, $\{V_1,V_2,...,V_R\}$ is the standard orthogonal basis, R is the rank of W , Λ is diagonal matrix, eigenvalues w are arranged at the diagonal, the projection of x_k on V_r is denoted as:

$$P_k^T=V_r^T x \quad (4)$$

$$Var(P_k)=E[(V_r^T x_k-V_r^T x)\cdot(V_r^T x_k-V_r^T x)^T]$$

$$=V_r^T E[(x_k-x)\cdot(x_k-x)^T]V_r$$

$$= V_r^T W V_r = \lambda_r \quad . \tag{5}$$

The projection's variance of x_k on V_r is the eigenvalue λ_r. Therefore, the extraction of feature vector is divided into the following seven steps:

Step 1: Convert face images in Training set to face vectors;

Step 2: Calculate the average face vector and subtract average face vector from each face vector ;

Step 3: Calculate the covariance matrix S;

Step4: Decrease the dimensionality of the data set ;

Step5: Calculate eigenvectors from the covariance matrix;

Step 6: Select some pieces of best eigenfaces to represent the whole data set;

Step 7: Represent each face image in a linear combination of all eigenvectors.

3. Classification methods for Face Recognition

3.1 Support Vector Machines

Support Vector Machine (SVM) is a novel pattern classification approach based on structural risk minimization. It is suitable to deal with magnitude feature problems with a given finite amount of training data [30]. SVM is very efficient in solving clustering problems that are not linearly separable. It has been successfully exploited to a number of applications ranging from handwritten digit recognition, face identification, text categorization, bioinformatics to database marketing.

The fundamental idea of SVM is to project (using a mapping function φ) the input vectors into a new feature space of greater dimension in which it is possible to find a hyperplane producing a linear separation between classes. For a two classes problem, assume there is a training data set $S : \{(x_i, y_i)\}_{i=1}^{W}$, where each input $x_i \in \Re^m$ and output $y_i \in \{-1,+1\}$. The goal of SVM is to map the input vector x into a feature space $z = \Phi(x)$ and find an optimal hyperplane $wz + b = 0$ in the feature space. The hyperplane can separate the training data into two classes with the maximum margin, where $w = \sum_{i=1}^{N} a_i y_i z_i$, a_i is a set of Lagrange multipliers to the following dual problem [31] :

$$\text{Maximize:} \quad W(a) = \sum_{i=1}^{N} a_i - \frac{1}{2}\sum_{i=j=1}^{N} a_i a_j y_i y_j (z_i . z_j)$$

Subject to: $C \geq a_i \geq 0$, $\sum_{i=1}^{N} a_i y_i = 0$, (6)

where C is a user-defined regularization hyperparameter which determines the trade-off between maximizing margin and minimizing the number of misclassified data examples. It is useful to handle non-separable problems and outliers. The kernel trick of SVM allows us to substitute the dot product of data points with just a kernel function:

$$K(x_i, x_j) = z_i . z_j \quad . \tag{7}$$

The decision function is made by computer:

$$f(x) = sign(w.z + b) = sign(\sum_{i=1}^{N} a_i y_i k(x, x_i) + b) \quad , \tag{8}$$

where $f(x)$ is the distance of a testing data x to the optimal hyperplane. If the distance is less than 0, then testing data belongs to the negative class. Otherwise, it is in the positive class. Several kernel functions have been used widely and successfully, such as, polynomial kernel with degree d :

$$K(x_i, x_j) == (1 + x_i x_j)^d \tag{9}$$

Gaussian RBF kernel with parameters σ :

$$K(x_i, x_j) = \exp(-\|x_i - x_j\|^2 / (2\sigma^2)) \ . \tag{10}$$

And sigmoid kernel with parameter θ :

$$K(x_i, x_j) = \tanh(x_i x_j - \theta) \ . \tag{11}$$

As discussed above, SVM is a two-classes classifier. For face recognition problem we need to extend it to a multi-class classifier [32]. The one-against-one strategy and the one-against-all strategy are the two most popular methods. For the one-against-one strategy, one support vector machine is trained to separate each pair of classes. So the method needs $c*(c-1)/2$ support vector machines trained, where c is the number of classes. During testing phase, each support vector machine votes for one class and the winning class is the one having the largest number of votes. For the one-against-all strategy, each support vector machine is trained to separate a single class from remaining classes. In other words, each class is associated to one hyperplane. So it needs c support vector machines trained. Each test vector is assigned to the class whose hyperplane is furthest from it. Since the one-against-all method is simpler and is as effective as the one-against-one method, we use it as our face classification method.

3.2 Linear Regression Classification

Consider a linear model:

$$y=X\beta+e\ , \tag{12}$$

where the dependent or response variable $y \in \mathbb{R}^{q\times 1}$, the regressor or predictor variable $X \in \mathbb{R}^{q\times 1}$, the vector of parameters $\beta \in \mathbb{R}^{q\times 1}$ and error term $e \in \mathbb{R}^{q\times 1}$. The problem of robust estimation is to estimate the vector of parameters so as to minimize the residual [33] :

$$r = y - \hat{y}; \hat{y} = X\hat{\beta}\ , \tag{13}$$

$\hat{y}$ is the predicted response variable. In classical statistics, the error term is usually treated as an average Gaussian noise. The most traditional method in the optimization of regression is to minimize the least square (LS) error:

$$\underbrace{\arg\min}_{\beta} \sum_{j=1}^{q} r_j^2(\hat{\beta})\ \ , \tag{14}$$

where $r_j(\beta)$ is the j-th component of the residual vector r. However, in the presence of outliers, the least squares estimator will perform poorly. Least squares method of this classic statistical methods, though very powerful, can not avoid the I type error. The I type error is actually the existence of the null hypothesis. But note that type I error rates are often lower than the nominal values due to the presence of outliers in the traditional method, which is the classical statistical conservatism. However, if the data is contaminated, the type II error will increase dramatically. The type II error is actually a false error when the null hypothesis exists. This problem is often referred to as the classical approach. In addition, the classical statistical methods are usually assumed to be well performed with the same variance data model. In many practical cases, however, this assumption is not true, to emphasize the need for robust estimation.

Methods such as R-estimators and L-estimators have been proposed and applied. However, due to its versatility and high decomposition point [34], M-estimator has gradually shown its superiority. This estimation method is mainly based on the minimization of the residual error of M-estimation function:

$$\hat{\beta} = \underbrace{\arg\min}_{\hat{\beta}\in\mathbb{R}^p} \left\{ F(\hat{\beta}) \equiv \sum_{j=1}^{q} \rho(r_j(\hat{\beta})) \right\}\ , \tag{15}$$

where $\rho(r)$ is a symmetric function with a unique minimum at zero:

$$\rho(r) = \begin{cases} \dfrac{1}{2\gamma} r^2 & for\, |r| \le \gamma \\ |r| - \dfrac{1}{2}\gamma & for\, |r| > \gamma \end{cases}\ , \tag{16}$$

γ being a tuning constant called the Huber threshold.

Consider N number of distinguished classes with p_i number of training images from the i-th class such that $i = 1, 2, \ldots, N$. Each grayscale training image is of an order $a \times b$ and is represented as $u_i^{(m)} \in \mathbb{R}^{a \times b}, i = 1, 2, \ldots, N$ and $m = 1, 2, \ldots, p_i$. Each gallery image is down sampled to an order c×d and transformed to a vector through column concatenation such that $u_i^{(m)} \in \mathbb{R}^{a \times b} \to W_i^{(m)} \in \mathbb{R}^{q \times 1}$, where q=cd, cd ≤ ab. Each image vector is normalized to a maximum pixel value of 1. Using the concept that patterns from the same class lie on a linear subspace [35], a class specific model Xi is employed by stacking the q-dimensional image vectors,

$$X_i = \left[\overset{(1)}{W_i} \overset{(2)}{W_i} \ldots\ldots \overset{p_i}{W_i} \right] \in \mathbb{R}^{q \times p_i}, i = 1, 2, \ldots, N \quad . \tag{17}$$

Each vector $\overset{(m)}{W_i}, m = 1, 2, \ldots, p_i$, spans a subspace of $\mathbb{R}^q$ also called the column space of Xi. Therefore at the training level each class i is represented by a vector subspace, Xi, which is also called the regressor or predictor for class i. Z is an unlabeled test image and we need to classify z as one of the classes $i = 1, 2, \ldots, N$. The grayscale image z is transformed and normalized to be an image vector $y \in \mathbb{R}^{q \times 1}$ as discussed for the gallery. If it is part of the i-th class, it can be represented by a linear combination of the training images from the i-th class i.e.

$$y = X_i \beta_i + e, i = 1, 2, \ldots, N \quad , \tag{18}$$

where $\beta_i \in \mathbb{R}^{pi \times 1}$. From the point of view of face recognition, the corresponding explanatory variables of the system training are usually carried out in a controlled and noise free environment. When the presence of faulty sensors and channel noise is contaminated, the problem of robustness will appear in a given test pattern. Given that q ≥ pi, the system of equations in Eq.17 is well-conditioned and βi is estimated using robust Huber estimation:

$$\hat{\beta}_i = \underbrace{\arg\min}_{\hat{\beta}_i \in \mathbb{R}^{p_i}} \left\{ F(\hat{\beta}_i) \equiv \sum_{j=1}^{q} \rho(r_j(\hat{\beta}_i)) \right\}, i = 1, 2, \ldots, N \quad , \tag{19}$$

where $r_j(\hat{\beta}_i)$ is the j-th component of the residual:

$$r_j\left(\hat{\beta}_i\right) = y - X_i \beta_i, i = 1, 2, \ldots, N \quad . \tag{20}$$

The estimated vector of parameters $\hat{\beta}_i$, along with Xi is used to predict the response vector:

$$. \quad \hat{y}_i = X_i \hat{\beta}_i, i = 1, 2, \ldots, N \tag{21}$$

The distance is calculated between the predicted response vector $\hat{y}_i, i = 1, 2, \ldots, N$ and the original response vector y,

$$d_i(y) = \|y - \hat{y}_i\|_2, i = 1, 2, \ldots, N \quad , \tag{22}$$

and a corresponding minimum distance calculated as:

$$\underbrace{\min}_{i} d_i(y), i = 1, 2, \ldots, N \quad (23)$$

4. Experiments

In this Section, we evaluate the performance of proposed method before based on three different databases: Yale B, CMU_PIE and JAFFE Database [36-39]. All the experiments are carried on a laptop equipped with an Intel Dual Core 2.6Ghz and 8GB RAM. All the algorithms are programmed in python language (version 3.6). These images are preprocessed before feature extraction: the original image is cropped, most of the facial features are retained. All images are normalized to size of 64 ×64, and illumination compensation and gray normalization are performed. All the experimental results are the average values of many experiments.

4.1 Yale B Database

4.1.1 Evaluation 1

Yale B database consists of 2,414 frontal face images of 38 subjects under various lighting conditions. The database was divided into five subsets according to the angle between the light source direction(see Fig.1): subset 1 consists of 266 images (7 images per subject) under nominal lighting conditions; subsets 2 and 3, consists of 12 images per subject with slight-to-moderate lighting variations; subset 4 (14 images per person) and subset 5 (19 images per person) with more adverse illumination variations [40].

All experiments on Yale B database were conducted with images down sampled to an order 64 x 64; training is conducted using subset1 and the testing is validated on the remaining subsets. PCA and IPCA are used to extract facial features respectively, and SVM is used as classifier. When the kernel function parameters of SVM are selected at different orders of magnitude, the recognition results vary considerably. Detailed experimental results are shown in the Fig. 2.

Fig. 1. Starting from the top, each row illustrates samples from subsets 1, 2, 3, 4, and 5, respectively.

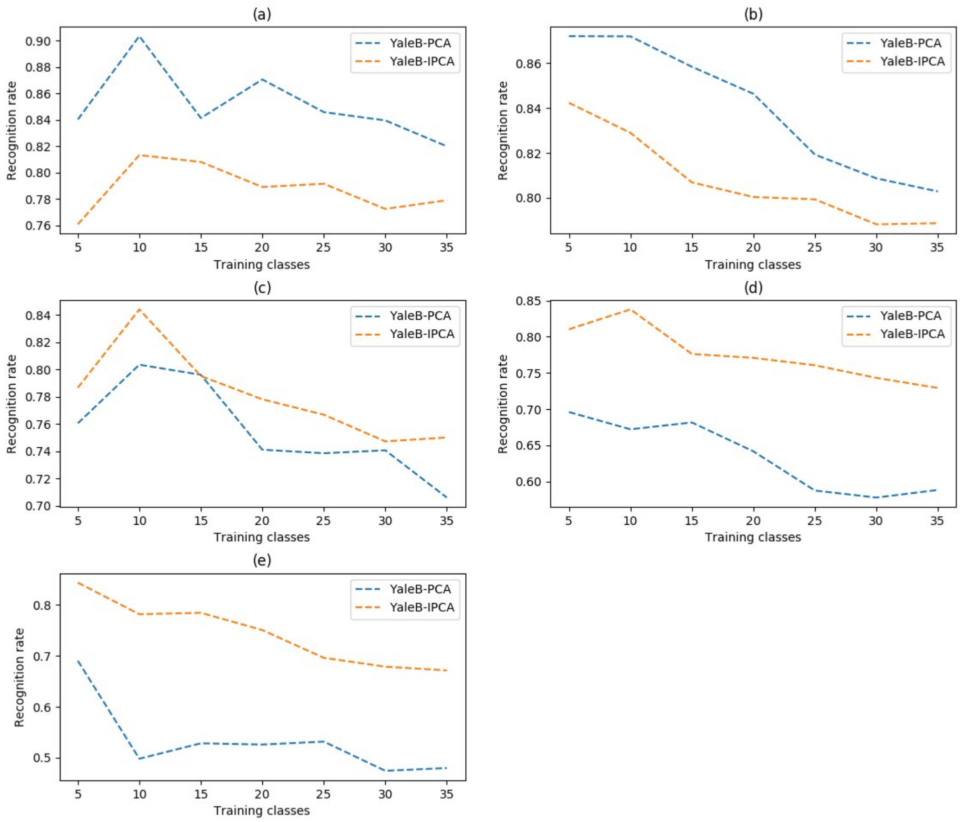

Fig. 2. Different face recognition results with different SVM Key function variances 0.01, 0.025, 0.05, 0.075 and 0.1, respectively.

For a comprehensive comparison of the recognition performance with different SVM kernel function variances, extensive verification experiments were conducted. Different face recognition results are showed in Fig. 2. Face recognition profiles for various degree of SVM kernel function variances on the scale [0.01, 0.1] show an excellent performance index for the proposed

IPCA approach. With the increasing SVM kernel function variances, the proposed approach shows good recognition rate as 82.71% compared to PCA. Specifically with 0.1 SVM kernel function variance, the proposed IPCA approach achieves an equal recognition rate of 82.71% comprehensively beating PCA recognition rate by 3.04%. Shown form experiment results, the face recognition rate will increase as the kernel parameter of SVM increases.

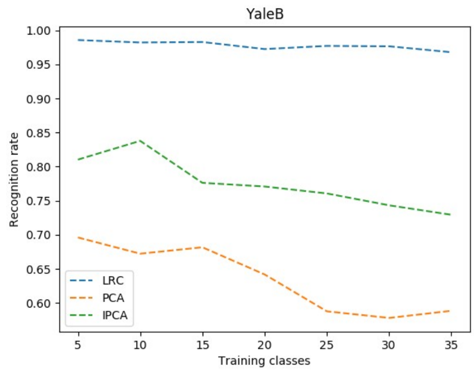

Fig. 3. Face recognition results with LRC

The experiment results of three recognition methods are shown in Fig.3. The diagram shows an excellent performance index for the proposed LRC. With the increasing training classes, LRC shows good average recognition rate as 98.29% compared to PCA and IPCA. With the increase of training classes, the face recognition rate of PCA and IPCA declined suddenly, however, the recognition rate of LRC changed little, and the recognition rate remained above 97.38%.

4.1.2 Evaluation Protocol 2

In evaluation Protocol 2, robust facial expression recognition experiments were performed on subsets1 of Yale B database. 152 facial expressions images of 38 people (4 images per person) were chose from subset 1. (as shown in Figure 4, left to right are happy, sad, surprised, normal) [41]. For different facial expressions, the recognition results are shown in table 1.

Fig. 4. Facial expression of Yale B database

Table 1 shows the facial expression recognition results on Yale B database. From the table we can see that, although the two different facial expressions, that is normal and sad, IPCA_SVM obtained an average recognition rate of 87.8%, which is not very high, but in four different expressions on the average recognition rate is 89.3% using our method, beat other methods PCA_NEAR [42] , FLD_NEAR [43] , PCA_SVM [44] by a large margin.

Table 1. Recognition results of different approaches for Yale B Database

Expression	Classifier			
	PCA_NEAR	*FLD_NEAR*	*PCA_SVM*	*IPCA_SVM*
happy	66.7%	73.3%	80.0%	85.6%
normal	73.3%	73.3%	80.0%	82.3%
sad	80.0%	86.7%	86.7%	93.3%
surprised	73.3%	80.0%	86.7%	95.8%
average	73.3%	78.3%	83.4%	89.3%

4.2 CMU_PIE Database

Extensive experiments were conducted on CMU-PIE database. We randomly selected a subset of database consisting of 65 subjects with 21 illumination variations per subject [45], all images are resized to 64*64. Fig. 5 represents 21 different alterations for a typical subject. We followed two experimental setups as proposed in the first set of experiments where the system is trained using images with near frontal lighting and validation is conducted across the whole database.

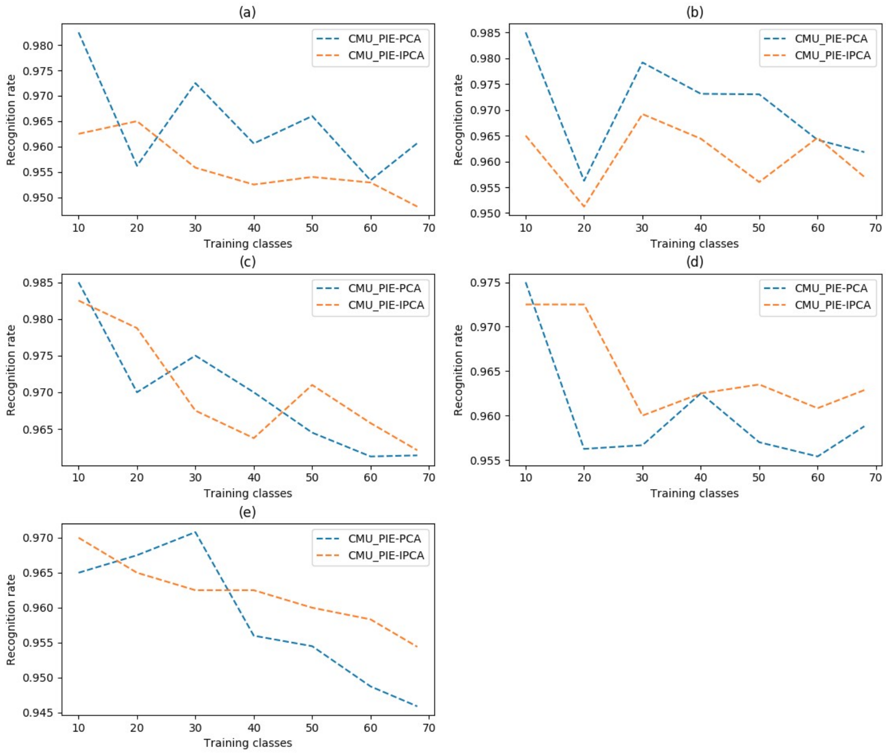

Fig. 5. Different face recognition results with different SVM Key function variances 0.01, 0.025, 0.05, 0.075 and 0.1, respectively.

For a comprehensive comparison of the recognition performance with different SVM kernel function variances, extensive verification experiments were conducted. Different face recognition results are showed in Fig. 5. Face recognition profiles for various degree of SVM kernel function variances on the scale [0.01, 0.1] show an excellent performance index for the proposed IPCA approach. With the increasing SVM kernel function variances, the proposed approach shows good recognition rate as 97.88% compared to PCA. Specifically with 0.05 SVM kernel function variance, the proposed IPCA approach achieves an equal recognition rate of 96.82% comprehensively beating PCA recognition rate by 96.7%. Shown form experiments results, the face recognition rate will increase as the kernel parameter of SVM increases.

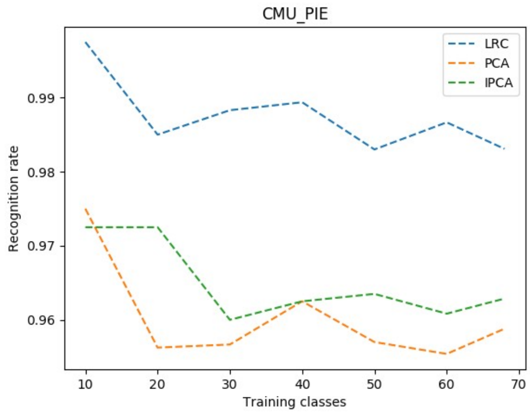

Fig. 6. face recognition results with LRC

The experiment results of three recognition methods are shown in Fig.6. The diagram shows an excellent performance index for the proposed LRC. With training classes increase, LRC shows a good average recognition rate as 98.9% compared to PCA and IPCA, while the face recognition rate of PCA and IPCA declined suddenly, however, the recognition rate of LRC changed little, and the recognition rate remained above 98.5%.

4.3 JAFFE Database

JAFFE Database are consisted of 165 pieces of 320×243 gray images. There are 7 kinds of Japanese women's facial expressions in the JAFFE database (as shown in Figure 7, due to the left to right are angry, disgusted, fearful, happy, normal, sad, surprised) [46]. We chose 120 images of all 10 database objects in the Japanese LAFFE facial expression database, which includes 7 kinds of facial expressions (angry, disgusted, terrified, happy, sad, surprised and normal). The database has been often used as a test database for the validation of face expression recognition algorithms. All results of experiments are shown in Table 2.

Fig. 7. Facial expression of JAFFE database

Table 2. Expression recognition results of different approaches for JAFFE Database

Expression	Classifier			
	PCA_NEAR	***FLD_NEAR***	***PCA_SVM***	***IPCA_SVM***
Angry	73.3%	80.0%	83.5%	88.7%
Disgusted	83.3%	83.3%	87.9%	89.6%
Terrified	76.7%	86.7%	85.7%	87.4%
Happy	84.6%	87.9%	88.9%	93.6%
Normal	74.6%	81.2%	85.6%	87.8%
Sad	78.1%	79.6%	80.2%	88.9%
Surprised	83.3%	86.7%	85,8%	98.6%
Average	79.1%	83.6%	85.4%	90.7%

Table 2 shows the recognition results on Japan JAFFE facial expression image, the average recognition rate of IPCA_SVM reached to 90.7%. In particular, for three kinds of facial expressions, such as anger, happiness and surprise, our method achieves an average recognition rate of 93.7%, far higher than other methods. This shows that IPCA_SVM expression recognition method is more effective than PCA_NEAR [42] , FLD_NEAR [43] and PCA_SVM [44], and obtain a satisfactory overall performance.

4.4 Evaluation of the algorithm efficiency

We compared IPCA_LRC with other methods in computational time on Yale B database and CMU_PIE Database. The results are given in Fig.8, which shows the average computational time of three algorithm on two testing database. We compute the recognition time based on PCA_SVM, IPCA_SVM and LRC.

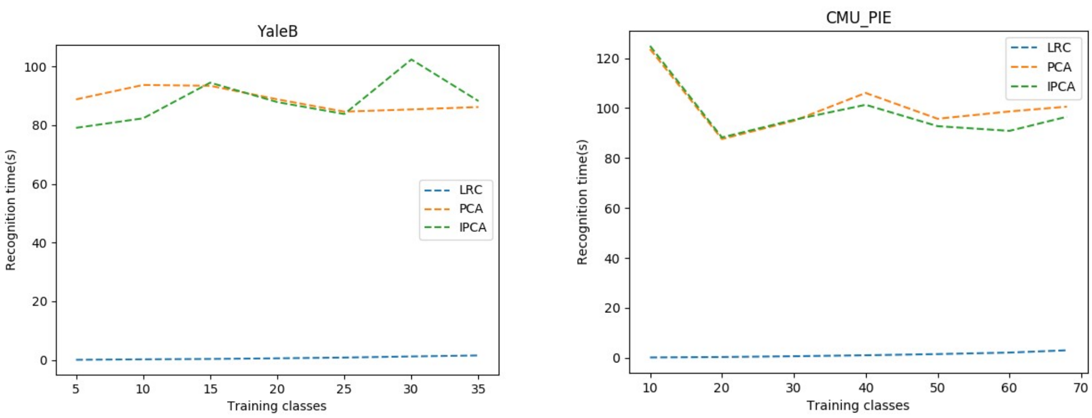

Fig.8. Recognition time of different approaches

The experiment results of three recognition time are shown in Fig.8. The diagram shows the recognition time used in the 3 experimental methods. As can be seen from the fig.8, with the increase of training category, the time of LRC algorithm is little changed, and the average time is the least in the 3 algorithms.

5. Conclusion

In this paper, a new robust face recognition algorithm based on robust estimation method is proposed. The comprehensive comparison with the state-of-art robust approaches indicates a comparable performance index for the proposed approach. Specifically, the challenges of varying facial expressions recognition are addressed. On the Yale B database and CMU_PIE database, the algorithms have achieved very good recognition effect. In addition, our report on the JAFFE face database also showed a good recognition performance. The proposed IPCA algorithm also reveals a number of interesting outcomes. Apart from the LRC approach for face recognition in the presence of noise, the LRC approach yields high recognition accuracy rate without image preprocessing. The LRC method comprehensively outperformed the benchmark method by different face patterns and the recognition rate reached a surprising 98.5%. Classifier LRC method is also very effective in the identification of face expressions. In this research, IPCA is used to extract facial expressions, SVM and LRC is used to realize pattern classification. The experimental results on Yale, CMU_PIE and JAFFE databases show that IPCA and LRC recognition algorithms are reliable and effective.

However, the LRC algorithm has not yet been shown to be effective in the case of severe lighting and random pixel corruption. In the future we will consider the construction of a hierarchical face recognition system without constraints, and try to solve the serious pollution problem in face recognition, such as noise interference and continuous occlusion existed in the practical situation.

Acknowledgment

This research was financially supported by the National Science Foundation of China(61402411).

References

1. V. N.Vapnik, The Nature of Statistical Learning Theory. New York: Springer-Verlag , 2000.
2. M.P. Roath, M. Winter, Survey of appearance-based methods for object recognition, Technical Report, Inst. For Computer Graphics and Vision, Graz University, Austria, 2008.
3. M. Turk, A. Pentland, Eigenfaces for recognition, Journal of Cognitive Neuroscience 3(1) (1991) pp. 71-86.

4. P. C. Yuen, J. H. Lai, Face representation using independent component analysis, Pattern Recognition 35(6) (2002) pp. 1247-1257.
5. D. D. Lee, H. S. Seung, Learning the parts of objects by non-negative matrix factorization, Nature 401(1999) pp. 788-791.
6. D. D. Lee, H. S. Seung, Algorithms for non-negative matrix factorization, Advances in Neural Information Processing Systems (2001) pp. 556-562.
7. V. Belhumeur et al., Eigenfaces vs fisherfaces: recognition using class specific linear projection, IEEE Transactions on PAMI 17 (7) (1997) pp. 711-720.
8. I. Naseem et al., Linear regression for face recognition, IEEE Transactions on PAMI 32(11) (2010) pp. 2106-2112.
9. I. Naseem, R. Togneri, M. Bennamoun. Face Identification using Linear Regression, in:IEEE ICIP, 2009.
10. J. Wright et al., Robust face recognition via sparse representation, IEEE Transactions on PAMI 31 (2) (2009) 210-227.
11. P. J. Huber, Robust Statistics, John Wiley, New York, 1981.
12. H. B. Nielsen, Computing a minimizer of a piecewise quadratic-implementation, Technical Report, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, 1998.
13. S. Fidler et al., Combining reconstructive and discriminative subspace methods for robust classification and regression by subsampling, IEEE Transactions on PAMI 28 (3) (2006) pp. 337-350.
14. C. V. Chorl, P. J. Rousseeuw, Integrating a high-breakdown option into discriminant analysis in exploration geochemistry, Journal of Geochemical Exploration 43 (1992) pp. 191-203.
15. M. Hubert, K. V. Driessen, Fast and robust discriminant analysis, Computational Statistics and Data Analysis 45 (2003) pp. 301-320.
16. J. J. Qian et al., Robust nuclear norm regularized regression for face recognition with occlusion, Pattern Recognition 48 (2015) pp. 3145-3159.
17. P. Huang et al., Adaptive linear discriminant regression classification for face recognition, Digital Signel Processing 55 (2016) pp. 78-84.
18. X. Jiang et al., "Eigenfeature Regularization and Extraction in Face Recognition," IEEE Trans. Pattern Analysis and Machine Intelligence, vol.30, no.3, pp. 383-394, Mar. 2008.
19. J. Yu et al., Multi-Task Deep Auto-encoder Model for Human Pose Recovery, IEEE TIE, DOI: 10.1109/TIE.2017.2739691, 2017.
20. J. Yu* et al., Deep Multi-modal Distance Metric Learning Using Click Constraints for Image Ranking, IEEE Transactions on Cybernetics, 2016, DOI: 10.1109/TCYB.2016.2591583.

21. D. C. Tao et al., Geometric Mean for Subspace Selection. IEEE Trans. Pattern Anal. Mach. Intelligence 31(2): 260-274 (2009).
22. W. Liu et al.， Multiview Hessian regularized logistic regression for action recognition， Signal Processing 110, pp.101-107.
23. W. Liu et al.， P-Laplacian Regularized Sparse Coding for Human Activity Recognition， IEEE Transactions on Industrial Electronics 63 (8), 5120-5129.
24. A. Alkandari, S. J. Aljaber, Principle Component Analysis Algorithm (PCA) for Image Recognition, The Second International Conference on Computing Technology and Information Management (ICCTIM), 2015, pp. 76-80.
25. A. A. Alorf, Performance Evaluation of the PCA Versus Improved PCA (IPCA) in Image Compression, and in Face Detection and Recognition, Future Technologies Conference 2016, pp. 537-546.
26. V. Franc, V. Hlavac, Multi-class support vector machine, The 16th International Conference on Pattern Recognition, IEEE Computer Society, vol. 2, 2002, pp. 236-239.
27. M. Turk, Pentland A (1991) Eigenfaces for recognition. J Cognitive Neurosci 13(1):71 – 86.
28. W. Zhao et al., Analyzing PCA-based Face Recognition Algorithms: Eigenvector, ACM Computing Surveys,Technical Report, Vol. 35, pp. 399-458, Dec 2003.
29. Y. Guo et al., Robust spectral regression for face recognition, IEEE Transactions on Neuro-computing 118 (2013) pp. 33-40.
30. A. Samal, P. A. Iyengar, Automatic recognition and analysis of human faces and facial expressions: a survey. Pattern Recognition, 1st ed. vol. 25, 1992, pp. 65-77.
31. P. Ekman, W. V.Friesen, Manual for the Facial Action Coding System. Palo Alto: Consulting Psychologists Press,1977.
32. V. N. Vapnik, The Nature of Statistical Learning Theory. New York :Springer - Verlag , 2000.
33. M. J. Lyons et al., Automatic classification of single facial images, IEEE Transactions on Pattern Analysis and Machine Intelligence, 12th ed.vol. 21,1999 , pp. 1357-1362.
34. Y. N. Zhu et al., Expression Recognition Based on Genetic Algorithm and SVM, Advances in Intelligent and Soft Computing, 2009, pp. 743-751.
35. Y. N. Zhu et al., Expression recognition based on EPCA and SVM, 7th World Congress on Intelligent Control and Automation, 2008, pp.8516-8520.
36. K. Madsen, H. B. Nielsen, Finite algorithms for robust linear regression, BIT Computer Science and Numerical Mathematics 30(4) 1990, pp. 682-699.
37. B. G. Park et al., Face recognition using face-ARG matching, IEEE Transactions on PAMI 27(12) (2005) pp. 1982-1988.

38. A. M. Pires, Robust linear discriminant analysis and the projection pursuit approach practical aspects, in: Proceeding soft he International Conference on Robust Statistics, 2001.
39. C. T. Liao, S. H. Lai, A novel robust kernel for appearance-based learning, in: International Conference on Pattern Recognition, 2008, pp. 1-4.
40. J. Wright et al., Sparse representation for computer vision and pattern recognition, in: IEEE International Conference of Computer Vision and Pattern Recognition, 2009.
41. X. Xu, M. Ahmadi, A human face recognition system using neural classifiers, International Conference on Computer Graphics, Imaging and Visualization(2007).
42. J. Yang et al., Two-dimensional PCA: a new approach to appearance-based face representation and recognition, IEEE Transactions on PAMI 26(1) pp.131-137.
43. X. He et al., Face recognition using Laplacian faces, IEEE Transactions on PAMI 27(3) 2005, pp. 328-340.
44. V. Franc, V. Hlavac, Multi-class support vector machine[C]. In: R Kasturi, D Laurendeau, Suen C eds. The 16th International Conference on Pattern Recognition, IEEE Computer Society, vol. 2, 2002, pp. 236-239.
45. H. B. Nielsen, Computing a minimizer of a piece wise quadratic implementation, Technical Report, Informatics and Mathematical Modelling, Technical University of Denmark, DTU, 1998.
46. S. Fidler et al., Combining reconstructive and discriminative subspace methods for robust classification and regression by sub-sampling, IEEET transactions on PAMI 28(3) 2006, pp.337－350.