

# Equidistant Prototypes Embedding for Single Sample Based Face Recognition with Generic Learning and Incremental Learning

Weihong Deng<sup>a</sup>, Jiani Hu<sup>a</sup>, Xiuzhuang Zhou<sup>\*,b</sup>, Jun Guo<sup>a</sup>

<sup>a</sup>*School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, 100876, Beijing, China*

<sup>b</sup>*College of Information Engineering, Capital Normal University, Beijing, China*

---

## Abstract

We develop a parameter-free face recognition algorithm which is insensitive to large variations in lighting, expression, occlusion, and age using a single gallery sample per subject. We take advantage of the observation that equidistant prototypes embedding is an optimal embedding that maximizes the minimum one-against-the-rest margin between the classes. Rather than preserving the global or local structure of the training data, our method, called linear regression analysis (LRA), applies least-square regression technique to map gallery samples to the equally distant locations, regardless of the true structure of training data. Further, a novel generic learning method, which maps the intra-class facial differences of the generic faces to the zero vectors, is incorporated to enhance the generalization capability of LRA. Using this novel method, learning based on only a handful of generic classes can largely improve the face recognition performance, even when the generic data are collected from a different database and camera set-up. The incremental learning based on the Greville algorithm makes the mapping matrix efficiently updated from the newly coming gallery classes, training samples, or generic variations. Although it is fairly simple and parameter-free, LRA, combined with commonly used local descriptors, such as Gabor representation and local binary patterns, outperforms the state-of-the-art methods for several standard experiments on the Extended Yale B, CMU PIE, AR, and

---

\*Corresponding author. Tel: +86 10 62283059 Fax: +86 10 62285019  
Email address: [whdeng@bupt.edu.cn](mailto:whdeng@bupt.edu.cn) (Weihong Deng)

FERET databases.

*Key words:* Face Recognition, One Sample Problem, Linear Regression, Feature Extraction, Generic Learning.

---

## 1. Introduction

In this paper, we consider the problem of face recognition with a single gallery sample per subject. This “one sample problem” is one of the major challenges in many real-world applications on law enforcement and homeland security. Theoretically, it is an extreme small sample size problem that deteriorates conventional pattern recognition techniques. As the supervised learning techniques are not applicable without intraclass information, unsupervised techniques, which find the low-dimensional embedding of the gallery data by ICA [1], PCA [2] or its variants [3][4][5], have widely been applied, but these methods are optimal only for face representation and effective only for the recognition under small variation. Recognizing the face by invariant features (e.g. Gabor representation [6][7] and local binary patterns [8]) increases robustness to facial variation. However, because holistic image feature based methods discards all information about the 3D layout of the faces, they have limited descriptive ability on unobserved variation. Unfortunately, overcoming these limitations to build effective 3D face models has proven to be quite challenging, especially when the recognition system must be made to work in present of uncontrolled lighting and occlusion. Approaches based on morphable model and virtual face rendering achieve robustness at significant computational expense. A more efficient approach is to recognize the face through a set of non-overlapping blocks [9][10], but existing implementation of this idea have yielded inconclusive results [11]. One other strategy for increasing robustness to facial variation, as suggested by MPEG-7 standard on face retrieval, is to learn discriminative statistics from a generic dataset, but the optimized discrimination between a close set of gallery faces has not been considered. Some recent works on sparse representation propose to integrate the information of gallery set and external generic dataset by new dictionary designs [12][? ]. For detailed surveys on the one sample problem see [13].

Different from existing methods that preserve the global or local structure of the training data [14], we exploit to solve the one sample problem by placing gallery samples at the equally distant locations in the feature space, regardless of the true structure of training data in the observation space.

This procedure, which subsequently is referred to as “*equidistant embedding*”, takes advantage of two observations:

- 1) Many gallery faces look very similar to each other, which may be corresponding to the close classes in the feature space. By mapping the close classes to be distant, equidistant embedding magnifies the subtle differences that are essential to discriminate similar faces, and, at the same time, reduces the risk of identity confusion in the present of varying lighting and occlusion.
- 2) If the gallery images are gathered under varying conditions, then, in the feature space, they may be tightly clustered according to the acquisition conditions, and the test sample might be inclined towards the wrong gallery samples with a similar condition. By dissipating the condition-based clusters, equidistant embedding reduces the bias caused by varying acquisition conditions.

Theoretically, the optimality of equidistant embedding can be interpreted by Vapnik’s statistical learning theory [15](page 353), which proofs that the minimum margin over all dichotomies of  $k \leq K$  points contained in a sphere in  $\mathbb{R}^{K-1}$  can be maximized by placing these points on a regular simplex whose vertices lie on the surface of the sphere. In the one sample problem, each sample represents the unique prototype of a class. In light of Vapnik’s theorem, considering all dichotomies of one class and the rest classes, equidistant prototypes is an optimal embedding that maximizes the minimum one-against-the-rest margin between the classes.

**Remark 1.** *In the feature space where there is a single sample per class, placing gallery samples at the equally distant locations is equivalent to maximizing the minimum margin between any class and the rest.*

Motivated by the Remark 1, we propose an highly efficient and parameter-free<sup>1</sup> algorithm, called linear regression Analysis (LRA), to address the one sample problem in face recognition. As illustrated in Fig. 1, LRA maps the three gallery images to  $[1, 0, 0]^T$ ,  $[0, 1, 0]^T$ ,  $[0, 0, 1]^T$ , using least-square regression. This setting of regression targets aims to avoid the time-consuming

---

<sup>1</sup>The property of parameter-free is crucial for practical usage, because the model selection procedures, such as cross validation, are not applicable with a single training sample per class.

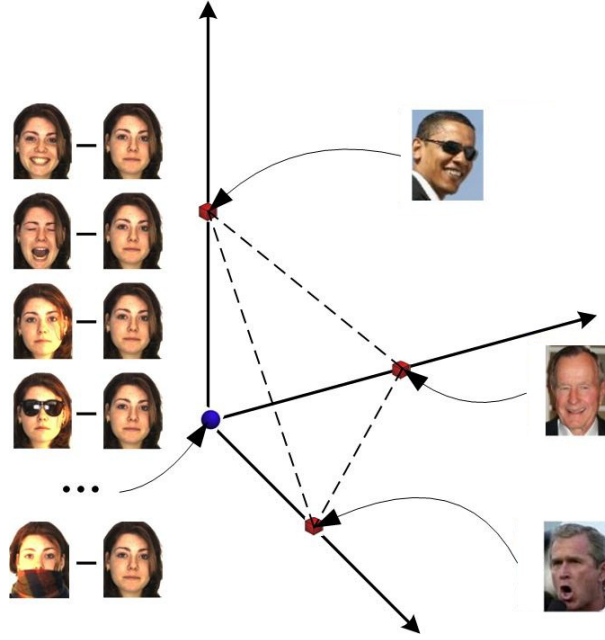


Figure 1: The geometric interpretation of our equidistant prototypes embedding space, where three gallery images are mapped to  $[1, 0, 0]^T$ ,  $[0, 1, 0]^T$ ,  $[0, 0, 1]^T$ , respectively, and the generic facial variations are all mapped to  $[0, 0, 0]^T$ . Incremental learning algorithm is formulated to make the model efficiently updated according to the newly coming gallery images, training samples, or generic variations.

search of nearest prototype, so that the test samples can be efficiently identified by finding the largest entry of the responding vector. Besides the equidistant prototype embedding, there are two additional advantages of our model as follows.

- *Generic Learning.* LRA maps the intra-personal differences of the generic faces to  $[0, 0, 0]^T$ , so that any similar variation on the gallery face would not affect the regression result. Empirical results confirm that this generic learning technique can largely improve the face recognition performance, even when the generic data are collected from a different database and camera set-up.
- *Incremental Learning.* The mapping matrix of LRA is formulated by the generalized inverse of the training data matrix, which can be incrementally calculated by the well-established Greville algorithm [16]. As a result, LRA model can be updated efficiently if new gallery classes, new training samples, or new generic variations is added. We will show that the incremental LRA achieves identical recognition results to the batch LRA.

It should be pointed out that least-square regression is a “classical” technique in pattern recognition, first applied by William G. Wee in 1968 for multiclass classification [17], and gained popularity recently in face recognition [18][19]. While previous works used linear regression to solve the problems with multiple samples per class, our proposed LRA method, for the first time, simply applies least-square regression in combination of common face descriptors to address the one-sample problem, and achieves superior results compared to the state-of-the-art methods. Theoretical and experimental comparisons between LRA and popular linear regression classification methods are also presented.

The rest of this paper is organized as follows: Section 2 details the mathematics of LRA model and explain how to apply the regression model to face recognition, Section 3 presents the experimental results, and Section 4 concludes this paper.

## 2. Linear Regression Analysis with Generic Learning

This section first formulates the equidistant prototypes based recognition in term of linear regression analysis (LRA), and then explores to enhance the generalization ability of the regressor by generic learning, incremental

learning, and invariant descriptors. The relations between LRA and other linear regression classification models are also discussed.

### 2.1. Linear Regression Analysis (LRA)

For the recognition of  $K$  subjects, countless equidistant embeddings are feasible by shifting and rotating a  $K - 1$  regular simplex. For the efficiency purpose, the class indicator vector  $y_i \in \mathbb{R}^K$  is applied to represent the  $i$ th subject, where  $y_i = [0, \dots, 1, \dots, 0]^T$  has a single 1, i.e. its  $i$ th component. This setting of equidistant embedding avoids the time-consuming nearest-neighbor search for recognition, since the nearest prototype can be efficiently found by the maximum element of the vector. Using these targets as multivariate outputs of the gallery samples  $X = [x_1, \dots, x_K] \in \mathbb{R}^{p \times K}$ , we can write the regression model in matrix notation

$$Y = WX + E \quad (1)$$

where  $Y = [y_1, y_2, \dots, y_K]$  is a  $K \times K$  target matrix,  $W$  is a  $K \times p$  mapping matrix and  $E = [e_1, e_2, \dots, e_K]$  is a  $K \times K$  matrix of errors. Without loss of generality, we assume that the gallery samples are ordered according to which class they are in, so that we can write  $Y$  as an  $K \times K$  identical matrix.

To minimize the sum of square error, i.e.  $\text{Tr}\{E^T E\}$ , the optimal transformation matrix can be computed as follows.

$$W = YX^\dagger = X^\dagger \quad (2)$$

where  $X^\dagger$  denotes the generalized inverse of  $X$ . When a novel test image is presented to the LRA based classifier, the feature vector of the image, denoted by  $x$ , is first extracted and then normalized to zeros mean and unit length. The response vector  $y \in \mathbb{R}^K$  is derived by a linear transformation:  $y = Wx$ . Finally, the recognition result is determined by the largest component of the response vector:

$$\omega = \arg \max_{i=1, \dots, K} y^{(i)} \quad (3)$$

where  $y^{(i)}$  denotes the  $i$ th element of the response vector  $y$ .

It should be mentioned that the mathematical formulation of LRA is identical to “linear regression of an indicator matrix [20]”, but it is renamed as “linear regression analysis” because of two following specialities. First, *LRA is derived as an efficient classifier for equidistant prototypes, which have a profound rationale for one-sample problem as indicated by Remark 1*. Second, by properly assigning the regression target, LRA is extended to include more information outside the gallery to further improve the performance.

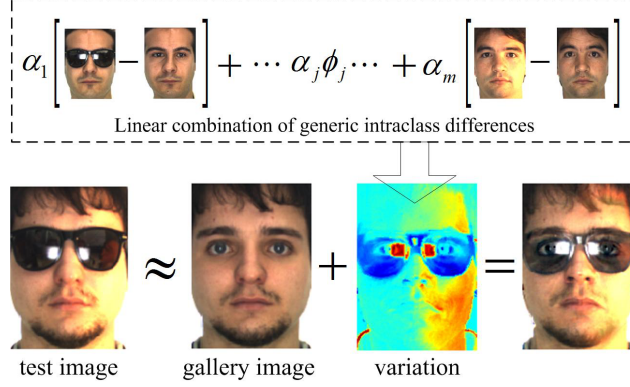


Figure 2: The intuition of generic learning of LRA: By mapping the generic intraclass differences to the zero vectors, the response vectors of the gallery and test images would be nearly identical.

## 2.2. LRA with Generic Learning (LRA-GL)

Given the learned mapping matrix  $W$  and a test sample  $x$  of subject  $i$ , the task of LRA is to recover the underlying  $y_i$  from the noisy responding vector  $y$ . Consider the face recognition problem as a linear regression model with measurement errors [21] as follows:

$$y = Wx + e \quad (4)$$

$$x = x_i + v_i \quad (5)$$

where the test sample  $x$  is represented as a combination of the correct gallery sample  $x_i$  and the associated intraclass variation  $v_i$ . In the training stage, the least-square solution in (2) assumes the error term  $e$  to be zero mean Gaussian noise. However, in the presence of unobserved variations  $v_i$  in expressions, lighting and occlusion, the test image  $x$  may largely deviate from the correct gallery image  $x_i$ , and the regression error is actually far from Gaussian distribution. Therefore, the least-square solution  $W$  is undesirable for the recognition problem with large variation  $v_i$ . Instead of using robust estimation [22][23] or local partition [9][18] based techniques to address outliers, we propose a novel learning method which makes full use of the similar shape and texture of human faces based on the following assumption:

**Assumption 1.** *The intra-class variation of any gallery face can be approximated by a linear combination of the intra-class differences from sufficient number of generic faces.*

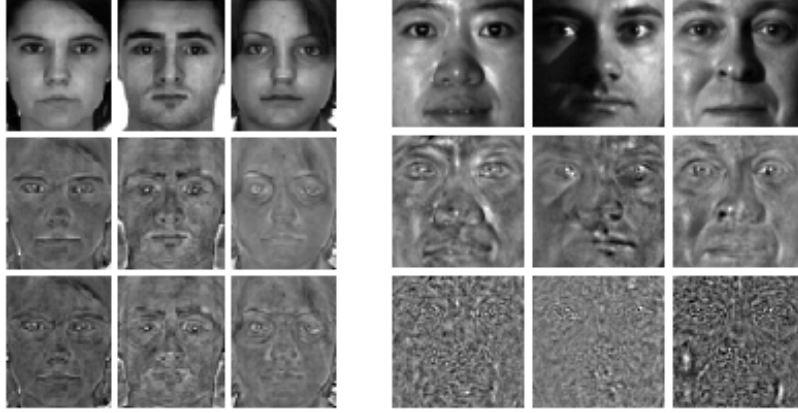


Figure 3: Some of the basis vectors obtained from the AR and Extended Yale B databases. The first row shows some gallery images, the second row shows corresponding basis vectors of LRA, and the final row shows the corresponding basis vectors of LRA-GL.

Given a sufficiently large set of  $m$  intraclass variant bases  $\{\phi_j\}_{j=1}^m$ , Assumption 1 guarantees that the unobserved variations between the test and gallery image of any subject can be approximated by a linear combination of “person-independent” bases, i.e.

$$x \approx x_i + \sum_{j=1}^m \alpha_j \phi_j \quad (6)$$

In our experiments, the intra-class variant bases are simply computed by subtracting the one sample from the other samples of the same class in a generic data set. Fig. 2 illustrates how Assumption 1 holds on facial images. The facial difference of the gallery face caused by a sunglasses and side light can be linearly approximated by some similar differences of other faces. Indeed, as the intra-class variation is not truly person independent, the linearly approximated variation can never be perfect in the representation sense (See the approximation error around the occluded eyes of Fig. 2 for example).

For the recognition purpose, however, canceling out such an approximated variation is sufficient to alleviate the large deviation between the gallery and test samples. Moreover, since the correlation between the interclass and intraclass variation of aligned human faces tend to be small [24][25], it is possible for the algorithm to project out the generic intraclass variations while preserving the separation of the gallery samples. Let  $W_{gl} \in \mathbb{R}^{K \times p}$  be the desirable mapping matrix of generic learning. By mapping all the intra-



class variant bases to the *zero vector*, i.e.  $W_{gl}\phi_j = z \in \mathbb{R}^K$ ,  $j = 1, \dots, m$ , LRA-GL reduces the unobserved intraclass variations by  $W_{gl}v_i \approx z$ . In this way, the underlying  $y_i$  can be accurately recovered by

$$y = W_{gl}(x_i + v_i) \approx W_{gl}x_i = y_i \quad (7)$$

To achieve this objective, the mapping matrix of LRA-GL is derived as follows:

$$W_{gl} = Y_{gl}X_{gl}^\dagger \quad (8)$$

$$Y_{gl} = \begin{bmatrix} Y, \underbrace{z, z, \dots, z}_m \end{bmatrix} \quad (9)$$

$$X_{gl} = [X, \phi_1, \phi_2 \dots, \phi_m] \quad (10)$$

where  $X_{gl}$  is the  $p \times (K + m)$  extended data matrix, and  $Y_{gl}$  is the  $K \times (K + m)$  extended target matrix. Under the typical condition on one sample problem<sup>2</sup>, the rank of  $X_{gl}$  is at most  $K + m$  and the rank of  $Y_{gl}$  is  $K$ , so that the rank of  $W_{gl}$  is  $K$ . Both LRA and LRA-GL derive a  $K \times p$  mapping matrix with the rank of  $K$ . Each of the  $K$  gallery images has a corresponding basis vector (the row vector of the mapping matrix) that spans the equidistant embedding space. The test sample is identified by choosing the best correlated basis vector. In order to observe how the basis vectors characterize the human faces, Fig. 3 shows some of the basis vectors (in image form) obtained from the AR and Extended Yale B databases. As we can see, the eyes, nose, part of the forehead above the eyebrows, and cheeks are dominant in basis vectors of LRA, and thus superior recognition performance could be achieved by emphasizing these invariant regions of each gallery face. By generic learning, the emphasized invariant regions become more localized, so that the robustness against sever variations could be further enhanced.

### 2.3. Incremental Learning of LRA-GL

The training set of a real-world face recognition system would change overtime. More training samples would be inserted to the system to enlarge the gallery or improve the accuracy. Thus, incremental learning is important

---

<sup>2</sup>We assume the gallery sample vectors and generic variant basis vectors are linearly independent, and the feature dimension is larger than the total number of gallery samples and generic variant bases. This condition is satisfied in all our experiments in this paper.

for the efficiency of the system. According to (8), the mapping matrix of LRA-GL is formulated as a multiplication of the target matrix and the generalized inverse matrix, which are efficiently updated as follows.

The update of the target matrix of LRA is straightforward according to three different situations.

- 1) If a sample of the  $i$ th class is added, the target matrix  $Y$  is updated according to  $Y_+ = [Y, y]$ , where  $y$  is a sparse vector with the  $i$ th element as 1;
- 2) If a generic variant vector is added,  $Y_+ = [Y, z]$ , where  $z$  is a zero vector with dimension equals to the number of rows of  $Y$ ;
- 3) If a sample of a new class is added,  $Y_+ = \begin{bmatrix} Y \\ z^T, y \end{bmatrix}$ , where  $z$  is a zero vector with dimension equals to the number of columns of  $Y$ , and  $y$  is a sparse vector with the last element as 1.

The update of the generalized inverse of the data matrix is conducted by the well-established Greville algorithm [16]. Denote the updated data matrix as  $X_+ = [X, x] \in \mathbb{R}^{p \times (N+1)}$ , where  $X \in \mathbb{R}^{p \times N}$  is the old data matrix and  $x \in \mathbb{R}^{p \times 1}$  is the added item (sample or generic variant vector). The Greville algorithm calculates the generalized inverse of matrix  $\hat{X}$  as follows:

$$X_+^\dagger = \begin{bmatrix} X^\dagger - X^\dagger x b^T \\ b^T \end{bmatrix} \quad (11)$$

where  $b^T$  is defined as:

$$b^T = \begin{cases} c^T, & \text{if } c \neq 0 \\ (1 + x^T X^\dagger x)^{-1} x^T X^\dagger, & \text{if } c = 0 \end{cases} \quad (12)$$

and

$$c = (I - X X^\dagger) x \quad (13)$$

When an item is added to the training data, Incremental LRA (ILRA) first updates  $Y$  and  $X^\dagger$  to  $\hat{Y}$  and  $\hat{X}^\dagger$ , respectively, and then computes the updated mapping matrix as

$$W_+ = (X_+^\dagger)^T Y_+^T \quad (14)$$

This update procedure consists of a small number of multiplication and addition of matrices and vectors, which is much more efficient than the direct

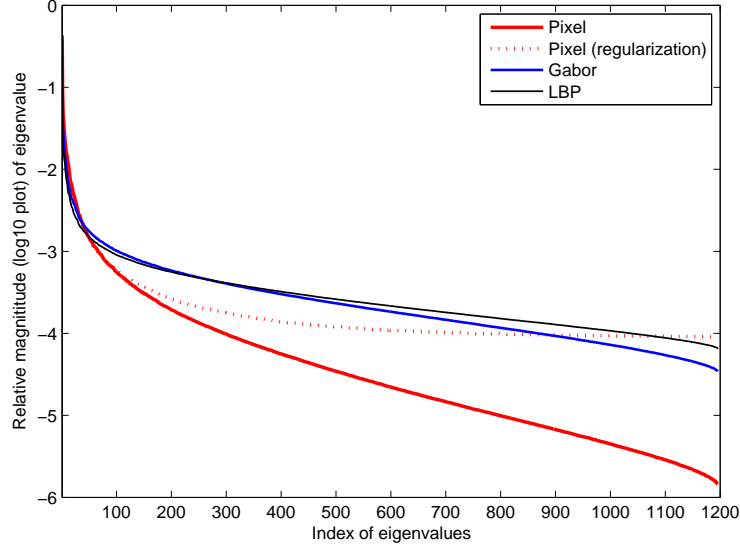


Figure 4: Relative magnitude ( $\lambda_i / \sum_{i=1}^n \lambda_i$ ) of the eigenvalues of the Gram matrix  $X^T X$  of three popular facial descriptors: pixel intensity, Gabor wavelet, and LBP. The data are from the 1196 gallery images of the FERET database.

computation of matrix inverse. With incremental learning, the gallery training and generic training of LRA can be conducted in the online manner. Moreover, the face-recognition systems can efficiently feedback the wrongly recognized samples to update the LRA model, avoiding re-training the whole system.

#### 2.4. Intrinsic Dimension, Regularization, and Face Descriptors

The number of pixels in a facial image is usually over  $10^4$ , but the intrinsic dimension of the image space is as low as  $O(10^2)$  [26] due to the similar shape and texture of human faces. The low intrinsic dimension causes the eigenspectrum of the pixel intensity feature decays rapidly, as shown in Fig. 4, which may deteriorate the recognition performance of LRA. Specifically, under the small sample size condition, the generalized inverse of  $X$  is computed as  $X^\dagger = (X^T X)^{-1} X^T$ , where the inverse Gram matrix  $(X^T X)^{-1}$  weights the face components by the inverse eigenvalues. Therefore, the small noise disturbance (of the test image) might cause large vibrations of the inverse eigenspectrum in the region of small eigenvalues, and thus largely effects the regression result. As evidence, in some cases used with pixel intensity, LRA

yields moderate accuracy similar to those of PCA and 2DPCA (See Table 2 for examples).

To stabilize the inverse eigenspectrum, the commonly used method is to regularize the eigenspectrum of the Gram matrix by adding a multiple of identity matrix, i.e.

$$X^\dagger = (X^T X + \lambda I)^{-1} X^T \quad (15)$$

Note that cross validation is not applicable to select the parameter for one-sample problem. Instead, we propose a LRA\* method<sup>3</sup> to control the strength of regularization by a statistic of the eigenvalue spectrum itself as

$$\lambda^* = \arg \max_{\forall \lambda_\delta} \left\{ \lambda_\delta \left| \frac{\sum_{i=1}^{\delta} \lambda_i}{\sum_{i=1}^n \lambda_i} \geq \theta \right. \right\} \quad (16)$$

where  $\lambda_i$  are the  $i$ th eigenvalue in decreasing order,  $n$  is the total number of the eigenvalues, and thus the threshold  $\theta$  represents the desirable proportion of energy captured in the first  $\delta$  eigenvectors. For the simplicity, we set  $\theta = 0.98$  for the pixel feature in our experiments. Fig. 4 shows an example of regularized eigenspectrum.

In practice, a more recommendable solution is to apply local descriptors, such as Gabor and LBP, to represent the faces instead of pixel intensity. In face recognition community, classification algorithms are mostly implemented and evaluated in pixel domain, as in [2][4][18][19], such that they ignore the underlying relationships between neighboring pixel points. In contrast, local descriptors consider the local relationships of the image pixel by predefined filter convolution or pattern matching, and therefore tend to have higher intrinsic dimension. As shown in Fig. 4, the eigenspectrums of Gabor and LBP features decay as slowly as the regularized one of the pixel intensity, and would be much stabler in the inverse form. Therefore, *local feature based representations are robust to face recognition because of not only their invariant property, but also their stable inverse eigenspectrums*. Implementing LRA on the local features may yield a more reliable performance compared with the pixel intensity. As evidence, local feature based classification using

---

<sup>3</sup>The proposed LRA method aims to solve one sample problem in a parameter-free manner. We apply regularization to LRA only when used with pixel feature, with the purpose of explaining why LRA cannot achieve desirable performance in some cases using pixel feature. In practice, we recommend to combine LRA with local features instead of pixel intensity such that regularization is not necessary.

LRA achieves consistently higher accuracies than competing methods under all test conditions of our experiments.

### 2.5. Relations to Linear Regression Classification

Linear regression classification (LRC) such as [18][19] has gained increased popularity lately in face recognition. LRC classifies the faces by finding the minimum residual between the test sample and the linear combination of the class-specific galleries. Although it is applicable to one sample problem, to the best of our knowledge, LRC has not been studied or tested with single sample per class. In the one sample problem, we find that LRC is reduced to a weighted nearest-neighbor classifier,

$$\omega = \arg \min_{i=1,\dots,K} \|x - w^{(i)}x_i\|_2 \quad (17)$$

Based on different manners to compute the scalar weight  $w^{(i)}$ , two algorithms have been proposed.

The first algorithm [18], which subsequently is referred to as LRC-1, computes the weight  $w^{(i)}$  as the projection of  $x$  on the  $i$ th gallery sample, i.e.

$$w_{\text{LRC-1}}^{(i)} = (x_i^T x_i)^{-1} x_i^T x \quad (18)$$

As illustrated in Fig. 5, the residual is determined only by the sine of  $\angle(x_i, x)$  for unit-length feature vectors, LRC-1 is actually an angle based nearest-neighbor classifier for one-sample problem. Further, if all of the samples are normalized to have zero mean and unit variance, then this procedure is equivalent to choosing the gallery sample that best correlates with the test sample. One should be aware that the correlation between feature vectors is very sensitive to the variation in lighting and occlusion.

The second algorithm [19], which subsequently is referred to as LRC-2, collaboratively computes the weights by the projection of  $x$  on the whole gallery set, i.e.

$$w_{\text{LRC-2}}^{(i)} = \delta_i (X^T X)^{-1} X^T x \quad (19)$$

where  $\delta_i = [0 \dots 1 \dots 0]$  is a vector extracting the coefficients corresponding to the  $i$ -th gallery sample. As illustrated in Fig. 5, the residual used in LRC-2 involves both the weight  $w^{(i)}$  and the angle  $\angle(x_i, x)$ . If there are some sever variations between the test image and correct gallery image, the angle  $\angle(x_i, x)$ , which essentially represents the feature correlation, might become a misleading factor in recognition. For instance, if a test image is captured

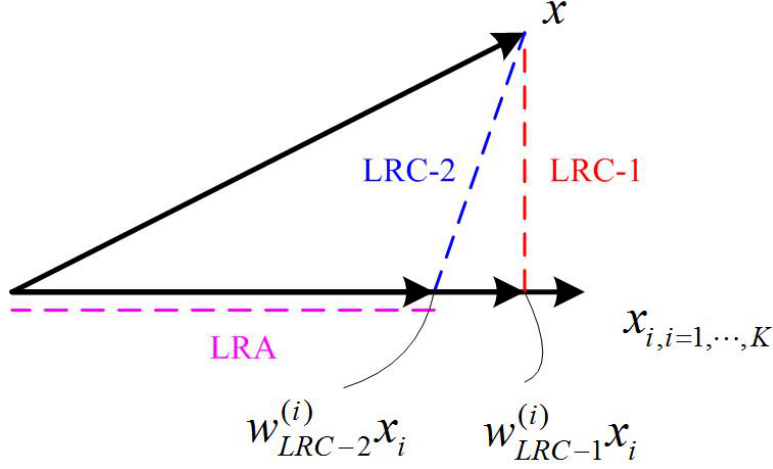


Figure 5: The relationships among popular linear regression based classification methods. The length of red line represents the residual used in LRC-1. The blue line represents the residual used in LRC-2. The length of magenta line represents the coefficient used in LRA.

under left lighting, all the gallery images with left light tend to have small angle with it. Once the correct gallery image is not under left light but some other gallery images are, mis-recognition of LRC-2 might occur.

It is interesting to point out that the weight  $w^{(i)}$  of LRC-2 in (19) is identical to the coefficient  $y^{(i)}$  of LRA in (3). In other words, LRA eliminates the misleading factor in LRC-2, and only use the weights of gallery samples for recognition. In this perspective, LRA is more robust than LRC-2 if the images in the gallery set is gathered under varying conditions (See Table 1 for evidences).

### 3. Experimental Results

Extensive experiments are carried out to illustrate the efficacy of the proposed approach using four standard databases, namely Extended Yale B [27], CMU PIE [28], AR [29] and FERET [30]. Several standard evaluation protocols reported in face recognition literature have been adopted and a comprehensive comparison of the proposed approach and the state-of-the-art techniques is presented using three widely used face descriptors, namely pixel intensity, Gabor, and LBP.

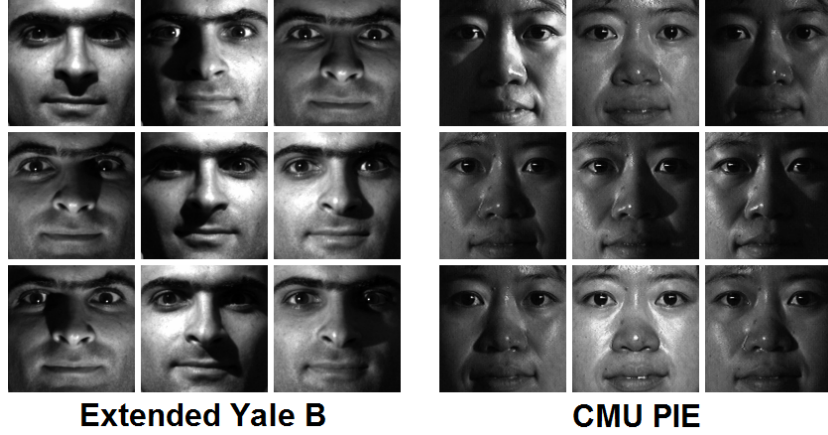


Figure 6: The cropped images of Extended Yale B, CMU PIE, and AR dataset. For each dataset, one training sample per subject is randomly selected for training and the rest samples are used for testing.

### 3.1. Extended Yale B, CMU PIE: Recognizing Faces Under Variable Illuminations

Extended Yale B (EYB) database contains frontal images of 27 individuals, each with 64 different illuminations. Two subsets (subset 2 and 3) with 24 light source directions from  $13^\circ$  to  $50^\circ$  are selected in our experiment. CMU PIE database [31] contains 68 subject with 41368 images as a whole. For each subject, the illumination subset (C27), which contains 21 distinct sources of lights, was used in our experiment. For the both datasets, all images are first normalized by a similarity transformation that sets the cen-

Table 1: The average recognition rate and standard deviation (std) across 10 tests on Extended Yale B, CMU PIE, and AR databases with single training sample per subject, using pixel intensity, Gabor, and LBP features respectively.

Methods	Extended Yale B			CMU PIE		
	Pixel	Gabor	LBP	Pixel	Gabor	LBP
2DPCA	23.0 $\pm$ 2.0	–	–	28.5 $\pm$ 1.1	–	–
$\chi^2$	–	–	46.9 $\pm$ 2.7	–	–	53.1 $\pm$ 1.0
PCA	21.9 $\pm$ 1.9	40.9 $\pm$ 3.6	38.7 $\pm$ 3.1	28.0 $\pm$ 1.0	40.0 $\pm$ 1.7	44.6 $\pm$ 1.4
LRC-1	23.0 $\pm$ 2.0	45.8 $\pm$ 3.2	44.5 $\pm$ 3.0	28.5 $\pm$ 1.1	42.9 $\pm$ 1.7	48.5 $\pm$ 1.2
LRC-2	38.7 $\pm$ 1.0	55.7 $\pm$ 2.9	51.1 $\pm$ 2.9	47.0 $\pm$ 0.5	59.0 $\pm$ 1.3	63.2 $\pm$ 1.0
LRA	<b>59.7<math>\pm</math>1.8</b>	<b>65.2<math>\pm</math>3.0</b>	<b>57.8<math>\pm</math>2.7</b>	<b>66.5<math>\pm</math>1.0</b>	<b>74.6<math>\pm</math>1.3</b>	<b>71.6<math>\pm</math>1.1</b>

tered inter-eye line horizontal and 70 pixel apart, and then cropped to the size of  $128 \times 128$  with the centers of the eyes located at (29, 34) and (99, 34) to extract the pure face region. No further preprocessing procedure is carried out in our experiments. The cropped images are used directly for Gabor and LBP feature extraction as described in [32], and resized to  $64 \times 64$  for pixel intensity feature extraction. Fig. 6 shows some cropped images of the three datasets.

For each dataset, one training sample per subject is randomly selected for training and the rest samples are used for testing. To make the reported results significant, we report the average recognition rate and standard deviation of 10 random training/testing partitions. Table 1 lists the comparative performance of six recognition methods on three data sets using three different features respectively. As expected, PCA yields very low accuracy on all datasets, which indicates that the feature correlation between the test image and single gallery image under variable illuminations is not reliable for recognition. Both 2DPCA [4] and LRC-1 [18] can improve the PCA baseline performance, but the improvement is rather marginal as they are essentially based on feature correlation (They yield identical accuracy in this experiment).  $\chi^2$  matching [8] yields notably better accuracy than the correlation based methods by taking advantage of the speciality of histogram matching, but its performance is still lower than LRC-2 [19] and LRA. Across all tested data sets and feature types, the proposed LRA method performs consistently better than LRC-2 followed by LRC-1, which clearly suggests that the superiority of the equidistant embedding to one-sample problem. By balancing the distances between gallery faces, the features around the invariant regions of each face would be emphasized, as illustrated in Fig. 3, which makes LRA significantly better than other methods. The relatively low accuracy of LRC-1 and LRC-2 may be because they are involved in the unreliable feature correlation in presence of variation in lighting. In particular, used with pixel intensity whose correlation is most misleading, LRA outperforms LRC-2 by margins of 20–30 percent, and outperforms other methods by margins of 30–60 percent.

Since the performance on EYB and PIE datasets are relatively low, we evaluate to further improve LRA by *cross-database* generic learning. Specifically, we use PIE data set as the generic training set for the recognition of EYB database, and use EYB data set as the generic training set for the recognition of PIE database. Fig. 7 shows the results of the LRA-GL method. Although these two databases are collected under different lighting and camera



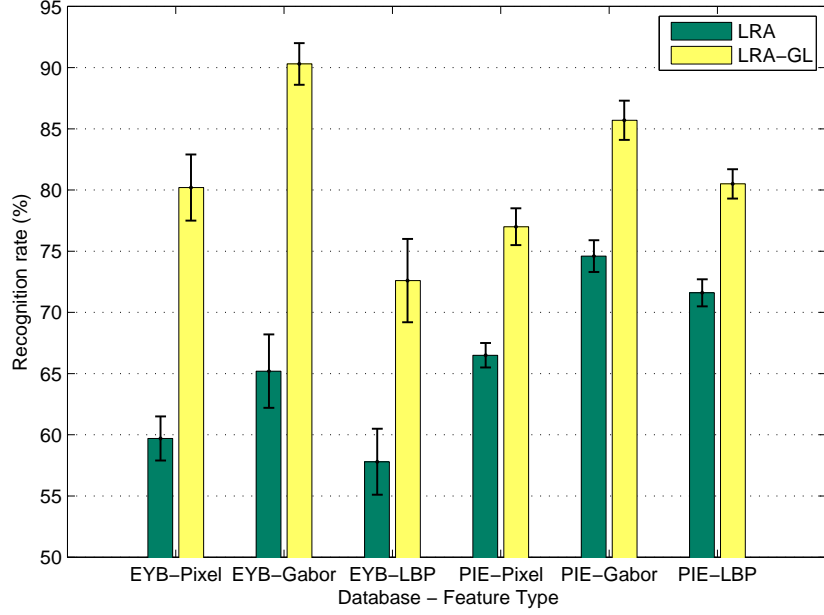


Figure 7: The average recognition rates of LRA and LRA-GL on the Extended Yale B (EYB) and CMU PIE data sets using three types of features, where the LRA-GL method uses a cross-database training procedure.

settings, generic learning between them can largely improve face recognition performance. In general, generic learning from PIE data set can boost the accuracy on EYB data set by 15–25 percent; Learning from EYB data set can improve the accuracy on PIE data set by about 10 percent. This result indicates that LRA-GL is feasible in wide-range real-world applications, even when the learning set is collected from different databases and camera set-ups.

### 3.2. AR Database: Recognizing Partially Occluded, Expression, and Lighting Variant Faces

The second set of experiments uses a large subset of AR database with 2500 images from 100 subjects. For each subject, the natural image of Session 1 is used for training, and the other 24 images (12 remaining images of Session 1, and 12 images of Session 2) are used for testing. The images are cropped with dimension  $165 \times 120$  and converted to gray scale as described in [9]. LBP feature is extracted on the cropped image with cells of  $11 \times 8$ . The cropped images are resized to  $128 \times 128$  for Gabor feature extraction, and resized to



Figure 8: The cropped images of one person in AR database. The single natural image is used for training, while the other 24 images with severe variation are used for testing.

82×60 for pixel feature. Fig. 8 shows the 25 cropped images of one person used in this experiment, and one can see from the figure that the test images are in different expressions, lighting, and occlusion (sunglass and scarf), and lighting+occlusion (L+O).

Table 2 reports the comparative accuracy of six different methods, and LRA performs differently using different types of features. Specifically, combined with Gabor feature, LRA achieves significantly higher accuracy than other methods; Used with LBP feature, LRA performs slightly better than LRC-2, and still much better than other methods; Using pixel feature, however, the accuracy is moderate, which is possibly caused by the unstable inverse eigenspectrum of the Gram matrix. Therefore, LRA\* method is applied to regularize the eigenspectrum, and significantly improved performance is observed. In general, compared with the results in Table 1, the advantage of LRA in this experiment is smaller. This may be because that the gallery set of this experiment is gathered in an unified condition, and the competing methods such as LRC-2 suffer less from the unreliable correlation.

It is also interesting to compare performance of different feature sets. In general, LBP feature performs much better than Gabor, followed by pixel intensity. Surprisingly, the accuracy difference between LBP and pixel could

Table 2: Comparative recognition rates on 100 subjects of AR database with single training sample and 24 test sample per subject

Methods	Expr.	Light	Occl.	L+O	Overall
Pixel+PCA	66.5	59.3	28.3	16.5	41.7
Pixel+2DPCA	67.5	60.8	29.5	17.4	42.8
Pixel+LRC-1	67.5	60.7	29.5	17.4	42.8
Pixel+LRC-2	70.0	<b>73.3</b>	37.0	<b>22.1</b>	<b>49.4</b>
Pixel+LRA	68.0	58.5	37.3	13.8	42.4
Pixel+LRA*	<b>72.8</b>	66.0	<b>40.8</b>	19.0	47.8
Gabor+PCA	65.0	73.3	33.0	23.6	47.9
Gabor+LRC-1	69.3	78.2	37.5	27.4	52.3
Gabor+LRC-2	77.2	90.8	61.0	42.6	66.4
Gabor+LRA	<b>79.2</b>	<b>93.5</b>	<b>70.3</b>	<b>52.5</b>	<b>72.4</b>
LBP+ $\chi^2$	90.5	93.8	89.3	77.3	86.7
LBP+PCA	84.0	88.5	79.3	66.5	78.5
LBP+LRC-1	88.0	90.8	83.5	73.4	83.1
LBP+LRC-2	<b>92.5</b>	<b>94.7</b>	<b>92.5</b>	83.0	89.9
LBP+LRA	92.3	<b>94.7</b>	<b>92.5</b>	<b>83.9</b>	<b>90.1</b>
LocPb [9]	82.3	N/A	65.8	N/A	N/A
SOM [10]	88.0	N/A	76.0	N/A	N/A

Table 3: Comparative average error rates of LRA and LRA with generic learning (LRA-GL) on 80 Subjects of AR Database using Single Training Sample. 20 Subjects outside the Gallery are used for generic learning. The notation  $\downarrow$  indicates the percentage of the recognition errors that are reduced by switching from LRA to LRA-GL.

Methods	Expression	Lighting	Occlusion	L+O	Overall
Pixel+LRA*	25.3	35.4	60.6	80.9	52.2
Pixel+LRA*-GL	19.8 ( $\downarrow$ 22%)	7.4 ( $\downarrow$ 79%)	18.0 ( $\downarrow$ 70%)	27.8 ( $\downarrow$ 66%)	19.1 ( $\downarrow$ 63%)
Gabor+LRA	20.3	6.9	30.3	47.7	27.7
Gabor+LRA-GL	12.7 ( $\downarrow$ 37%)	2.5 ( $\downarrow$ 64%)	9.1 ( $\downarrow$ 70%)	15.4 ( $\downarrow$ 68%)	10.4 ( $\downarrow$ 62%)
LBP+LRA	7.2	5.4	7.6	16.1	9.8
LBP+LRA-GL	4.5 ( $\downarrow$ 38%)	2.7 ( $\downarrow$ 50%)	3.0 ( $\downarrow$ 61%)	6.5 ( $\downarrow$ 60%)	4.4 ( $\downarrow$ 55%)

be up to 70 percent on the test set in presence of occlusion. The excellent performance on occluded faces may be because the uniform LBPs seldom appear in the occluded parts of the face, and the occluded parts are automatically discounted for recognition. We also compare our method with the state-of-the-art methods [10][9] that address the expression and occlusion by local partition with prior information. The SOM-face method needs to know in advance the occluded regions of the image, and LocPb method uses generic images to learn the weights for local region. Compared with these methods, LBP based classification using LRA provides a fairly simple but more robust solution to recognize partially occluded, expression, and lighting variant faces.

Next, we examine the effectiveness of generic learning on improving the accuracy of LRA. In this experiment, we chose a random subset of 80 subjects from the AR database. Following previous setting, the natural image of each subject in Session 1 is used for training, and the other 24 images are used for testing. To construct the intra-class variant bases for generic learning, another 20 subjects (not overlap with the 80 gallery subjects) are selected, with 13 images of Session 1 per subject. To make the reported results significant, we report the average error rate of 10 random gallery/generic partitions. Table 3 enumerates the recognition error rates for this experiment, in which we define an Error Reduction Rate (ERR), denoted by a notion  $\downarrow$ , to measure the effectiveness of generic learning. For instance, since the LBP feature based LRA-GL reduces the overall error rate from 52.2 to 19.1 percent, the ERR is

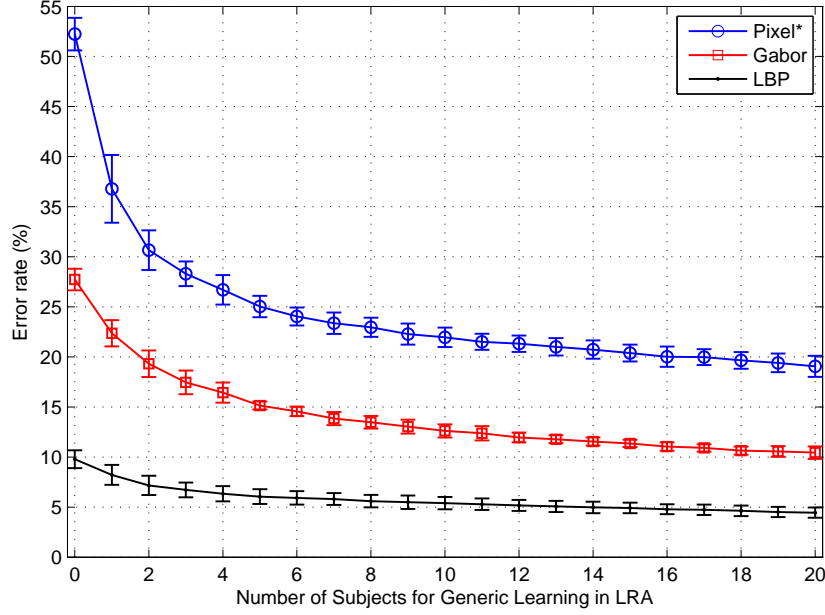


Figure 9: The average error rate (and standard deviation) of LRA-GL as a function of the number of subjects used for generic learning.

↓63 percent  $[(52.2-19.1)/52.2]$ , suggesting that 63 percent recognition errors can be avoided by switching LRA to LRA-GL.

For the three types of tested features, the overall ERR is about ↓55–↓63 percent, strongly proving the universal effectiveness of the LRA-GL method on various kinds of features. In particular, only 4.4 percent overall error rate is achieved by LBP feature based classification using LRA-GL. To better understand the effects of LRA-GL, Table 2 also enumerates the error rates of the four test variabilities respectively. Across all three tested feature types, LAR-GL is feasible on all types of variations, but the effectiveness appears to be different. The error reduction rates for lighting and occlusion (↓50–↓79 percent) are notably higher than those for expression (↓22–↓38 percent). The relatively low ERRs for expression indicates that the variation in expression is most sensitive to the specific facial shape of individual faces, and it is most difficult to be linearly approximated.

Even with this excellent performance, an interesting question remains: how many generic subjects are needed to construct the generic learning set. Fig. 9 shows a plot of the average error rate (and standard deviation) vs. the number of generic subjects used for learning. For all kinds of features,

the intra-class variant bases of a small number of subjects are sufficient to largely reduce error rate. For instance, leaning from five subjects reduces the error rate from 52 percent to 26 percent, using pixel feature based LRA-GL. This finding suggest that, once the intra-class variant bases are properly designed according to the testing condition, generic learning from a handful of subjects is enough to dramatically boost face recognition performance.

### 3.3. FERET Database: Large-Scale Face Recognition Under Complex Variations

The third set of experiments is designed to test the robustness of LRA against complex facial variation on the large-scale face database. The experiment follows the standard data partitions of the FERET database:

- **Generic training set** contains 1,002 images of 429 people, which are listed in the FERET standard training CD.
- **Gallery training set** contains 1,196 images of 1,196 people.
- **fb probe set** contains 1,195 images taken with an alternative facial expression.  
**fc probe set** contains 194 images taken under different lighting conditions.
- **dup1 probe set** contains 722 images taken in a different time.
- **dup2 probe set** contains 234 images taken at least a year later, which is a subset of the dup1 set.

All images are first normalized and cropped according to the eye coordinates, and the pixel, Gabor, and LBP features are extracted as described in the first set of experiments. Fig. 8 shows some cropped images which are used in our experiments, and one can see from the figure that the intraclass probe-to-gallery variability of this database is complex, as it is gathered in many sessions during several years.

Table 4 lists the comparative performance on the FERET database and the results show again that the LRA method is the best at simultaneously handling variation in expression, lighting, and ages, although the LRC-2 is a close second when LBP feature is used. Besides, we have compared LRA to other one-sample methods, namely the adaptive LDA [33] and the weighted  $\chi^2$  LBP matching [8]. Adaptive LDA uses a generic learning set to construct the within-class scatter matrix and the gallery set to construct the between-class scatter matrix, which makes its performance better than standard PCA.



Figure 10: The cropped images of some gallery images and corresponding probe images.

Table 4: Comparative recognition rates on 1196 subjects of FERET database with single training sample per subject

Features	Methods	fb	fc	dup1	dup2	Overall
<b>Pixel</b>	PCA	72.3	15.5	38.4	23.5	55.5
	2DPCA	73.1	16.0	39.2	24.4	56.3
	LRC-1	72.6	15.5	38.4	23.5	55.7
	LRC-2	74.6	71.1	56.2	51.3	68.0
	LRA	69.0	74.2	51.8	46.2	63.6
	LRA*	<b>79.5</b>	<b>82.5</b>	<b>62.6</b>	<b>59.8</b>	<b>74.0</b>
<b>Gabor</b>	AFLD [33]	<b>97.9</b>	92.3	70.8	54.7	88.1
	PCA	73.1	89.2	44.3	55.1	64.7
	LRC-1	76.4	91.2	47.6	57.7	67.9
	LRC-2	93.0	<b>98.5</b>	74.4	81.2	87.1
	LRA	96.4	<b>98.5</b>	<b>80.1</b>	<b>82.9</b>	<b>91.0</b>
<b>LBP</b>	$\chi^2$ [8]	97	79	66	64	84.7
	PCA	84.8	61.3	61.5	60.7	74.7
	LRC-1	89.7	69.6	67.3	67.5	80.2
	LRC-2	96.7	<b>97.4</b>	87.7	<b>84.6</b>	93.7
	LRA	<b>97.2</b>	96.9	<b>88.4</b>	<b>84.6</b>	<b>94.2</b>

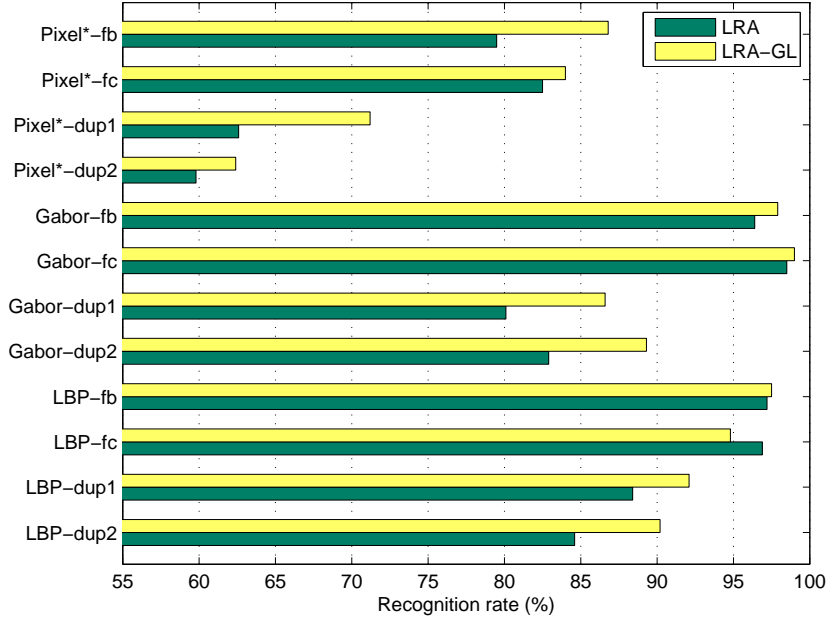


Figure 11: The comparative FERET recognition rates of LRA and LRA-GL on four probe sets using three types of features.

Weighted  $\chi^2$  method characterizes the prior knowledge on the importance of different face regions. However, their performance is still lower than LRA which uses only the gallery set for training. This comparison indicates that the equidistant embedding of LRA provides optimized way to utilize the discrimination information in the gallery set.

Although the accuracy of LRA is already very high, incorporating information in the generic training set can further improve performance. Fig. 11 shows that LRA-GL raises the recognition rate of LRA in 11 cases out of the 12 test cases (3 features  $\times$  4 probe sets). In particular, Gabor feature-based LRA-GL, with 99 percent accuracy on the fc set, is the best to handle the illumination changes, while LBP feature-based LRA-GL, with over 90 percent accuracy on the dup1 and dup2 set, is expert in addressing aging effects. Across all tested features, the boost on recognition accuracy is significant on the dup1 and dup2 sets, which are acquired in uncontrolled settings that are close to real-world conditions, indicating that the intraclass variability of face is sharable even in a complex situation.



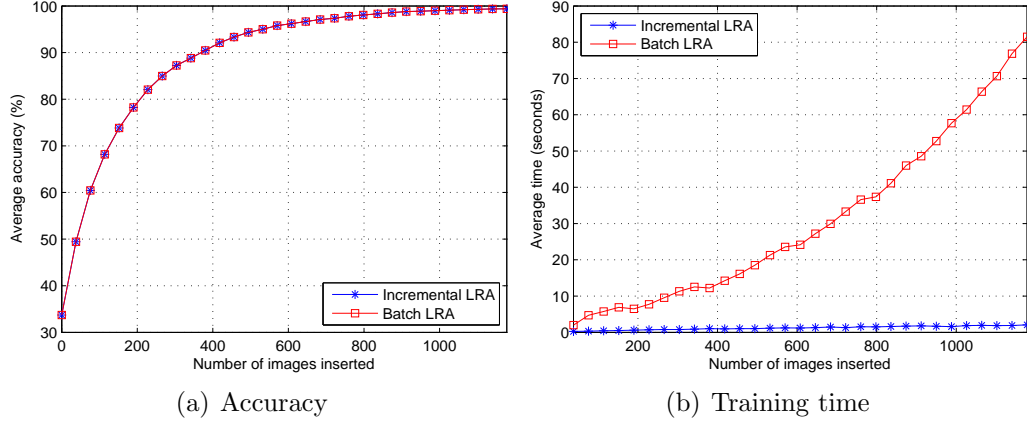


Figure 12: The comparison between batch-LRA and incremental-LRA on Extended Yale B database. (a) The accuracy as a function of the number of the insert training samples; (b) The training time as a function of the insert training samples.

#### 3.4. Incremental Learning from Sufficient Samples

To test the incremental learning of LRA, we have re-conducted the above four experiments in an incremental manner. Specially, we randomly select a small proportion of training samples as initialization, insert the remaining samples one by one, and update the mapping matrix incrementally using Eq. (11–14). We evaluate the mapping matrix achieved after all the training samples are inserted, and find that the recognition accuracies of Incremental LRA (ILRA) and (batch) LRA are identical on all experiments.

Previous experiments have demonstrated that LRA is effective to one sample problem. In practice, it is possible that *the system is started by a single image per person, and then enhanced by incoming samples in the runtime*. The last experiment therefore evaluates this situation using the Extended Yale B database [27]. The database consists of 2,414 frontal-face images of 38 individuals. The cropped and normalized  $192 \times 168$  face images were captured under various laboratory-controlled lighting conditions. For each subject, we randomly select 32 images per subject for training and the remaining for testing. To make the reported results significant, we report the average result of 10 random training/testing partitions.

A single training images per subject are randomly selected as initial data, and the remaining training images are added chunk by chunk for training in a random order. There are 31 chunks in total, and each chunk contains 38 samples with one training sample per subject. Fig. 12(a) shows the recog-

Table 5: Comparative recognition rates of SRC and LRA on the Extended Yale B database using 32 training images per subject.

Method	Accuracy
Fisherfaces+SRC	96.91% [34]
Laplacianfaces+SRC	96.52% [34]
Randomfaces+SRC	98.09% [34]
Downsample+SRC	97.10% [34]
LRC-1	96.87 $\pm$ 2.12% [19]
LRC-2	98.91 $\pm$ 1.37% [19]
Batch LRA	<b>99.61<math>\pm</math>0.08%</b>
Incremental LRA	<b>99.61<math>\pm</math>0.08%</b>

nition accuracy of ILRA as a function of the number of training samples inserted, and one can see from the figure that the accuracy keeps increasing as the training images inserting. We also implement (batch) LRA using the training set after the insertion of each chunk, and find that ILRA achieves identical accuracy to the LRA in all cases. Fig. 12(b) shows that, for each update, the execution time of ILRA is much less than the LRA, and the superiority of ILRA becomes more and more notable as the number of insertions increasing. Finally, when all 32 training samples are used, both methods achieve a nearly perfect accuracy of 99.61 $\pm$ 0.08% with 32 training images per subject.

SRC (sparse representation classifier) [34] is an up-to-date classification method using  $\ell_1$  minimization, and achieves impressive performance on Extended Yale B database. We therefore quote its performance for comparison in Table 5, which enumerates the performance of SRC associated with various feature extraction methods such as Fisherfaces [35], Laplacianfaces [14], Randomfaces [34], downsampling [34]. Compared with SRC, LRA/ILRA classifier yields a higher accuracy using the raw pixel feature, without any additional feature extraction procedure. Fig. 13 show the comparative performance of SRC (with 504 dimensional randomfaces as in [34]) and LRA with 1,4,8,16,32 training samples per subject. We randomly permute the training/test samples 10 times, and the average accuracy and its standard deviation are recorded. The results show that our LRA method outperforms SRC given equal numbers of training samples in all tests, and the accuracy

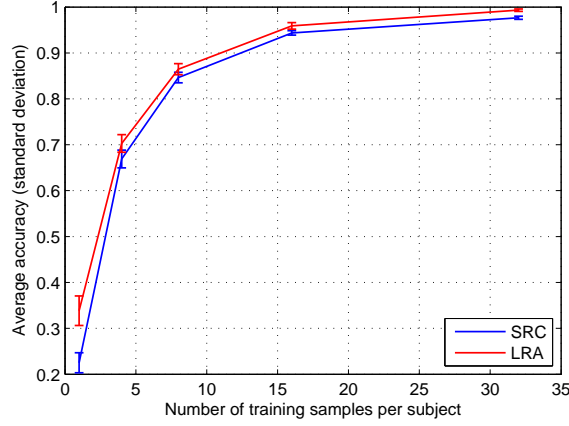


Figure 13: The comparative recognition rates of SRC and LRA on Extended Yale B database using different number of training samples per subject.

gap seems larger in the case with a less number of training samples.

#### 4. Conclusion

The experiments suggest a number of conclusions on one sample problem in face recognition:

- 1) PCA, 2DPCA and LRC-1 perform similarly on varying datasets, because they are essentially based to the correlation between the test sample and the gallery samples, which suffers under variation in lighting and occlusion.
- 2) The LRA method appears to be the best at simultaneously handling variation in lighting, expressions, occlusion, and ages, although the LRC-2 is sometimes a close second. This suggests that equidistant prototypes is an optimal embedding for one sample problem from a discrimination standpoint.
- 3) In the challenging applications where the gallery images are gathered under varying conditions, LRA outperforms the other methods by a large margin, since it is not involved in the unreliable correlation between the test sample and the gallery samples.
- 4) Local feature based LRA appears to be more effective than the one based on pixel intensity, partially because of the stable inverse eigen-spectrum of the Gram matrix.

- 5) Given the images sufficiently cover the testing conditions, no matter what kind of feature is used, generic learning from the intraclass variant bases of 5–10 subjects improves the recognition rate of LRA significantly. Moreover, LRA-GL is still effective when the generic data are collected from a different database and camera set-up.

Additionally, although the idea of equidistant embedding is effective to various types of feature, the choice of the face descriptor is critical to address the complex face variability. For instance, LBP descriptor is found to be the best at simultaneously handling variation in lighting, expression, occlusion, and age. A similar observation has been made by Wolf et al. [36], who combined several novel LBP-like and Gabor-like descriptors to achieve best results ever reported for the uncontrolled face verification problem on “Labeled Faces in the Wild” (LFW). Although achieving the best results on several benchmark data sets, LRA by itself may not be sufficient to address the one sample problem in the real world. Thus, we plan to design new face-image descriptors to capture more identity-specific facial features.

## References

- [1] W. Deng, Y. Liu, J. Hu, and J. Guo, “The small sample size problem of ica: A comparative study and analysis,” *Pattern Recognition*, vol. 45, no. 12, pp. 4438–4450, 2012.
- [2] M. Turk and A. Pentland, “Eigenfaces for recognition,” *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.
- [3] W. Deng, J. Hu, and J. Guo, “Gabor-eigen-whiten-cosine: A robust scheme for face recognition,” in *Analysis and Modelling of Faces and Gestures*. Springer, 2005, pp. 336–349.
- [4] J. Yang and D. Zhang, “Two-dimensional pca: A new approach to appearance-based face representation and recognition,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 1, pp. 131–137, 2004.
- [5] W. Deng, J. Hu, J. Lu, and J. Guo, “Transform-invariant pca: A unified approach to fully automatic face alignment, representation, and recognition,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 99, no. PrePrints, p. 1, 2013.

- [6] C. Liu and H. Wechsler, “Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition,” *IEEE Trans. Image Processing*, vol. 11, no. 4, pp. 467–476, 2002.
- [7] W. Deng, J. Hu, J. Guo, W. Cai, and D. Feng, “Emulating biological strategies for uncontrolled face recognition,” *Pattern Recognition*, vol. 43, no. 6, pp. 2210–2223, 2010.
- [8] T. Ahonen, A. Hadid, and M. Pietikinen, “Face description with local binary patterns: Application to face recognition,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, no. 12, pp. 2037–2041, 2006.
- [9] A. Martínez, “Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class,” *IEEE Trans. Pattern Anal. Machine Intell.*, pp. 748–763, 2002.
- [10] D. Zhang, Z.-H. Zhou, and S. Chen, “Recognizing partially occluded, expression variant faces from single training image per person with som and soft knn ensemble,” *IEEE Transactions on Neural Networks*, vol. 16, no. 4, pp. 875–886, 2005.
- [11] J. Wang, K. Plataniotis, J. Lu, and A. Venetsanopoulos, “On solving the face recognition problem with one training sample per subject,” *Pattern Recognition*, vol. 39, no. 9, pp. 1725–1745, 2006.
- [12] W. Deng, J. Hu, and J. Guo, “Extended src: Undersampled face recognition via intraclass variant dictionary,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 34, no. 9, pp. 1864–1870, 2012.
- [13] X. Tan, S. Chen, Z.-H. Zhou, and F. Zhang, “Face recognition from a single image per person: a survey,” *Pattern Recognition*, vol. 39, no. 9, pp. 1746–1762, 2006.
- [14] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang, “Face recognition using laplacianfaces,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 3, pp. 328–340, 2005.
- [15] V. Vapnik, “Statistical learning theory. 1998,” 1998.
- [16] A. Ben-Israel and T. Greville, *Generalized inverses: Theory and applications*. Springer Verlag, 2003, vol. 15.

- [17] W. Wee, “Generalized inverse approach to adaptive multiclass pattern classification,” *Computers, IEEE Transactions on*, vol. 100, no. 12, pp. 1157–1164, 1968.
- [18] I. Naseem, R. Togneri, and M. Bennamoun, “Linear regression for face recognition,” *IEEE Trans. Pattern Anal. Machine Intell.*, pp. 2106–2112, 2010.
- [19] Q. Shi, A. Eriksson, A. van den Hengel, and C. Shen, “Is face recognition really a compressive sensing problem?” in *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. IEEE, 2011, pp. 553–560.
- [20] J. Friedman, T. Hastie, and R. Tibshirani, *The Elements of Statistical Learning*, 1st ed. Springer-Verlag, 2001.
- [21] P. Bekker, “Comment on identification in the linear errors in variables model,” *Econometrica: Journal of the Econometric Society*, pp. 215–217, 1986.
- [22] P. Huber, “Robust regression: asymptotics, conjectures and monte carlo,” *The Annals of Statistics*, vol. 1, no. 5, pp. 799–821, 1973.
- [23] R. He, W. Zheng, and B. Hu, “Maximum correntropy criterion for robust face recognition,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 33, no. 8, pp. 1561–1576, 2011.
- [24] A. M. Martinez and M. Zhu, “Where are linear feature extraction methods applicable?” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 12, pp. 1934–1944, 2005.
- [25] B. Moghaddam, W. Wahid, and A. Pentland, “Beyond eigenfaces: probabilistic matching for face recognition,” *Automatic Face and Gesture Recognition, 1998. Proc. Third IEEE Int. Conf.*, pp. 30–35, Apr 1998.
- [26] M. Meytlis and L. Sirovich, “On the dimensionality of face space,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, no. 7, pp. 1262–1267, 2007.
- [27] A. Georghiades, P. Belhumeur, and D. Kriegman, “From few to many: Illumination cone models for face recognition under variable lighting and

- pose,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 23, no. 6, pp. 643–660, 2001.
- [28] T. Sim, S. Baker, and M. Bsat, “The cmu pose, illumination, and expression database,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 12, pp. 1615–1618, 2003.
  - [29] A. M. Martinez and R. Benavente, “The ar face database,” *CVC Technical Report #24*, June 1998.
  - [30] P. J. Phillips, H. Moon, P. Rizvi, and P. Rauss, “The feret evaluation method for face recognition algorithms,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, pp. 0162–8828, 2000.
  - [31] T. Sim, S. Baker, and M. Bsat, “The cmu pose, illumination, and expression (pie) database,” in *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on.* Ieee, 2002, pp. 46–51.
  - [32] W. Deng, J. Hu, J. Guo, W. Cai, and D. Feng, “Robust, accurate and efficient face recognition from a single training image: A uniform pursuit approach,” *Pattern Recognition*, vol. 43, no. 5, pp. 1748–1762, 2010.
  - [33] Y. Su, S. Shan, X. Chen, and W. Gao, “Adaptive generic learning for face recognition from a single sample per person,” in *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on.* IEEE, 2010, pp. 2699–2706.
  - [34] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, “Robust Face Recognition via Sparse Representation,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 31, no. 2, pp. 210–227, 2009.
  - [35] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, “Eigenfaces vs. fisherfaces: Recognition using class specific linear projection,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 19, no. 7, pp. 711–720, 1997.
  - [36] L. Wolf, T. Hassner, and Y. Taigman, “Effective unconstrained face recognition by combining multiple descriptors and learned background statistics,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 33, no. 10, pp. 1978 –1990, oct. 2011.