The CAS-PEAL Large-Scale Chinese Face Database and Baseline Evaluations

Wen Gao, Bo Cao, Shiguang Shan, Delong Zhou, Xiaohua Zhang, Debin Zhao

JDL-TR-04-FR-001

May 2004

ICT-ISVISION
Joint Research&Development Laboratory for Face Recognition
Chinese Academy of Sciences

The CAS-PEAL Large-Scale Chinese Face Database and Baseline Evaluations

Wen Gao, Bo Cao, Shiguang Shan, Delong Zhou, Xiaohua Zhang, Debin Zhao
ICT-ISVISION Joint Research&Development Laboratory for Face Recognition, Chinese Academy of Sciences, P.O.Box 2704, Beijing, China, 100080

Abstract

The ICT-ISVISION Joint Research & Development Laboratory (JDL) for Face Recognition has constructed the CAS-PEAL face database, supported by National Hi-Tech Program and ISVISION Technologies Co., Ltd. The goals to create the PEAL face database include (1) providing the worldwide researchers of FR community a large-scale face database for training and evaluating their algorithms; (2) facilitating the development of FR by providing large-scale face images with different sources of variations, especially Pose, Expression, Accessories, and Lighting (PEAL); (3) advancing the state-of-the-art face recognition technologies aiming at practical applications especially for the oriental. Currently, the CAS-PEAL face database contains 99,594 images of 1040 individuals (595 males and 445 females) with varying Pose, Expression, Accessory, and Lighting (PEAL). For each subject, 9 cameras spaced equally in a horizontal semicircular shelf are used to simultaneously capture images across different poses in one shot. Each subject is also asked to look up and down to capture 18 images in another two shots. We also considered 5 kinds of expressions, 6 kinds of accessories (3 pairs of glasses, and 3 caps), and 15 lighting directions. This face database is now partly made available (a subset named by CAS-PEAL-R1, containing 30,871 images of 1040 subjects) for research purpose only on a case-by-case basis only. JDL is serving as the technical agent for distribution of the database and reserves the copyright of all the images in the database.

CONTENTS

1. INTRODUCTION

Automatic Face Recognition (AFR) has been studied for over 30 years [1], [2]. Especially in recent years, it has become one of the most active research areas in pattern recognition, computer vision and psychology, owing to the extensive public expectation of its wide potential applications in public security, financial security, entertainment, intelligent human-computer interaction, etc. And much progress does have been made in the past few years. However, one has to admit that AFR remains a research area far from mature, partly due to the non-ideal imaging conditions and the subjects' non-cooperation in practical applications, though a great number of algorithms, frameworks, and systems are being proposed every year. Therefore, evaluating and comparing the potential AFR technologies exhaustively and objectively, discovering the real choke points and the valuable future research topics are becoming more and more significant.

Aiming at these goals, large-scale face database is obviously one of the basic requirements. Internationally, FERET [3], [4] and FRVT [9] have pioneered both evaluation protocols and database construction. FRVT2002 had performed evaluation on a larger scale face database with more than 30 thousands of faces. Unfortunately, it seems that their public distribution is impossible. FERET has publicly distributed its database containing thousands of images that has now been the standard testing set of AFR community. However, we believe that the FERET face database needs a complement considering the following reasons:

(1) The sources of variation covered by FERET face database are not sufficient enough and not controlled systematically. Especially for pose case, it contains only data of 200 subjects of several poses. Lighting, expression, and accessories are also not systematically controlled. Therefore, it cannot be used to evaluate these factors respectively.

(2) The performance of the state-of-the-art on FERET probe sets (especially FB) has approaching the possible limit. Therefore, it is really hard for us to judge which one is better statistically.

(3) Most of the subjects in the FERET database are western. How a system can be generalized to peoples from other races will be unknown.

Of course, there are other face databases internationally including CMU PIE [5], AR [6], XM2VTS [7], ORL, MPI, UMIST, MIT, Yale and Yale B, BANCA [8], KFDB [10], etc. Among them, the PIE face database has well controlled the sources of variation, especially the pose and illumination. However, there are only 68 subjects in the database, which may not satisfy the practical requirement for a large-scale face searching and provide enough training images for some algorithms. KFDB has been recently reported that it contains more than 52,000 images of 1,000 Koreans with sources of variation of pose, expression and illumination. Unfortunately, it has not been publicly distributed by now.

To sum up, we believe that the AFR community needs a large-scale face database for the abovementioned purposes, especially a large-scale Chinese face database, which covers the variations under varying pose, lighting, expression, accessory, backgrounds etc. This has been our motivation to construct the CAS-PEAL Chinese face database and distribute the CAS-PEAL-R1 publicly.

The CAS-PEAL face database is constructed by the ICT-ISVISION Joint Research &

Development Laboratory (JDL) for Face Recognition, supported by National Hi-Tech Program and ISVISION Technologies Co., Ltd. The images are all collected in Beijing, China between August 2002 and April 2003. The goals to create the PEAL face database include (1) providing the worldwide researchers of FR community a large-scale face database for training and evaluating their algorithms; (2) facilitating the development of FR by providing large-scale face images with different sources of variations, especially Pose, Expression, Accessories, and Lighting (PEAL); (3) advancing the state-of-the-art face recognition technologies aiming at practical applications especially for the oriental.

Currently, the CAS-PEAL face database contains 99,594 images of 1040 individuals (595 males and 445 females) with varying Pose, Expression, Accessory, and Lighting (PEAL). For each subject, 9 cameras spaced equally in a horizontal semicircular shelf are used to simultaneously capture images across different poses in one shot. Each subject is also asked to look up and down to capture 18 images in another two shots. We also considered 5 kinds of expressions (excluding the neutral expression), 6 kinds accessories (3 pairs of glasses, and 3 hats), and 15 lighting directions. This face database is now partly made available (a subset named by CAS-PEAL-R1, containing 30,871 images of 1040 subjects) for research purpose only on a case-by-case basis.

The remaining part of the paper is organized as follows: the equipment setup is described in Section 2. The following section presents the design of the CAS-PEAL face database. The CAS-PEAL-R1 is described in details in Section 4. Some primary evaluation based on the CAS-PEAL-R1 is pending in Section 5. Finally, we give out how to get a copy of the CAS-PEAL-R1 face database from us.

2. THE JDL PHOTOGRAPHIC ROOM

In order to capture face images conveniently and efficiently, a special photographic room is set in the Joint Research & Development Lab of Chinese Academy of Sciences. The dimensions of the room is about 4m*5m*3.5m. To capture faces with different poses, expression, accessories, and lighting, some special equipment are configured in the room including multiple digital cameras, all kinds of lamps, accessories (glasses, hats).

2.1 Camera System

In our photographic room, a camera system consisting of nine digital cameras and a computer is elaborately designed. All the nine cameras are placed in a horizontal semicircular shelf with radius and height being 0.8 meters and 1.1 meters respectively. The type of the cameras is web-eye PC631 with 370,000 pixels CCD. They are all pointed to the center of the semicircular shelf and labeled as C0 to C8 from the subject’s right to left. The planform of the cameras distributed on the semicircle shelf is illustrated in Figure 1.

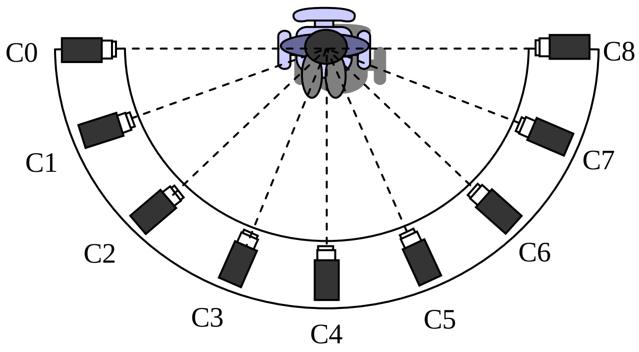


Fig. 1 Planform of our camera system

All of the nine cameras are connected to and controlled by the same computer through USB interface. The computer has been specially designed to support nine USB ports. We have designed software to control the nine cameras and capture images from them nearly simultaneously in one shot. In each shot, the software can obtain nine images of the subject across different poses within no more than 2 seconds and store these images in the hard disk using an uniform naming conventions.

Each subject is asked to sit down in a height-adjustable chair. Before taking photos, the chair will be adjusted to make the head of the subject be located at the center of the circle, and to make the subject face horizontally to the camera C4 that locates at the middle of the semicircular arch of the shelf (as Fig. 1 shows). Fig. 2 shows the situation that one subject sat in the chair and was ready for the photographing procedure.

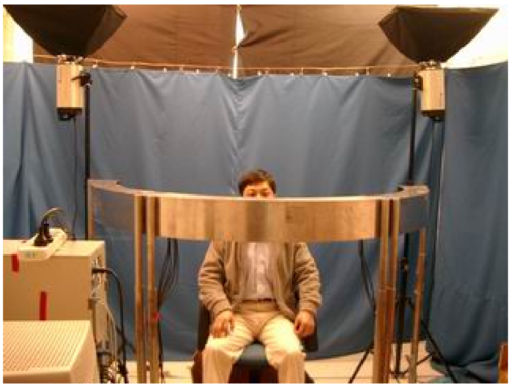


Fig. 2. Setup of the JDL Photographic room

2.2 Lighting System

To cover varying lighting conditions, we set a lighting system in our photographic room using multiple lamps and lanterns. To simulate the ambient illumination, two photographic sunlamps of high power covered with ground glass are used to mimic the indoor lighting environment. Actually, to obtain uniform lighting, they are arranged to irradiate to the matte white wall.

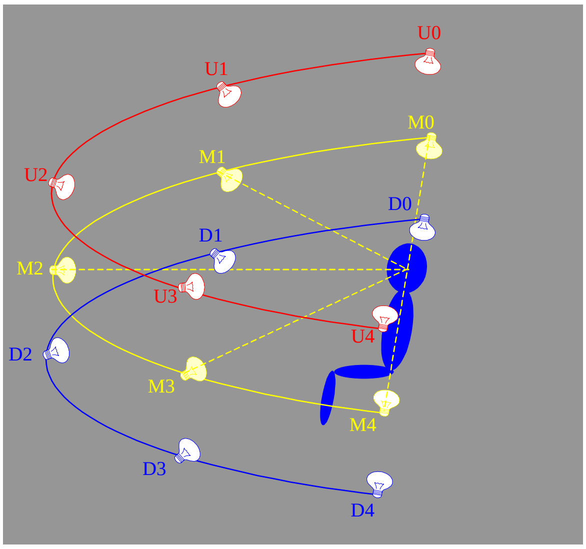


Fig. 3. Configuration of the lamps

Then, some fluorescent lamps are coarsely arranged as “lighting sources” to form the varying lighting conditions. 15 fluorescent lamps are placed at the “lamp” positions as shown in Figure 3. In the spherical coordinate system whose origin is the center of the circle which coincided with the semicircular shelf (the x axes was the middle camera’s axis and the y axes was horizontal), these positions are at 5 azimuths (-90^o, -45^o, 0^o, +45^o, +90^o) and 3 elevations (-45^o, 0^o, +45^o). By turning on/off each lamp, different lighting conditions are simulated. In order to decrease the labor, currently, we are exploiting a multiplex switch circuit to control the on/off of these lamps. Note that, in all cases, the ambient lamps are kept on. And for the purpose of mimicking practicality simply, the flash systems like CMU or YALE are not exploited in our case. Therefore, these images with varying lighting conditions are recommended for the purpose of image processing and face recognition under natural illumination.

2.3 Accessories: Glasses and Hats

Several kinds of glasses and hats are prepared in the room used as accessories to further increase the diversity of the database. The glasses consisted of dark frame glasses, thin and white frame glasses, glasses without frame. The hats also have brims of different size and shape.

2.4 Backgrounds

Without special statement, we are capturing face images with a sheet of blue cloth as the default background. However, in practical applications, many cameras are working under the auto-white balance mode, which may change the face appearance much in different scenes. Therefore, it is necessary to mimic this situation in the database. In the current version of the CAS-PEAL, we just consider the cases when the background color has been changed. Concretely, five different unicolor (blue, white, black, red and yellow) sheets of cloth are used.

3. DESIGN OF THE DATABASE

Utilizing the equipments described in Section 2, we defined seven combined variations to construct the CAS-PEAL face database: pose variation, pose and expression variation, pose and lighting variation, pose and accessory variation, pose and background variation, pose and session variation and pose and distance variation (the nine cameras are always working simultaneously). Table 1 has listed all the possible sources of variations we have considered when constructing the CAS-PEAL face database. Note that, except the looking right into camera case, we also ask the subject to look up (about 30 degree) and look down (about 30 degree) as another two pose sessions, which is listed in Table 1 as the variation of facing directions.

Table 1. All possible sources of variation considered in the CAS-PEAL face database

#Viewpoints	9						
# Variations	Facing directions	Expression	Lighting	Accessory	Background	Aging	Distance
	3	6	15	6	4	2	2
# Combined	27	54	135	54	36	18	18
#Total	342						

However, it is almost impossible to ask all the subjects concerned to finish all the sessions because of the different cooperation degree of the subjects. Therefore, sometimes, we have to abnegate some of the variations. However, any subject will be captured under at least two of these combined variations.

The following subsections describe each the variations and demonstrate some example face images. Note that, in order to simplify the description of some of the combinations, not all of the images from these nine cameras are demonstrated.

3.1 Pose Variation

To capture images with pose variation, the subject is asked to look upwards, look right into the camera C4 (the middle one), and look downwards, respectively. In each pose, nine images will be obtained from the nine cameras at one shot. So, totally 27 images of the subject will be obtained. Fig. 4 shows an example of the 27 images of one subject (ID=2).

U0 U1 U2 U3 U4 U5 U6 U7 U8

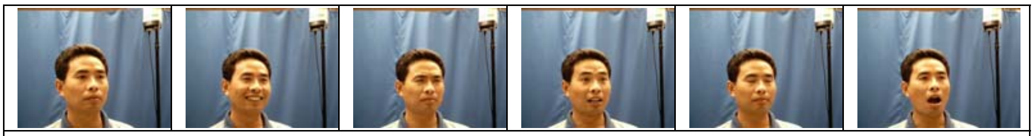
F0 F1 F2 F3 F4 F5 F6 F7 F8

D0 D1 D2 D3 D4 D5 D6 D7 D8

Fig. 4. The 27 images of one subject under pose variation in the CAS-PEAL database. The nine cameras were spaced equally in the horizontal semicircular shelf, each about 22.5^{o} apart. The subject was asked to look upwards, right into the camera C4 (the middle camera) and look downwards. Then, the 27 poses were named after the subject's pose (Up, Frontal, Down) and the number of the corresponding camera (from 0 to 8). The name of each pose was beneath its corresponding image.

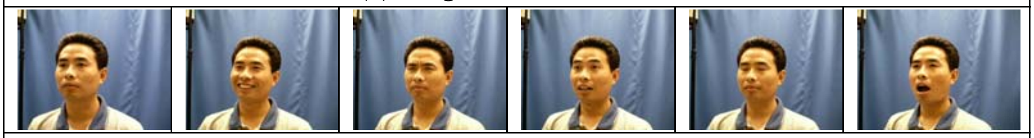
3.2 Pose and Expression Variation

In addition to the neutral expression, cooperative subjects will be asked to smile, to frown, to surprise, to close eyes and to open mouth. For each expression, 9 images of the subject are obtained using the 9 cameras. Figure 5 shows some example images of the six expressions (including neutral one) across three poses.



(a) Images from camera C3

(b) Images from Camera C4



(c) Images from Camera C5

Neutral	Smile	Frown	Surprise	Close eyes	Open mouth

Fig. 5. Example images of one subject with six expressions across 3 poses F3, F4, and F5

3.3 Pose and Lighting Variation

Lighting changes the face appearance greatly. Using the lighting system described in Section 2.2, we capture multiple images of each face. Some example images are illustrated in figure 6. Note that in all cases, the ambient lighting lamps are turned on.

(-90^{o}, +45^{o}) (-45^{o}, +45^{o}) (0^{o}, +45^{o}) (+45^{o}, +45^{o}) (+90^{o}, +45^{o})

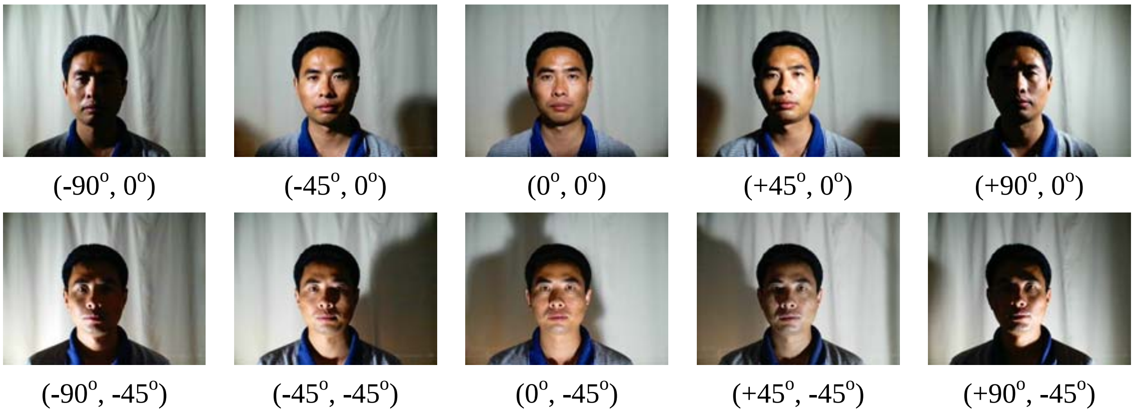


Fig.6. Some example images from one subject illuminated by fluorescent light source located at different azimuth and elevation coordinates. The (α, θ) beneath the image designates the azimuth and elevation.

3.4 Pose and Accessory Variation

For subjects that are willing to have this session, the prepared accessories, 3 hats and 3 pairs of glasses, are weared one by one. And nine images are captured using the camera systems. Fig. 7 illustrates the example images of one subject recorded by the camera C4.

Fig. 7. Example image of one subject with 6 different accessories

3.5 Pose and Background Variation

As has been mentioned above, sheets of different unicolor background cloth have been changed manually to capture the effect of the auto white-balance. Some example images are shown in Figure 8. We can see that for our camera case the white-balance has changed the face appearance a great deal.

Fig. 8. Example images of one subject with different background

3.6 Aging

In FERET and other tests, aging was always another important factor decreasing the recognition rates. In most face databases, images of one subject captured under different sessions are insufficient or absent because the subjects are hardly traced. In CAS-PEAL database, 66 subjects have been captured in two sessions half a year apart. Fig. 9 illustrates six images captured in the two sessions.

Fig. 9. Example images captured with time difference. The images in the bottom row are captured after half a year.

3.7 Different Distance

One can easily have a look at the effect of the distance changing between the camera and the face in figure 9. Some example images of the same face captured with half a year apart have been shown in figure 10.

Fig. 10. Example images of different distances from the camera

4. DESCRIPTION OF THE RELEASED CAS-PEAL FACE DATABASE: CAS-PEAL-R1

4.1 Contents of the CAS-PEAL-R1

The CAS-PEAL face database has been cut to form the first distribution: CAS-PEAL-R1. This distribution contains 30,871 images of 1,040 subjects. These images belong to two main subsets: frontal subset and pose subset.

1) In the frontal subset, all images are captured from camera C4 (see Fig. 1) with the subject looking right into this camera. Among them, 377 subjects have images with 6 different expressions. 438 subjects have images wearing 6 different accessories. 233 subjects have images under at least 9 lighting changes. 297 subjects have images against 2 to 4 different backgrounds. 296 subjects have images with different distances from cameras. Furthermore, 66 subjects have images recorded in two sessions at a 6-month interval.
2) In the pose subset, images of 1040 subjects across 21 different poses (subset of those described in Section 3.1) without any other variations are included.

Table 2 summarizes the contents CAS-PEAL-R1.

Table 2. The contents of CAS-PEAL-R1

Subset		# Variations	# Subjects	# Images
Frontal	Normal	1	1040	1,040
	Expression	5*	377	1,884
	Lighting	>= 9	233	2,450
	Accessory	6	438	2,616
	Background	2-4	297	650
	Distance	1-2	296	324
	Aging	1	66	66
	Total:			9,031
Pose		21 (3*7)	1040	21,840
Total:				30,871

5* : Neutral expression is not counted in.

4.2 Image Naming Convention

In CAS-PEAL face database, the filename of each image encodes the majority of the ground truth information of that image. Its format is described as follows:

xx_nnnnnn_Ixx±nn_Px±nn_Ex_An_Dn_Tn_Bx_Mn_Rn_Sn

1 2 3 4 5 6 7 8 9 10 11 12

It consists of 12 fields and is 46 characters long. The fields are separated by underline marks as shown above. In fields, “x”s and “n”s represent character type sequence and digital number sequence respectively, which vary with the properties of each image. The meaning, character type sequences and number sequences of each field are described in turn as follows:

1) Gender and age field. Its two character type sequence are defined as follows:

“xx”	FY	FM	FO	MY	MM	MO
Meaning	Female, Young	Female, Middle-aged	Female, Old	Male, Young	Male, Middle-aged	Male, Old

2) ID field. Its six digital number sequence indicates the identification of the subject in the image, increasing from 000001 to 001042 (000833 and 000834 are absent.).

3) Lighting variation field. The character “I” represents illumination variation. The first “x” (E, F, L) indicates the lighting source. The second “x” (U, M, D) indicates the elevation of the lighting source. The “±nn” indicates the azimuth of the lighting source.

Symbol	E	F	L	U	M	D
Meaning	Ambient lighting	Fluorescent lighting	Incandescent lighting	Elevation: $+45^{o}$	Elevation: 0^{o}	Elevation: -45^{o}

4) Pose field. The character “P” represents pose variation. The “x” (U, M, D) indicates the subject’s pose. The “±nn” indicates the azimuth of the camera from which the image is obtained.

Symbol	U	M	D
Meaning	Looking up	Looking into camera C4	Looking down

5) Expression field. The character “E” indicates that this field relates to expression variation. The following “x” has value from “N”, “L”, “F”, “S”, “C” and “O”. Its meaning is as follows:

Symbol	N	L	F	S	C	O
Meaning	Neutral	Laughing	Frowning	Surprising	eyes Closed	mouth Open

6) Accessory field. The character “E” indicates that this field relates to expression variation. The following “n” has value ranging from 0 to 6.

Value	0	1	2	3	4	5	6
Meaning	Without accessories	Hat 1	Hat 2	Hat 3	Glasses 1	Glasses 2	Glasses 3

7) Distance field. The character “D” represents distance variation. The following “n” has value ranging from 0 to 2, indicating different distance from the subject to the camera C4.

8) Session field. The character “T” indicates aging sessions. The following “n” has values denoting different sessions.

Value	0	1	2
Session	First session	Second session (3 months later)	Third session (6 months later)

9) Background field. The character “B” represents background variations. The “x” has the following values:

Value	B	R	D	Y	W
Background	Blue	Red	Dark	Yellow	White

10) Reserved for later use.

11) Privacy field. (Refer to the CAS-PEAL Database Release Agreement for details on this field.)

12) Resolution field. The character “S” represents resolution. The “n” has two values: 0 and 1, denoting two different resolutions of the image.

Value	0	1
Meaning	Size: 640*480	Size: 320*240

4.3 Image Format and Directory Structure

The original 30,871 RGB color images of size 640 × 480 in CAS-PEAL-R1 require about 26.6 GB storage space. To facilitate the distribution, all the images were converted to grey-scale images. Then, each grey-scale image is cropped to size 360 × 480 excluding most of the background without any transformation to the pixel values. The cropped images are stored as lossless Lempel-Ziv-Welch compression TIFF files. Several cropped images are shown in Fig. 11.

Fig. 11. Several examples of the cropped face images.

The directory tree of the CAS-PEAL-R1 is as follows:

```
<CAS-PEAL>                          //The root directory of CAS-PEAL-R1
    |
    +-  <Frontal>                   //Frontal subset
          |-  <Normal>
          |-  <Expression>
          |-  <Lighting>
          |-  <Accessory>
          |-  <Background>
          |-  <Distance>
    |     |-  <Aging>
    +-  <Pose>                      //Pose subset
          |-  <000001>                  //All images of one person are stored in one directory.
          |-  <000002>                  //Each directory's name is one person's ID.
          |-  <000003>
                …
                …
```

Because the filename of each image describes the property of that image in great detail, images in the database can be retrieved and reorganized easily to fulfill specific requirements. In addition, each leaf directory contains a text file (named FaceFP_2.txt) which provides the eyes locations of all the images in that directory.

5. EVALUATION RESULTS OF BASELINE ALGORITHMS ON THE CAS-PEAL-R1 DATABASE

The main objectives of the evaluation of baseline algorithms on the CAS-PEAL-R1 database are to 1) elementarily assess the difficulty of the database for face recognition algorithms, 2) provide an example evaluation protocol on the database, and 3) identify the strengths and weakness of commonly used algorithms.

In this section, the partition of datasets which include the training set, the gallery set and the probe sets is described. Then, three baseline algorithms are briefly introduced. The preprocessing process of the face images has been demonstrated that it can effectively affect the performance of the face recognition algorithms, so the details of the preprocessing process are also provided. Finally, the evaluation results are presented.

5.1 Datasets Used in the Evaluation

To compare different algorithms convincingly, two additional aspects should be considered: 1) the scale of the common datasets which are used in the training and testing of a specific algorithm, 2) the statistical significance of the differences between different algorithms.

These two aspects are closely related. If the scale of the test sets is very small, the performance scores may be highly stochastic, that is, it is hard to say that algorithm A is better than algorithm B just because the performance score of A is several percent higher than that of B on a small test set. Beveridge et al. [12] used a permutation methodology to compare two algorithms statistically. Also, the scale of the training set can influence the comparison of two algorithms. Martinez et al. [13] demonstrates that PCA can outperform LDA when the training data set is small. Considering these problems, we compose the test sets and the training set from the CAS-PEAL database as large as possible. And the test sets are categorized to restrict the images in one probe set to undergo one main variation as described in Section 3. These partitions can be used to identify the strengths and weakness of a specific algorithm, and to address the variations associated with changes in the probe sets.

In the evaluation, three kinds of datasets are composed from the CAS-PEAL-R1 database: training set, gallery set and probe sets. Their definition and descriptions are as follows:

1. Training set.

A training set is a collection of images which are used to generate a generic representation and to tune parameters for an algorithm. In the protocols, the training set contains 1,200 images (300 subjects randomly selected from the 1,040 subjects in the CAS-PEAL-R1 database and each subject contains four images randomly selected from the frontal subset of the CAS-PEAL-R1 database). (More details about the images can be added here…)

2. Gallery set.

A gallery set is a collection of images of known individuals against which testing images are matched. In the protocols, the gallery set contains 1,040 images of 1,040 subjects (each subject has one image under normal condition). Actually, the gallery set consists of all the normal images described in Table 2.

3. Probe sets.

A probe set is a collection of probe images of unknown individuals that need to be recognized. In the protocols, nine probe sets are composed from the CAS-PEAL-R1 database. Among them, six probe sets correspond to the six subsets in the frontal subset: expression, lighting, accessory, background, distance and aging, as described in Table 2. The other three probe sets correspond to the images of subjects in the pose subset: looking upwards, looking right into the camera C4 (the middle one), and looking downwards. All the images that appear in the training set are excluded from these probe sets.

Table 3. The datasets used in the evaluation protocols

Datasets	Training set	Gallery set	Probe sets (frontal)					
			expression	lighting	accessory	background	distance	aging
Num. of images	1,200	1,040	1,570	2,243	2,285	553	275	66

Datasets	Probe sets (pose)		
	looking upwards (PU)	looking right into the camera C4 (PM)	looking downwards (PD)
Num. of images	4,998	4,993	4,998

5.2 Baseline Face Recognition Algorithms

The three baseline algorithms evaluated are Principle Components Analysis (PCA), a.k.a Eigenfaces, a combined Principle Component Analysis and Linear Discriminant Analysis (PCA+LDA), and PCA+LDA algorithm based on Gabor features (G PCA+LDA). PCA and PCA+LDA based face recognition algorithms are both fundamental and well studied [14], [15], [16], [17]. Recently, 2D Gabor wavelets are extensively used for local feature representation and extraction, and demonstrate their success in face recognition [18], [19], [20]. So, the PCA+LDA algorithm based on Gabor features is also used as a baseline algorithm to reflect this trend.

5.2.1 Principle Components Analysis (PCA)

Principal Components Analysis (PCA) is commonly used for dimensionality reduction in face recognition, also known as Eigenfaces. PCA chooses projection directions W_{opt} that maximize the total scatter across all images of all faces in the training set.

For a training set that contains N sample images $\{x_1, \quad x_2, \quad \cdots, \quad x_N\} \in R^n$, the total scatter matrix S_T is defined as follows:

$$S_T = \sum_{k=1}^{N} (x_k - \mu)(x_k - \mu)^T$$

where μ is the mean vector of all sample images in the training set.

Then, the projection matrix W_{opt} can be chosen as follows:

$$\begin{aligned} W_{opt} &= \arg\max_W \left| W^T S_T W \right| \\ &= \left[w_1 \quad w_2 \quad \dots \quad w_m \right] \end{aligned}$$

$\{w_i \mid i = 1, 2, \dots, m\}$ is the set of n-dimensional eigenvectors of S_T corresponding to the m largest eigenvalues. In most circumstances, m can be chosen far less than n without significantly decrease the recognition rates.

PCA extracts features from face images that capture the main directions along which face images differ the most. So it may achieve reasonable recognition results. The main drawback of this method is that the scatter being maximized is due not only to the between-class scatter, but also to the within-class scatter. The within-class scatter should be eliminated in the extraction of face image features for classification. In most cases, the within-class variations due to illumination changes, varying facial expressions and so on are larger than the between-class variations due to the change in face identity. Thus, PCA may not be effective in some face recognition applications, especially when lighting conditions change significantly during the acquisition of face images. It has been suggested that by discarding the three most significant principal components, the variation due to lighting is reduced [17]. But at the same time, information contained in these three components that is useful for classification is lost.

5.2.2 Combined PCA and LDA (PCA+LDA)

Linear Discrimiant Analysis (LDA) is a widely used method for feature extraction and dimensionality reduction in pattern recognition and has been proposed in face recognition. LDA tries to find the "best" project direction in which training samples belonging to different classes are best separated. Mathematically, it selects the projection matrix W_{fld} in such a way that the ratio of the determinant of the between-class scatter matrix of the projected samples and the within-class scatter matrix of the projected samples is maximized.

For a c-class problem, the between-class scatter matrix is defined as follows:

$$S_b = \sum_{i=1}^{c} \Pr(\Omega_i)(\mu_i - \mu)(\mu_i - \mu)^T$$

where $\Pr(\Omega_i)$ is the prior class probability, μ_i is the mean vector of class Ω_i and μ is the grand mean vector of all samples from all classes.

The within-class scatter matrix is defined as follows:

$$S_w = \sum_{i=1}^{c} (\Pr(\Omega_i) \times \frac{1}{N_i} \sum_{y_k \in \Omega_i} (y_k - \mu_i)(y_k - \mu_i)^T)$$

where N_i is the number of samples in class Ω_i and y_k is one of the sample vectors of class Ω_i.

If S_w is non-singular, the projection matrix W_{fld} can be chosen as follows:

$$W_{fld} = \arg\max_W \frac{|W^T S_b W|}{|W^T S_w W|}$$

Then, W_{fld} can be calculated by solving the generalized eigenvalue problem:

$$S_b W = S_w W \Lambda$$

In face recognition application, if N is the number of images in the training set and n is the number of pixels in each image, the rank of $S_w \in R^{n \times n}$ is at most $N-c$. Typically, the number of training images is much smaller than the dimension n. In this case, S_w is singular. To overcome this difficulty, PCA is first used to reduce the dimension of the images from n to $N-c$ or less, then the recalculated S_w will be non-singular and LDA can be used to find the projection matrix W_{fld}.

5.2.3 PCA+LDA Algorithm based on Gabor Features (G PCA+LDA)

Instead of using the grey-scale images as the original features in the above two algorithms, the representation of the original features in this algorithm is based on the Gabor wavelet transform of the original images. Gabor wavelets are biologically motivated convolution kernels which are plane waves restricted by a Gaussian envelope function, and those kernels demonstrate spatial locality and orientation selectivity. In face recognition, Gabor wavelets exhibit robustness to moderate lighting changes, small shifts and deformations [18].

A family of Gabor wavelets (kernels, filters) can be defined as follows:

$$\psi_{u,v}(z) = \frac{\| k_{u,v} \|^2}{\sigma^2} e^{(-\|k_{u,v}\|^2 \|z\|^2 / 2\sigma^2)} \left[e^{i\vec{k}_{u,v} z} - e^{-\sigma^2/2} \right] \tag{1}$$

where $k_{u,v} = k_v e^{i\phi_u}$; $k_v = \frac{k_{\max}}{f^v}$ gives the frequency, and $\phi_u = \frac{u\pi}{8}, \phi_u \in [0,\pi)$ gives the orientation, and $z = (x, y)$.

$$k_{u,v} = k_v e^{i\phi_u}$$

where $e^{i \vec{k}_{u,v} z}$ is the oscillatory wave function, whose real part and imaginary part are cosine function and sinusoid function respectively.

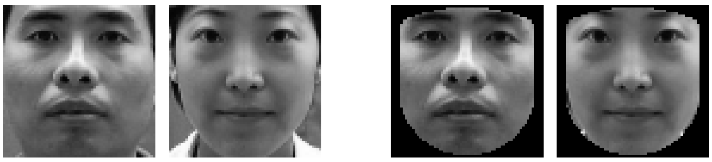
In this algorithm, we use the Gabor wavelets with the following parameters: five scales $v \in \{0,1,2,3,4\}$, eight orientations $u \in \{0,1,2,3,4,5,6,7\}$, $\sigma = 2\pi$, $k_{\max} = \pi$, and $f = \sqrt{2}$. These parameters can be adjusted according to the size of the normalized faces.

At each image pixel, a set of convolution coefficients can be calculated using a family of Gabor kernels as defined by equation (1). The Gabor wavelet transform of an image is the collection of the coefficients of all the pixels. To reduce the dimensionality, the pixels are sampled and their convolution coefficients are concatenated to form the original features of the PCA+LDA algorithm described above. These concatenated coefficients are also called the augmented Gabor feature vector in [19]. In the experiments, the size of the normalized faces is 64×64 and the pixels are sampled every four pixel both in row and in column, so the dimensionality of the features is 9000 ($15 \times 15 \times 40$). It should be noted that each feature is normalized to zero mean and unit variance to compensate for the scale variance of different Gabor kernels.

5.3 Preprocessing

In the evaluation, the preprocessing process of the face images is divided into three steps: geometric normalization, masking, and illumination normalization. The first two steps are to provide features that are invariant to geometric transformations of the face images, such as the location, the rotation and the scale of the face in an image, and remove irrelevant information for the purpose of face recognition, such as the background and the hair of a face. Illumination normalization is to decrease the variations of images of one face induced by lighting changes while still keeping distinguishing features, which is generally much more difficult than the first two steps. The details of the three steps are described as follows:

In geometric normalization step, each face image is scaled and rotated so that the eyes are positioned in line and the distance between them equals a predefined length. Then, the face image is cropped to include only the face with little hair and background as Fig. 12 (a) shows (the size of the cropped face image is 64×64). In masking step, a predefined mask is put on each cropped face image to further reduce the effect of different hair styles and backgrounds which are not the intrinsic characteristics, as Fig. 12 (b) shows. Typically, the hair style of a specific subject and the background are constant in a face database, so better performance can be obtained with larger face regions.



(a) (b)

Fig. 12. Example normalized face images in Step 1 and Step 2. (a) Geometrically normalized face images. (b) Masked face images.

In illumination normalization step, four illumination normalization methods are evaluated: Histogram Equalization (HE), Gamma Intensity Correction (GIC), Region-based Histogram Equalization (RHE) and Region-based Gamma Intensity Correction (RGIC) [21].

5.3.1 Gamma Intensity Correction (GIC)

The Gamma Intensity Correction method is to correct the overall brightness of the face images to a pre-defined "canonical" face images. It is formulated as following:

Predefine a canonical face image, I_0, which should be lighted under some normal lighting condition. Then, given any face image, I, captured under some unknown lighting condition. Its canonical image is computed by a Gamma transform pixel by pixel over the image position x, y:

$$I'_{xy} = G(I_{xy}; \gamma^*) \tag{2}$$

where the Gamma coefficient γ^* is computed by the following optimization process, which aims at minimizing the difference between the transformed image and the pre-defined normal face image I_0:

$$\gamma^* = \arg\min_{\gamma} \sum_{x,y} [G(I_{xy}; \gamma) - I_0(x, y)]^2 \tag{3}$$

where I_{xy} is the gray-level of the image position x,y; and

$$G(I_{xy}; \gamma) = c \cdot I_{xy}^{\frac{1}{\gamma}},$$

is the Gamma transform; c is a gray stretch parameter, and γ is the Gamma coefficient.

From equation 2 and 3, intuitively, the GIC is expected to make the overall brightness of the input images best fit that of the pre-defined normal face images. Thus, its intuitive effect is that the overall brightness of all the processed face images is adjusted to the same level as that of the common normal face I_0. Fig. 12 (b) illustrates the effect of GIC on an example face image.

5.3.2 Region-based HE and GIC (RHE and RGIC)

Both HE and GIC are global transforms over the whole image area. Therefore, they are doomed to fail when side lighting exists. Since the possible side lighting mainly cause the non-symmetry between the left and right part of the face, as well as the intensity variance between the top region and the bottom region, we partition the face into four regions according to the given eye centers as shown in Fig. 13 (a). Then, performing HE or GIC in these pre-defined face regions to better alleviate the highlight, shading and shadow effect caused by the unequal illumination. Fig. 13 (b) illustrates the effect of RHE and RGIC on an example face image.

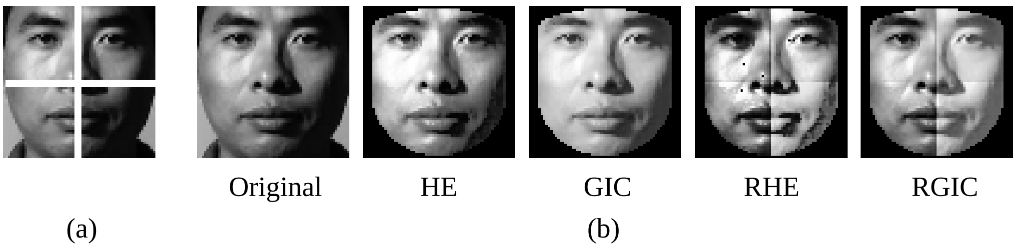
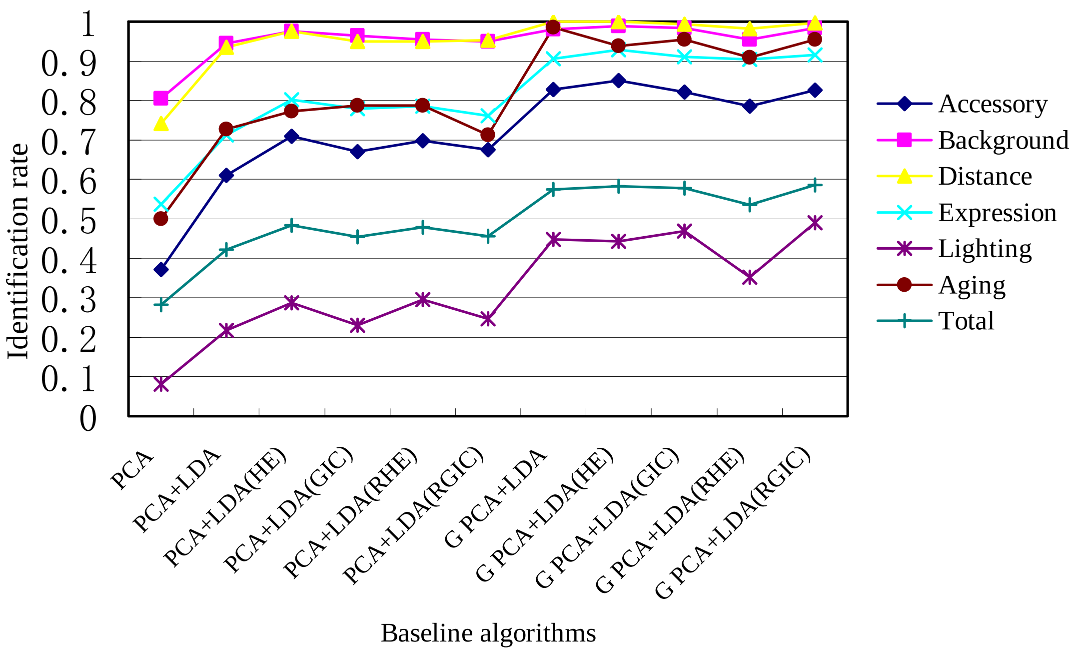


Fig. 13. Partition of face region and example images processed by different illumination normalization methods. (a) Partition of face region for region-based illumination normalization methods. (b) Images processed by different illumination normalization methods.

5.4 Evaluation Results on Frontal Face Images

The three baseline face recognition algorithms (PCA, PCA+LDA and G PCA+LDA) are trained on the training set, and evaluated on the six frontal probe sets as described in Section 5.1. Before training and testing, all the images are preprocessed as described in Section 5.3, using the four illumination normalization methods or no illumination normalization respectively. Fig. 14 shows the performance of these algorithms on the frontal probe sets.



(a)

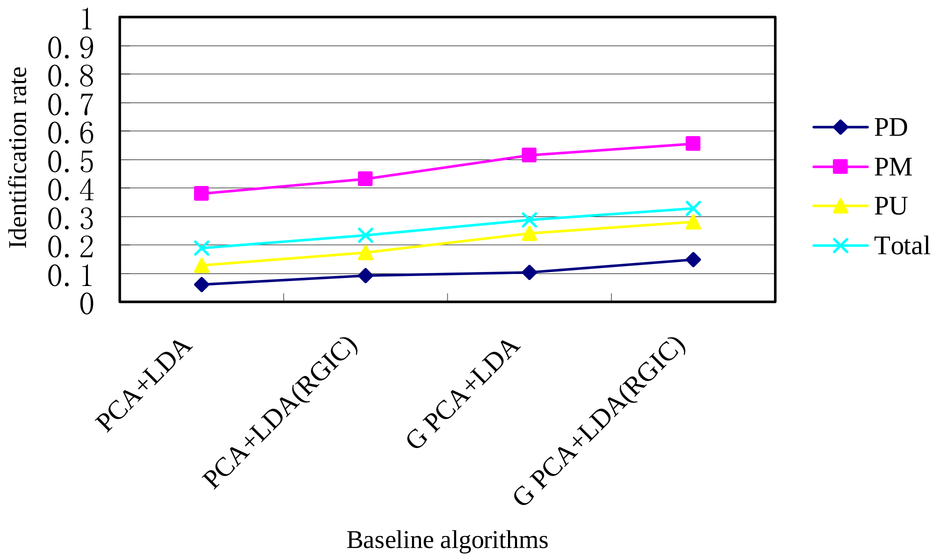
Probe sets / Algorithms	Accessory	Background	Distance	Expression	Lighting	Aging	Total
PCA	0.371	0.805	0.742	0.537	0.082	0.500	0.282
PCA+LDA	0.610	0.944	0.935	0.713	0.218	0.727	0.422
PCA+LDA(HE)	0.710	0.975	0.975	0.802	0.288	0.773	0.484
PCA+LDA(GIC)	0.670	0.964	0.949	0.780	0.230	0.788	0.454
PCA+LDA(RHE)	0.698	0.955	0.949	0.785	0.296	0.788	0.478
PCA+LDA(RGIC)	0.675	0.949	0.953	0.762	0.247	0.712	0.455
G PCA+LDA	0.828	0.980	1.00	0.906	0.448	0.985	0.574
G PCA+LDA(HE)	0.851	0.989	1.00	0.929	0.443	0.939	0.583
G PCA+LDA(GIC)	0.821	0.984	0.993	0.911	0.469	0.955	0.578
G PCA+LDA(RHE)	0.785	0.955	0.982	0.904	0.352	0.909	0.537
G PCA+LDA(RGIC)	0.827	0.984	0.996	0.916	0.490	0.955	0.586

(b)

Fig. 14. Identification performance of the three baseline algorithms on the six frontal probe sets and the union (Total) set of these sets.

5.5 Evaluation Results on Face Images under Different Poses

Two baseline face recognition algorithms (PCA+LDA and G PCA+LDA) are trained on the training set, and evaluated on the three pose probe sets as described in Section 5.1. Before training and testing, all the images are preprocessed as described in Section 5.3, using the RGIC illumination normalization method or no illumination normalization respectively. Fig. 15 shows the performance of these algorithms on the pose probe sets.



(a)

Algorithms \ Probe sets	PD	PM	PU	Total
PCA+LDA	0.061	0.380	0.128	0.190
PCA+LDA(RGIC)	0.092	0.432	0.174	0.233
G PCA+LDA	0.104	0.515	0.241	0.287
G PCA+LDA(RGIC)	0.149	0.556	0.28	0.328

(b)

Fig. 15 Identification performance of two baseline algorithms on the three pose probe sets and the union (Total) set of these sets.

6. OBTAINING THE CAS-PEAL-R1

To get a copy of the CAS-PEAL-R1 face database, please download the release agreement (http://www.jdl.ac.cn/peal/index.html), print and fill in the agreement appropriately, and fax it back to +86 10 8264 9298. Then we will contact you on how you can get a copy either by posting a CD package (some CD fee and postage will be charged though the database itself is free.) or downloading through the Internet.

7. CONCLUSION AND FUTURE WORK

This technical report has described the CAS-PEAL face database, a large-scale face images with different sources of variations. Currently, the CAS-PEAL face database contains 99,594 images of 1040 individuals (595 males and 445 females) with varying Pose, Expression, Accessory, and Lighting (PEAL). This face database is now partly made available (a subset named by CAS-PEAL-R1, contain 30,871 images of 1040 subjects) for research purpose only on a case-by-case basis.

ACKNOWLEGEMENTS

This research is partially sponsored by National Hi-Tech Program of China (No. 2001AA114010and No. 2001AA114190) and NSFC under contract No. 60332010. This work is also partially sponsored by ISVISION Technologies Co., Ltd.

REFERENCES

[1] M. H. Yang, D. Kriegman, and N. Ahuja, “Detecting Faces in Images: A Survey,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 24, No. 1, pp. 34-58, 2002.

[2] W. Zhao, R. Chellappa, A. Rosenfeld and P. J. Phillips, “Face Recognition: A Literature Survey,” Technical Report, CS-TR4167, University of Maryland, 2000. Revised 2002, CS-TR4167R.

[3] P. J. Phillips, H. Wechsler, and P. Rauss, “The FERET Database and Evaluation Procedure for Face-Recognition Algorithms,” Image and Vision Computing, Vol. 16, No. 5, pp. 295-306, 1998.

[4] P. J. Phillips, H. Moon, S. Rizvi, and P. Rauss, “The FERET Evaluation Methodology for Face-Recognition Algorithms,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, No. 10, pp. 1090-1104, 2000.

[5] T. Sim, S. Baker, and M. Bsat, “The CMU Pose, Illumination, and Expression Database,” IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.25, No. 12, pp. 1615-1618, 2003.

[6] A. R. Martinez and R. Benavente, “The AR Face Database,” Technical Report 24, Computer Vision Center(CVC) Technical Report, Barcelona, Spain, June 1998.

[7] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre, “XM2VTSDB: The Extended M2VTS Database,” In Second International Conference on Audio and Video-based Biometric Person Authentication, March 1999.

[8] E. B. Bailliere, S. Bengio, F. Bimbot, M. Hamouz, J. Kittler, J. Mariethoz, J. Matas, K. Messer, V. Popovici, F. Poree, B. Ruiz, and J. P. Thiran, “The BANCA Database and Evaluation Protocol,” In Audio- and Video-Based Biometric Person Authentication (AVBPA), pp. 625-638, 2003.

[9] P. J. Phillips, P. Grother, J. Ross, D. Blackburn, E. Tabassi, and M. Bone, “Face Recognition Vendor Test 2002: Evaluation Report,” March 2003.

[10] B. Hwang, H. Byun, M. Roh, and S. Lee, “Performance Evaluation of Face Recognition Algorithms on the Asian Face Database, KFDB,” In Audio- and Video-Based Biometric Person Authentication (AVBPA), pp. 557-565, 2003.

[11] P. J. Phillips and E. M. Newton, “Meta-Analysis of Face Recognition Algorithms,” In 5th IEEE Conf. on Automatic Face and Gesture Recognition, Washington, D.C., May 2002.

[12] J. R. Beveridge, K. She, B. A. Draper, and G. H. Givens, “A Nonparametric Statistical Comparison of Principal Component and Linear Discriminant Subspaces for Face Recognition,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 535 – 542, December 2001.

[13] A. M. Martinez and A. C. Kak, “PCA versus LDA,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 23, no. 2, pp. 228-233, 2001.

[14] M. Turk and A. Pentland, “Eigenfaces for Recognition,” Journal of Cognitive Neuroscience, vol. 3, pp. 72-86, 1991.

[15] D. L. Swets and J. Weng, “Using Discriminant Eigenfeatures for Image Retrieval,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 18, no. 8, pp. 831-836, 1996.

[16] W. Zhao, R. Chellappa and P. J. Phillips, “Subspace Linear Discriminant Analysis for Face Recognition,” Technical Report, CS-TR4009, University of Maryland, 1999

[17] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, “Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 711-720, 1997.

[18] N. Kruger, C. Malsburg, L. Wiskott, and J. M. Fellous, “Face Recognition by Elastic Bunch Graph Matching,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 775-779, 1997.

[19] C. Liu and H. Wechsler, “Gabor Feature Based Classification Using the Enhanced Fisher Linear Dsicriminant Model for Face Recognition,” IEEE Trans. Image Processing, vol. 11, no. 4, pp. 467-476, 2002.

[20] B. Gokberk, M. O. Irfanoglu, L. Akarun, and E. Alpaydin, “Optimal Gabor Kernel Location Selection for Face Recognition,” International Conference on Image Processing, pp. 677-680 2003.

[21] S. Shan, W. Gao, B. Cao, D. Zhao, “Illumination Normalization for Robust Face Recognition against Varying Lighting Conditions,” IEEE International Workshop on Analysis and Modeling of Faces and Gestures, pp. 157-164, Nice, France, 2003.