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

Wen Gao, *Senior Member, IEEE*, Bo Cao, Shiguang Shan, *Member, IEEE*, Xilin Chen, *Member, IEEE*, Delong Zhou, Xiaohua Zhang, and Debin Zhao

***Abstract*—In this paper, we describe the acquisition and contents of a large-scale Chinese face database: the CAS-PEAL face database. The goals of creating the CAS-PEAL face database include the following: 1) providing the worldwide researchers of face recognition with different sources of variations, particularly pose, expression, accessories, and lighting (PEAL), and exhaustive ground-truth information in one uniform database; 2) advancing the state-of-the-art face recognition technologies aiming at practical applications by using off-the-shelf imaging equipment and by designing normal face variations in the database; and 3) providing a large-scale face database of Mongolian. Currently, the CAS-PEAL face database contains 99 594 images of 1040 individuals (595 males and 445 females). A total of nine cameras are mounted horizontally on an arc arm to simultaneously capture images across different poses. Each subject is asked to look straight ahead, up, and down to obtain 27 images in three shots. Five facial expressions, six accessories, and 15 lighting changes are also included in the database. A selected subset of the database (CAS-PEAL-R1, containing 30 863 images of the 1040 subjects) is available to other researchers now. We discuss the evaluation protocol based on the CAS-PEAL-R1 database and present the performance of four algorithms as a baseline to do the following: 1) elementarily assess the difficulty of the database for face recognition algorithms; 2) preference evaluation results for researchers using the database; and 3) identify the strengths and weaknesses of the commonly used algorithms.**

***Index Terms*—Accessory, evaluation protocol, expression, face databases, face recognition, lighting, pose.**

Manuscript received June 29, 2005; revised January 9, 2006. This work was supported in part by the National Natural Science Foundation of China under Grants 60332010, 60772071, and 60533030 by the 100 Talents Program of the Chinese Academy of Sciences. The work of B. Cao was also supported in part by the Natural Science Foundation of China under Grant 60473043. This paper was recommended by Associate Editor M. Celenk.

W. Gao was with the Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080, China. He is now with the School of Electronics Engineering and Computer Science, Peking University, Beijing 100871, China (e-mail: wgao@jdl.ac.cn).

B. Cao is with the ICT–ISVISION Joint Research and Development Laboratory for Face Recognition, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080, China, and also with the Graduate School of the Chinese Academy of Sciences, Beijing 100049, China (e-mail: bcao@jdl.ac.cn).

S. Shan and X. Chen are with the ICT–ISVISION Joint Research and Development Laboratory for Face Recognition, Institute of Computing Technology and the Key Laboratory of Intelligent Information Processing, Chinese Academy of Sciences, Beijing 100080, China (e-mail: sgshan@jdl.ac.cn; xlchen@jdl.ac.cn).

D. Zhou was with the ICT–ISVISION Joint Research and Development Laboratory for Face Recognition, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080, China. He is now with the College of Information Engineering, Zhejiang University of Technology, Hangzhou 310027, China (e-mail: zdl@zjut.edu.cn).

X. Zhang was with the ICT–ISVISION Joint Research and Development Laboratory for Face Recognition, Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080, China. He is now with the Department of Electrical and Computer Engineering, University of Rochester, Rochester, NY 14627 USA (e-mail: xhzhang79@yahoo.com).

D. Zhao is with the Department of Computer Science, Harbin Institute of Technology, Harbin 15001, China, and also with the Institute of Computing Technology, Chinese Academy of Sciences, Beijing 100080, China (e-mail: dbzhao@jdl.ac.cn).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSMCA.2007.909557

I. Introduction

Automatic face recognition (AFR) has been studied for over 30 years [1]–[3]. Especially in recent years, it has become one of the most active research areas in pattern recognition, computer vision, and psychology due to the extensive public expectation of its wide potential applications in public security, financial security, entertainment, intelligent human–computer interaction, etc. In addition, much progress has been made in the past few years. However, AFR remains a research area far from maturity, and its applications are still limited in controllable environments. Therefore, it is becoming more and more significant to discover the bottleneck and the valuable future research topics by evaluating and comparing the potential AFR technologies exhaustively and objectively.

Aiming at these goals, large-scale and diverse face databases are obviously one of the basic requirements. Internationally, face recognition technology (FERET) [4], [5], face recognition vendor test (FRVT) [6], [7], and face recognition grand challenge (FRGC) [8] have pioneered both evaluation protocols and database construction. Furthermore, FERET has released its database that contains 14 051 face images of over 1000 subjects and has variations in expression, lighting, pose, and acquisition time. Despite its success in the evaluations of face recognition algorithms, the FERET database has limitations in the relatively simple and unsystematically controlled variations of face images for research purposes. FRGC has released its training and validation partitions. The training partition consists of two training sets: the large still training set (6388 controlled and 6388 uncontrolled still images from 222 subjects) and the 3-D training set (3-D scans, and controlled and uncontrolled still images from 943 subject sessions). The validation partition contains images from 466 subjects collected in 4007 subject sessions. Other publicly available face databases include the CMU PIE [9], AR [10], XM2VTSDB [11], ORL [12], UMIST [13], MIT [14], Yale [15], (Extended) Yale Face Database B [16], [17], BANCA [18], etc. Among them, both the CMU PIE and the (Extended) Yale Face Database B have well-controlled variations in pose and illumination. The CMU PIE contains

TABLE I
OVERVIEW OF THE RECODING CONDITIONS IN SOME FACE DATABASES

Database	#Subjects	Pose	Expression	Accessory	Lighting	Time	Background	Distance	Race
Yale B	10	C: 9	-	-	C: 64	-	-	-	WM
E-Yale B	28	C: 9	-	-	C: 64	-	-	-	WEM
CMU PIE	68	C: 13	C: 3	-	C: 43	-	-	-	WEM
AR	126	-	C: 4	C: 2	C: 4	S: 2	-	-	W
BANCA	208	-	Speech sequence	-	U	G: 12	U: 3	-	WM
XM2VTS	295	2 rotation sequences	Speech sequence	-	C: 2	S: 4	-	-	WEM
FRGC(Released part)	569	3D	U: 2	-	U	G: 1-22	U	U	WEM
CAS-PEAL-R1	1040	C: 21	C: 1-6	C: 1-6	C: 9-15	S: 1-2	C: 2-4	C: 1-2	M
FERET	1199	C: 9-20	U: 2	-	U: 1-2	G: 1-2	U: 1-3	-	WEM

For all the recording conditions: '-' means no variations or slight variations which are not specially designed.
For the Time condition: 'S' means other conditions, such as lighting, pose, camera, etc., were consistent across different sessions. 'G' means that these other conditions may be changed.
For the other recording conditions: 'C' means the variation was specially controlled. 'U' means significant uncontrolled variation.
The number after the characters, if any, denotes the number of variations under that recording condition.
For Race: 'W' means Caucasian. 'E' means Ethiopian. 'M' means Mongolian.

68 subjects, whereas the Yale Face Database B contains ten subjects, which may not satisfy the practical requirements for training and evaluating most face recognition algorithms.

To complement the existing face databases, we design and construct a large-scale Chinese face database—the CAS-PEAL face database which covers variations in pose, expression, accessory, lighting, backgrounds, etc. Currently, it contains 99 594 images of 1040 individuals (595 males and 445 females). A selected subset CAS-PEAL-R1, which contains 30 863 images of 1040 subjects, is now made available for other researchers. Table I gives a brief overview of these databases to help researchers choose the most appropriate one for their specific needs. Some older databases are not included in the table (for a complete reference, refer to [19]). It is obvious that the CAS-PEAL-R1 database has advantages both in the quantity of subjects and in a number of controlled variations of the recording conditions, which facilitate the training and evaluation of face recognition algorithms, particularly those statistical-based learning techniques. Furthermore, most of the current face databases mainly consist of Caucasian people, whereas the CAS-PEAL database consists of Mongolian people. Such difference makes it possible to study on the "cross-race" effect in face recognition algorithms [20]–[22].

This paper describes the design, collection, and categorization of the CAS-PEAL database in detail. In addition, we present an evaluation protocol to regulate the potential future evaluation on the CAS-PEAL-R1 face database, based on which we then evaluate the performance of several typical face recognition methods including the eigenface [principle components analysis (PCA)] [14], the fisherface [PCA+linear discriminant analysis (LDA)] [15], [23], [24], the Gabor-based PCA+LDA (G PCA+LDA) [25], [26], and the local Gabor binary-pattern histogram sequence (LGBPHS) [27] in combination with the different preprocessing methods. The evaluation results have assessed the difficulty of the database for face recognition algorithms on the basis of individual probe sets containing different variations. By analyzing their performance, some insights to the commonly used algorithms and preprocessing methods are obtained.

The remaining part of this paper is organized as follows. The setup of the photographic room is described in Section II. Then, the design of the CAS-PEAL face database is detailed

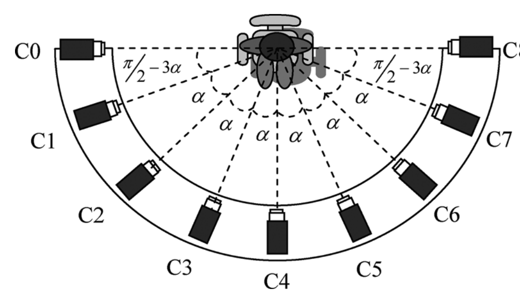


Fig. 1. Illustration of the camera configuration. Note that, in our face database, α is equal to 22.5 degree for the subjects #001 $\sim$ #101, while for other subjects, i.e., #102 $\sim$ #1042, α is equal to 15 degree.

in Section III. The publicly released CAS-PEAL-R1 and its accompanying evaluation protocol are described in Section IV. The evaluation results of four baseline algorithms on the CAS-PEAL-R1 database are presented in Section V. Finally, some conclusions are drawn in the last section with some further discussions.

II. PHOTOGRAPHIC ROOM

To capture face images with varying poses, expressions, accessories, and lighting conditions, a special photographic room with the dimension of 4. 0 $\times$ 5. 0 m and 3.5-m height is set in our laboratory, and the necessary apparatuses are configured in the room including a camera system, a lighting system, accessories, and various backgrounds. The details are described in the following sections.

A. Camera System

In our photographic room, a camera system consisting of nine digital cameras and a computer is elaborately designed. The cameras we used are Web-Eye PC631 with 640 $\times$ 480 pixels charge-coupled device (CCD). All nine cameras are mounted on a horizontal semicircular arm of 0.8-m radius and 1.1-m height. They all point to the center of the semicircular arm and are labeled as C0–C8 from the subject's right to left. The sketch map of the cameras' distribution on the semicircle arm is shown in Fig. 1.

All nine cameras are connected to a computer through USB interface. The computer is specially configured to support up

Fig. 2. Setup of the photographic room.

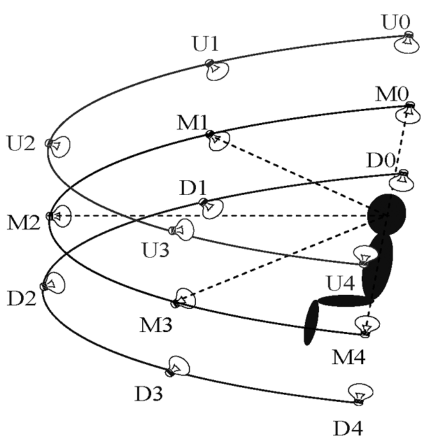


Fig. 3. Configuration of the lamps and their serial numbers. “U,” “M,” and “D” denote the rough positions of the lamps: “upper,” “middle,” and “down,” respectively.

to 12 USB ports. We developed software to control the nine cameras and capture the images in one shot. In each shot, the software can obtain nine images of the subject across the nine poses and store these images in the hard drive using a uniform naming convention.

Each subject is asked to sit down in a height-adjustable chair. Before photographs are taken, the chair is adjusted to keep the head of the subject at the center of the arm, and the subject is asked to look directly into the camera C4 that locates at the middle of the semicircular arm (as Fig. 1 shows). Fig. 2 shows the scene that one subject sat on the chair and was ready for the photography procedure.

B. Lighting System

To simulate the ambient illumination, two photographic sunlamps of high power covered with a ground glass are used to irradiate to the rough white ceiling, which can obtain more uniform lighting and mimic the normal indoor-lighting environment (overhead lighting sources).

To generate various lighting conditions needed, we set a lighting system in the photographic room using multiple lamps and lampshades. Fifteen fluorescent lamps are placed at the “lamp” positions, as shown in Fig. 3, to form varying directional lighting environments. In a spherical coordinate system whose origin is the center of the circle that coincides with the semicircular shelf (the x axes is the middle camera’s optical axis, and the y axes is horizontal), these positions are located at the crossover of five azimuths ($-90°$, $-45°$, $0°$, $+45°$, and $+90°$) and three elevations ($-45°$, $0°$, and $+45°$). By turning on/off each lamp while the aforementioned ambient lamps are kept on, different directional lighting conditions are simulated. A switch matrix is exploited to control the on/off conditions of these lamps. It should be noted that the flash systems like the CMU PIE [9] or YALE Face Database B [16] are not exploited in our system. Therefore, the illumination variations are not as strictly controlled as those in the PIE or Yale. However, these illumination variations are more natural and complicated.

TABLE II
ALL POSSIBLE SOURCES OF VARIATIONS COVERED IN THE CAS-PEAL FACE DATABASE

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

C. Accessories: Glasses and Hats

In the tasks of face detection, landmark localization, and face recognition, wearing accessories such as glasses and hats may cause great difficulty because they sometimes result in lighting change or occlusion or both. However, it is hardly evitable in the practical applications such as video surveillance. In the existing face databases, the accessory variations are not adequate. Therefore, we have carefully used several types of glasses and hats as accessories to further increase the diversity of the CAS-PEAL database. The glasses we used include dark frame glasses, glasses without frame, and sunglasses. There are also several hats with brims of different sizes and shapes. In the image collection, some of the subjects are asked to wear these accessories.

Another purpose to evaluate face recognition systems with some heads wearing different hats is to emphasize the variability of hairstyles. Typically, the hairstyle of a specific subject is constant in a face database, which was captured in a single session and thus may be used as discriminating features, whereas it is changeable in daily life.

D. Backgrounds

The background variations, in theory, may not influence the performance of face recognition algorithms provided that the face region is correctly segmented from the background. However, in real-world applications, many cameras are working under the mode of automatic white balance or automatic intensity gain, which may change the face appearance evidently under different imaging conditions, particularly for those consumer video cameras. 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) blankets are used.

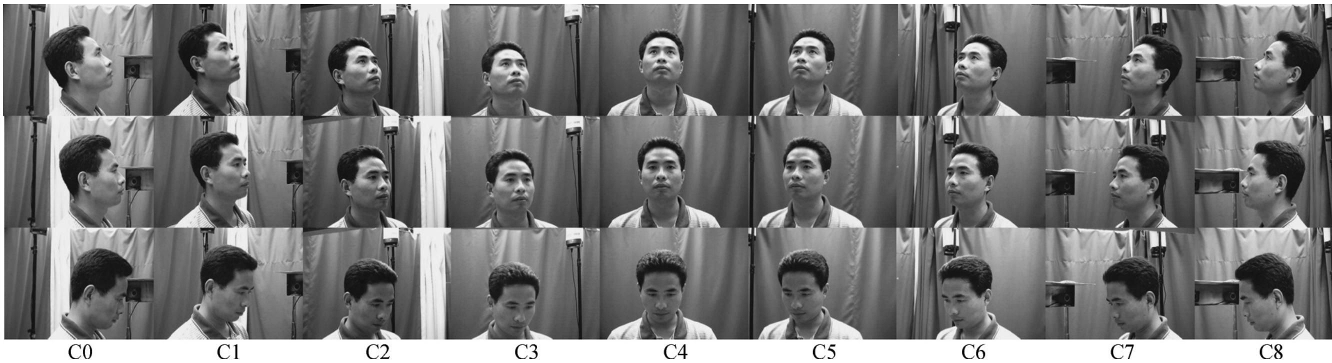


Fig. 4. The 27 images of one subject under pose variation in the CAS-PEAL database. The nine cameras (C0–C8) are mounted on the horizontal semicircular arm, (see Fig. 1 for the camera locations). The subject was asked to look upward, look right into the camera C4, and look downward.

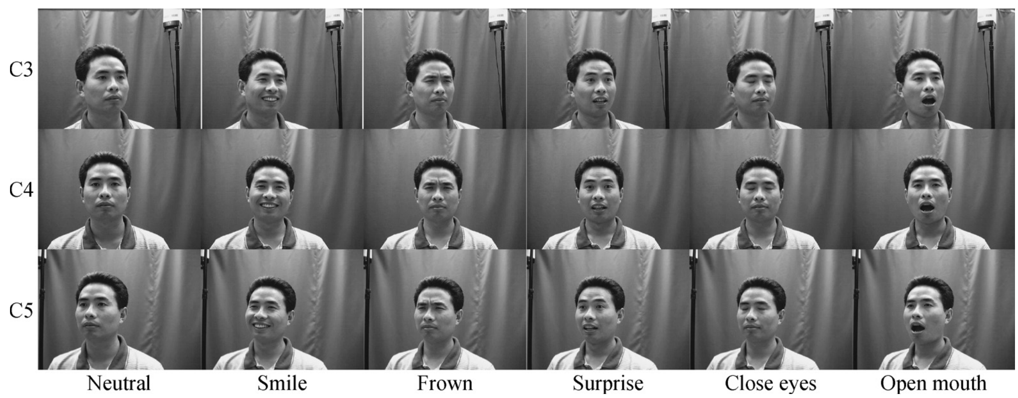


Fig. 5. Example images of one subject with six expressions across three poses (from cameras C3, C4, and C5).

III. Design of the CAS-PEAL Database

By utilizing the devices described in Section II, seven variations are applied to construct the CAS-PEAL face database: pure-pose, expression, lighting, accessory, background, time, and distance variations. Due to the fact that nine cameras from different directions are used to capture each subject simultaneously, all the variations are automatically combined with nine pose (viewpoint) changes. Table II lists all the possible sources of variations. For some subjects in the database, not all the variations are captured. However, any subject is captured under at least two kinds of these variations. The following sections describe each of the variations and demonstrate some example face images.

A. Pure-Pose Variation

To capture images with varying poses, the subject is asked to look upward (about 30°), look right into the camera C4 (the middle one), and look downward (about 30°). In each facing direction, nine images are obtained from the nine cameras in one shot. Thus, a total of 27 images of the subject will be obtained. Fig. 4 shows the 27 images of one subject.

B. Expression Variation

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

C. Lighting Variation

Using the lighting system described in Section II-B, we capture the images of a number of subjects under 15 different illumination conditions. Example images of one subject under these conditions are shown in Fig. 6. Note that, in all cases, the ambient lighting lamps are turned on.

D. Accessory Variation

For those subjects who are willing to perform this session, the prepared accessories, three hats and three pairs of glasses, are adorned one by one. Fig. 7 shows the example images of one subject recorded by the camera C4.

E. Background Variation

As mentioned in Section II-D, the background is changed by using different unicolor blankets. Example images under five different backgrounds are shown in Fig. 8. It can be found that the exposures of these images are highly dependent on the backgrounds.

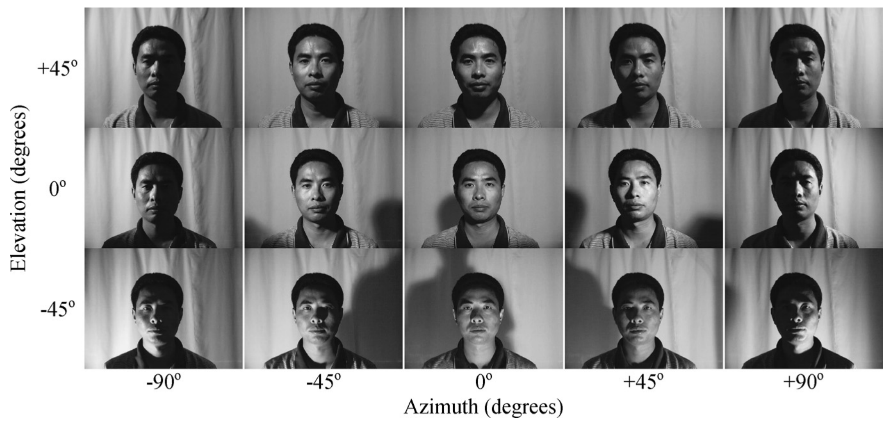


Fig. 6. Example images of one subject illuminated by fluorescent light source located at different azimuth and elevation coordinates from camera C4.

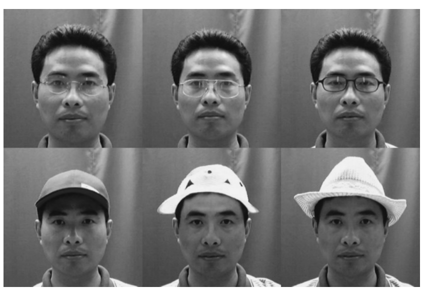


Fig. 7. Example images (cropped) of one subject with six different accessories.

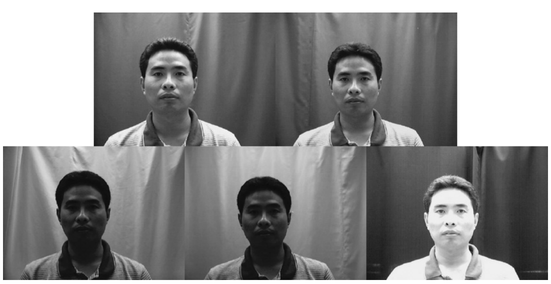


Fig. 8. Example images of one subject with different backgrounds.

F. Time Difference

In FERET, FRVT, and other face recognition competitions, time difference is another important factor decreasing the accuracy. In most face databases, images of one subject captured in different times are insufficient or absent because the subjects are hard to be traced. In CAS-PEAL database, 66 subjects have been captured in two sessions half a year apart. Fig. 9 shows six images captured in the two sessions. We are further extending this part of the database.

G. Different Distance

In real-world applications, the distance between the subject and the camera is subject to changing, which may not be simply treated as a scale problem. To make possible the evaluation of this problem's effect on face recognition, we collect some images at different distances for some subjects. In our system, the focal length of the cameras is equal to 36 mm. Three distances are used: 0.8, 1.0, and 1.2 m. Fig. 10 shows three images of one subject at these distances from the camera.

Fig. 9. Example images captured with time differences. The images in the bottom row were captured half a year after those in the top row.

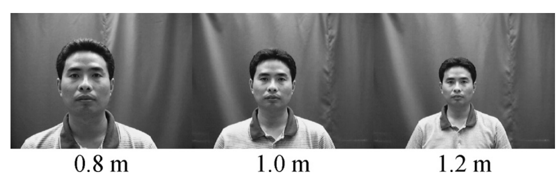


Fig. 10. Example images at different distances from the camera.

IV. Publicly Released CAS-PEAL-R1 and Corresponding Evaluation Protocol

A subset of the CAS-PEAL face database, named CAS-PEAL-R1, has been made publicly available to researchers working on AFR. This section describes the CAS-PEAL-R1 as well as its accompanying evaluation protocol.

A. Publicly Released CAS-PEAL-R1 Face Database

Contents of the CAS-PEAL-R1: CAS-PEAL-R1 is a subset of the entire CAS-PEAL face database. It contains 30 863 images of 1040 subjects. These images belong to two main subsets: the frontal and nonfrontal subsets.

1) In the frontal subset, all images are captured by the camera C4 (see Fig. 1), with the subjects looking right into this camera. Among them, 377 subjects have images with six different expressions. Some 438 subjects have images wearing six different accessories. Some 233 subjects have

TABLE III
CONTENTS OF CAS-PEAL-R1

Subset		# Variations	# Subjects	# Images
Frontal	Normal	1	1,040	1,040
	Expression	5 [a]	377	1,884
	Accessory	6	438	2,616
	Lighting	>= 9	233	2,450
	Time	1	66	66
	Background	2-4	297	651
	Distance	1-2	296	324
	Total:			9,031
Non-frontal		21 (3*7)	1,040	21,832 [b]
Total:				30,863

[a]Neutral expression is not counted in.
[b]With 8 images corrupted.

TABLE IV

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

TABLE V

FY	FM	FO	MY	MM	MO
Female, Young	Female, Middle-aged	Female, Old	Male, Young	Male, Middle-aged	Male, Old

Young: 18 - 44 years old; Middle-aged: 45 - 59 years old;
Old: 60 - 74 years old

TABLE VI

E	F	L	U	M	D
Ambient lighting	Fluorescent lighting	Incandescent lighting	Elevation: +45°	Elevation: 0°	Elevation: -45°

TABLE VII

U	M	D
Looking up	Looking into camera C4	Looking down

TABLE VIII

N	L	F	S	C	O
Neutral	Laughing	Frowning	Surprising	Eyes closed	Mouth open

TABLE IX

0	1	2	3	4	5	6
None	Hat 1	Hat 2	Hat 3	Glasses 1	Glasses 2	Glasses 3

TABLE X

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

images under at least nine lighting changes. The 297 subjects have images against two to four different backgrounds. The 296 subjects have images with different distances from the camera. Furthermore, 66 subjects have images recorded in two sessions at a six-month interval.
2) In the nonfrontal subset, the images of 1040 subjects across 21 different poses (subset of those described in Section III-A) without any other variations are included.

Table III summarizes the contents of CAS-PEAL-R1.

Image Naming Convention: In the CAS-PEAL face database, the filename of each image encodes the majority of the ground-truth information of that image. Its format is described in Table IV. It consists of 12 fields of 46 characters long in total. The fields are separated by underline marks as shown above. In these fields, x's and n's represent the letter and digit sequences, respectively, which vary with the properties of each image. The meaning of each field, letter sequence, and digit sequence is described in turn as follows.

1) Gender and age field. Its two-character type sequence is defined in Table V.
2) ID field. Its six-digit sequence indicates the serial number of the subject in the image, ranging from 000001 to 001042 (000833 and 000834 are absent.).
3) Lighting-variation field. The initial character "I" represents illumination variation. The first "x" (E, F, L) indicates the kind of lighting source. The second "x" (U, M, D) indicates the elevation of the lighting source. The "±nn" indicates the azimuth of the lighting source. See Table VI.
4) Pose field. The initial character "P" represents pose variation. The "x" (U, M, D) indicates the subject's pose (see Table VII). The "±nn" indicates the azimuth of the camera by which the image is obtained. Please refer to Fig. 1 for the configuration of the cameras.
5) Expression field. The initial character "E" represents expression variation. The following "x" can be "N," "L," "F," "S," "C," or "O." Its meaning is as shown in Table VIII.
6) Accessory field. The initial character "A" represents accessory variation. The following "n" can be a value ranging from 0 to 6 (see Table IX).
7) Distance field. The initial character "D" represents distance variation. The following "n" has a value ranging from 0 to 2, indicating different distances from the subject to the camera C4.
8) Time field. The initial character "T" indicates time variation. The following "n" has value denoting different sessions (see Table X).
9) Background field. The initial character "B" represents background variation. Table XI gives the value for "x".
10) This field is reserved for future use.
11) Privacy field. Only images whose ID is less than 100 and with an "R1" label in this field will be published or released in technical reports and papers in the face recognition research area only.
12) Resolution field. The initial character "S" represents resolution. The "n" has two values: 0 and 1, denoting two different resolutions of the image (see Table XII).

Because the filename of each image describes the property of the subject in that image, the images in the database can be retrieved and reorganized easily to meet any specific requirement. In addition, the ground-truth eye locations of all the images are provided in a text file (named FaceFP_2.txt).

Image Format: The original 30 863 RGB color images of size 640 × 480 in CAS-PEAL-R1 require about 26.6 GB storage space. To facilitate the release, all the images were

TABLE XI

B	R	D	Y	W
Blue	Red	Dark	Yellow	White

TABLE XII

0	1
640*480	320*240

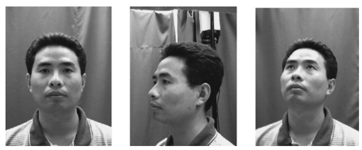


Fig. 11. Several examples of the cropped face images in CAS-PEAL-R1.

converted to grayscale images and cropped to size 360×480 excluding most of the background. The cropped images are stored as TIFF files with lossless Lempel–Ziv–Welch compression. Several cropped images are shown in Fig. 11.

B. Evaluation Protocol

Given a face database, there are many possible methods to evaluate a specific AFR method. To facilitate the comparisons among the results of different methods, we have specified a standard evaluation protocol accompanying the database, and expect the potential users of the database to evaluate their methods according to the protocol. In the following part of this section, we describe the proposed evaluation protocol by presenting the definition, design, and some underlying design philosophies of the data sets as well as the evaluation methods.

1) Data Sets for Evaluation: In the proposed evaluation protocol, three kinds of data sets are composed from the CAS-PEAL-R1 database: one training set, one gallery set, and several probe sets. Their definitions and descriptions are as follows.

a) Training set: A training set is a collection of images used to build a recognition model or to tune the parameters of the model or both. We construct a training set containing 1200 images of 300 subjects, which are randomly selected from the 1040 subjects in the CAS-PEAL-R1 database, with each subject contributing four images randomly selected from the frontal subset of the CAS-PEAL-R1 database.

b) Gallery set: A gallery set is a collection of images of known individuals against which a probe image is matched. In the evaluation protocol, we formed a gallery set containing 1040 images of the 1040 subjects, with each subject having one image under a normal condition. The gallery set consists of all the normal images mentioned in Table III.

c) Probe sets: A probe set is a collection of probe images of unknown individuals that need to be recognized. In the evaluation, nine probe sets are composed from the CAS-PEAL-R1 database, and each probe set contains images restricted to one main variation, as described in Section III. These partitions can be used to identify the strengths and weaknesses of a specific algorithm and to address the performance variations associated with the changes in the probe sets. Among them, six probe sets correspond to the six subsets in the frontal subset: expression, lighting, accessory, background, distance, and time, as described in Table III. The other three probe sets correspond to the images of subjects in the nonfrontal subset: looking upward, looking right into the camera C4 (the middle one), and looking downward. All the images that appear in the training set are excluded from these probe sets.

The data sets used in the evaluation are summarized in Table XIII.

2) Evaluation Methods: Based on the aforementioned data sets, one may set up many meaningful evaluation methods for a specific face recognition algorithm. Basically, we believe that how an evaluation method is configured depends on the following three criteria.

a) Is the training set for constructing and tuning the face model restricted or open?: For most statistics- or learning-based face recognition algorithms, their performance on the designated testing sets heavily depends on the composition of the training set, such as the size (the number of subjects and the number of images per subject) of the training set, the variations (lighting, pose, expression, etc.) contained in it, and so on. Generally speaking, the training images with similar attributes to those in the testing set would lead to superior performance. Therefore, in most literature works, the performance comparison of different algorithms is conducted based on the same training set to achieve justice. On the other hand, the proposed training set may not be appropriate for a specific face recognition method or not be adequate to fully utilize the learning capability of the method; thus, the evaluation results are still biased. Considering these aspects, we have defined two training modes to construct face models: one is the restricted mode using and only using the TS training set specified in Section IV-B1; the other is the open mode with no restriction on the training set except that no testing images are included. Hereinafter, these two modes are denoted as “R” and “O,” respectively.

b) Does the face recognition algorithm work in a fully automatic mode or a partially automatic one?: A fully automatic mode means that the presented face recognition algorithm completes face detection, facial landmark localization, and identification without any interaction. On the other hand, in a partially automatic mode, the precise facial landmark locations are provided to the algorithm beforehand. In most cases, the coordinates of the two eye centers are given. The partially automatic mode has been exploited by the FERET and most of the academic publications so far since it facilitates a “clean” comparison for researchers. However, perfect automatic eye localization is impossible, and many face recognition algorithms would degrade abruptly with the increase of the eye location error [28]. Therefore, it is necessary to compare different algorithms in the fully automatic mode to investigate its practicability in real-world applications. Hereinafter, these two modes are denoted as “F” and “P,” respectively.

c) What task does the algorithm complete: identification or verification?: In practical applications, there are typically three different tasks: identification, verification, and watch list [7]. While identification and verification are the special cases of the watch-list task, they are still the most fundamental and different tasks. For an identification task, one needs to determine the identity of the given face image by matching

TABLE XIII
DATA SETS USED IN THE EVALUATION PROTOCOL

Datasets	Training set	Gallery set	Probe sets (frontal)						Probe sets (non-frontal)		
			Expression	Accessory	Lighting	Time	Background	Distance	Looking upwards	Looking right into the camera C4	Looking downwards
Abbr.	TS	GS	PE	PA	PL	PT	PB	PS	PU	PM	PD
#Images	1,200	1,040	1,570	2,285	2,243	66	553	275	4,998	4,993	4,998
#Subjects	300	1,040	377	438	233	66	248	247	1,040	1,039	1,040

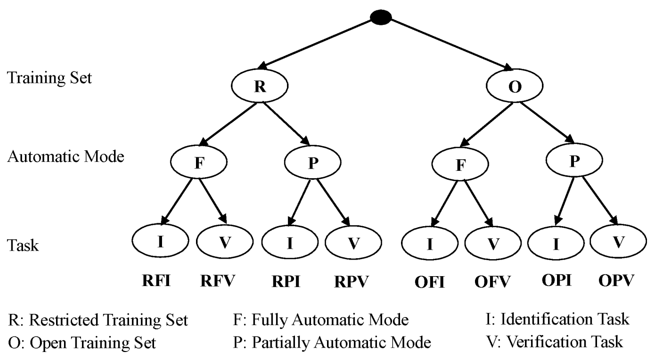


Fig. 12. Eight distinct evaluation methods.

it against all the prototype images in the gallery set, whereas for a verification task, one needs to tell whether the claimed identity is that of the input face image by matching it against the prototypes of the claimed identity. For the identification task in our evaluation protocol, the 1040 images in the GS set are used as the prototypes to enroll the 1040 subjects, and the images in a probe set (PE, PA, PL, PT, PB, PS, PU, PM, or PD) should be matched against those in the GS set to get the recognition performance scores (cumulative matching curve) for that probe set. For the verification task in our evaluation protocol, each image in a probe set is matched against all the images in the GS set. By accumulating the false positives and false negatives for a specific threshold, the false-reject and false-accept rates can be estimated for that probe set. By moving the threshold over all possible values, the receiver operating characteristic curve for each probe set will be generated.

According to the earlier criteria, in this paper, we explicitly define eight distinct evaluation methods, which are shown in Fig. 12 as a binary tree. The eight leaves represent the eight evaluation methods, which are denoted by RFI, RFV, RPI, RPV, OFI, OFV, OPI, OPV, respectively. The details of each method can be inferred easily from its path from the boot node to the leaf node in the binary tree. For example, method RFI stands for the fully automatic identification using the restricted training set, i.e., TS. By defining the eight methods, we expect that all the potential users of the database can adopt one or more appropriate methods to evaluate their algorithms according to the application or the characteristics of the algorithms, which will also facilitate the comparisons of various algorithms developed by worldwide researchers.

V. 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 as follows: 1) to elementarily assess the difficulty of the database for face recognition algorithms; 2) to provide reference evaluation results for researchers using the database; and 3) to identify the strengths and weaknesses of the commonly used algorithms.

Four baseline algorithms are briefly introduced. The preprocessing process of the face images has been demonstrated that it can effectively affect the performance of face recognition algorithms; thus, the details of the preprocessing process are also provided. Finally, the evaluation results are presented using the RPI evaluation method.

A. Baseline Face Recognition Algorithms

The four baseline algorithms evaluated are the PCA, also known as the eigenfaces, the combined PCA and LDA (PCA+LDA, a variant of fisherfaces), the PCA+LDA algorithm based on Gabor features (G-PCA+LDA), and the LGBPHS. The PCA- and the PCA+LDA-based face recognition algorithms are both fundamental and well studied [14], [15], [23], [24], [29]. Recently, 2-D Gabor wavelets and local binary pattern (LBP) [30] are extensively used for local feature representation and extraction and demonstrate their success in face recognition [25], [27], [31]–[33]. Therefore, the PCA+LDA algorithm based on Gabor features and the LGBPHS are also used as baseline algorithms to reflect this trend. Furthermore, in contrast to the other three baseline algorithms, the LGBPHS is not a statistical learning method. As a result, it will not be tuned to a specific training set and not suffer from the generalizability problem.

1) Principal Component Analysis (PCA): PCA is commonly used for dimensionality reduction in face recognition. PCA chooses projection directions $\mathbf{W}_{\text{opt}}$ that maximize the total scatter across all images of all faces in the training set.

For a training set that contains N sample images $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\} \in R^n$, the total scatter matrix $\mathbf{S}_T$ is defined as follows:

$$\mathbf{S}_T = \frac{1}{N}\sum_{k=1}^{N}(\mathbf{x}_k - \boldsymbol{\mu})(\mathbf{x}_k - \boldsymbol{\mu})^T$$

where $\boldsymbol{\mu}$ is the mean vector of all sample images in the training set.

Then, the projection matrix $\mathbf{W}_{\text{opt}}$ can be chosen as follows:

$$\mathbf{W}_{\text{opt}} = \arg\max_{\mathbf{W}} \left|\mathbf{W}^T\mathbf{S}_T\mathbf{W}\right| = [\,\mathbf{w}_1 \quad \mathbf{w}_2 \quad \ldots \quad \mathbf{w}_m\,]$$

where $\{\mathbf{w}_i | i = 1, 2, \ldots, m\}$ is the set of n-dimensional eigenvectors of $\mathbf{S}_T$ corresponding to the m largest eigenvalues. In most circumstances, m can be chosen far less than n without significantly decreasing the recognition rates.

2) Combined PCA and LDA (PCA+LDA): 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 $\mathbf{W}_{\text{fld}}$ in such a way that the ratio of the determinant of the between-class scatter matrix $\mathbf{S}_{\text{b}}$ of the projected samples and the within-class scatter matrix $\mathbf{S}_{\text{w}}$ of the projected samples is maximized.

$\mathbf{W}_{\text{fld}}$ can be calculated by solving the generalized eigenvalue problem

$$\mathbf{S}_{\text{b}}\mathbf{W} = \mathbf{S}_{\text{w}}\mathbf{W}\mathbf{\Lambda}.$$

Typically, in face recognition application, the number N of training images is much smaller than the dimension n. In this case, $\mathbf{S}_{\text{w}}$ 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 $\mathbf{S}_{\text{w}}$ will be nonsingular, and LDA can be used to find the projection matrix $\mathbf{W}_{\text{fld}}$.

3) PCA+LDA Algorithm Based on Gabor Features (G-PCA+LDA): Instead of using the original grayscale image as the input in the previous two algorithms, the input in this algorithm is the Gabor wavelet transformed image from the original one. 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 [31].

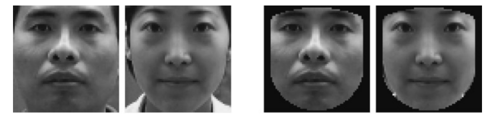
A family of Gabor wavelets (kernels and filters) can be defined as follows:

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

where $k_{u,v} = k_v e^{i\phi_u}$; $k_v = k_{\max}/f^v$ gives the frequency; $\phi_u = u\pi/8$, where $\phi_u \in [0,\pi)$, gives the orientation; $\mathbf{z} = (x,y)$; and $e^{i\vec{k}_{u,v}\mathbf{z}}$ is the oscillatory wave function, whose real and imaginary parts are the cosine and sinusoid functions, 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 (1). The Gabor wavelet transform of an image is the collection of the coefficients of all the pixels. To reduce the dimensionality, the convolution coefficients are sampled and concatenated to form the original features of the PCA+LDA algorithm described previously. These concatenated coefficients are also called the augmented Gabor feature vector in [25]. In the experiments, the size of the normalized faces is 64×64, and the coefficients are sampled every four pixels in both row and column; therefore, 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.



(a) (b)

Fig. 13. Example normalized face images in steps 1 and 2. (a) Geometrically normalized face images. (b) Masked face images.

4) Local Gabor Binary Pattern Histogram Sequence (LGBPHS): LGBPHS is a representation approach based on multiresolution spatial histogram combining local feature distribution with the spatial information. In addition, the Gabor wavelet transforms of the original image are used as the features, followed by the LBP operator to form the LGBP maps. Then, an image is modeled as a "histogram sequence" by concatenating the histograms of all the local nonoverlapping regions of the LGBP maps.

In face recognition, histogram intersection is used to measure the similarity of the LGBPHSs of two face images, and the nearest neighborhood is exploited for final classification. In contrast to the previous three algorithms, the modeling procedure of LGBPHS does not involve any learning process, i.e., it is non-learning method and need no training set.

B. 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 to remove irrelevant information for the purpose of face recognition, such as the background and the hair of a subject. 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 the geometric normalization step, each face image is scaled and rotated so that the eyes are positioned in a horizontal line, and the distance between them equals a predefined length. Then, the face image is cropped to include only the face region with little hair and background as Fig. 13(a) shows (the size of the cropped face image is 64×64). In the masking step, a predefined mask is put on each cropped face image to further reduce the effect of different hairstyles and backgrounds which are not the intrinsic characteristics as Fig. 13(b) shows. Typically, the hairstyle of a specific subject and the background are constant in a face database; thus, better performance can be obtained with larger face regions.

In the illumination normalization step, four illumination normalization methods are evaluated: histogram equalization (HE), gamma intensity correction (GIC), region-based HE (RHE), and region-based GIC (RGIC) [34].

1) Gamma Intensity Correction (GIC): The GIC method is to correct the overall brightness of the face images to a predefined "canonical" face image. It is formulated as follows.

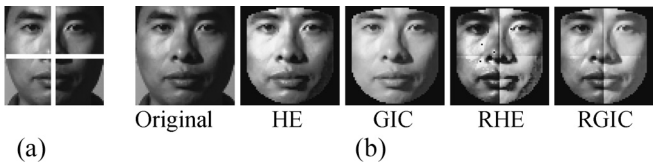


Fig. 14. Partition of face region and example images processed by different illumination normalization methods. (a) Partition of face region for region-based illumination normalization methods. (b) Images processed by different illumination normalization methods.

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

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

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

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

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

From (2) and (3), intuitively, the GIC is expected to make the overall brightness of the input images best fit that of the predefined normal face image. 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 image I_0. Fig. 14(b) shows the effect of GIC on an example face image.

2) Region-Based HE and GIC (RHE and RGIC): Both HE and GIC are global transforms over the whole image area. Therefore, they are likely to fail when side lighting exists. Since the possible side lighting mainly causes the nonsymmetry between the left and right parts of the face, as well as the intensity variance between the top and bottom regions, we partition the face into four regions according to the given eye centers, as shown in Fig. 14(a). Then, HE or GIC is performed in these predefined face regions to better alleviate the highlight, shading, and shadow effect caused by the unequal illumination. Fig. 14(b) shows the effect of RHE and RGIC on an example face image.

C. Evaluation Results on Frontal Face Images

The four baseline face recognition algorithms (PCA, PCA+LDA, G-PCA+LDA, and LGBPHS) are evaluated on the six frontal probe sets according to the RPI method described in Section IV-B, except that LGBPHS need not be trained on the training set. Before training and testing, all the images

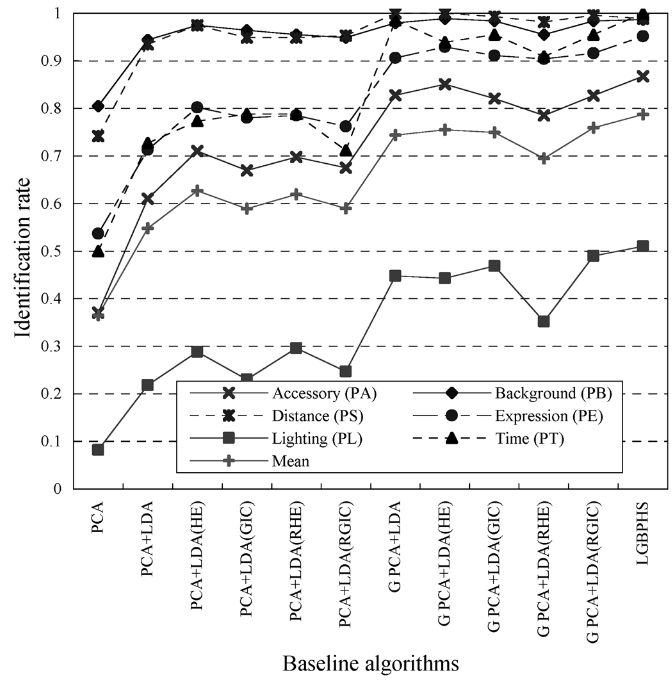


Fig. 15. Identification performance of the four baseline algorithms on the six frontal probe sets and the union (Mean) set of these sets. The identification rate of each algorithm on a probe set varies with the dimensionalities of the PCA and LDA subspace, and the best result is presented. The result on the "Mean" set is the weighted summation of the results of the algorithm.

are preprocessed, as described in Section V-B, using the four illumination normalization methods or no illumination normalization, respectively. Fig. 15 shows the performance of these algorithms on the frontal probe sets.

As it can be seen from Fig. 15, using Gabor features of the original images has evidently improved the performance of the algorithms based on PCA+LDA. On average, the "Mean" identification rate of the algorithms using Gabor features is 14.6% higher than that of the algorithms directly using grayscale images. Furthermore, it is clear that, among the variations (excluding pose variation), the lighting variation is still the most challenging problem in face recognition, and the accessory variation is the second one. In the CAS-PEAL database, different types of glasses and caps are used to form the accessory variation. We believe that accessories change the face appearance in two aspects: occlusion and shadow. For glasses, its frame and the light reflection on the eyeglass can both be considered as a kind of occlusion. For caps, its shadow plays a dominant role in changing the face appearance (as can be seen in the second-row images of Fig. 7), particularly when a tight mask is used, as shown in Fig. 13(b).

In recent years, some important contributions have been made to tackle the face recognition problems under varying illumination conditions [16], [17], [35]–[38], expression variations [31], [39]–[43], and face occlusions [40], [44]–[46]. Improved results were obtained using these methods under certain assumptions and tested on one of the databases mentioned in Section I. However, further efforts for a practical face recognition system must still be made to combine some of these interacted methods in a positive way and to test them on a large, diverse, and complicated database.

It should be noted that, in the experiment, time difference has not decreased the identification rate in such a degree as in FERET evaluation [5] and FRVT2000 [6], especially when Gabor features are used. The images in the Duplicate I and Duplicate II probe sets of FERET evaluation, which are generally emphasized as images taken on different days, comprise compound variations: time, pose, lighting, etc. In the FERET test, these variations are coupled, and it is hard to say which factor has a dominant impact on the identification rate. In the CAS-PEAL face database, the capture environment was strictly consistent when the images of the same subject were captured in different sessions, even half a year apart. As a result, only time difference exists in these images, which is not significant so far, referring to the identification rate. Therefore, for many applications, uncontrollable lighting and pose variations are still the real challenges. However, time difference becomes more significant when the elapsed time between the gallery and probe images gets much longer. According to FRVT2002 [7], identification performance decreases at 5% points per year.

The impact of different illumination normalization methods on identification performance is tricky. Although the RGIC method applied to the G-PCA+LDA algorithm achieves the best "Mean" identification rate among all the G-PCA+LDA algorithms, its superiority is not maintained across all the probe sets. Moreover, the best choice of normalization method varies with each baseline face recognition algorithm. Therefore, before we can conclude the superiority of a specific face-image preprocessing method, its generality should be tested against different recognition algorithms and different image variations, or we can select the most suitable one according to the recognition algorithm adopted and the problem to be solved.

D. Evaluation Results on Nonfrontal Face Images

Three baseline face recognition algorithms (PCA+LDA, G-PCA+LDA, and LGBPHS) are evaluated on the three nonfrontal probe sets based on the RPI method described in Section IV-B, except that LGBPHS need not be trained on the training set. Before training and testing, all the images are preprocessed, as described in Section V-B, using the RGIC illumination normalization method or no illumination normalization, respectively. Fig. 16 shows the performance of these algorithms on the nonfrontal probe sets.

From the identification rate curves, it can be observed that the probe set PD is the most difficult one. These results coincide with the observation that when a person lowers his head, more details of his face are occluded or distorted than under other pose variations. Furthermore, Gabor features consistently prove their superiority in the identification rate over the original grayscale features as they do for the frontal face images. On average, the "Mean" identification rate improves by 9.6%, comparing the PCA+LDA with the G-PCA+LDA algorithms. The algorithm based on LGBPHS dramatically outperforms the other two algorithms, especially on the probe set PD. However, even the identification rates of the best performing baseline algorithm are very low. It should be noted that face recognition algorithms relatively insensitive to pose variations have been

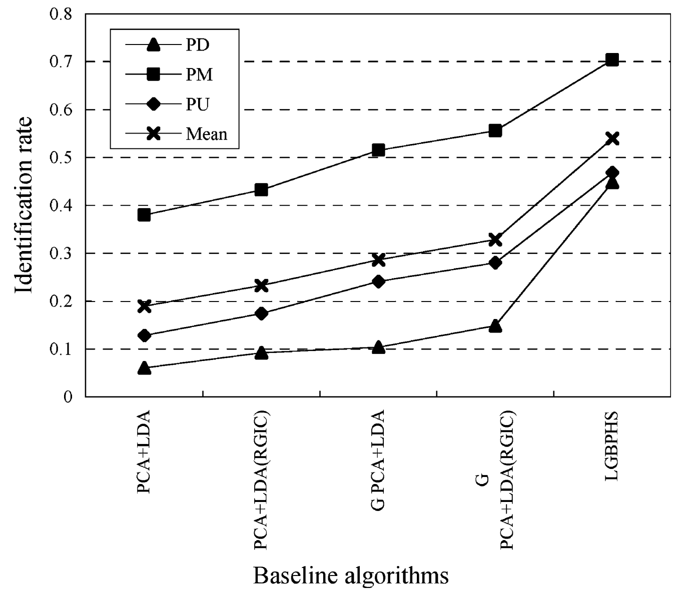


Fig. 16. Identification performance of the three baseline algorithms on the three nonfrontal probe sets and the union (Mean) set of these sets. The identification rate of each algorithm on a probe set varies with the dimensionalities of the PCA and LDA subspace. For each probe set, the best result is presented, and the result on the "Mean" set is the weighted summation of these results.

studied extensively in recent years [16], [31], [37], [42], [47], [48] and that the integration of 3D face model techniques to the existing face recognition algorithms has also been investigated [49], [50]. Better results can be expected using these algorithms. In practical face recognition systems, however, pose variation is still a challenging problem as FRVT2002 indicated.

VI. Obtaining the CAS-PEAL-R1 Database

The CAS-PEAL-R1 face database has been distributed so far to more than 150 research groups. The information on how to obtain a copy of the database can be found on the project Web site (http://www.jdl.ac.cn/peal/index.html).

VII. Conclusion

In this paper, we describe the photographic room design and setup and the contents of the CAS-PEAL face database. We also present detailed descriptions of the released CAS-PEAL-R1 database, including the contents, image naming convention, and image format of the database. The main characteristics of the CAS-PEAL face database lie in three aspects: 1) the large-scale face images, consisting of 99 594 images of 1040 subjects; 2) the diversity of the variations, including pose, expression, accessory, lighting, and the combined variations; and 3) the detailed ground-truth information and well-organized structure. Furthermore, the database enriches the existing face databases by providing images of Mongolian.

As an important complement to the database, we have proposed the evaluation protocol and provided evaluation results of four baseline algorithms on the database. From these results, the difficulty of the database and the strengths and weaknesses of the commonly used algorithms can be inferred. The partition method of the data sets used in the evaluation is also included in the distribution of the CAS-PEAL-R1 database and can be used as standard training and testing sets.

ACKNOWLEDGMENT

The authors would like to thank the Associate Editor and the anonymous reviewers for their insightful comments. The authors would also like to thank Dr. J. Yang from Carnegie Mellon University for helping revise and improve the manuscript. W. Zhang provided the evaluation results of the face recognition method based on LGBPHS.

REFERENCES

[1] M.-H. Yang, D. J. Kriegman, and N. Ahuja, "Detecting faces in images: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 1, pp. 34–58, Jan. 2002.

[2] W. Zhao, R. Chellappa, J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Comput. Surv.*, vol. 35, no. 4, pp. 399–458, Dec. 2003.

[3] R. Chellappa, C. L. Wilson, and S. Sirohey, "Human and machine recognition of faces: A survey," *Proc. IEEE*, vol. 83, no. 5, pp. 705–741, May 1995.

[4] P. J. Phillips, H. Wechsler, J. Huang, and P. J. Rauss, "The FERET database and evaluation procedure for face recognition algorithms," *Image Vis. Comput.*, vol. 16, no. 5, pp. 295–306, Apr. 1998.

[5] P. J. Phillips, M. Hyeonjoon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face recognition algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 10, pp. 1090–1104, Oct. 2000.

[6] D. M. Blackburn, J. M. Bone, and P. J. Phillips, "Face recognition vendor test 2000: Evaluation report," Nat. Inst. Standards Technol., Gaithersburg, MD, Tech. Rep. A269514, 2001.

[7] P. J. Phillips, P. Grother, R. J. Micheals, D. M. Blackburn, E. Tabassi, and J. M. Bone, "Face recognition vendor test 2002: Evaluation report," Nat. Inst. Standards Technol., Gaithersburg, MD, Tech. Rep. NISTIR 6965, 2003.

[8] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," in *Proc. IEEE Comput. Vis. Pattern Recog.*, 2005, vol. 1, pp. 947–954.

[9] 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, Dec. 2003.

[10] A. M. Martinez and R. Benavente, "The AR face database," CVC, Barcelona, Spain, Tech. Rep. 24, 1998.

[11] K. Messer, J. Matas, J. Kittler, J. Luettin, and G. Maitre, "XM2VTSDB: The extended M2VTS database," in *Proc. Int. Conf. Audio Video-Based Biometric Person Authentication*, 1999, pp. 72–77.

[12] F. S. Samaria and A. C. Harter, "Parameterisation of a stochastic model for human face identification," in *Proc. IEEE Workshop Appl. Comput. Vis.*, 1994, pp. 138–142.

[13] D. B. Graham and N. M. Allinson, "Characterizing virtual eigensignatures for general purpose face recognition," in *Face Recognition: From Theory to Applications, NATO ASI Series F, Computer and Systems Sciences*, vol. 163. Berlin, Germany: Springer-Verlag, 1998, pp. 446–456.

[14] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recog.*, 1991, pp. 586–591.

[15] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.

[16] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 643–660, Jun. 2001.

[17] K.-C. Lee, J. Ho, and D. J. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 5, pp. 684–698, May 2005.

[18] 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 *Proc. Int. Conf. Audio Video-Based Biometric Person Authentication*, 2003, pp. 625–638.

[19] R. Gross, "Face databases," in *Handbook of Face Recognition*, S. Z. Li and A. K. Jain, Eds. New York: Springer-Verlag, 2005, pp. 301–327.

[20] R. K. Bothwell, J. C. Brigham, and R. S. Malpass, "Cross-racial identification," *Pers. Soc. Psychol. Bull.*, vol. 15, no. 1, pp. 19–25, 1989.

[21] N. Furl, P. J. Phillips, and A. J. O'Toole, "Face recognition algorithms and the other-race effect: Computational mechanisms for a developmental contact hypothesis," *Cogn. Sci.: A Multidiscip. J.*, vol. 26, no. 6, pp. 797–815, 2002.

[22] X. Lu and A. K. Jain, "Ethnicity identification from face images," in *Proc. SPIE—Defense and Security Symp.*, 2004, vol. 5404, pp. 114–123.

[23] D. L. Swets and J. J. Weng, "Using discriminant eigenfeatures for image retrieval," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 8, pp. 831–836, Aug. 1996.

[24] K. Etemad and R. Chellappa, "Discriminant analysis for recognition of human face images," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 14, no. 8, pp. 1724–1733, 1997.

[25] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition," *IEEE Trans. Image Process.*, vol. 11, no. 4, pp. 467–476, Apr. 2002.

[26] S. Shan, W. Gao, Y. Chang, B. Cao, and P. Yang, "Review the strength of Gabor features for face recognition from the angle of its robustness to mis-alignment," in *Proc. IEEE Int. Conf. Pattern Recog.*, 2004, vol. 1, pp. 338–341.

[27] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang, "Local Gabor binary pattern histogram sequence (LGBPHS): A novel non-statistical model for face representation and recognition," in *Proc. IEEE Int. Conf. Comput. Vis.*, 2005, vol. 1, pp. 786–791.

[28] S. Shan, Y. Chang, W. Gao, B. Cao, and P. Yang, "Curse of mis-alignment in face recognition: Problem and a novel mis-alignment learning solution," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recog.*, 2004, pp. 314–320.

[29] X. Wang and X. Tang, "A unified framework for subspace face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 9, pp. 1222–1228, Sep. 2004.

[30] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971–987, Jul. 2002.

[31] L. Wiskott, J. M. Fellous, N. Kuiger, and C. von der Malsburg, "Face recognition by elastic bunch graph matching," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 775–779, Jul. 1997.

[32] B. Gokberk, M. O. Irfanoglu, L. Akarun, and E. Alpaydm, "Optimal Gabor kernel location selection for face recognition," in *Proc. IEEE Int. Conf. Image Process.*, 2003, vol. 1, pp. 677–680.

[33] T. Ahonen, A. Hadid, and M. Pietikainen, "Face recognition with local binary patterns," in *Proc. Eur. Conf. Comput. Vis.*, 2004, vol. 1, pp. 469–481.

[34] S. Shan, W. Gao, B. Cao, and D. Zhao, "Illumination normalization for robust face recognition against varying lighting conditions," in *Proc. IEEE Int. Workshop Anal. Model. Faces Gestures*, 2003, pp. 157–164.

[35] Y. Adini, Y. Moses, and S. Ullman, "Face recognition: The problem of compensating for changes in illumination direction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 721–732, Jul. 1997.

[36] A. Shashua and T. Riklin-Raviv, "The quotient image: Class-based re-rendering and recognition with varying illuminations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 2, pp. 129–139, Feb. 2001.

[37] R. Gross, I. Matthews, and S. Baker, "Appearance-based face recognition and light-fields," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 26, no. 4, pp. 449–465, Apr. 2004.

[38] R. Ramamoorthi, "Analytic PCA construction for theoretical analysis of lighting variability in images of a Lambertian object," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 10, pp. 1322–1333, Oct. 2002.

[39] M. Lades, J. C. Vorbruggen, J. Buhmann, J. Lange, C. von der Malsburg, R. P. Wurtz, and W. Konen, "Distortion invariant object recognition in the dynamic link architecture," *IEEE Trans. Comput.*, vol. 42, no. 3, pp. 300–311, Mar. 1993.

[40] A. M. Martinez, "Recognizing imprecisely localized, partially occluded, and expression variant faces from a single sample per class," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 6, pp. 748–763, Jun. 2002.

[41] D. Beymer and T. Poggio, "Image representations for visual learning," *Science*, vol. 272, no. 5270, pp. 1905–1909, Jun. 1996.

[42] G. J. Edwards, T. F. Cootes, and C. J. Taylor, "Face recognition using active appearance models," in *Proc. Eur. Conf. Comput. Vis.*, 1998, vol. 2, pp. 581–595.

[43] A. M. Martinez, "Matching expression variant faces," *Vis. Res.*, vol. 43, no. 9, pp. 1047–1060, Apr. 2003.
[44] C.-Y. Huang, O. I. Camps, and T. Kanungo, "Object recognition using appearance-based parts and relations," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recog.*, 1997, pp. 877–883.
[45] T. Kurita, T. Takahashi, and Y. Ikeda, "A neural network classifier for occluded images," in *Proc. IEEE Int. Conf. Pattern Recog.*, 2002, vol. 3, pp. 45–48.
[46] X. Tan, S. Chen, Z.-H. Zhou, and F. Zhang, "Recognizing partially occluded, expression variant faces from single training image per person with SOM and soft k-NN ensemble," *IEEE Trans. Neural Netw.*, vol. 16, no. 4, pp. 875–886, Jul. 2005.
[47] D. J. Beymer, "Face recognition under varying pose," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recog.*, 1994, pp. 756–761.
[48] T. Vetter and T. Poggio, "Linear object classes and image synthesis from a single example image," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 733–742, Jul. 1997.
[49] V. Blanz and T. Vetter, "Face recognition based on fitting a 3D morphable model," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 9, pp. 1063–1074, Sep. 2003.
[50] V. Blanz, P. Grother, P. J. Phillips, and T. Vetter, "Face recognition based on frontal views generated from non-frontal images," in *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recog.*, 2005, vol. 2, pp. 454–461.

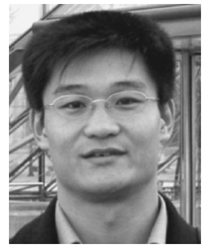
Wen Gao (S'87–M'88–SM'05) received the B.S. degree in computer science from the Harbin University of Science and Technology, Harbin, China, in 1982, the M.S. degree in computer science from the Harbin Institute of Technology (HIT), Harbin, in 1985, and the Ph.D. degree in electronics engineering from the University of Tokyo, Tokyo, Japan, in 1991.

He was with the HIT from 1985 to 1995, serving as a Lecturer, a Professor, and the Head of the Department of Computer Science. He was with the Institute of Computing Technology (ICT), Chinese Academy of Sciences (CAS), Beijing, from 1996 to 2005. During his career as a Professor at CAS, he was the Director of the ICT, the Executive Vice President of the Graduate School, as well as the Vice President of the University of Science and Technology of China. He is currently a Professor with the School of Electronics Engineering and Computer Science, Peking University, Beijing. He has published four books and over 300 technical articles in refereed journals and proceedings in the areas of multimedia, video compression, face recognition, sign-language recognition and synthesis, image retrieval, multimodal interface, and bioinformatics.

Dr. Gao is the Editor-in-Chief of the *Journal of Computer* (in Chinese), the Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, and the Editor of the *Journal of Visual Communication and Image Representation*. He received the China's State Scientific and Technological Progress Awards in 2000, 2002, 2003, and 2005.

Bo Cao received the B.S. degree in computer science from Xi'an Jiaotong University, Xi'an, China, in 1999. He is currently working toward the Ph.D. degree at the Institute of Computing Technology (ICT), Chinese Academy of Sciences, Beijing, where he is currently with the ICT–ISVISION Joint Research and Development Laboratory for Face Recognition and also with the Graduate School.

His research interests include pattern recognition, face recognition system evaluation, and biometrics. He has also pursued the development of practical face recognition systems and designed prototype systems for surveillance and border control.



Shiguang Shan (M'04) received the M.S. degree in computer science from the Harbin Institute of Technology, Harbin, China, in 1999, and the Ph.D. degree in computer science from the Institute of Computing Technology (ICT), Chinese Academy of Sciences (CAS), Beijing, in 2004.

He is currently an Associate Professor and the Vice Director of the Digital Media Center, ICT, CAS. He is also the Vice Director of the ICT–ISVISION Joint Research and Development Laboratory for Face Recognition, ICT, CAS. His research interests cover image analysis, pattern recognition, and computer vision. He is focusing specifically on face recognition-related research topics.

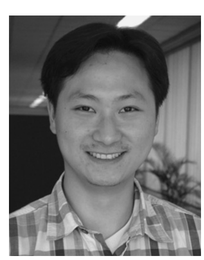
Xilin Chen (M'00) received the B.S., M.S., and Ph.D. degrees in computer science from the Harbin Institute of Technology (HIT), Harbin, China, in 1988, 1991, and 1994, respectively.

He was a Professor with the HIT from 1999 to 2005 and was a Visiting Scholar with Carnegie Mellon University, Pittsburgh, PA, from 2001 to 2004. He has been with the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, since August 2004. His research interests include image processing, pattern recognition, computer vision, and multimodal interface.

Dr. Chen is a member of the IEEE Computer Society. He has received several awards, including the China's State Scientific and Technological Progress Award in 2000, 2003, and 2005 for his research work. He has served as a program committee member for more than 20 international and national conferences.

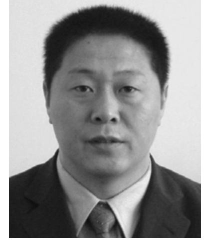
Delong Zhou received the B.S. degree in thermal energy and power engineering from Zhejiang University, Hangzhou, China, and the M.S. and Ph.D. degrees in control theory and control engineering from Northwestern Polytechnical University, Xi'an, China, in 1997 and 2001, respectively.

From 2001 to 2003, he was a Postdoctoral Research Fellow with the Joint Development Laboratory, Institute of Computing Technology, Chinese Academy of Sciences, Beijing. He is currently an Associate Professor with the College of Information Engineering, Zhejiang University of Technology. His research interests focus on pattern recognition, image fusion, and neural networks.



Xiaohua Zhang received the B.S. degree in computer science from the Beijing Institute of Technology, Beijing, China, in 2001, and the M.S degree from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, in 2004. He is currently working toward the M.S. degree in the Department of Electrical and Computer Engineering, University of Rochester, Rochester, NY.

He worked as a Software Engineer from 2004 to 2006 with Broadcom Corporation, Irvine, CA.



Debin Zhao received the B.S., M.S., and Ph.D. degrees in computer science from the Harbin Institute of Technology (HIT), Harbin, China, in 1985, 1988, and 1998, respectively.

After he graduated, he joined the Department of Computer Science, HIT, where he is currently a Professor of computer science. He was a Research Fellow with the Department of Computer Science, City University of Hong Kong, Kowloon, Hong Kong from 1989 to 1993. His research interests include video and still image compression, image processing, and pattern recognition. He has authored or coauthored over 100 publications.