

CAS-PEAL Large-Scale Chinese Face Database and Evaluation Protocols

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 the 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 setup 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 accessories (3 glasses, and 3 caps), and 15 lighting directions. This face database is now partly made available (a subset named by CAS-PEAL-R1, contain 30,900 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.

1. INTRODUCTION

Automatic Face Recognition (AFR) has been studied for over 30 years. 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 and FRVT 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 felt that the FERET face database commonly used now by AFR researchers 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 probes set (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 the western. How an system can be generalized to peoples from other races will be unknown.

Of course, there are other face databases internationally including CMU PIE, AR, XM2VTS, ORL, MPI, UMIST, MIT, Yale and Yale B, MIT, KFDB, 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. 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 at least by now. Therefore,

To sum up, we felt that the AFR community needs a large-scale face database for the abovementioned purposes, especially a large-scale Chinese face database, which covers the sources of 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 under the sponsors of 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 (599 males and 443 females) with varying Pose, Expression, Accessory, and Lighting (PEAL). For each subject, 9 cameras spaced equally in a horizontal semicircular shelf are setup 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 accessories (3 glasses, and 3 caps), and 15 lighting directions. This face database is now partly made available (a subset named by CAS-PEAL-R1, contain 30,900 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 content of the construction 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 setup in the Joint Research & Development Lab of Chinese Academy of Sciences. The space size 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 0 to 8 from the subject's right to left. The planform of the cameras distributed on the semicircle shelf is illustrated in Figure 1.

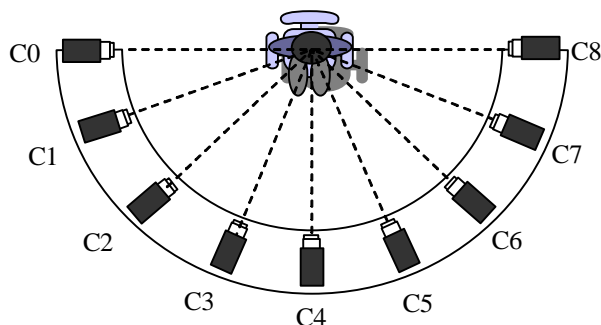


Fig. 1 Planform of our camera system

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 ourselves to control the nine cameras and capture images from them 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 a uniform naming conventions.

Each subject is asked to seat 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 fifth camera that locates at the middle of the semicircular arch

of the shelf. Fig. 2 shows the situation that one subject has seated in the chair and ready for the photographing procedure.

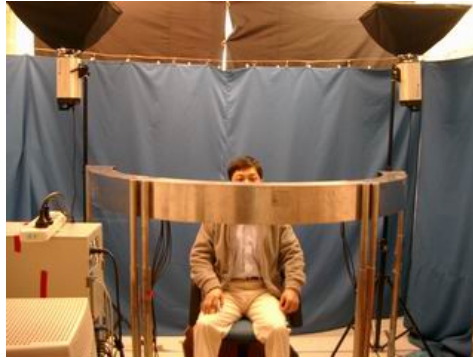


Fig. 2. Setup of the Camera Systems

2.2 Lighting System

To cover varying lighting conditions, we setup 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.

Then, some fluorescent lamps are coarsely arranged as “lighting sources” to form the varying lighting conditions. The lamps are configured in a spherical coordinate as shown in Figure 3, whose origin is the center of the circle, which coincided with the semicircular shelf. Fifteen fluorescent lamps are placed at the “lamp” positions as shown in Figure 3, which are uniformly located at specific 5 azimuths (-90° , -45° , 0° , $+45^\circ$, $+90^\circ$) and 3 elevations (-45° , 0° , $+45^\circ$). 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.

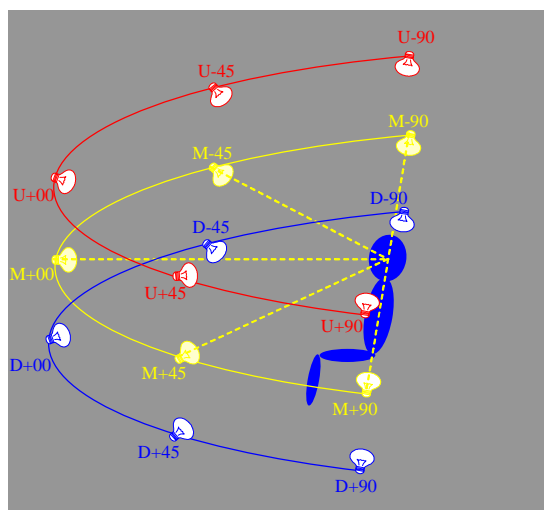


Fig. 3. Configuration of the lamps

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 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. 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 sheets of cloth with five different unicolors (blue, white, black, red and yellow) are used.

3. DESIGN OF THE DATABASE

Utilizing the equipments described in Section 2, we defined six 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 and pose and session 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
# 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).

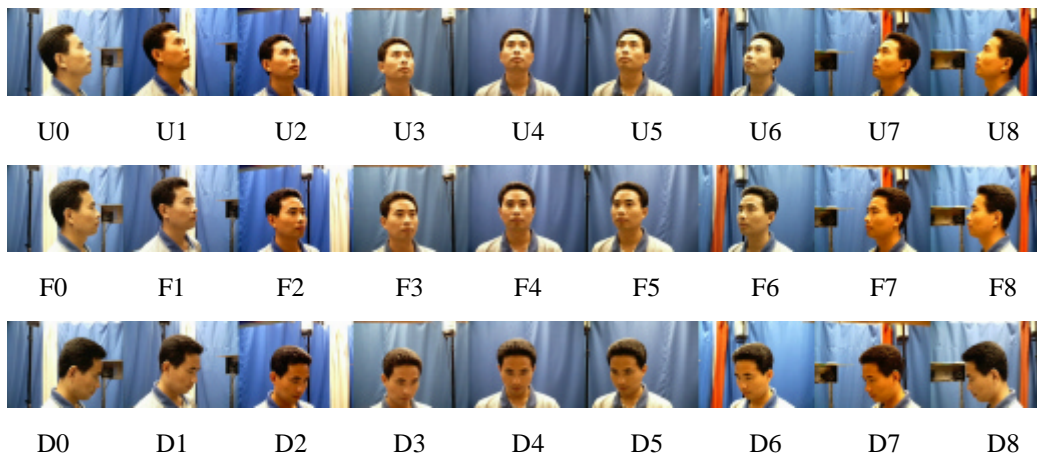


Fig. 4. The 27 images of one subject under pose variation in the CAS PEAL database. The nine cameras were spaced equally in the horizontal semicircular shelf, each about 22.5° apart. The subject was asked to look upwards, right into the camera C4 (the middle camera) and look downwards. Then, the 27 poses 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, frown, 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.

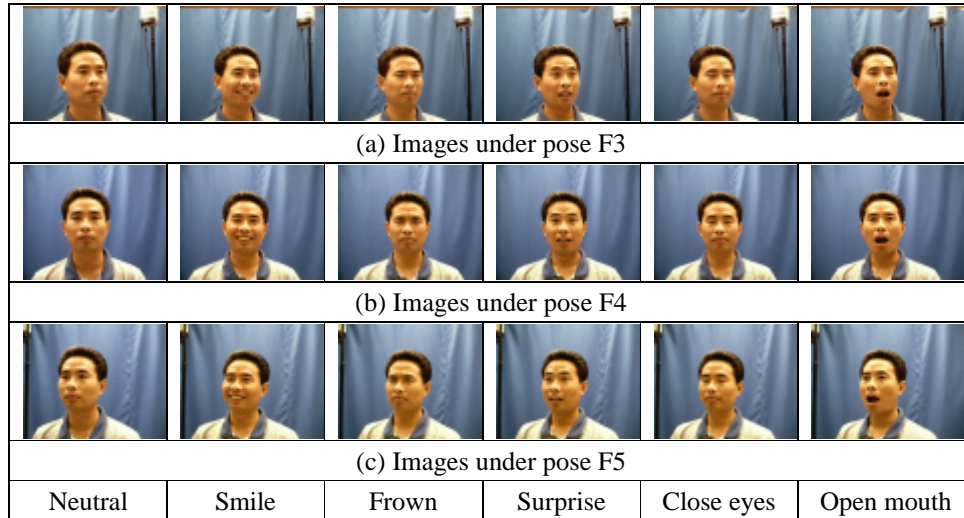


Fig. 5. Example images of one subject with six expressions across 3 poses F3, F4, and F5

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.

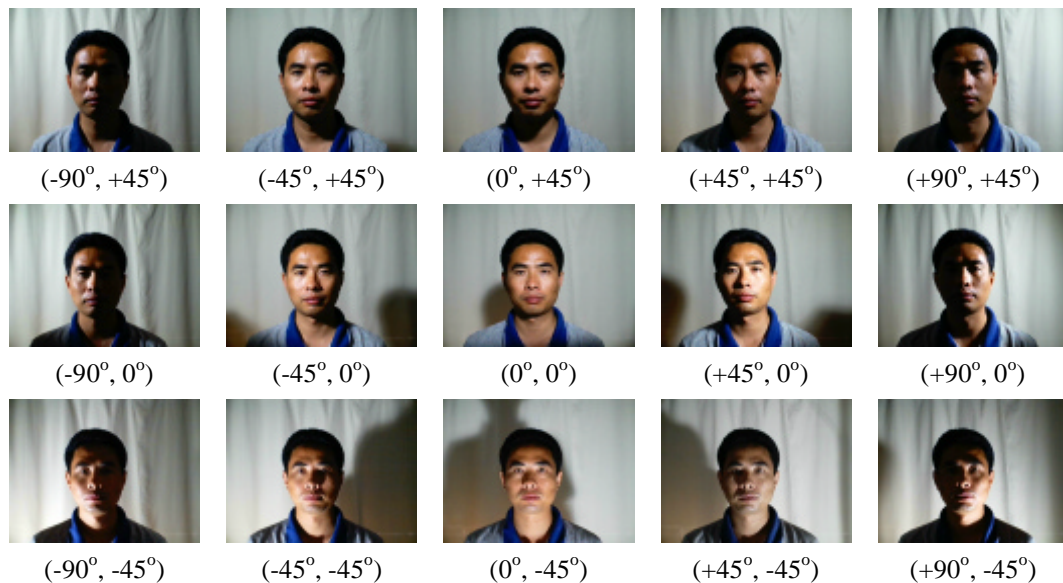


Fig.6. Some example images from one subject illuminated by fluorescent light source located at

different azimuth and elevation coordinate. The (\mathbf{a}, \mathbf{q}) 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 glasses, are asked to wear one by one. And nine images are captured using the camera systems. Figure 6 illustrates the example images of one subject recorded by the camera C4.

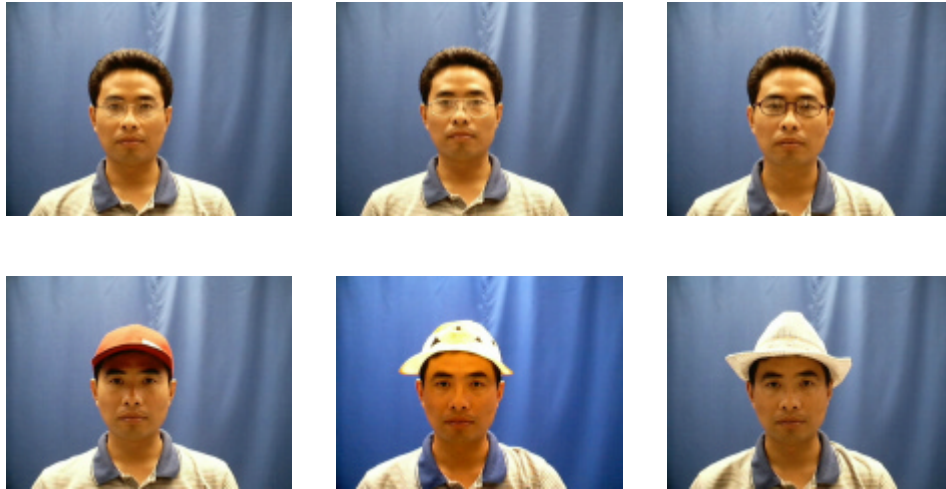


Fig. 7. Example image of one subject with 6 different accessories

3.5 Pose and Background Variation

As has been mentioned above, different unicolor background cloth has 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 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.9 Example images of different distances from the camera



Fig.10 Example images captured with time difference of one year. The images in the bottom row are captured after half a year.

4. DESCRIPTION OF THE RELEASED CAS-PEAL FACE DTABASE:

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 30,900 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	2,646
	Background	2-4	297	650
	Distance	1-2	296	324
	Aging Session	1	66	66
	Total:			9,060
Pose		21 (3*7)	1040	21,840
Total:				30,900

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°	Elevation: 0°	Elevation: -45°

- 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 30,900 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 BMP files. Several cropped images are shown in Fig. 6.



Fig. 6. Several example of the cropped face images.

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

```
<CAS-PEAL>                //The root directory of CAS-PEAL database
|
+- <Frontal>                //Frontal subset
  |- <Normal>
  |- <Expression>
  |- <Lighting>
  |- <Accessory>
  |- <Background>
  |- <Distance>
  |   |- <Aging>
+- <Pose>                    //Pose subset
```

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, the eyes locations of all the images are provided in a text file which is stored in the root directory of the database.

5. EVALUATION PROTOCOLS

We will soon supplement the evaluation protocols based on the CAS-PEAL large Chinese face database, and report some evaluation results of some benchmark AFR methods, if possible, as well as the evaluation results of some most successful commercial systems.

6. OBTAINING THE CAS-PEAL FACE DATABASE

To get a copy of the CAS-PEAL face database, please download the release agreement, 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 30,900 images of 1040 subjects) for research purpose only on a case-by-case basis.

ACKNOWLEDGMENTS

This research is partially sponsored by National Hi-Tech Program of China (No.

2001AA114010 and No. 2001AA114190) and NSFC under contract No. 60332010. This work is also partially sponsored by ISVISION Technologies Co., Ltd.

REFERENCE

1. T. Sim, S. Baker, and M. Bsat. The CMU pose, illumination, and expression database. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12):1615–1618, 2003.
2. A. R. Martinez and R. Benavente. The AR face database. Technical Report 24, Computer Vision Center(CVC) Technical Report, Barcelona, Spain, June 1998.
3. K. Messer, J. Matas, J. Kittler, J. Luetttin, and G. Maitre. XM2VTSDB: The extended M2VTS database. In *Second International Conference on Audio and Video-based Biometric Person Authentication*, March 1999.
4. M.-H. Yang, D. Kriegman, and N. Ahuja. Detecting faces in images: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(1):34–58, 2002.
5. W. Zhao, R. Chellappa, A. Rosenfeld, and P. J. Phillips. Face recognition: A literature survey. Technical Report CS-TR4167, University of Maryland, 2000.
6. E. Bailly-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)*, pages 625–638, 2003.
7. P. J. Phillips, P. Grother, J. Ross, D. Blackburn, E. Tabassi, and M. Bone. Face recognition vendor test 2002: Evaluation report, March 2003.
8. A. Georghiades, D. Kriegman, and P. Belhumeur. From few to many: Generative models for recognition under variable pose and illumination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):643–660, 2001.
9. B.-W. Hwang, H. Byun, M.-C. Roh, and S.-W. Lee. Performance evaluation of face recognition algorithms on the asian face database, KFDB. In *Audio- and Video-Based Biometric Person Authentication (AVBPA)*, pages 557–565, 2003.
10. 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*, 22(10):1090–1104, 2000.
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. P. J. Phillips, H. Wechsler, and P. Rauss. The FERET database and evaluation procedure for face-recognition algorithms. *Image and Vision Computing*, 16(5):295–306, 1998.