Baseline Evaluations on the CAS-PEAL-R1 Face Database

Bo Cao[1,2], Shiguang Shan[1], Xiaohua Zhang[1], Wen Gao[1,2]

[1]ICT-ISVISION Joint Research&Development Laboratory for Face Recognition, Chinese Academy of Sciences, P.O.Box 2704, Beijing, China, 100080
{bcao, sgshan, wgao}@ict.ac.cn
[2]Graduate School of the Chinese Academy of Sciences

Abstract. In this paper, three baseline face recognition algorithms are evaluated on the CAS-PEAL-R1 face database which is publicly released from a large-scale Chinese face database: CAS-PEAL. The main objectives of the baseline evaluations are to 1) elementarily assess the difficulty of the database for face recognition algorithms, 2) provide an example evaluation protocol on the database, and 3) identify the strengths and weakness of some popular algorithms. Particular description of the datasets used in the evaluations and the underlying philosophy are given. The three baseline algorithms evaluated are Principle Components Analysis (PCA), a combined Principle Component Analysis and Linear Discriminant Analysis (PCA+LDA), and PCA+LDA algorithm based on Gabor features (G PCA+LDA). Four face image preprocessing methods are also tested to emphasize the influences of the preprocessing methods on the performances of face recognition algorithms.

1 Introduction

Automatic Face Recognition (AFR) has become one of the most active research areas in pattern recognition, computer vision and psychology and much progress has been made in the past few years [1], [2]. However, AFR remains a research area far from mature and its application is still limited in controllable environments. Therefore, evaluating and comparing the potential AFR technologies exhaustively and objectively to discover the real choke points and the valuable future research topics, and developing algorithms robust to variations in pose, expression, accessories, lighting, etc. are becoming more and more significant.

Aiming at these goals, large-scale and diverse face databases are obviously one of the basic requirements. Internationally, FERET [3] and FRVT [4] have pioneered both evaluation protocols and database construction. Especially, the FERET database is widely used in the research field. Besides the significant FERET tests, its success can also be contributed to 1) the public availability of the database, 2) the large number of subjects and diverse images, and 3) the explicit and categorized partition of the gallery sets and probe sets. Actually, these sets are used by many different researchers to compare their algorithms and test the performances under different image variations, such as illumination, facial expression and aging variations.

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. Considering these limitations, we design and construct a large-scale Chinese face database (the CAS-PEAL face database), which is now partly made available as a released subset named by CAS-PEAL-R1.

In the paper, the recommended partition of the images in the CAS-PEAL-R1 database is proposed to compose the training set, the gallery set and the probe sets which can be used to evaluate a specific face recognition algorithm. Also, the evaluation results of three baseline face recognition algorithms combining different preprocessing methods are provided. This paper can be an informative complement to the documents on the database itself.

Table 1. The contents of CAS-PEAL-R1

Subset		# Variations	# Subjects	# Images
Frontal	Normal	1	1040	1,040
	Expression	5*	377	1,884
	Lighting	>= 9	233	2,450
	Accessory	6	438	2,616
	Background	2-4	297	651
	Distance	1-2	296	324
	Aging	1	66	66
	Total:			9,031
Pose		21 (3*7)	1040	21,832
Total:				30,863

5*: Neutral expression is not counted in.

2 Contents of the Released CAS-PEAL Face Database: CAS-PEAL-R1

Currently, the CAS-PEAL face database contains 99,594 images of 1040 individuals (595 males and 445 females) with controlled Pose, Expression, Accessory, and Lighting variations. The details of the database can be referred to [5], [6].

The CAS-PEAL face database has been cut, arranged and labeled to form the first distribution: CAS-PEAL-R1. This distribution contains 30,863 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 with the subject looking right into the 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 the camera.

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 without any other variations are included.

The content of CAS-PEAL-R1 is summarized in Table 1.

3 Datasets Used in the Evaluations

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

These two aspects are closely related. If the scale of the test sets is very small, the performance scores may be highly stochastic and become incomparable. Though some methods do exist to tackle this problem, such as the permutation methodology proposed in [7], a large-scale test set is still helpful. Also, the scale of the training set can influence the comparison of two algorithms. Martinez et al. [8] demonstrates that PCA can outperform LDA when the training data set is small, while LDA is normally considered superior than PCA in face recognition. Considering these problems, we compose the test sets and the training set from the CAS-PEAL-R1 database as large as possible. And the test sets are categorized to restrict the images in one probe set to undergo one main variation, which can be used to identify the strengths and weakness of a specific algorithm and to address the variations associated with changes in the probe sets.

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

1. **Training set**. A training set is a collection of images which are used to generate a generic representation of faces and/or to tune parameters of the classifier. In the evaluation, the training set contains 1,200 images (300 subjects randomly selected from the 1,040 subjects in the CAS-PEAL-R1 database and each subject contains four images randomly selected from the frontal subset of the CAS-PEAL-R1 database).
2. **Gallery set.** A gallery set is a collection of images of known individuals against which testing images are matched. In the evaluation, the gallery set contains 1,040 images of 1,040 subjects (each subject has one image under normal condition). Actually, the gallery set consists of all the normal images mentioned in Table 1.
3. **Probe set.** A probe set is a collection of probe images of unknown individuals to be recognized. In the evaluation, nine probe sets are composed from the CAS-PEAL-R1 database. Among them, six probe sets correspond to the six subsets in the frontal subset: expression, lighting, accessory, background, distance and aging, as described in Table 1. The other three probe sets correspond to the images of subjects in the pose subset: looking upwards, looking right into the camera, and looking downwards. All the images that appear in the training set are excluded from these probe sets.

The datasets used in the evaluation are summarized in Table 2.

Table 2. The datasets used in the evaluation protocols

<table>
<tr><td rowspan="2">Datasets</td><td rowspan="2">Training set</td><td rowspan="2">Gallery set</td><td colspan="6">Probe sets (frontal)</td></tr>
<tr><td>expression</td><td>lighting</td><td>accessory</td><td>background</td><td>distance</td><td>Aging</td></tr>
<tr><td>Num. of images</td><td>1,200</td><td>1,040</td><td>1,570</td><td>2,243</td><td>2,285</td><td>553</td><td>275</td><td>66</td></tr>
<tr><td rowspan="2"></td><td colspan="8">Probe sets (pose)</td></tr>
<tr><td colspan="3">looking upwards (PU)</td><td colspan="3">looking right into the camera (PM)</td><td colspan="2">looking downwards (PD)</td></tr>
<tr><td>Num. of images</td><td colspan="3">4,998</td><td colspan="3">4,993</td><td colspan="2">4,998</td></tr>
</table>

4 Baseline Face Recognition Algorithms

The three baseline algorithms evaluated are Principle Components Analysis (PCA) [9], (also known as Eigenfaces), a combined Principle Component Analysis and Linear Discriminant Analysis (PCA+LDA, a variant of Fisherfaces) [10], and PCA+LDA algorithm based on Gabor features (G PCA+LDA). PCA and PCA+LDA based face recognition algorithms are both fundamental and well studied. Recently, 2D Gabor wavelets are extensively used for local feature representation and extraction, and demonstrate their success in face recognition [11], [12]. So, the PCA+LDA algorithm based on Gabor features is also used as a baseline algorithm to reflect this trend.

Instead of using the grey-scale images as the original features in the PCA and PCA+LDA algorithms, the representation of the original features in the third algorithm is based on the Gabor wavelet transform of the original images. Gabor wavelets are biologically motivated convolution kernels which are plane waves restricted by a Gaussian envelope function, and those kernels demonstrate spatial locality and orientation selectivity. In face recognition, Gabor wavelets exhibit robustness to moderate lighting changes, small deformations [11].

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

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

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

$$k_{u,v} = k_v e^{i\phi_u} \quad \textbf{(2)}$$

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

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

At each image pixel, a set of convolution coefficients can be calculated using a family of Gabor kernels as defined by equation (1). The Gabor wavelet transform of an image is the collection of the coefficients of all the pixels. To reduce the dimensionality, the pixels are sampled and their convolution coefficients are concatenated to form the original features of the PCA+LDA algorithm. These concatenated coefficients are also called the augmented Gabor feature vector in [12]. In the experiments, the size of the normalized faces is 64×64 and the pixels are sampled every four pixel both in row and in column, so the dimensionality of the features is 9000 ($15 \times 15 \times 40$). It should be noted that each feature is normalized to zero mean and unit variance to compensate for the scale variance of different Gabor kernels.

5 Preprocessing

In the evaluation, the preprocessing of the face images is divided into three steps: geometric normalization, masking, and illumination normalization. The first two steps are to provide features that are invariant to geometric transformations of the face images, such as the location, the rotation and the scale of the face in an image, and remove irrelevant information for the purpose of face recognition, such as the background and the hair of a 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 geometric normalization step, each face image is scaled and rotated so that the eyes are positioned in line and the distance between them equals a predefined length. Then, the face image is cropped to include only the face region with little hair and background as Fig. 1(a) shows (the size of the cropped face image is 64×64). In masking step, a predefined mask is put on each cropped face image to further reduce the effect of different hair styles and backgrounds which are not the intrinsic characteristics, as Fig. 1(b) shows. Typically, the hair style of a specific subject and the background are constant in a face database, so better performance can be obtained with larger face regions. However, this bias should be avoided as much as possible by restricting the above cropping and masking procedures.

In illumination normalization step, four illumination normalization methods are evaluated: Histogram Equalization (HE), Gamma Intensity Correction (GIC), Region-based Histogram Equalization (RHE) and Region-based Gamma Intensity Correction (RGIC) [13], [14]. Fig. 2(b) illustrates the effect of these four methods on an example face image.

(a) (b)

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

Original HE GIC RHE RGIC

(a) (b)

Fig. 2. 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

6 Evaluation Results

6.1 Frontal Face Images

The three baseline face recognition algorithms (PCA, PCA+LDA and G PCA+LDA) are trained on the training set, and evaluated on the six frontal probe sets as described in Section 3. Before training and testing, all the images are preprocessed as described in Section 5, using the four illumination normalization methods or no illumination normalization respectively. Table 3 and Fig. 3 show the performance of these algorithms on the frontal probe sets.

Table 3. Identification performance of the three baseline algorithms on the six frontal probe sets and the union (Total) set of these sets

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

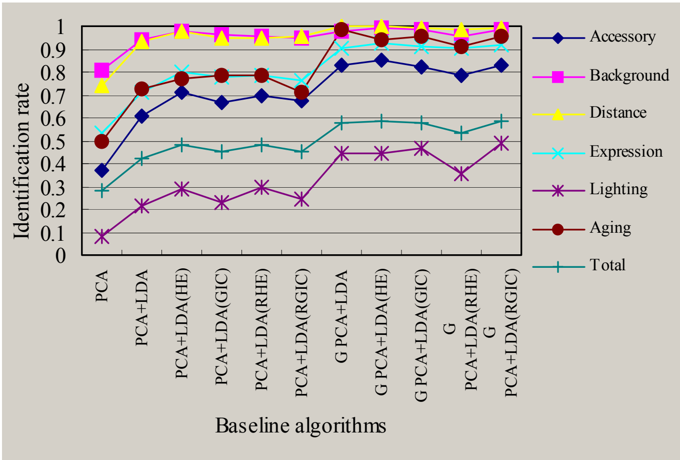

Fig. 3. Identification performance on frontal images

6.2 Face Images under Different Poses

Two baseline face recognition algorithms (PCA+LDA and G PCA+LDA) are trained on the training set, and evaluated on the three pose probe sets as described in Section 3. Before training and testing, all the images are preprocessed as described in Section 5, using the RGIC illumination normalization method or no illumination normalization respectively. Table 4 and Fig. 4 show the performance of these algorithms on the pose probe sets.

7 Conclusion

In this paper, the evaluation results of three baseline algorithms with different preprocessing methods on the CAS-PEAL-R1 database are presented. Also, we describe the contents of the CAS-PEAL-R1 database which is a released version of the CAS-PEAL database, and the partition of the datasets used in the evaluation. From these results, the difficulty of the database, and the strengths and weakness of commonly used algorithms can be inferred, which may be some good references for the potential users of the database.

We believe that not only the characteristics of a face database itself, such as the scale, the diversity of the variations, the detailed ground-truth information and well organized structure, but also a standard evaluation protocol including the partition of training sets, gallery sets and probe sets, performance measures, etc. contributes to the success of the database. The paper sets an example of the evaluations on the CAS-PEAL-R1 database and provides baseline evaluation results to the research community.

Table 4. Identification performance of two baseline algorithms on the three pose probe sets and the union (Total) set of these sets

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

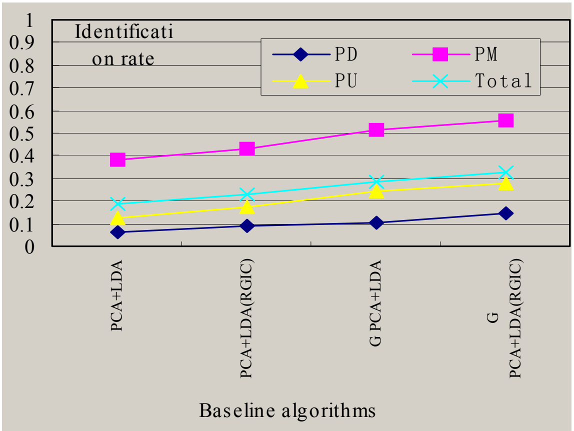

Fig. 4. Identification performance on images under poses

Acknowledgements

This research is partially sponsored by Natural Science Foundation of China under contract No.60332010, National Hi-Tech Program of China (No. 2001AA114190 and 2002AA118010), and ISVISION Technologies Co., Ltd.

References

1. W. Zhao, R. Chellappa, A. Rosenfeld and P. J. Phillips, "Face Recognition: A Literature Survey," Technical Report, CS-TR4167, University of Maryland, 2000. Revised 2002, CS-TR4167R.
2. R. Chellappa, C. Wilson, and S. Sirohey, "Human and Machine Recognition of Faces: A Survey," Proc. IEEE, vol. 83, no. 5, pp. 705-740, 1995.
3. P. J. Phillips, H. Moon, S. Rizvi, and P. Rauss, "The FERET Evaluation Methodology for Face-Recognition Algorithms," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 22, No. 10, pp. 1090-1104, 2000.
4. P. J. Phillips, P. Grother, J. Ross, D. Blackburn, E. Tabassi, and M. Bone, "Face Recognition Vendor Test 2002: Evaluation Report," March 2003.

5. D. Zhou, X. Zhang, B. Liu, and W. Gao, “Introduction of the JDL Large-scale Facial Database,” The 4th Chinese Conference on Biometrics Recognition (Sinobiometrics’03), pp. 118-121, 2003.
6. http://www.jdl.ac.cn/peal/index.html
7. J. R. Beveridge, K. She, B. A. Draper, and G. H. Givens, “A Nonparametric Statistical Comparison of Principal Component and Linear Discriminant Subspaces for Face Recognition,” Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 535 – 542, December 2001.
8. A. M. Martinez and A. C. Kak, “PCA versus LDA,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 23, no. 2, pp. 228-233, 2001.
9. M. Turk and A. Pentland, “Face Recognition Using Eigenfaces,” IEEE Conference on Computer Vision and Pattern Recognition, 1991.
10. P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, “Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 711-720, 1997.
11. N. Kruger, C. Malsburg, L. Wiskott, and J. M. Fellous, “Face Recognition by Elastic Bunch Graph Matching,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 775-779, 1997.
12. C. Liu and H. Wechsler, “Gabor Feature Based Classification Using the Enhanced Fisher Linear Dsicriminant Model for Face Recognition,” IEEE Trans. Image Processing, vol. 11, no. 4, pp. 467-476, 2002.
13. S. Shan, W. Gao, B. Cao, and D. Zhao, “Illumination Normalization for Robust Face Recognition against Varying Lighting Conditions,” IEEE International Workshop on Analysis and Modeling of Faces and Gestures, pp. 157-164, Nice, France, 2003.
14. S. Shan, W. Gao, and L. Qing, “Generating Canonical Form Face Images for Illumination Insensitive Face Recognition,” The 4th Chinese Conference on Biometrics Recognition (Sinobiometrics’03), pp. 29-35, 2003.