Ethnicity Identification from Face Images

Xiaoguang Lu and Anil K. Jain

Department of Computer Science & Engineering, Michigan State University
East Lansing, MI 48824
{lvxiaogu, jain}@cse.msu.edu

ABSTRACT

Human facial images provide the demographic information, such as ethnicity and gender. Conversely, ethnicity and gender also play an important role in face-related applications. Image-based ethnicity identification problem is addressed in a machine learning framework. The Linear Discriminant Analysis (LDA) based scheme is presented for the two-class (Asian vs. non-Asian) ethnicity classification task. Multiscale analysis is applied to the input facial images. An ensemble framework, which integrates the LDA analysis for the input face images at different scales, is proposed to further improve the classification performance. The product rule is used as the combination strategy in the ensemble. Experimental results based on a face database containing 263 subjects (2,630 face images, with equal balance between the two classes) are promising, indicating that LDA and the proposed ensemble framework have sufficient discriminative power for the ethnicity classification problem. The normalized ethnicity classification scores can be helpful in the facial identity recognition. Useful as a "soft" biometric, face matching scores can be updated based on the output of ethnicity classification module. In other words, ethnicity classifier does not have to be perfect to be useful in practice.

Keywords: Ethnicity classification, LDA, ensemble, face recognition

1. INTRODUCTION

The human face is a highly rich stimulus that provides diverse information for adaptive social interaction with people. Humans are able to process a face in a variety of ways to categorize it by its identity, along with a number of other demographic characteristics, including ethnicity* (or race), gender, and age. Over the past few decades, a lot of effort has been devoted in the biological, psychological, and cognitive sciences areas, to discover how the human brain perceives, represents, and remembers faces. Computational models have also been developed to gain some insight into this problem.

Anthropometrical statistics show the racial and ethnic morphometric differences in the craniofacial complex.[1,2] In Ref. 1, based on carefully defined facial landmarks, 25 measurements on head and face were taken to examine three racial groups (i.e., North American Caucasian, African-American, and Chinese). Farkas identified several difference in these three groups. For example, the Chinese group had the widest faces; the main characteristics of the orbits of the Chinese group were the largest intercanthal width. Further, the soft nose is less protruding and wider in the Chinese group and it had the (relatively) highest upper lip in relation to mouth width, etc. Enlow[2] also conducted research on the structural basis for ethnic variations in facial form.

The demographic features, such as race and gender, are involved in human face identity recognition. Humans are better at recognizing faces of their own ethnicity/race than faces of other races.[3,4] Golby et al. show that same-race faces elicit more activity in brain regions linked to face recognition.[5] They use functional magnetic resonance imaging (fMRI) to examine if the same-race advantage for face identification involves the fusiform face area (FFA), which is known to be important for face recognition.[6] O'Toole et al.[7]investigate the differences in the way people perceive own- versus other-race faces. They found that the perceived typicality of own-race faces was based both on global shape information and on small distinctive feature markers, whereas the typicality of other-race faces related more to the presence/absence of local distinctive features. O'Toole et al.[8] have shown that people categorize faces of their own-race by sex more efficiently than they categorize faces of another race

*In this paper, we do not make distinction between the terms 'ethnicity' and 'race', which are used to refer to people who share common facial features that perceptually distinguish them from members of other ethnic groups.

by sex. The identification of race and gender can help the face recognition system to focus more on the identity-related features, and limit the number of entries to be searched in a large database, improving the search speed and efficiency of the retrieval systems. The ethnicity and gender are also useful for demographic statistics in many social applications. Unlike the identity, the ethnic categories are loosely defined classes. In this paper, we reduce the ethnicity classification into a two-category classification problem, Asian and non-Asian, each of which have relatively distinct anthropometrical features.

Image-based face recognition has been drawing a lot of attention over the past decade. A number of face recognition algorithms have been investigated[9] and several commercial face recognition products (e.g., Viisage,[10] Identix[11]) are available. In real applications, cross-race, cross-gender and large-scale face recognition tasks need to be solved.

Among the face recognition algorithms, appearance-based approaches,[12, 13] which utilize the intensity or intensity-derived features of the input images, have been successfully used.[9, 14] Figure 1 shows the principal component analysis[13, 15] results on the dataset used in our experiments. Comparing Figs. 1(a) and 1(c), we observe that the "average" non-Asian face appears to be different from the "average" Asian face. In Ref. 12, a 'PCA+LDA' scheme was proposed to reduce the dimensionality of the input space; LDA is used to extract the discriminant projection directions by taking into account the class label information. Moghaddam and Yang[16] used support vector machine to enhance the appearance-based framework for gender classification.

Images at different scales provide different levels of information as the visual stimuli. Multiscale analysis is widely used in Ref. 17. Face images with different resolutions construct different manifolds in the input space of different dimensionalities. A classifier at each scale can provide confidence of the assigned class membership for each test face image. The final decision may be enhanced by integrating the confidence from different scales. Kittler[18] provides a theoretical framework to combine various classifiers at the decision level. Many practical applications of combining multiple classifiers have been developed. Brunelli and Falavigna[19] presented a person identification system by combining outputs from classifiers based on audio and visual cues. Jain et al.[20] integrated multiple fingerprint matchers to develop a robust fingerprint verification system. Hong and Jain[21] designed a decision fusion scheme to combine face and fingerprint for personal identification. Although images with high resolutions tend to provide detailed information, their dimensionality is very high relative to the limited number of training samples, which makes it difficult to correctly estimate the statistical distribution. While the low resolution images may not contain enough clues for the individual identity recognition, it can still be used for other human trait identification, such as race and gender.

Compared to ethnicity identification, the gender classification has received more attention.[16, 22, 23] Gutta et al.[23] proposed a hybrid classifier based on RBF networks and inductive decision trees for classification of gender and ethnic origin, using a 64×72 image resolution. They achieved an average accuracy rate of 92% for the ethnic classification part of the task. Experimental results for gender classification in Moghaddam and Yang[16] are based on 21×12 image resolution. Shakhnarovich et al.[24] presented a real-time face detection and recognition system based on a boosted classifier. The same structure is used for demographic information extraction, including gender and ethnicity. Two categories of ethnicity are defined, Asian and non-Asian. Again, their system is focused on low resolution (24×24) images with face data weakly aligned. Their reported accuracy is about 80%.

We address the problem of race identification based on gray-scale human face images. Because the robust facial landmarks localization is still an open problem due to the complex facial appearance in the real-world environment, we do not utilize the anthropometrical measurements based classification scheme. Instead, we explore the appearance-based scheme, which has demonstrated its power in facial identity recognition. The task is formulated as a two-category classification problem, to classify the subject as an Asian or non-Asian. The input images are resized to different scales. At each scale, a classic appearance-based face recognizer based on the LDA representation is developed under the Bayesian statistical decision framework. An ensemble is then constructed by integrating the classification results to arrive at the final decision. The product rule is used as the integration strategy.

Section 2 presents the LDA and the framework of ensembles at multiple scales. Section 3 provides the experimental results and discussion. Conclusions and future work are summarized in section 4.

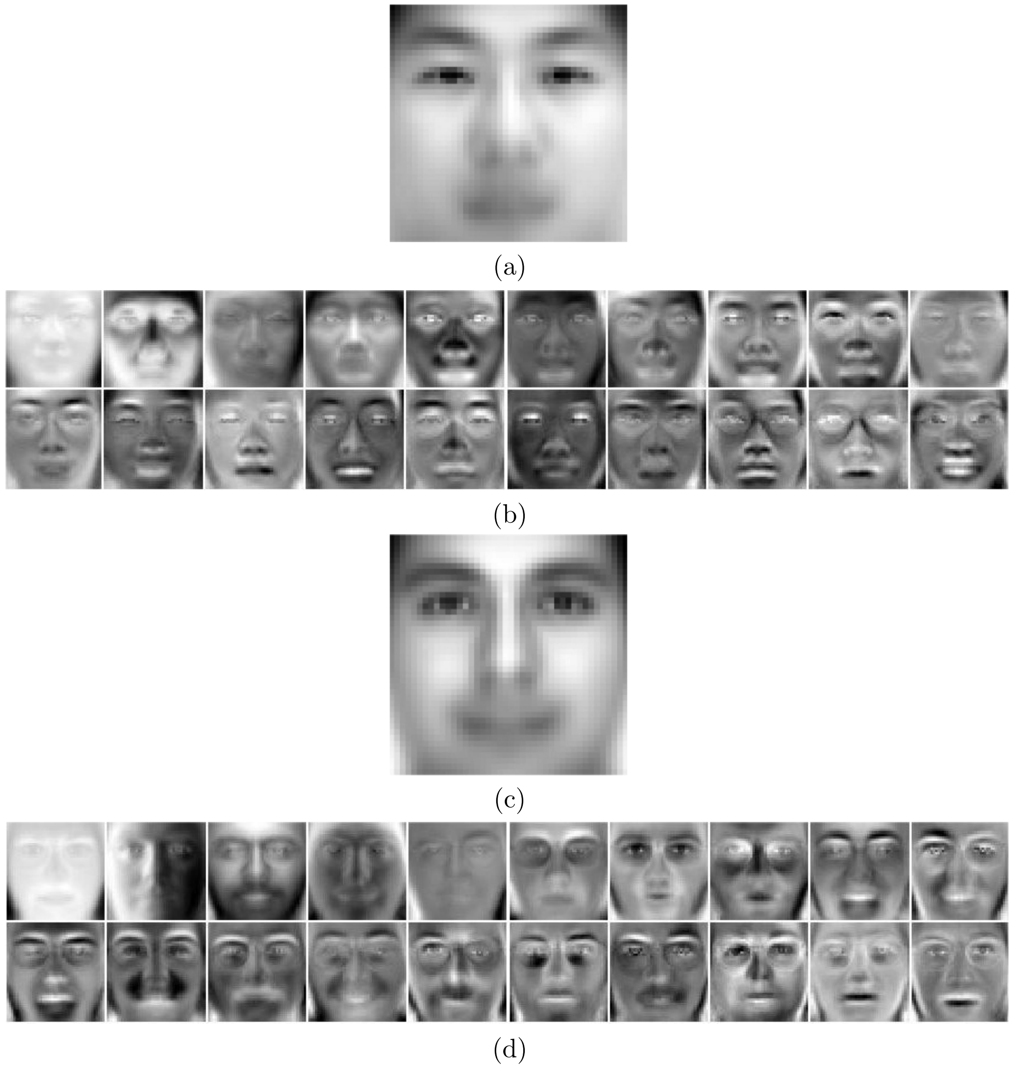


Figure 1. PCA[13, 15] on Asian and non-Asian datasets. (a) "average" Asian face; (b) top 20 eigenfaces of Asian dataset; (c) "average" non-Asian face; (d) top 20 eigenfaces of non-Asian dataset.

2. LDA ENSEMBLES AT MULTIPLE SCALES

2.1. Linear Discriminant Analysis

Linear Discriminant Analysis (LDA) is a well-known statistical method to project the given multidimensional data to a lower dimension such that the ratio of between-class scatter to within-class scatter is maximized.[25] A two-dimensional face image is considered as a one-dimensional vector, by concatenating each row (or column) of the image. Let $X = (x_1, x_2, \dots, x_i, \dots, x_N)$ denote the data matrix, where n represents the total number of pixels in the face image and N is the number of face images in the training set. Each x_i is a face vector of dimension n, concatenated from a $p \times p$ face image, where $n = p \times p$. The LDA representation is a linear transformation of the original data matrix X, to a projected data matrix, Y, i.e.

$$Y = W^T X, \tag{1}$$

where Y is the $d \times N$ feature vector matrix, d is the dimension of the feature vector y_i, $d \ll n$. The transformation matrix W is derived by

$$W_{LDA} = \arg\max_W \frac{W^T S_B W}{W^T S_W W}, \tag{2}$$

where S_B is the between-class scatter matrix and S_W is the within-class scatter matrix,

$$S_B = \sum_{i=1}^{c} N_i(\mu_i - \mu)(\mu_i - \mu)^T, \tag{3}$$

$$S_W = \sum_{i=1}^{c} \sum_{x_k \in X_i} (x_k - \mu_i)(x_k - \mu_i)^T. \tag{4}$$

In the above expression, N_i is the number of training samples in class i; c is the number of distinct classes; μ is the mean vector of all the samples, i.e., $\mu = \sum_{i=1}^{N} X_i$; μ_i is the mean vector of samples belonging to class i and X_i represents the set of samples belonging to class i.

In the face recognition problem, if the within-class scatter matrix S_W is singular, due to the fact that the rank of S_W is at most $N - c$ and the number of training samples is generally less than the dimensionality of the face image (number of pixels), PCA transform[13] can be used to reduce the dimensionality of the face image space to $N - c$[12] prior to applying LDA.

Fisher linear discriminant analysis is the LDA in two-class classification cases. LDA derives a low dimensional representation of a high dimensional face feature vector space. The face vector is projected by the transformation matrix W. The projection coefficients are used as the feature representation of each face image. In our implementation, a single Gaussian is used to model the data distribution (density estimation) in the feature space for each class, i.e. the class-conditional probability density as denoted in Eq. 5. Note that for a 2-class problem, the linear discriminant is a line. Figure 2 shows the empirical distribution in our experiments described in Sec. 3, indicating that Gaussian assumption is reasonable.

$$p(y|c) = \frac{1}{\sqrt{2\pi}\sigma} \exp \frac{-(y-m)^2}{2\sigma^2}, \tag{5}$$

where y is the projection feature of a test sample, c is class label (Asian or non-Asian), the parameters m and σ can be estimated by the sample mean and the standard deviation. The maximum a posteriori probability (MAP) is used to determine the category to which a test image belongs. It is also utilized as a matching score fed to the following classifier ensemble. Let c_1 and c_2 denote the two classes. The decision rule is to decide $y \in c_1$ if $\frac{p(y|c_1)}{p(y|c_2)} > \frac{P(c_2)}{P(c_1)}$, otherwise $y \in c_2$, where $P(c_1)$ and $P(c_2)$ denote the prior probabilities of class c_1 and c_2.

2.2. LDA Ensembles at Multiple Scales

Images at different scales provide different levels of information as the visual stimuli. In our implementation, each face image is resized to three different scales. At each scale, a LDA based classifier is constructed. Therefore, the number of classifiers in our ensemble is equal to the number of scales. The final ensemble consists of the LDA based classifiers at different scales. The confidence values of the class membership for each test image derived from different scales are combined in the ensemble. The system framework is illustrated in Fig. 3.

In our scheme, the product rule[18] strategy is applied. At each scale, the matching scores (confidence) of the test face image belonging to each of the two classes are computed as the a posteriori probability. Let $MS(i,j)$ be the matching score between the test image and the j^{th} class ($j \in 1, 2$), calculated by the classifier at the i^{th} scale. Let K denote the number of different scales.

- Product rule

  Calculate $MS_j = \prod_{i=1}^{K} MS(i,j)$. Assign the test image to the J^{th} class, such that $J = \arg\max_j MS_j$.

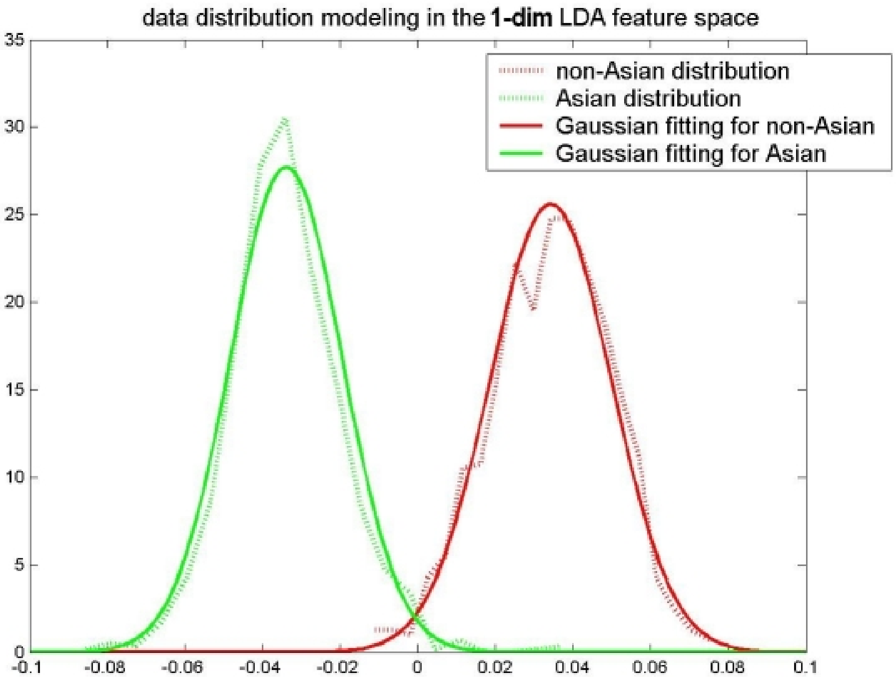


Figure 2. Modeling data distributions in the feature space by a single Gaussian.

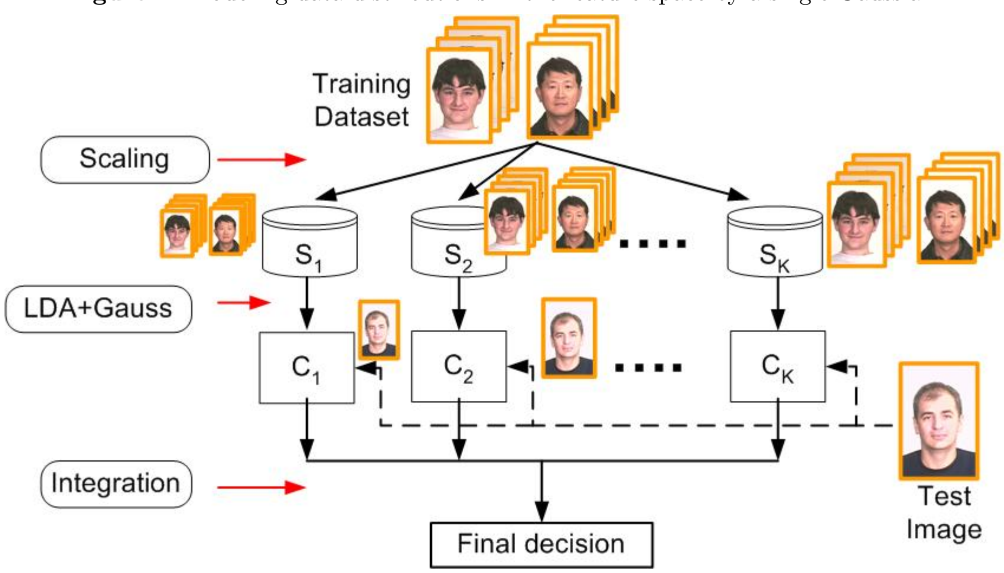


Figure 3. LDA ensembles at multiple scales for face recognition. S_1 to S_K are datasets with different image resolutions constructed from the original images. C_1 to C_K are LDA (with single Gaussian model) based classifiers trained on the corresponding subsets.

3. EXPERIMENTS AND DISCUSSION

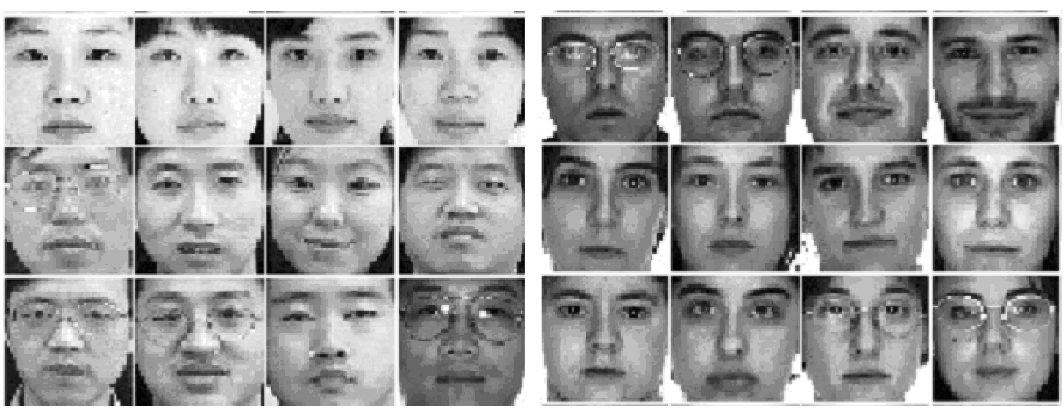
Our database is a union of four different face databases, three of which are available in the public domain[†] (see Table 1). It contains 2,630 face images of 263 subjects, with 10 images per subject. The dataset is separated

[†]Currently, the NLPR database is not available in the public domain. Less than 10% of the data in our experiments came from this database.

into two ethnic groups, Asian and non-Asian. The non-Asian database is composed of Yale and AR databases. The Asian database is composed of AsianPF01, NLPR databases and part of Yale database[‡]. Most of the Asian faces are of Korean and Chinese origins. These face images contain variations in (slight) pose, illumination and expression. Some images in the original databases were not selected for our experiments because they either had out-of-plane rotation by more than 45 degrees or were occluded due to sunglasses or a scarf. Sample images from the databases are shown in Fig. 4.

Table 1. Database description.

Face database		No. of Subjects	Size of the Face Area	Variations included
Asian (132 subjects)	AsianPF01[26]	107	$\sim 180 \times 180$	Illum. and expr.
	NLPR	21	$\sim 160 \times 160$	Pose and expr.
	Yale[27]	4	$\sim 170 \times 170$	Illum. and expr.
Non-Asian (131 subjects)	AR[28]	120	$\sim 300 \times 300$	Illum. and expr.
	Yale[27]	111	$\sim 170 \times 170$	Illum. and expr.



(a) (b)

Figure 4. Representative faces in the database. (a) Asian; (b) Non-Asian.

Face images are cropped to include part of the face contour, the internal facial structures such as the eyebrows, eyes, nose and mouth. Since the LDA is an appearance-based scheme whose performance is sensitive to the alignment quality, all cropped face images are aligned by the centers of the two eyes. The intensity contrast is stretched to normalize for ambient lighting variations. Each image vector is normalized to be of unit length.

In our experiments, for each aligned face image, three different resolutions were utilized. The highest resolution is 42×42. It is decreased by $\sim 25\%$ to 32×32, and then further reduced to 24x24. The image is resized by using the nearest neighbor interpolation method. Figure 5 demonstrates an example. Shakhnarovich et al. used only the resolution of 24×24.

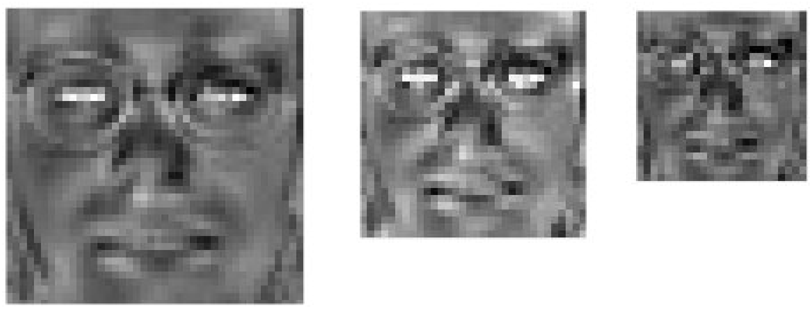
Each ethnic face database is randomly divided into two parts, two-thirds for the training set and one-third for the test set. Images from the same subject are grouped into the same set in order to eliminate the identity factor. This random partition is repeated 20 different times. The recognition accuracy reported here is the average of these 20 different test sets. In order not to make the training process biased to one ethnicity, the datasets are balanced, 132 subjects for Asian and 131 subjects for non-Asian. This also suggests that the class prior probabilities of two classes are almost equal (0.5).

[‡]Asian subjects in the Yale database were moved into the Asian database.

(a) (b) (c)

Figure 5. A face image in our database processed at three different resolutions. (a) 42×42; (b) 32×32; (c) 24×24.

Figure 6 shows the derived linear discriminant for Asian/Non-Asian classification from one of the randomly selected dataset partitions.



(a) (b) (c)

Figure 6. Fisher linear discriminant used for Asian/Non-Asian Classification, derived from the database used in our experiments at three resolutions. (a) 42×42; (b) 32×32; (c) 24×24.

For comparison purposes, the nearest neighbor rule is applied. The cosine value of the angle between the test image vector and the training image vector is calculated as the matching score. The training image with the best match is used to determine the classification result of the test image. The recognition performance of different ethnicity classification schemes implemented by us is shown in Fig. 7 for comparison, and listed in Table 2 for details.

Based on the performance evaluation, two major points are illustrated.

1. The LDA classifier on each individual scale outperforms the corresponding nearest neighbor classifier in terms of the total accuracy rate.
2. LDA ensembles outperform each LDA classifier at individual scales in terms of the total accuracy rate.

Note that: The nearest neighbor classifier tends to be biased to the Asian category. The NN classifier's accuracy rate for the non-Asian class is lower than that for the Asian class. However, the performance of LDA based classifier tends to be balanced between the two classes. The correctly classified and misclassified examples by the LDA ensemble are shown in Fig. 8 and Fig. 9. We also conducted the experiments, in which the LDA ensemble is applied to only two of the three different scales. The classification accuracy is slightly lower than LDA ensemble at three scales.

As different data preprocessing and alignment schemes are used,[23, 24] it is difficult to compare their performance with our proposed scheme. But these systems are still worth mentioning for completeness. In Ref. 23,

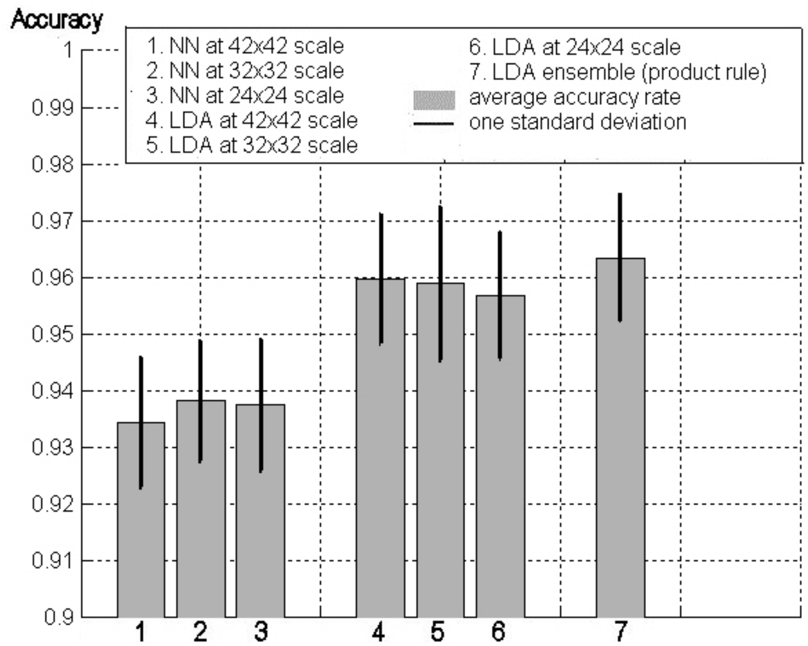


Figure 7. Performance comparison. (The accuracy represents the Total Accuracy in Table 2.)

Table 2. Classification performance (20-fold cross-validation). NN: nearest neighbor rule. There are three different resolutions, 42×42, 32×32, and 24×24. Each cell in the table has two numbers, the average accuracy rate and the standard deviation (in parentheses).

	Asian Accuracy	Non-Asian Accuracy	Total Accuracy
NN at 42×42 scale	97.7% (0.014)	89.3% (0.021)	93.5% (0.011)
NN at 32×32 scale	97.6% (0.015)	90.0% (0.017)	93.8% (0.011)
NN at 24×24 scale	97.2% (0.015)	90.3% (0.021)	93.8% (0.012)
LDA at 42×42 scale	95.8% (0.025)	96.1% (0.019)	96.0% (0.012)
LDA at 32×32 scale	95.6% (0.027)	96.2% (0.017)	95.9% (0.014)
LDA at 24×24 scale	95.7% (0.022)	95.6% (0.017)	95.7% (0.011)
LDA ensemble (product rule)	96.0% (0.025)	96.6% (0.014)	96.3% (0.011)

the experimental results yielded an average accuracy rate of 92% for the ethnic classification task, involving 4 categories (Caucasian, South Asian, East Asian, and African). Their database consisted of 3006 frontal face images. Asian subjects (South Asian + East Asian) comprised only $\sim 27\%$ of the dataset, which is not balanced to the non-Asian category. The system developed by Shakhnarovich et al.[24] is focused at low image resolution (24x24) with faces weakly aligned. Therefore, their reported accuracy is only about 80%. The datasets collected by them are not available in the public domain.

4. CONCLUSIONS AND FUTURE WORK

This paper has addressed the ethnicity identification problem based on facial images. The Linear Discriminant Analysis based scheme has been developed for the two-class (Asian vs. non-Asian) ethnicity classification task. An ensemble framework, which integrates the LDA analysis for the input face images at multiple scales, is proposed to further improve the classification performance of the ethnicity identification system. The product rule is implemented as the integration strategy to combine the outputs of individual classifiers at different scales.

(a) (b)

Figure 8. Correctly classified examples by the proposed LDA ensemble at three scales. (a) Asian; (b) Non-Asian.

(a) (b)

Figure 9. Misclassified examples by the proposed LDA ensemble at three scales. (a) Asian faces misclassified as non-Asians; (b) Non-Asian faces misclassified as Asians.

Extensive cross-validation was conducted on the available database to estimate the classification accuracy of the proposed scheme. Experimental results based on a face database containing 263 subjects (2,630 face images, with equal number between the two classes) are encouraging, illustrating that LDA has the discriminative power for the presented ethnicity classification problem. Further, both the combination strategies for the LDA based ensemble lead to more accurate classification than that made by any one of the individual classifiers. Ethnicity classification from face images can not be solved completely by linear classifiers due to the complexity of facial appearance in real-world situations. More advanced classifiers, such as support vector machine,[29] and higher-level feature representation, such as landmark points, may be necessary to improve the performance of ethnicity classification problem.

The normalized classification scores can be used as the confidence with which each image belongs to an ethnic class. This confidence is helpful to the image-based face recognition, and cross-race face recognition. Separating the race factor from the other factors can help the recognition system to extract more identity-sensitive features, thereby enhancing the performance of the current face identity recognition systems. The proposed ensemble scheme can also be applied to other image-based classification tasks.

Acknowledgments

This research was supported by the Office of Naval Research, contract grant No. N00014-01-1-0266. The authors would like to thank Dr. Yunhong Wang for providing us the NLPR database.

REFERENCES

1. L. Farkas, *Anthropometry of the Head and Face*, Raven Press, 2nd ed., 1994.

2. D. Enlow, *Facial Growth*, W.H. Saunders, 3rd ed., 1990.
3. R. Malpass and J. Kravitz, “Recognition for faces of own and other race,” *J. Perc. Soc. Psychol.* **13**, pp. 330–334, 1969.
4. J. Brigham and P. Barkowitz, “Do ’they all look alike?’ the effect of race, sex, experience and attitudes on the ability to recognize faces,” *J. Appl. Soc. Psychol* **8**, pp. 306–318, 1978.
5. A. Golby, J. Gabrieli, J. Chiao, and J. Eberhardt, “Differential responses in the fusiform region to same-race and other-race faces,” *Nature Neuroscience* **4**(8), pp. 845–850, 2001.
6. A. Puce, T. Allison, J. Gore, and G. McCarthy, “Face-sensitive regions in human extrastriate cortex studied by functional mri,” *J. Neurophysiol.* **74**, pp. 1192–1199, 1995.
7. A. O’Toole, K. Deffenbacher, D. Valentin, and H. Abdi, “Structural aspects of face recognition and the other race effect,” *Memory & Cognition* **22**, pp. 208–224, 1994.
8. A. O’Toole, A. Peterson, and K. Deffenbacher, “An other-race effect for classifying faces by sex,” *Perception* **25**, pp. 669–676, 1996.
9. W. Zhao, R. Chellappa, A. Rosenfeld, and P. Phillips, “Face recognition: A literature survey,” *CVL Technical Report, University of Maryland* , Oct. 2000. <ftp://ftp.cfar.umd.edu/TRs/CVL-Reports-2000/TR4167-zhao.ps.gz>.
10. Viisage, *<http://www.viisage.com/>*. Littleton, MA.
11. Identix, *<http://www.identix.com/>*. Minnetonka, MN.
12. 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* **19**, pp. 711–720, Jul. 1997.
13. M. Turk and A. Pentland, “Eigenfaces for recognition,” *Journal of Cognitive Neuroscience* **3**, pp. 71–86, Mar. 1991.
14. P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, “The feret evaluation methodology for face-recognition algorithms,” *IEEE Trans. Pattern Analysis and Machine Intelligence* **22**(10), pp. 1090–1104, 2000.
15. M. Kirby and L. Sirovich, “Application of the Karhunen-Loéve procedure for the characterization of human faces,” *IEEE Trans. Pattern Analysis and Machine Intelligence* **12**, pp. 103–108, Jan. 1990.
16. B. Moghaddam and M. Yang, “Learning gender with support faces,” *IEEE Trans. Pattern Analysis and Machine Intelligence* **24**, pp. 707–711, May. 2002.
17. T. Lindeberg, *Scale-Space Theory in Computer Vision*, Kluwer Academic Publishers, Dordrecht, Netherlands, 1994.
18. J. Kittler, M. Hatef, R. Duin, and J. Matas, “On combining classifiers,” *IEEE Trans. Pattern Analysis and Machine Intelligence* **20**(3), pp. 226–239, 1998.
19. R. Brunelli and D. Falavigna, “Person identification using multiple cues,” *IEEE Trans. Pattern Analysis and Machine Intelligence* **17**, pp. 955–966, Oct. 1995.
20. A. K. Jain, S. Prabhakar, and S. Chen, “Combining multiple matchers for a high security fingerprint verification system,” *Pattern Recognition Letters* **20**(11-13), pp. 1371–1379, 1999.
21. L. Hong and A. Jain, “Integrating faces and fingerprint for personal identification,” *IEEE Trans. Pattern Analysis and Machine Intelligence* **20**(12), pp. 1295–1307, 1998.
22. B. Glolomb, D. Lawrence, and T. Sejnowski, “Sexnet: A neural network identifies sex from human faces,” in *Advances in Neural Information Processing Systems (NIPS)*, **3**, pp. 572–577, 1990.
23. S. Gutta, J. Huang, P. Phillips, and H. Wechsler, “Mixture of experts for classification of gender, ethnic origin, and pose of human faces,” *IEEE Trans. Neural Networks* **11**, pp. 948–960, Jul. 2000.
24. G. Shakhnarovich, P. A. Viola, and B. Moghaddam, “A unified learning framework for real time face detection and classification,” in *Proc. IEEE International Conference on Automatic Face and Gesture Recognition*, 2002.
25. R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*, Wiley, New York, 2nd ed., 2000.
26. *Asian face database PF01.* <http://nova.postech.ac.kr/archives/imdb.html>.
27. *Yale University face database.* <http://cvc.yale.edu/projects/yalefaces/yalefaces.html>.
28. A. Martinez and R. Benavente, “The ar face database,” *CVC Tech. Report # 24* , Jun. 1998.
29. V. Vapnik, *Statistical Learning Theory*, Wiley, NY, 1998.