Resampling for Face Recognition*

Xiaoguang Lu and Anil K. Jain

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

Abstract. A number of applications require robust human face recognition under varying environmental lighting conditions and different facial expressions, which considerably vary the appearance of human face. However, in many face recognition applications, only a small number of training samples for each subject are available; these samples are not able to capture all the facial appearance variations. We utilize the resampling techniques to generate several subsets of samples from the original training dataset. A classic appearance-based recognizer, LDA-based classifier, is applied to each of the generated subsets to construct a LDA representation for face recognition. The classification results from each subset are integrated by two strategies: majority voting and the sum rule. Experiments conducted on a face database containing 206 subjects (2,060 face images) show that the proposed approaches improve the recognition accuracy of the classical LDA-based face classifier by about 7 percentages.

1 Introduction

Human face recognition has been drawing a lot of attention in the past decade. A number of face recognition algorithms have been investigated [21] and several commercial face recognition products [9][20] are available. However, robust face recognition in unconstrained environments is still a very challenging problem.

A face recognition system has two stages, training and test. In real applications, current face recognition systems encounter difficulties due to the small number of available training face images and complicated facial variations during the testing stage. In other words, available training samples are not very representative. Human face appearance has a lot of variations resulting from varying lighting conditions, different head poses and facial expressions. Examples of these variations for one subject are illustrated in Fig. 1. In real-world situations, only a small number of samples for each subject are available for training. These samples cannot capture all the possible facial variations.

Among the face recognition algorithms, appearance-based approaches [2][19], which utilize the intensity or intensity-derived features of original images, have been successfully developed [21][13]. The dimensionality of the feature vector

* This research was supported by NSF IUC on Biometrics (CITeR), at West Virginia University.

used by these methods is often very high while the training sample size is relatively small. The classifier based on such training data may be biased and have a large variance, resulting in a poor performance [10][17]. To improve the performance of the weak classifiers, a number of approaches have been presented [4][6][8]. Breiman [4] proposed a bootstrapping-aggregating (bagging) method. The training set is randomly resampled with replacement to generate independent bootstrap replicates. A classifier is developed based on each replicate. Finally, the majority voting is applied to integrate results of all the classifiers. Freund and Schapire [6] have developed a boosting algorithm, which trains a series of classifiers based on the reweighted training set in a sequential mode. The final decision is made by majority voting. In the random subspace method [8], classifiers are constructed in the random subspaces of the feature space. Simple majority voting is used as the final decision rule. Skurichina and Duin analyzed these methods for linear classifiers [17][16]. Bolle et al. [3] used the bootstrap techniques for evaluating authentication systems. The boosting has been used to several applications, such as text categorization [15] and image retrieval [18]. Guo and Zhang [7] applied boosting for fast face recognition.

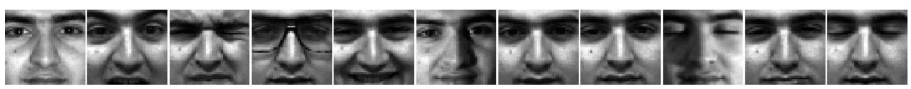

Fig. 1. Facial variations under different lighting conditions and facial expressions for the same subject [1]

We propose a resampling-integration scheme for face recognition. A resampling technique is utilized to generate a number of subsets from the original training dataset. A classic appearance-based face recognizer based on the LDA representation is constructed on each of the generated subsets. Two integration strategies, majority voting and the sum rule, are used to combine the classification results to arrive at the final decision.

In section 2, the resampling and integration scheme is presented. Section 3 provides the experimental results and discussion. Conclusions are summarized in section 4.

2 Resampling and Integration

2.1 System Overview

Our resampling-integration scheme is illustrated in Fig. 2. The training dataset contains a small number of sample face images. A number of subsets are generated by resampling the training set. Each subset S_i is used to train a classifier C_i. In the test stage, the test image is loaded into each component classifier. Two strategies, (i) simple majority voting, and (ii) the sum rule, are used to integrate the outputs of component classifiers; the classifier outputs can be either the classification labels or the matching scores. Currently, the face recognizer based on

LDA representation is used as the component classifier, but this framework does not limit the component classifiers to be of the same type.

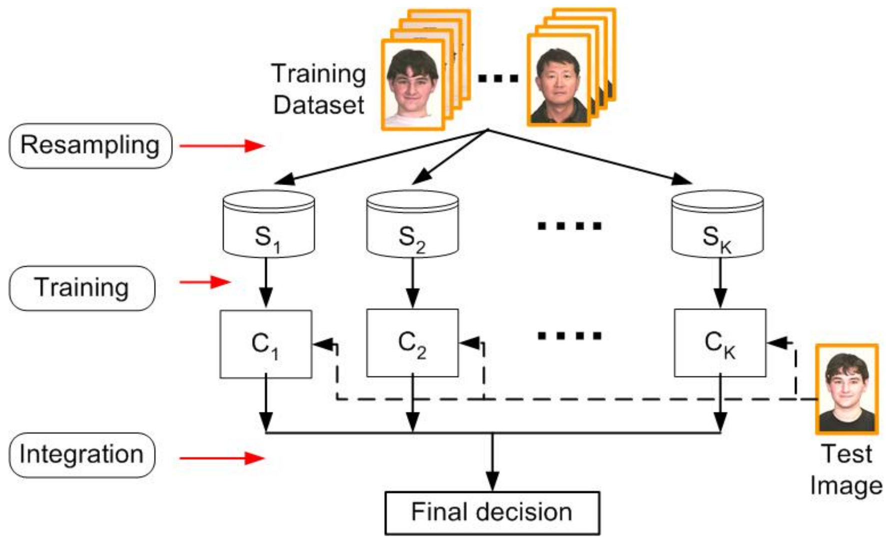

Fig. 2. The Resampling-Integration scheme for face recognition. S_1 to S_K are the subsets resampled from the original training dataset. C_1 to C_K are classifiers trained using the corresponding subsets. Here, K is the total number of subsets

2.2 LDA-based Face Classifier

A two-dimensional face image is considered as a vector, by concatenating each row (or column) of the image. Let $X = (x_1, x_2, \ldots, x_i, \ldots, x_N)$ denote the data matrix, where 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 represents the total number of pixels in the face image and $n = p \times p$. The Linear Discriminant Analysis (LDA) [5][2] representation is a linear transformation from the original image vector to a projection feature vector, i.e.

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

where Y is the $d \times N$ feature vector matrix, d is the dimension of the feature vector, $d \ll n$ and W_{LDA} is the transformation matrix, 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,

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

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

In the above expression, N_i is the number of training samples in class i; c is the number of distinct classes; m is the mean vector of all the samples, i.e., $m = \sum_{i=1}^{N} x_i$; m_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 facts 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 [19] can be used to reduce the dimensionality of the original face image space [2] prior to applying LDA.

LDA derives a low dimensional representation of a high dimensional face feature vector space. The face vector is projected by the transformation matrix W_{LDA}. The projection coefficients are used as the feature representation of each face image. The matching score between the test face image and the training image is calculated as the cosine value of the angle between their coefficients vectors. A larger matching score means a better match.

2.3 Resampling

The resampling module generates a number of subsets from the original training set. A number of resampling methods have been proposed in the literatures. For instance, in classic bagging [4], a random sampling with replacement is used to generate independent bootstrap replicates where the size of the subset is the same as that of the original set. In the LDA based face recognition, both intra- and inter-class information (between-class scatter matrix and within-class scatter matrix) are utilized, so our sampling strategy does not randomly sample the whole training set, but does randomly sampling within each class (subject), subject to the following conditions:

1. The number of sample images for each subject in the subset is equal or as equal as possible.
2. Sampling within each class is achieved based on a uniform distribution.

The requirements listed above may not be the optimal ones, but work well as demonstrated by the empirical evaluation.

2.4 Integration

After resampling, several LDA-based classifiers are constructed. The matching scores between the test face image and the training images are computed by each component classifier. Let $MS(i, j)$ be the matching score between the test image and the j^{th} training image, calculated by the i^{th} component classifier. For the i^{th} component classifier, the classification result for the test image is the subject label, denoted by $Label(i)$. This classification can be achieved by the nearest neighbor rule.

Two strategies for integration are applied, namely the simple majority voting and the sum rule.

1. Simple majority voting
   Assign the test image with the label which appears most frequently in $Label(i)$, where $i = 1 \dots K$.
2. The sum rule
   Calculate $MS_j = \sum_{i=1}^{K} MS(i,j)$. Assign the test image with the label of the J^{th} training image, such that

$$J = \arg\max_{j} MS_j. \tag{5}$$

The integration rules may not give desired results when the number of component classifiers (K) is too small. But due to the resampling scheme presented, here K could be as large as needed.

3 Experiments and Discussion

Our database is a union of four different face databases, which are available in the public domain (see table 1). It contains 2,060 face images of 206 subjects, with 10 images per subject. The set of face images contains variations in pose, illumination and expression. Some images in the individual 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 sun glasses or a scarf. Sample images from the databases are shown in Fig. 3. Face images are closely cropped to include only the internal facial structures such as the eyebrows, eyes, nose and mouth, and aligned by the centers of the two eyes. All cropped images are resized to 42×42 pixels. Each image vector is normalized to be of unit length.

Table 1. Database description

Face database	number of subjects	Variations included
ORL [14]	40	Slight pose and expression
Yale [1]	15	Illumination and expression
AR [12]	120	Illumination and expression
NLPR+MSU (collected by the authors)	31	Slight pose and expression

The entire face database is divided into two parts. Nine images of each subject are used to construct the original training data and the remaining one is used for testing. This partition is repeated 10 different times so that every image of the subject can be used for testing. The recognition accuracy is the average of these ten different test sets. In resampling, 8 of 9 images for each subject are randomly selected according to the uniformly distributed seeds between 1 and 9. The sampling is without replacement. Each subject has different random generated

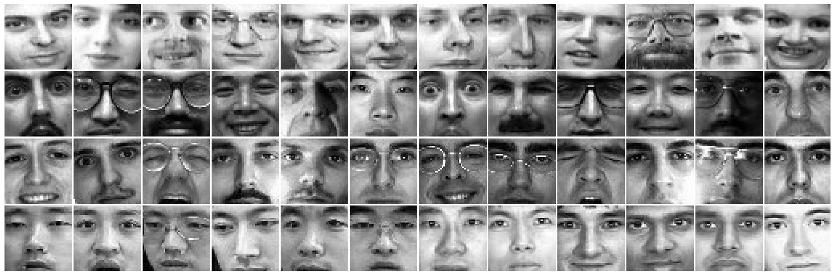

Fig. 3. Representative face images in the database. (a) ORL, (b) Yale, (c) AR and (d) NLPR+MSU

seeds. Consequently, each resampled subset contains 8x206=1,648 images. The LDA-based classifier is trained on this subset.

The component classifiers compute the cosine value of the angle between the two projection coefficients vectors (one from the test image and the other from the database image) as the matching score. Database image with the best match is used to determine the classification result of the input image from the component classifier. The recognition accuracy of different face recognition schemes is listed in table 2. Figure 4 shows some images which were misclassified by the classic LDA-based face recognizer but correctly classified using the proposed scheme.

Table 2. Recognition accuracy (The number of resampled subsets, $K = 20$.)

Without resampling	Resampling + Majority Voting	Resampling + Sum rule
81.0%	88.7%	87.9%

Fig. 4. Examples which are misclassified by classic LDA-based face recognizer but correctly classified using the proposed scheme

The number of subsets, K, is decided empirically. In order to analyze the influence of K in our scheme, we conducted experiments with different settings of K values, from 1 to 20. Figure 5 demonstrates the recognition accuracy of the proposed recognition schemes as the number of subsets changes. These results show that the proposed resampling-integration scheme generally improves the performance of the LDA-based face classifier as K increases up to 20.

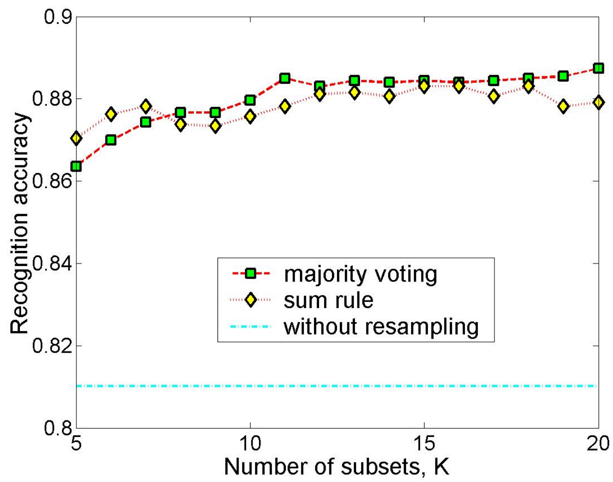

Fig. 5. Recognition accuracy with respect to the number of subsets

4 Conclusions and Future Work

The resampling-integration scheme is proposed to improve the recognition accuracy of face classification. The resampling is designed to generate a number of subsets, which are used to train the component classifiers (parameter adjustment). The integration rules are applied to combine the outputs of component classifiers for the final decision. Two integration rules are presented and corresponding experiments are carried out. The LDA-based face classifier is integrated into the scheme with the corresponding resampling design. Experiments conducted on a face database containing 206 subjects (2,060 face images) show that the recognition accuracy of the classical LDA-based face classifier is improved by applying the proposed scheme. The system framework is scalable in terms of the number of subsets, the type of component classifiers and resampling techniques.

Different resampling techniques can be explored in this scheme. Since the subsets are resampled randomly, the resulting component classifiers may have different weights in the final decision. Some classifier selection technique can be applied. Although in our experiments, all component classifiers are LDA-based, the presented scheme does not limit the type of the component classifier. However, currently there is no guarantee that the proposed scheme always works for any type of classifiers. From the perspective of classifier combination, many other integration rules can be tried out [11].

References

1. *Yale University face database.* <http://cvc.yale.edu/projects/yalefaces/yalefaces.html>.
2. 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(7):711–720, Jul. 1997.
3. R. Bolle, N. Ratha, and S. Pankanti. Evaluating authentication systems using bootstrap confidence intervals. In *Proc. 1999 IEEE Workshop on Automatic Identification Advanced Technologies*, pages 9–13, Morristown NJ, 1999.
4. Leo Breiman. Bagging predictors. *Machine Learning*, 24(2):123–140, 1996.
5. R. O. Duda, P. E. Hart, and D. G. Stork. *Pattern Classification.* Wiley, New York, 2nd edition, 2000.
6. Yoav Freund and Robert E. Schapire. Experiments with a new boosting algorithm. In *Proc. International Conference on Machine Learning*, pages 148–156, 1996.
7. Guo-Dong Guo and Hong-Jiang Zhang. Boosting for fast face recognition. In *Proc. IEEE ICCV Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems*, pages 96–100, 2001.
8. T. K. Ho. The random subspace method for constructing decision forests. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 20(8):832–844, 1998.
9. Identix. *<http://www.identix.com/>*. Minnetonka, MN.
10. A. K. Jain and B. Chandrasekaran. Dimensionality and sample size considerations in pattern recognition practice. *Handbook of Statistics, P. R. Krishnaiah and L. N. Kanal (eds.)*, 2:835–855, 1987.
11. J. Kittler, M. Hatef, R. Duin, and J. Matas. On combining classifiers. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 20(3):226–239, 1998.
12. A.M. Martinez and R. Benavente. The ar face database. *CVC Tech. Report # 24*, Jun. 1998.
13. P. Jonathon Phillips, Hyeonjoon Moon, Syed A. Rizvi, and Patrick J. Rauss. The feret evaluation methodology for face-recognition algorithms. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 22(10):1090–1104, 2000.
14. Ferdinando Samaria and Andy Harter. Parameterisation of a stochastic model for human face identification. In *Proc. 2nd IEEE Workshop on Applications of Computer Vision*, Sarasota FL, Dec. 1994.
15. R.E. Schapire and Y. Singer. Boostexter: A boosting-based system for text categorization. *Machine Learning*, 39(2-3):135–168, May/June 2000.
16. M. Skurichina and R.P.W. Duin. Bagging for linear classifiers. *Pattern Recognition*, 31(7):909–930, 1998.
17. M. Skurichina and R.P.W. Duin. Bagging, boosting and the random subspace method for linear classifiers. *Pattern Analysis and Applications*, 5(2):121–135, 2002.
18. K. Tieu and P. Viola. Boosting image retrieval. In *Proc. CVPR*, 2000.
19. M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, Mar. 1991.
20. Viisage. *<http://www.viisage.com/>*. Littleton, MA.
21. W. Zhao, R. Chellappa, A. Rosenfeld, and P.J. Phillips. Face recognition: A literature survey. *CVL Technical Report, University of Maryland*, October 2000. <ftp://ftp.cfar.umd.edu/TRs/CVL-Reports-2000/TR4167-zhao.ps.gz>.