Face Recognition – Combine Generic and Specific Solutions

Jie Wang, Juwei Lu, K.N. Plataniotis, and A.N. Venetsanopoulos

Department of Electrical and Computer Engineering, University of Toronto,
10 King's College Road, Toronto, Canada M5S 3G4
{jwang, juwei, kostas, anv}@dsp.utoronto.ca

Abstract. In many realistic face recognition applications, such as surveillance photo identification, the subjects of interest usually have only a limited number of image samples a-priori. This makes the recognition a difficult task, especially when only one image sample is available for each subject. In such a case, the performance of many well known face recognition algorithms will deteriorate rapidly and some of the algorithms even fail to apply. In this paper, we introduced a novel scheme to solve the one training sample problem by combining a specific solution learned from the samples of interested subjects and a generic solution learned from the samples of many other subjects. A multi-learner framework is firstly applied to generate and combine a set of generic base learners followed by a second combination with the specific learner. Extensive experiments based on the FERET database suggests that in the scenario considered here, the proposed solution significantly boosts the recognition performance.[1]

1 Introduction

Face recognition (FR) which has many realistic applications such as forensic identification, access control and human computer interface receives more and more attentions in both the academic and industrial areas. However it is still a difficult problem far from well solved since face objects usually exhibit various appearance due to aging, illumination and pose variations. Furthermore, image samples available for training are usually limited. Particularly, if only one image sample per subject is available, the problem becomes even more challenging.

In literature, many state-of-the-art FR algorithms have been proposed and the recent surveys could be found in[1] [2]. Among various face recognition techniques, appearance based approach which treats the face image as a holistic pattern is one of the most attractive methodologies [3]. A 2D face image is treated as a vector in the high dimensional image space and the subject identification

[1] This work is partially supported by a Bell University Lab research grant and CITO Student Internship Program. The authors would like to thank the FERET Technical Agent, the U.S. National Institute of Standards and Technology for providing the FERET database.

M. Kamel and A. Campilho (Eds.): ICIAR 2005, LNCS 3656, pp. 1057–1064, 2005.

is performed by applying statistical classification methodologies, among which principle component analysis (PCA)[4], an unsupervised technique, and linear discriminant analysis (LDA)[5][6][7], a supervised technique, are most commonly used. It is generally believed that the supervised techniques are superior to those unsupervised ones for classification purposes. However, such techniques are more susceptible to the so-called "small sample size" problem, where the number of the training samples is much smaller than the dimensionality of the samples. The problem will be particularly severe when only one training sample is available for each subject. In such a case, the intra-subject information cannot be estimated which makes the supervised learning technique such as LDA based algorithms fail to apply. Thus training an unsupervised learner seems to be the only choice. However, unsupervised learning techniques are not optimal for classification tasks [5], furthermore, due to the fact that only limited number of samples are available, the estimation of the statistical model is not reliable, resulting in a poor performance.

In this paper, we proposed a scheme to solve the one sample problem by combining a generic and a specific solution. A generic FR system is built on a generic database. It is assumed that the subjects contained in the generic database do not overlap those to be identified in a specific FR task. Therefore, a generic FR system which is built to classify the generic subjects could be generalized to identify the unseen subjects in a specific FR task. This is based on a reasonable assumption, that human faces share similar intra-subject variations[8]. Thus discriminant information of the specific subjects (those to be identified) can be learned from other. It is also a realistic solution since a reasonably sized generic database is always existed. Therefore, without the one sample limitation, supervised learning techniques can be applied on the generic database. It is well known that supervised techniques are class specific and the learner which is optimal for the trained subjects may not work well with those specific subjects which are not included in the training session. In order to improve the generic behavior of the supervised algorithms and enhance the generalization power, a multi-learner framework is introduced. Generic FR system is formed by combining a set of base generic FR subsystems which are trained on different generic subsets. Since the generic learner does not target at the specific subjects, it provides a bias solution for a specific FR task. In order to further improve the recognition performance, a specific FR system is built on those specific subject images (1 image per subject) by using an unsupervised leaning algorithm. The final identification is performed by aggregating the output from both the generic and specific FR systems. Extensive experimentations on the FERET database [9] indicate that the proposed algorithm significantly improves the performance under the considered scenario which is often encountered in practice.

The rest of the paper is organized as follows: Section 2 introduces the system framework. The generic and the specific learners are described in section 3 and section 4 respectively while their combination is discussed in section 5. Experimental results obtained by using the FERET database are given in section 6 followed by the conclusion drawn in section 7.

2 System Framework

In order to facilitate the presentation, some terminologies are defined. Let $GalD$ be the gallery set containing the subjects of interest with the identity labels, one frontal image sample per subject. Let $PrbD$ be the probe set which includes the face images to be identified. It is assumed that there is no overlap between gallery samples and probe samples. Thus the task of a FR system is to determine which gallery subject the probe image belongs to. A generic database, denoted as $GenD$ is collected elsewhere. The subjects included in the $GenD$ do not overlap with those in the gallery set and the probe set.

In the training session, a number of generic subsets are generated from the generic database. Each training subset contains the image samples of T subjects which are selected randomly from the total subjects in generic database without replacement. With each training subset, a corresponding base generic learner, denoted as $H_G^k, k = 1, ..., M$, is built which includes a feature extractor and a classifier. Similarly, the specific learner is generated from the gallery images, denoted as H_S. While in the operation session, both the probe $\mathbf{p}$ and the gallery samples are inputed to the base generic learners and the specific learner. A generic recognition result is obtained by aggregating the results from each base learners which is denoted as the level 1 combination. The final determination is performed by combining of the generic result and the specific result, which is denoted as level 2 combination. The system framework is depicted in Fig.1

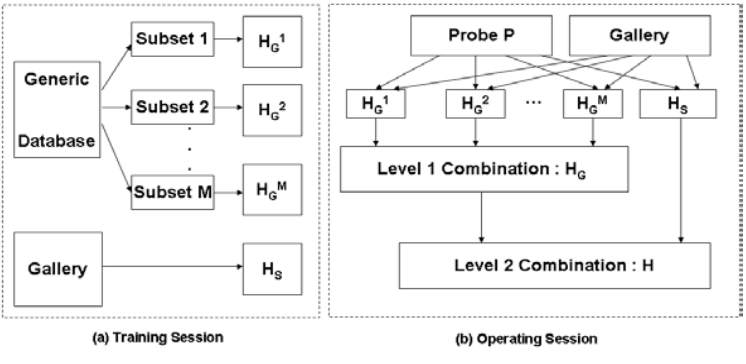

Fig. 1. System Framework

3 Generic Learner

3.1 Multiple Base Generic Learners

Let $GenD$ be the generic set of size $C \times L$ containing C subjects L images each. $\mathbf{t}_{i,j}$ is the jth image of subject i, $i = 1, ..., C, j = 1, ..., L$. M generic training subsets are generated from $GenD$, each of which contains T subjects

randomly selected from all C subjects in the $GenD$ without replacement. Let $SGenD_k$ be the kth training subset containing T subjects, L images each, where $k = 1, ..., M$. Therefore, a base generic learner, denoted as H_G^k, is trained on the subset $SGenD_k$.

In appearance approach, a learner is generally formed by a feature extractor and a classifier. Since the generic database is collected elsewhere, it is reasonable to assume that at lease two image samples are available for each generic subject. Therefore, supervised techniques can be applied. In this paper, direct linear discriminant analysis (DLDA)[10] is selected as generic feature extractor due to its good performance. Linear discriminant analysis (LDA) and its variants[5][6][10] provide class specific solutions by maximizing the so called Fisher's criterion, i.e., the ratio of the between- and within-class scatters are maximized, $A = \arg\max_A \frac{|A^T S_b A|}{|A^T S_w A|}$, where S_b and S_w are the between- and within-class scatter matrices of the training samples respectively and A is the optimal transformation matrix from the original image space to the feature space. Direct LDA procedure solves the above optimization problem by firstly diagonalizing the between-class scatter followed by diagonalizing the within-class scatter. However, in the SSS scenario, the variance of the estimation of the small eigenvalues of S_w increases significantly resulting in exaggerating the importance of the corresponding eigenvectors. Therefore, a modified but equivalent criterion is utilized, i.e., $A = \arg\max_A \frac{|A^T S_b A|}{|A^T S_w A + A^T S_b A|}$[7]. Following the feature extractor, nearest center classifier is selected to determine the probe identity by calculating the distance between the probe and each gallery subject in the extracted feature subspace. The identity of the probe is therefore determined as the one with the smallest distance.

Let A_G^k be the transformation matrix obtained from the generic training subset $SGenD_k$. Let $GalD$ be the gallery set, containing of H image samples $\mathbf{s}_i, i = 1, ..., H$, one per subject, thus the generic base learner H_G^k outputs the the probe identity as follows:

$$H_G^k(\mathbf{p}) = \arg\min_i D_G^k(\mathbf{p}, \mathbf{s}_i) \qquad D_G^k(\mathbf{p}, \mathbf{s}_i) = ||(A_G^k)^T\mathbf{p}, (A_G^k)^T\mathbf{s}_i|| \tag{1}$$

where D_G^k denotes the distance of the probe and the gallery subject in the feature subspace specified by A_G^k, and $||.||$ is the distance metric. In this paper, Euclidean distance is selected for DLDA extracted feature space.

In addition to the probe label, each base learner also makes a soft decision by providing a membership score $R_G^k(\mathbf{p}, \mathbf{s}_i)$ which indicates how the probe $\mathbf{p}$ belongs to the gallery subject $\mathbf{s}_i$. The larger the score, the higher possibility the probe belongs to the subject $\mathbf{s}_i$. Therefore, we define the membership score as follows, i.e.,

$$R_G^k(\mathbf{p}, \mathbf{s}_i) = (D_{Gmax}^k - D_G^k(\mathbf{p}, \mathbf{s}_i))/(D_{Gmax}^k - D_{Gmin}^k) \tag{2}$$

$$D_{Gmax}^k = max(\{D_G^k(\mathbf{p}, \mathbf{s}_i)\}_{i=1}^H) \qquad D_{Gmin}^k = min(\{D_G^k(\mathbf{p}, \mathbf{s}_i)\}_{i=1}^H)$$

With such definition, small distance results in high membership score and vice versus. Therefore, the identity of the probe is equivalent to that with the highest membership score, i.e., $H_G^k(\mathbf{p}) = \arg\max_i R_G^k(\mathbf{p}, \mathbf{s}_i)$.

3.2 Combine Base Learners – Level 1 Combination

In order to combine multiple learners, many combination policies are developed in literature[11]. In this paper, sum rule is selected to combine the generic base learners for its simplicity and robust performance.

The final score, denoted as R_G, is therefore the summation of the scores obtained by all base learners and the identity is the one with the highest value,i.e., $R_G(\mathbf{p}, \mathbf{s}_i) = \sum_{k=1}^{M} R_G^k(\mathbf{p}, \mathbf{s}_i) \qquad H_G(\mathbf{p}) = \arg\max_i R_G(\mathbf{p}, \mathbf{s}_i)$.

4 Specific Learner

The specific learner, denoted as H_S, is trained on the gallery set, where each subject only has one image sample. Therefore unsupervised learning techniques are selected. In this paper, PCA is adopted as the specific feature extractor resulting in a specific feature space specified by A_S, while the classifier is again the nearest center classifier. The membership score provided by the specific learner, $R_S(\mathbf{p}, \mathbf{s}_i)$, is defined in a similar way, i.e., $R_S(\mathbf{p}, \mathbf{s}_i) = (D_{Smax} - D_S(\mathbf{p}, \mathbf{s}_i))/(D_{Smax} - D_{Smin})$, $D_{Smax} = max(\{D_S(\mathbf{p}, \mathbf{s}_i)\}_{i=1}^{H})$ and $D_{Smin} = min(\{D_S(\mathbf{p}, \mathbf{s}_i)\}_{i=1}^{H})$, where $D_S(\mathbf{p}, \mathbf{s}_i)$ is the distance between probe $\mathbf{p}$ and gallery subject $\mathbf{s}_i$ in the specific feature space A_S. Here, Mahalanobis distance is selected for the PCA based feature subspace due to its good performance. Correspondingly, the probe identity is determined as: $H_S(\mathbf{p}) = \arg\max_i R_S(\mathbf{p}, \mathbf{s}_i)$.

5 Combine Generic and Specific Learners – Level 2 Combination

The generic leaner, trained on the samples of generic subjects, is usually bias the optimal one for a specific recognition task, since it does not target at the subjects of interest. On the other hand, the specific learner is exactly trained on the subjects of interest, however, due to the limited sample size, the estimation relies heavily on the gallery samples, giving rise to high variance. Therefore it is reasonable to combine these two learners by using a regularization factor η to balance the bias and variance. Here, we propose to combine the generic and specific learners with the following form:

$$R(\mathbf{p}, \mathbf{s}_i) = \eta R_G(\mathbf{p}, \mathbf{s}_i) + (1 - \eta) R_S(\mathbf{p}, \mathbf{s}_i) \qquad H(\mathbf{p}) = arg \max_i R(\mathbf{p}, \mathbf{s}_i) \quad (3)$$

where η is the regularization factor, $0 \leq \eta \leq 1$, $R_G(.)$ and $R_S(.)$ are the membership scores provided by the generic and the specific learner and $R_G(.)$ has been normalized to 0-1. It is observed that if $\eta = 0$, the final learner results in the specific leaner which exhibits large variance. When $\eta = 1$, only generic learner affects the performance resulting in a biased solution.

6 Experiments

6.1 Experiment Setup

A set of experiments are performed on the FERET database. In the current FERET database, 3817 face images of 1200 subjects are provided with the eye coordinates information which is required to align and normalize the images. In all experiments reported here, images are preprocessed following the FERET protocol guidelines: (1) images are rotated and scaled so that the centers of the eyes are placed on specific pixels and the image size is normalized to 150×130; (2) a standard mask is applied to remove non-face portions; (3) histogram equalization is performed and image intensity values are normalized to zero mean and unit standard deviation; (4) each image is finally represented, after the application of mask, as a vector of dimensionality 17154.

Among these 1200 subjects, there exist 226 subjects with 3 images per subject. These 678 images are used to form the generic training database. In addition, there are 1097 images of 207 subjects each of which has 4-9 images. Of these images, we randomly select 207 frontal images, one per subject, to form the gallery set while the remaining 890 images are treated as probes.

For specific learner, PCA is applied for feature extraction denoted as H_{S-PCA}. As for DLDA based generic learner,H_{G-DLDA}, it is formed by the combination of 50 base learners generated from 50 different generic training subsets, each of which has H subjects, where H is varied from 30 to 110 with the interval of 10. For comparison purposes, two single generic learners trained on the whole generic training set are also generated by using PCA and DLDA respectively, denoted as $H_{G-PCA-Single}$ and $H_{G-DLDA-Single}$.

6.2 Results and Analysis

The comparison of the correct recognition rate (CRR) obtained by the single generic learners ($H_{G-PCA/DLDA-Single}$) and the combination of multiple base generic learners (output of level 1 combination,H_{G-DLDA}) is depicted in Fig.2(a). As for the single generic learners, the best CRRs are utilized for comparison. It is well-known that CRR is a function of feature number and the best found CRR is the one with the peak value corresponding to the optimal feature number (M^*) which is obtained by exhaustively searching all possible feature numbers. In addition, the comparison of the CRRs obtained by the specific learner (H_{S-PCA}), generic learner (H_{G-DLDA}) and their combination ($H_{S-PCA} + H_{G-DLDA}$) is depicted in Fig.2(b). It can be observed from Fig.2(a) that the introduced multi-learner framework improves the recognition performance with respect to the single generic learner. Fig.2(b) indicates that the combination of the generic and specific solution further boosts the recognition performance, outperforming either of them.

Fig.3(a) depicts the effect that the regularization factor η has on the recognition performance. It can be observed that the best performance is between $\eta = 0$ and $\eta = 1$. The result is consistent with our claim that balancing the

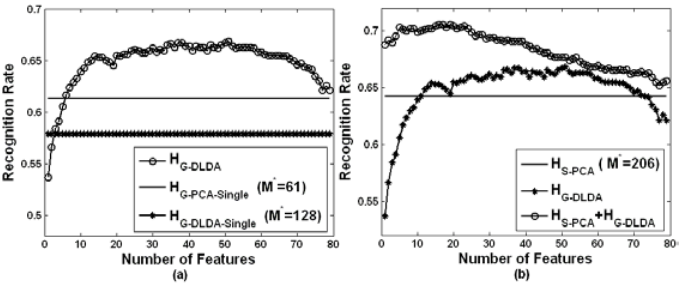

Fig. 2. (a) CRRs obtained by the single generic learners and the combination of multiple base learners v.s. feature number for each base learner; (b) CRRs obtained by the generic, specific learners and their combination v.s. feature number for each base learner, $\eta = 0.6$; Each base learner is trained with 80 subjects

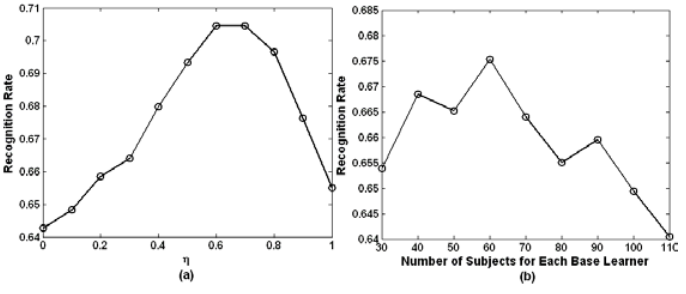

Fig. 3. (a) CRR obtained by the combination of generic and specific learners $H_{G-DLDA} + H_{S-PCA}$ v.s. η; Each base learner is trained with 80 subjects and retain 20 features. (b) CRR obtained by the generic learner H_{G-DLDA} with 20 features v.s. number of subjects including in each training subset.

biased generic solution and the specific solution with high estimation variance can provide better performance.

The last experiment deals with the influence of the subject number in each training subset. Fig.3(b) demonstrates the relationship of the CRR obtained by H_{G-DLDA} and the number of subjects used to train each base generic learner. The results indicate that the performance initially improves as the number of the training subjects increases. However, if too many subjects are included, the performance will degrade. It is well known that a necessary and sufficient condition for combining a set of learners to be more accurate than any of its individual members is if these base learners are accurate and diverse[12]. When the number of subjects are small, including more subjects and more samples could increase the learning capacity of the base learner which makes the base learner more accurate. However, since the number of the total generic subjects is fixed, continuing increasing the subjects in each training subset leads to heavier overlapping between different subsets, thereafter, the base learners trained on which

become more similar. The decreasing of the base learner diversity leads to the combination effect degraded.

7 Conclusion

In this paper, we proposed a novel framework to combine the generic solution and the specific solution for face recognition applications when only one image sample for each subject of interest is available. A set of base generic learners trained on the generic subject samples are firstly combined to provide a generic solution followed by a combination with the specific solution obtained from the subject samples of the interest. Experimentations on the FERET database indicate that the proposed scheme significantly improves the recognition performance.

References

1. Chellappa,R., Wilson,C.L. and Sirohey,S.,: Human and machine recognition of faces: A survey. Proceedings of the IEEE Vol.83,(1995) 705-740.
2. Zhao,W.Y., Chellappa,R., Rosenfeld,A. and Phillips,P.J.,: Face recognition: A literature survey. ACM Computing Surveys Vol.35(4),(2003) 399-458.
3. Brunelli,R., and Poggio,T.,: Face recognition: feature versus templates. IEEE Transactions on Pattern Analysis and Machine Intelligence Vol.15(10),(1993) 1042-1052.
4. Turk,M.A and Pentland,A.P.,: Eigenfaces for recognition. Journal of Cognitive Neuroscience Vol.3(1),(1991) 71-86.
5. Belhumeur,P.N., Hespanha, J.P. and Kriegman, D.J.,: Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence Vol.19(7),(1997) 711-720.
6. Lu,J., Plataniotis,K.N. and Venetsanopoulos, A.N.,: Regularization Studies of Linear Discriminant Analysis in Small Sample Size Scenarios with Application to Face Recognition. Pattern Recognition Letter Vol.26(2),(2005) 181-191.
7. Lu,J., Plataniotis,K.N. and Venetsanopoulos,A.N.,: Face recognition using LDA-based algorithms. IEEE Transactions on Neural Networks 14(1)(2003) 195-200.
8. Wang,X. and Tang,X.,: Unified subspace analysis for face recognition. Proceedings of the Nineth IEEE International Conference on Computer Vision (2003) 679-686.
9. Phillips,P.J., Moon,H., Rizvi,S.A and Rauss,P.,: The FERET evaluation method for face recognition algorithms. IEEE Transactions on Pattern Analysis and Machine Intelligence Vol.22(10),(2000) 1090-1104.
10. Yu,H. and Yang,J.,: A direct LDA algorithm for high dimensional data - with application to face recognition. Pattern Recognitio Vol.34, (2001) 2067–2070
11. Kittler,J., Hatef,M., Duin, R. and Matas,J.,: On combining classifiers. IEEE Transactions on Pattern Analysis and Machine Intelligence Vol.20, (1998) 226–239
12. Hansen,L. and Salamon,P.,: Neural network ensembles. IEEE Transactions on Pattern Analysis and Machine Intelligence Vol.12, (1990) 993–1001