Combining Classifiers in Rotated Face Space

Shaokang Chen [a,b], Ting Shan [a,b] and Brian C. Lovell [a,b]
[a] NICTA, 300 Adelaide Street, Brisbane QLD 4000, Australia
[b] ITEE, University of Queensland, Brisbane QLD 4072, Australia
{Shaokang.Chen, Ting.Shan, Brian.Lovell}@nicta.com.au

Abstract

Face recognition is a very complex classification problem due to nuisance variations in different conditions. Normally no single classifier can discriminate patterns well when unpredictable variations and a huge number of classes are involved. Combining multiple classifiers can improve discriminability over the best single classifier. In this paper, we present a way to combine classifiers for face recognition problem based on APCA classifiers. The proposed combinator generates various classifiers by rotating various face spaces and fusing them by applying a weighted distance measure. The combined classifier is tested on the Asian Face Database with 856 images. Experiments show a 30% reduction in classification error rate of our combined classifier and illustrates that combining classifiers from different face spaces may perform better than those based on a single face space.

1 Introduction

Face recognition is a very challenging task because differences between images of the same face due to nuisance variations, such as lighting condition, view point, pose, expression, and age are often greater than those between different faces. This problem has attracted considerable attention from psychophysicists, neuroscientists and engineers who wish to deal with face recognition in different conditions. Various techniques has been applied for automatic face recognition, such as Principal Component Analysis (PCA) [19], Linear Discriminant Analysis (LDA) [2, 14], Hidden Markov Models (HMMs) [16], Neural Networks [11, 5] and Support Vector Machines (SVMs) [6]. Nevertheless, currently most systems work well only under constrained conditions where lighting, pose, and camera parameters are strictly controlled. An ideal face recognition system should recognize new images of a known face and be insensitive to nuisance variations. However, no single classifier can discriminate patterns well enough especially for complex pattern classification problems like face recognition, where the number of classes are huge and the variation in the classes is large [10].

One way to improve system performance is to combine multiple classifiers. The motivation of combining classifiers is due to the observation that patterns misclassified by different classifiers may be different, even if one of them may achieve the overall best performance. Each classifier defines a mapping from a feature space to outputs (generally class labels). Because there exist differences between classifiers, the mappings are often disparate from each other, which may result in different performance of each classifier and dissimilar classification errors as well. Therefore, different classifiers may contain complementary information on pattern classification from other classifiers [8, 10]. Proper combination of these classifiers may solve the dilemma of bias and variance and enhance the performance [7].

We proposed a method based on Adaptive Principal Component Analysis (APCA) [3, 13] for face recognition, which is robust to face image variations in illumination and expression. We then extended it to pose invariant face recognition [17] in 2006. In this paper, we introduce a new method to combine APCA classifiers to build a stronger classifier to further improve recognition accuracy. In Section 2, we briefly explain the APCA method. Then we discuss in detail a method for generating different complementary base classifiers based on APCA and design a framework to fuse the classifiers in Section 3. Section 4 is devoted to empirical evaluation. Finally, we draw conclusions and indicate future work in Section 5.

2 Adaptive Principal Component Analysis

Adaptive Principal Component Analysis [3, 13] is a linear pattern classification algorithm that inherit merits from both PCA and LDA by warping the face subspace according to the within-class and between-class covariance of samples. We first apply PCA on face images to extract eigenfaces. Consequently, every face image is projected into a face subspace with reduced dimensionality to form a m-dimension feature vector $s_{j,k}$ with $k = 1, 2, ..., K_j$ denoting the k^{th} sample of the class S_j . Then the face subspace is warped by the following three steps:

- Space Rotation: The feature space is rotated according to the overall within-class covariance. The rotation matrix R is a set of eigen vectors obtained by applying singular value decomposition to the overall within-class covariance matrix. By space rotation, the representativeness of features are enhanced and features are either more discriminative or generative after rotation.

- Whitening Transformation: The subspace is whitened according to the eigenvalues $\lambda_i (i = 1, 2, ..., m)$ of the PCA extracted face subspace with a whitening power p. Each eigenface u_i is whitened according to the corresponding eigen-value λ_i with the power p. Consequently, the whitening matrix is:

$$Z = diag\{\lambda_1^p, \lambda_2^p, ..., \lambda_m^p\} \tag{1}$$

  Whitening transformation is used to control the overall scatter of all samples and compensate for the overweighing of low frequency components.

- Eigenface Filtering: Eigen-features are weighted according to the identification-to-variation value $ITV_i (i = 1, 2, ..., m)$ with a filtering power q. The ITV is a ratio measuring the correlation with a change in person versus a change in variation for each of the eigenfaces. It is defined as the following:

$$\begin{aligned} ITV_i &= \frac{\frac{1}{M}\sum_{j=1}^{M}\frac{1}{K}\sum_{k=1}^{K}|s_{i,j,k} - \varpi_{i,k}|}{\frac{1}{M}\sum_{j=1}^{M}\frac{1}{K}\sum_{k=1}^{K}|s_{i,j,k} - \mu_{i,j}|}, \\ \varpi_{i,k} &= \frac{1}{M}\sum_{j=1}^{M} s_{i,j,k}, \\ \mu_{i,j} &= \frac{1}{K}\sum_{k=1}^{K} s_{i,j,k}, i = [1, ..., m], \end{aligned} \tag{2}$$

where $s_{i,j,k}$ denotes the i_{th} element of the face vector of the k_{th} sample for class (person) S_j. Then the filtering matrix Υ is defined by:

$$\Upsilon = diag\{ITV_1^q, ITV_2^q, ...ITV_m^q\}, \tag{3}$$

The aim of eigenface filtering is to diminish the contribution of eigenfaces that are strongly affected by illumination and expression variations and enhance those features that capture the main differences between the classes.

The whitening power p and filtering power q are to be determined empirically by searching in a two dimensional space by the following cost function. Define the distance between two face vectors $s_{j,k}$ and $s_{j',k'}$ as the Euclidean distance of their transformed vectors:

$$d_{jj',kk'} = \|Z\Upsilon(s_{j,k} - s_{j',k'})\|_2. \tag{4}$$

The cost function OPT is a combination of error rate and the ratio of between-class distance to within-class distance as the following:

$$\begin{aligned} &OPT = \textstyle\sum_{j=1}^{M}\sum_{k=1}^{K}\sum_m\left(\frac{d_{jj,k0}}{d_{jm,k0}}\right), \\ &\forall m \in d_{jm,k0} < d_{jj,k0}, m \in [1...m]. \end{aligned} \tag{5}$$

where $d_{jj,k0}$ is the within-class distance between the variant sample $s_{j,k}$ and the standard reference image $s_{j,0}$ (typically the normally illuminated neutral image) for class S_j. Correspondingly, $d_{jm,k0}$ is the between-class distance between sample $s_{j,k}$ and the reference image $s_{m,0}$ for class S_m. The experimental results on face images in Asian Face Database [18] with both illumination and expression variations show that APCA performs much better than PCA and LDA. For more details of the APCA algorithm please refer to paper [3].

3 Combining APCA Classifiers

Though the APCA classifier deals with illumination and expression well, there is still some room for further improvement in performance. In 2005, we developed a hybrid learning algorithm to combine generative learning and discriminative learning by finding a trade-off between these two approaches [4]. Experiments demonstrated that hybrid learning may outperform both generative and discriminative paradigms in terms of classification error rate. In this section, we are going to propose an alternative method to improve the performance by combining different classifiers.

3.1 Classifier combination

Unlike hybrid learning, which finds a balance between generative and discriminative learning to build one classifier, combining classifiers tends to find the optimal solution by taking advantages of the differences between different base classifiers. Therefore, a major factor that has strong influence on the performance of the combination of classifiers is the different behaviours of individual classifiers in pattern classification [8, 10]. The base classifiers of a classifier ensemble may be different in the following aspects: parameter choice, architecture, category, training sets, and feature sets [15]. Hence, by changing some aspects of a classifier we can generate multiple base classifiers for combination. Another important factor that influences the final performance of the combination is the combination strategy. This includes the combiner architecture (serial, parallel or hybrid), fusing method (combination at abstract-level,

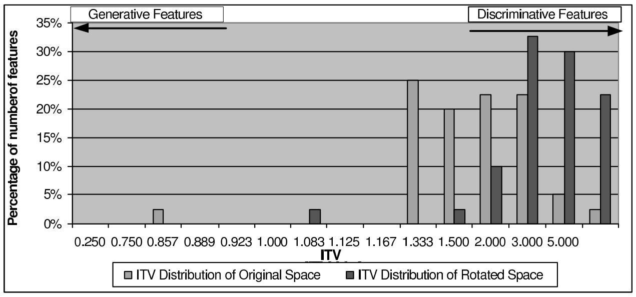

Figure 1. ITV distribution in original and rotated spaces for face images with expression variations.

ranked-level or measurement-level), and combination rules (fixed rules or trained rules) [8, 15, 10, 12, 9]. These factors vary immensely and it is very difficult to estimate their influence on classifier combination. Accordingly, the design of a multiple classifier system involves two stages: 1) classifier ensemble design to generate a set of complementary and diverse classifiers and 2) combinator design to create a fusion mechanism that optimally combines those base classifiers and benefits from them [15]. We will describe our proposed method for combining classifiers based on these two stages.

3.2 Space rotation for classifier ensemble

The classifier ensemble needs to be carefully designed so that patterns misclassified by different classifiers do not overlap. In order to generate suitable complementary and diverse base classifiers, we control space rotation of our APCA approach to control misclassification errors. Figure 1 shows that after rotation, features of the new face subspace are more representative.

The x-axis in Figure 1 is the ITV value and the y-axis is the percentage of the number of features with the corresponding ITV value. The higher the ITV value of a feature, the more discriminative it is for classification. After space rotation, most features are more discriminative with an ITV value greater than 2. Hence, whitening and eigen-filtering are more efficient which leads to improved discriminability of the warped space and higher classification accuracy of the classifier. This implies that space rotation can affect classifier performance. In other words, different rotations of the face subspace result in different classification errors. Hence, by controlling space rotation we can create numerous base classifiers with complementary information.

We illustrate this effect in Figure 2. Sub-figure A shows two class distributions in the original space. The solid axes are the main axes in the original space and the dotted axes and dotted-broken axes are the two pairs of axes in the two rotated spaces respectively. Sub-figures B and D show the distribution of the two classes in the corresponding rotated spaces. Points a1 and a2 are two samples for class A and similarly points b1 and b2 are two samples belong to class B. We can see that with only space rotation, classification performance in different spaces with the nearest neighbor rule does not change — sample a1 in class A and b1 in class B are very likely to be misclassified. While after whitening and eigen filtering, classification errors may vary. Misclassified samples change from a1 of class A and b1 of class B in one rotated space to a2 and b2 in another rotated space. Consequently, classification errors of two classifiers with different space rotations are complementary to each other. That is, there always exists a classifier that can recognize certain samples correctly. Appropriate combination of these classifiers would improve the performance.

3.3 Combinator design

After creating base classifiers, we need to combine these classifiers appropriately. We choose to combine classifiers at the measurement-level for two reasons. First, our proposed APCA method can measure posterior possibilities of a sample by calculating the distance between the sample and different classes. Second, we can utilize the *a priori* knowledge of classifiers for combination, which is proved to be helpful when highly correlated classifiers are combined [1, 12, 20]. Suppose, we have built up two classifiers C_1 and C_2 by warping the original space with different rotation, whitening and filtering. ρ_1 and ρ_2 are the classification accuracy of the two classifiers respectively. In order to fuse two classifiers correctly, measurement of the distance in warped space for the two classifiers should be in the same scale. Therefore, the space needs to be normalized to be uniform with a scale ς as the following:

$$\varsigma = \frac{1}{\sqrt{\Sigma_{i=1}^{m} \lambda_i^{2p} ITV_i^{2q}}} \tag{6}$$

Then, we construct the following fusion function to measure the overall distance between two samples:

$$D = \rho_1 * \varsigma_1 * d_1 + \rho_2 * \varsigma_2 * d_2, \tag{7}$$

where d_1 and d_2 are the weighted Euclidean distance between these two samples in the rotated space for classifier C_1 and C_2 respectively and can be calculated by equation 4. Thus the input I is classified to the closest class by the following rule:

$$\textit{assign } I \longrightarrow S_n, if D_n = \min_j D_j. \tag{8}$$

This method treats two classifiers individually and assume their prior possibilities as ρ_1 and ρ_2 respectively. Space rotation together with corresponding whitening and filtering

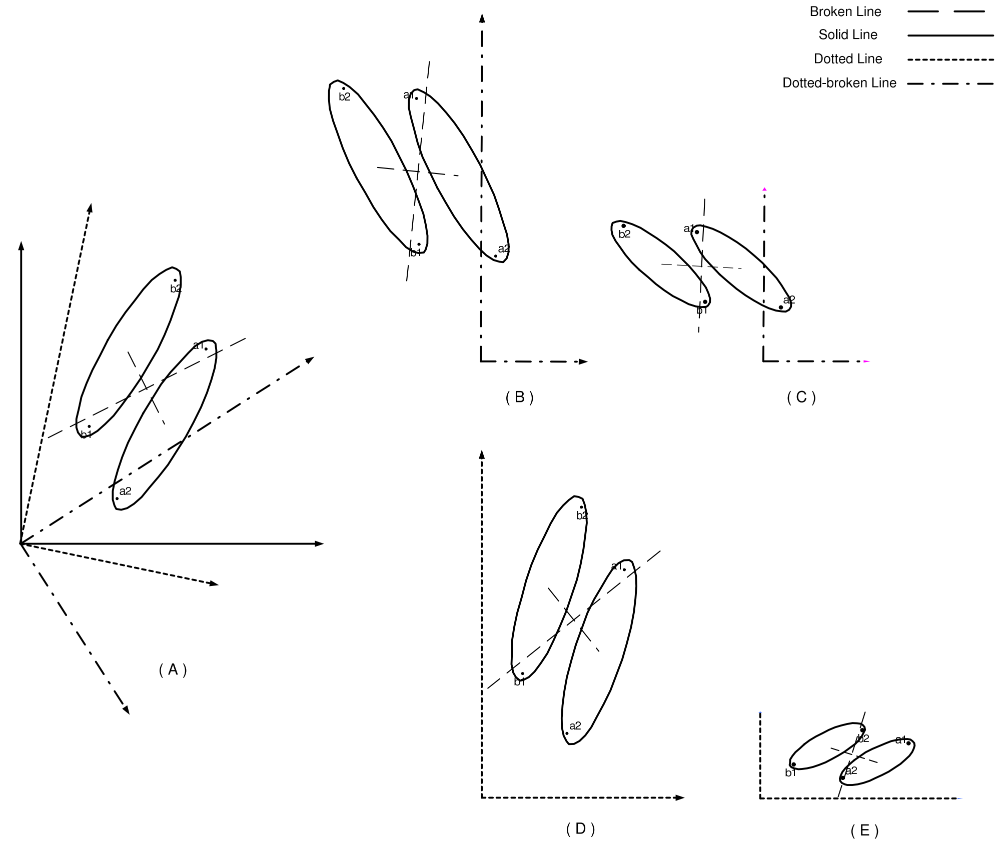

Figure 2. Effect of space rotation on classification errors.

can be considered as defining a projection from the original feature space to the rotated space. Because of the differences in rotation, mappings from the original space to rotated spaces are different, which results in diversity in pattern classification.

4 Empirical Evaluation

We test our proposed classifier combination method on a subset of the Asian Face Database [18]. It consists of 856 facial images corresponding to 107 subjects with each subject 5 images taken under different standardized illuminations and 4 images taken with variant facial expressions. The face region of each image is extracted and normalized to 171 by 171 pixels with 256 gray levels per pixel. Face images are aligned according to their eye positions. Figures 3 and 4 show some sample face images with different lighting conditions and expressions from Asian Face Database. We design two sets of tests as follows in an attempt to fully investigate the capability of our combined classifiers.

Figure 3. Sample face images with different lighting conditions from Asian Face Database.

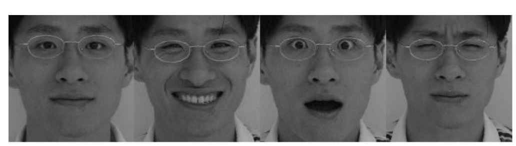

Figure 4. Sample face images with different expressions from Asian Face Database.

4.1 Combining classifiers in a single space

Firstly, we experiment with combining base classifiers in a single face space. We divide the data base into three

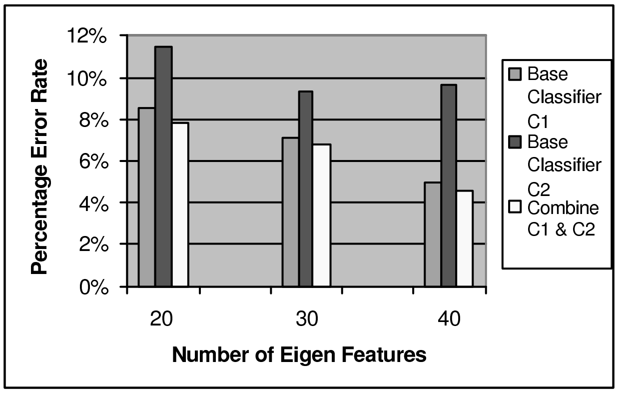

Figure 5. Classification error rate of base classifiers and combined classifier in a single space by fusing distance measure.

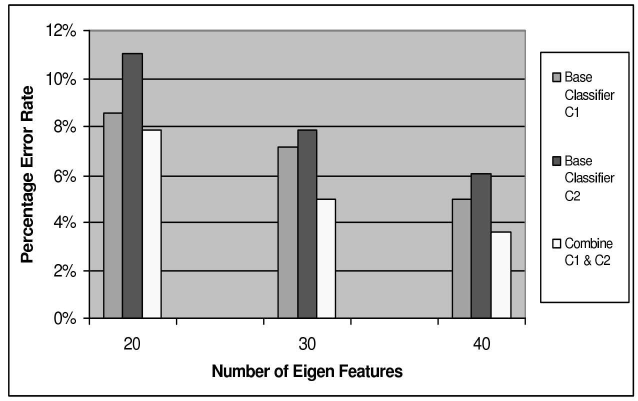

Figure 6. Classification error rate of base classifiers and combined classifier in two face spaces.

non-overlapping data sets with nearly the same size. We use data-set 1 (288 images corresponding to 36 subjects) to construct the eigenface space by applying PCA on the sample images. We then generate two different base classifiers C_1 and C_2 by rotating the face spaces according to the within-class variance of sample images from data-set 1 and data-set 2 separately. Finally, we use data-set 3, which contains all images of unseen subjects, to test our base classifiers and the combined classifier. We only register one normal lighting neutral image (the first image in Figure 3 and 4) of each unseen subject into database and use the rest of the unseen images with variable lighting conditions and expressions for testing. This experiment is done based on a three-fold cross validation rule by altering the training data and testing data and the results are the average of all three tests. Figure 5 plots the percentage error rate of the base classifiers and the combined classifier in a single space. The combined classifier of C1 and C2 is produced according to equation 7. We can see that among the three classifiers, the combined classifier always performs best and achieves the lowest error rate. Classifier C1 is preferable to classifier C2 regardless of the number of eigen features, because classifier C2 is rotated and optimized on the same data set, it tends to be more overfitted to the training data than classifier C1. This is also the reason why with an increase in the number of features used for classification, the error rate of classifier C2 does not decrease steadily. On the contrary, performance of classifier C1 and the combined classifier improved monotonically with the number of eigen features increasing. Moreover, even though performance of the classifier C2 is not as stable as classifier C1, the combined classifier still outperforms both classifiers C1 and C2 consistently regardless of the number of features used. This supports the hypothesis that by controlling space rotation we can generate complementary and diverse base classifiers and that our classifier fusion scheme is successful.

4.2 Combining classifiers from different spaces

Secondly, we explore the combination method by combining classifiers from different face spaces — that is to combine classifiers with different features. Each base classifier is derived from a different face space which is constructed by a completely different set of eigen features. After producing base classifiers in different face spaces, we attempt to combine those classifiers by fusing their distance measures. In other words, we divide the database into three sections as was done in the previous test: the first partition is required to derive eigen features to construct the face space; the second partition is used for training to warp the space to enhance pattern classification performance; and the third partition is required for testing. In order to generate various base classifiers in different face spaces for combination, we exchange the roles of the data partitions so that two face spaces can be created and warped independently. For example, if for classifier C1, we use dataset 1 for constructing the face space and dataset 2 for optimization. Then for classifier C2, dataset 2 is used for constructing face space and dataset 1 is used for optimization. Finally, the combined distance is calculated to combine the two classifiers by equation 7 and classification is performed according to equation 8.

Experimental results of combining the classifiers from two different face spaces are shown in Figure 6.

The misclassification percentage of base classifiers C1 and C2 and the combined classifier reduce monotonically with the number of features increasing. Again, the combined classifier achieves the lowest error rate compared to

Table 1. Comparison of classification error rate for combined classifiers in single face space and different face spaces.

NO. of Features	Single Space	Different Spaces
20	7.86%	7.86%
30	6.79%	5.00%
40	4.64%	3.57%

base classifiers regardless of the number of features used. The error rate of the best base classifier reduced from 7.14% to 5% for the combined classifier with 30 features involved and from 5% to 3.57% with 40 features individually, which is a nearly 30% reduction.

Table 1 shows the comparison of classification error rate of combined classifiers in a single face space and from different face spaces. We can see that by combining classifiers from different face spaces the classifier performs better than using a single space. That is because base classifiers constructed from different face spaces are more complementary and diverse than those created in the same space. In a single face space, base classifiers vary by different space rotation and optimization due to the variation in training data. While for multiple spaces, images are projected into different face spaces (hyperplanes) to form image vectors with different reconstruction error. Hence, patterns distribute completely different in those face spaces. Moreover, each face space is warped differently by whitening and filtering which results in a diverse distribution of classes. Thus, base classifier C1 and C2 have totally different sets of features and different training data that result in more complementary misclassification patterns. It is interesting that when using less than 20 features, the performance of two combination strategies are similar. We suspect that the reason is that for lower dimensionality, the set of eigen features extracted in two different spaces are the low frequency components which are similar to each other. Hence, the two spaces are similar and combining classifiers from different spaces is very close to combing classifiers in a single space with different rotation.

5 Conclusion and Future Work

In this paper, we developed a method to combine different classifiers for face recognition based on APCA. We generate a base classifier ensemble by controlling space rotation based on the observation that different rotations will result in different classification errors. We then proposed a combinator using face space normalization which involved normalising the distance measures and weighted distances to balance between the base classifiers. The experimental results on face recognition show that combined classifiers outperform corresponding base classifiers. In addition, combination of different face spaces is better than the combination of single spaces because the eigen features which form the basis for the space are more diverse. Currently we control space rotation based on choosing different training datasets and combine only two base classifiers. Our future work may involve in generating several base classifiers based on rotation angles or classification errors and investigating the efficacy of the corresponding combination.

Acknowledgements

The authors thank Abbas Bigdeli, Conrad Sanderson and Erik Berglund for useful suggestions. This project is supported by a grant from the Australian Government Department of the Prime Minister and Cabinet and by the Australian Research Council through the Research Network for Securing Australia. NICTA is funded by the Australian Government's *Backing Australia's Ability* initiative, in part through the Australian Research Council.

References

[1] A. Baykut and A. Ercil. Towards automated classifier combination for pattern recognition. In *Multiple Classifier Systems, 4th International Workshop*, Guilford, UK, 2003.

[2] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, 1997.

[3] S. Chen and B. C. Lovell. Illumination and expression invariant face recognition with one sample image. In *Proceedings of International Conference on Pattern Recognition*, Cambridge, UK, 2004.

[4] S. Chen, B. C. Lovell, and T. Shan. Combining generative and discriminative learning for face recognition. In *Proceedings of the Digital Image Computing on Techniques and Applications*, Cairns, Australia, 2005.

[5] M. J. Er, S. Wu, and J. Lu. Face recognition using radial basis function neural networks. In *Proceedings of the Conference on Decision and Control*, Phoenix, USA, 1999.

[6] G. Guo, S. Z. Li, and K. Chan. Face recognition by support vector machines. In *Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 196–201, Grenoble, France, 2000.

[7] S. S. Haykin. *Neural Networks: A Comprehensive Foundation*. Prentice Hall, 3rd edition, 2007.

[8] J. Kittler, M. Hatef, R. P. Duin, and J. Matas. On combining classifiers. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3):226 – 239, 1998.

[9] L. I. Kuncheva. A theoretical study on six classifier fusion strategies. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(2):281 – 286, 2002.

[10] L. I. Kuncheva. *Combining Pattern Classifiers: Methods and Algorithms*. Wiley Interscience, 2004.

[11] S. Lawrence, C. L. Giles1y, A. C. Tsoi, and A. D. Back. Face recognition: A convolutional neural network approach. *IEEE Transactions on Neural Networks, Special Issue on Neural Networks and Pattern Recognition*, 8(1):98–113, 1997.

[12] V. Lecce, G.Dimauro, A.Guerriero, S.Impedovo, G.Pirlo, and A.Salzo. Classifier combination: the role of a-priori knowledge. In *The 7th International Workshop on Frontiers in Handwriting Recognition*, Amsterdam, Netherlands, 2000.

[13] B. C. Lovell and S. Chen. *Robust Face Recognition for Data Mining in Encyclopedia of Data Warehousing and Mining*. Idea Group, 2005.

[14] A. M. Martinez and A. C. Kak. PCA versus LDA. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(2):228 – 233, 2001.

[15] R. P.W.Duin. The combining classifier: to train or not to train. In *Proceedings of the 16th International Conference on Pattern Recognition*, Quebec, Canada, 2002.

[16] F. Samaria and F. Fallside. Automated face identification using hidden markov models. In *Proceedings of the International Conference on Advanced Mechatronics*, Tokyo, Japan, 1993.

[17] T. Shan, B. C. Lovell, and S. Chen. Face recognition robust to head pose from one sample image. In *Proceedings of the 18th International Conference on Pattern Recognition*, pages 515–518, 2006.

[18] Intelligent Multimedia Lab, Department of Computer Science and Engineering, Pohang University of Science and Technology. Asian face image database PF01. http://nova.postech.ac.kr/.

[19] M. A. Turk and A. P. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71–86, 1991.

[20] K. Woods, W. P. Kegelmeyer, and J. K. Bowyer. Combination of multiple classifiers using local accuracy estimates. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(4):405 – 410, 1997.