S. Chen, T. Shan, B. C. Lovell, 'Robust Face Recognition in Rotated Eigenspaces', *Proceedings of Image and Vision Computing New Zealand 2007*, pp. 1–6, Hamilton, New Zealand, December 2007.

Robust Face Recognition in Rotated Eigenspaces

Shaokang Chen [1,2], Ting Shan [1,2] and Brian C. Lovell [1,2]

[1] NICTA, 300 Adelaide Street, Brisbane QLD 4000, Australia
[2] ITEE, University of Queensland, Brisbane QLD 4072, Australia
Email: {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. Most face recognition approaches either assume constant lighting condition or standard facial expressions, thus cannot deal with both kinds of variations simultaneously. Principal Component Analysis (PCA) cannot handle complex pattern variations such as illumination and expression. Adaptive PCA rotates eigenspace to extract more representative features thus improving the performance. In this paper, we present a way to extract various sets of features by different eigenspace rotations and propose a method to fuse these features to generate nonorthogonal mappings for face recognition. The proposed method is tested on the Asian Face Database with 856 images from 107 subjects with 5 lighting conditions and 4 expressions. We register only one normally lit neutral face image and test on the remaining face images with variations. Experiments show a 95% classification accuracy and a 20% reduction in error rate. This illustrates that the fused features can provide significantly improved pattern classification.

Keywords: face recognition, eigen features, space rotation, nonorthogonal features

1 Introduction

Face recognition has attracted considerable attention from psychophysicists, neuroscientists and engineers for more than 50 years. Current research on face recognition has been focused on dealing with face image variations in lighting conditions, facial expressions, and pose. Two main approaches have been proposed for illumination invariant face recognition. One is to abstract features that are less sensitive to illumination change, such as edge maps of an image [1, 2]. Yet edge features generated from shadows are highly related to illumination changes and have a signification impact on recognition. The other approach is to construct a low dimensional subspace for face images taken under different lighting conditions [3, 4]. This approach is based on an assumption that surface of human face is Lambertian reflected and convex and the method requires several images of the same face taken under specific lighting source directions to construct the model of the given face. Thus, it is hard for these systems to deal with cast shadows and they cannot handle face recognition problems when only one gallery image is available per person.

As for expression invariant face recognition, one approach is to morph images to be the same shape as the one used for training [5]. Unfortunately, it is not guaranteed that all images can be morphed correctly. For example an image of a face wearing sunglasses cannot be morphed to a neutral face image because of the lack of texture information near the eyes. Another approach is to use optical flow to estimate pixel displacement between two images [6, 7]. However, it is difficult to learn the local motions within feature space to determine the expression changes of each face, since different persons express a certain expression in different ways. Martinez [8] proposed a weighting method that weights independently those local areas which are less sensitive to expressional changes. But features that are insensitive to expression changes may be sensitive to illumination changes as noted in [9].

We proposed a method Adaptive Principal Component Analysis (APCA) [10] for robust face recognition with lighting and expression variations. We then extended it to pose invariant face recognition [11] in 2006. In this paper, we introduce a new method to further improve the performance of APCA by extracting various sets of features through different space rotations and by fusing them to generate a set of nonorthogonal features for face recognition. In section 2, we briefly explain the APCA method. Then we discuss the details of space rotation for generating different sets of features. We then discuss feature fusion for nonorthogonal mapping in section 3. Section 4 is devoted to the experimental results. Finally, we present conclusions and future work in section 5.

2 Adaptive Principal Component Analysis

APCA is a linear pattern classification algorithm that inherit merits from both Principal Component Analysis (PCA) and FLD (Fisher Linear Discriminant) 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- dimensional 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.
- Whitening Transformation: The subspace is whitened according to the eigen values $\lambda_i (i = 1, 2, ..., m)$ of the features in rotated face subspace with a whitening power p. Consequently, the whitening matrix is:

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

- Eigenface Filtering: Eigen-features are weighed 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 follows:

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

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 Γ is defined by:

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

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

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

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

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

The experimental results on face images in the Asian Face Database [12] with both illumination and expression variations show that APCA performs significantly better than PCA, PRM [13] and FLD. For more details of the APCA algorithm please refer to [10].

3 Fusing Features in Rotated Eigen Space

Although the APCA classifier can deal with illumination and expression, there is still scope to further improve performance. In this section, we propose another method to improve the APCA classifier by fusing features after controlled feature space rotation.

3.1 Space Rotation for Complementary Misclassification

In the APCA algorithm, space rotation is very important because it improves the representativeness of features. Figure 1 shows the effect of rotation on the ITV distribution of features. 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

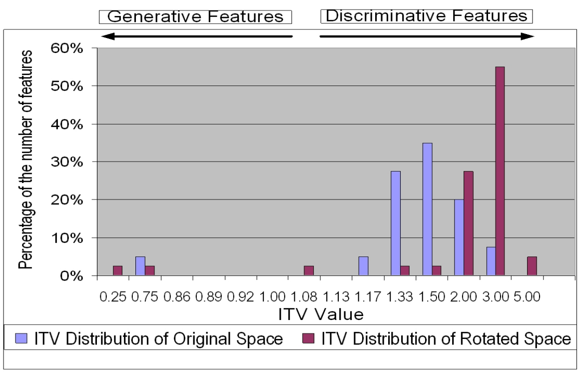

Figure 1: ITV distribution in original and rotated spaces for face images with both illumination and expression variations.

discriminative with an ITV value greater than 2. Hence, whitening and eigen filtering become more efficient leading to improved discriminability of the warped space and higher classification accuracy of the classifier. The experimental results in [10] show that after rotation, the performance of both APCA and PRM are significantly enhanced.

However, for complex face recognition problems with different face variations, a single space rotation may not completely distinguish all classes correctly even though it achieves better performance. Different rotations of the face subspace may result in different classification errors. If we use the complementary pattern classification information in different space rotations, we are able to improve the performance. This idea is similar to combining classifiers for complex pattern classification when unpredictable variations and numerous classes are involved [14]. We illustrate this effect in figure 2. Sub-figure A shows two class distributions in the original space. Axes in black are the main axes in the original space and axes in dotted lines and dotted-broken lines are two pairs of the main axes of two rotated spaces respectively. Points a1 and a2 are two samples for class A and similarly points b1 and b2 are two samples belonging to class B. Sub-figure B and D show the distribution of two classes in the corresponding rotated spaces. We can see that with only space rotation, classification performance in different spaces with nearest neighbor rules 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 rotation are complementary to each other. That is, there always exists a classifier that can recognize a certain sample correctly. Appropriate combination of these classifiers would improve the performance.

3.2 Reverse Space Rotation

After generating different classifiers by space rotation, we introduce reverse space rotation so that features extracted from different rotated spaces can be merged. It is meaningless to add two features from separate spaces directly, because features in different rotated spaces represent different patterns. However, these features are not completely isolated. They are highly correlated by the rotation of the same eigen space. Reverse rotation can transform features of the rotated space back to the corresponding features in the original eigen space, which can be easily fused. Moreover, reverse rotation can induce nonorthogonal mapping between the original space and the rotated space, which is preferable for complex pattern classification.

Given a vector v and corresponding rotation matrix R, whitening matrix Z, and filtering matrix Γ, the projected vector $\widetilde{v}$ in warped space is:

$$\widetilde{v} = Z\Gamma Rv, \tag{6}$$

$$R^T R = E, \tag{7}$$

where E is a unit matrix.

After reverse rotation of the warped space, the vector in the new space is described by:

$$v' = R^{-1}Z\Gamma Rv. \tag{8}$$

The transformation matrix $\Lambda = R^{-1}Z\Gamma R$ is very likely to be nonlinear. Assume the basis vectors of Λ are orthogonal, then $\Lambda^T\Lambda$ should be a diagonal matrix. However, the following equation shows that Λ is diagonal only when $Z\Gamma = kE$.

$$\begin{aligned} \Lambda^T\Lambda &= (R^{-1}Z\Gamma R)^T R^{-1}Z\Gamma R \\ &= R^T\Gamma^T Z^T RR^T Z\Gamma R \\ &= R^T\Gamma^T Z^T Z\Gamma R \\ &= R^T(Z\Gamma)^2 R. \end{aligned} \tag{9}$$

Hence, rotating each warped space in reverse can bring about linearly dependant features.

3.3 Fusing Features

Reverse rotation does not change the performance of the classifiers which are trained to minimize classification errors. However, classifiers are optimized individually and patterns are clustered diversely in each subspace. We needs to exploit complementarity and diversity of those classifiers to combine nonorthogonal features from different subspace together to derive improved discriminative features.

By reverse space rotation, we can merge features directly in the original space. But different rotated spaces are warped separately, we need to normalize the space to achieve uniform gain. Because we are using the nearest neighbor rule for classification, measurement of the distance in rotated warped space should be in the same scale. That is space should be normalized with a scale ς as the following:

$$\varsigma = \frac{1}{\sqrt{\Sigma_{i=1}^{m}\lambda_i^{2p}ITV_i^{2q}}}. \tag{10}$$

After this space normalization, the sum of the coefficients square for all the features is equal to one, that is

$$\varsigma||Z\Gamma||_2 = 1. \tag{11}$$

Then we fuse features for the normalized spaces. Without losing generality, we only consider fusing

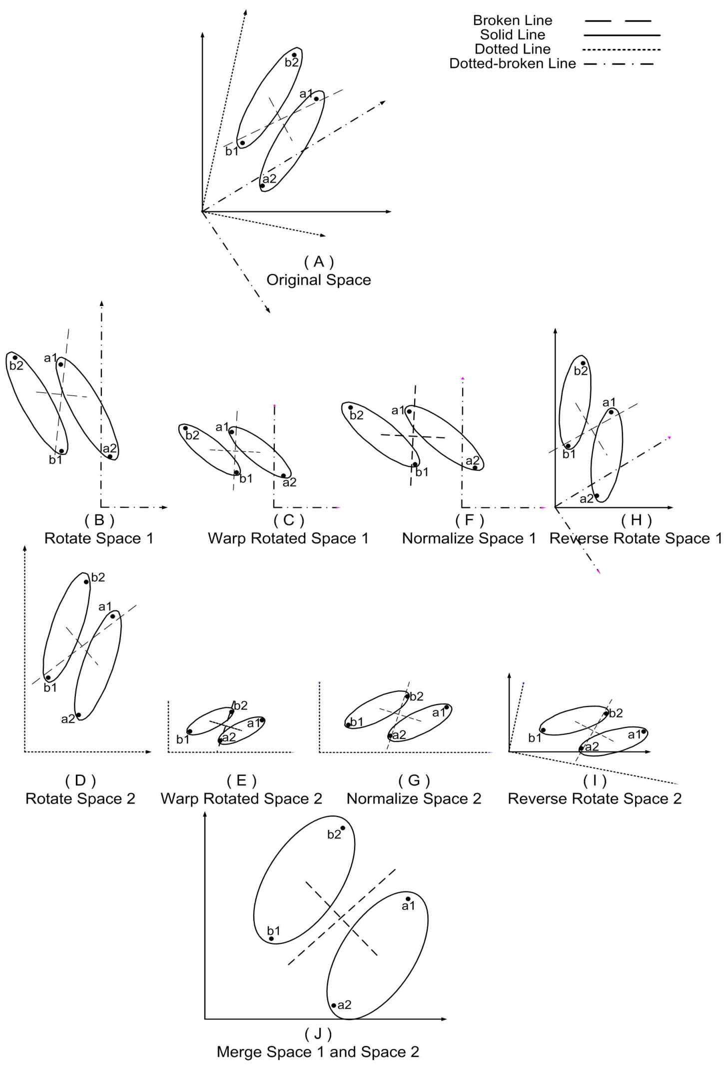

Figure 2: Merge features by space rotation.

features with two different rotations. Suppose we rotate the original feature space differently for two classifiers C_1 and C_2 according to rotation matrix R_1 and R_2 respectively. The matrices Z_1 and Z_2 are two whitening matrices for corresponding rotated space. The matrices Γ_1 and Γ_2 are the filtering matrices. Then vector v in original space is projected into corresponding spaces by:

$$\begin{aligned} \widetilde{v}_1 &= Z_1\Gamma_1 R_1 v \\ \widetilde{v}_2 &= Z_2\Gamma_2 R_2 v. \end{aligned} \tag{12}$$

We then rotate the warped space in reverse as follows:

$$\begin{aligned} v'_1 &= R_1^{-1}\widetilde{v}_1 \\ v'_2 &= R_2^{-1}\widetilde{v}_2. \end{aligned} \tag{13}$$

Now, elements of v'_1 and v'_2 represent the projection on the same features and can be added directly.

$$\begin{aligned} v' &= v'_1 + v'_2 \\ &= R_1^{-1}\widetilde{v}_1 + R_2^{-1}\widetilde{v}_2 \\ &= R_1^{-1}Z_1\Gamma_1 R_1 v + R_2^{-1}Z_2\Gamma_2 R_2 v \\ &= R_1^T Z_1\Gamma_1 R_1 v + R_2^T Z_2\Gamma_2 R_2 v \\ &= (R_1^T Z_1\Gamma_1 R_1 + R_2^T Z_2\Gamma_2 R_2)v \\ &= \Delta v \end{aligned} \tag{14}$$

The final transformation matrix is nonorthogonal, since

$$\begin{aligned} \Delta^T\Delta &= (R_1^T Z_1\Gamma_1 R_1 + R_2^T Z_2\Gamma_2 R_2)^T \\ &\quad *(R_1^T Z_1\Gamma_1 R_1 + R_2^T Z_2\Gamma_2 R_2) \\ &= [(R_1^T Z_1\Gamma_1 R_1)^T + (R_2^T Z_2\Gamma_2 R_2)^T] \\ &\quad *(R_1^T Z_1\Gamma_1 R_1 + R_2^T Z_2\Gamma_2 R_2) \\ &= (R_1^T\Gamma_1 Z_1 R_1 + R_2^T\Gamma_2 Z_2 R_2) \\ &\quad *(R_1^T Z_1\Gamma_1 R_1 + R_2^T Z_2\Gamma_2 R_2) \\ &= R_1^T\Gamma_1 Z_1 R_1 R_1^T Z_1\Gamma_1 R_1 \\ &\quad +R_1^T\Gamma_1 Z_1 R_1 R_2^T Z_2\Gamma_2 R_2 \\ &\quad +R_2^T\Gamma_2 Z_2 R_2 R_1^T Z_1\Gamma_1 R_1 \\ &\quad +R_2^T\Gamma_2 Z_2 R_2 R_2^T Z_2\Gamma_2 R_2 \\ &= R_1^T(Z_1\Gamma_1)^2 R_1 + R_1^T\Gamma_1 Z_1 R_1 R_2^T Z_2\Gamma_2 R_2 \end{aligned} \tag{15}$$

$$+R_2^T\Gamma_2 Z_2 R_2 R_1^T Z_1 \Gamma_1 R_1$$
$$+R_2^T(Z_2\Gamma_2)^2 R_2$$

is not a diagonal matrix. Therefore, by fusing features by rotating the warped space back into the original space, we bring about nonorthogonal features which have great functional significance in biological sensory systems [15].

Figure 2 illustrates the procedure and effect of our proposed fusing strategy. Sub-figure A to E show the effect of rotation and corresponding whitening and filtering on pattern clustering in two different rotated spaces. Sub-figure F and G show the distribution after normalizing warped space. Sub-figure H and I are the results by reverse rotating the subspaces back. Comparing H, I and A, we can see how different the distribution of classes is in the new space and the original space. Sub-figure J plots the distribution by combining features of two new spaces depicted in sub-figure H and I together. It is apparent that in this new distribution derived from nonorthogonal mappings of the original space, exemplars tends to be more scattered, which leads to higher classification accuracy.

4 Experimental Results

We test our proposed classifier combination method on the Asian Face Database [12]. It consists of 856 facial images under 5 different standardized illuminations and 4 variant facial expressions corresponding to 107 subjects. The size of each image is 171 by 171 pixels with 256 gray levels per pixel. Face images are aligned according to their eye positions. We divide the database into three nearly equal-sized data sets. For test 1, we use data set 1 (288 out of 856 images) to construct the eigen face space by applying PCA on the sample images. We then generate two different base classifiers C_1 and C_2 by rotating the face space 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 unseen images, to test our base classifiers and the combined classifier. We only register normal lighting neutral images and use the rest images with lighting and expression variations for testing. This experiment is done based on a three-fold cross validation rule and the results are the average of all three tests.

Figure 3 plots the percentage of error rate of base classifiers and fused classifiers in different number of features. We can see that among three classifiers, the merged classifier always performs the 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 merged classifier improved monotonically with the number of eigen features increasing. Moreover, even though performance of classifier C2 is not as stable as classifier C1, the merged classifier still achieves the best performance with an error rate around 20% less than that of the better base classifier C1 constantly regardless of the number of features used.

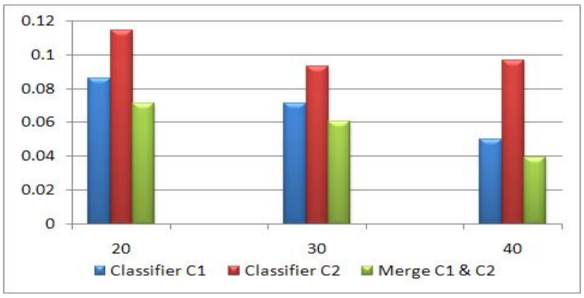

Figure 3: Classification error rate of individual classifiers and combined classifier in a single space by fusing features.

Table 1: Comparison of classification accuracy

# of Features	PCA	FLD	PRM	APCA	Merged APCA
20	59.7%	75.5%	77.1%	91.4%	92.9%
30	62.7%	83.1%	83.1%	92.9%	93.9%
40	64.3%	91.5%	86.5%	95.0%	96.1%

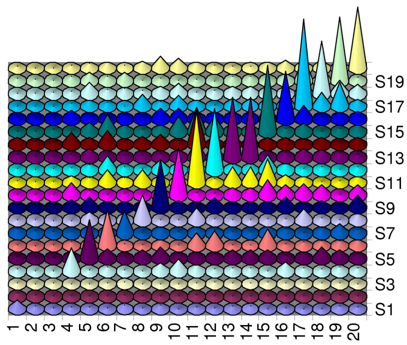

Figure 4: Nonorthogonality of the features derived by fusing features from different rotated spaces.

Figure 4 shows the nonorthogonality of those features generated by fusing features from two base classifiers. The cone located at the mth row and nth column in the figure represents the dot product of feature m and feature n. We can see that along the diagonal there exist very high cones and the height of these cones is proportional to the weight of corresponding features. Other cones represent the dot product between two different features. There exist non-zero dot product besides the diagonal, so these features are not orthogonal. In order to show this effect clearly, we set the dot products to be their absolute values. Hence, all the dot products are greater than or equal to zero. The result verifies that by the reverse rotation of spaces to fuse features we can introduce nonorthogonal features, which are helpful for face classification.

5 Conclusion and Future Work

In this paper, we developed a method to fuse features from different APCA classifiers for face recognition. We rotate the feature spaces differently based on the observation that various rotations will result in diverse classification errors. We then normalize the face space to normalize the distance measure and reverse rotate the space to merge the features. The experimental results on face recognition show that merged classifiers outperform corresponding base classifiers by 20%. In addition, reverse space rotation and feature fusion can induce nonorthogonal mapping that benefits pattern classification. However, currently we control the space rotations based on choosing different training data and only merge features from two rotated spaces. Our future work may involve controlling rotation by angle or classification error to generate multiple subspaces and then fuse them successfully.

6 Acknowledgements

This project is supported by a grant from the Australian Government Department of the Prime Minister and Cabinet. NICTA is funded by the Australian Government's *Backing Australia's Ability* initiative, in part through the Australian Research Council.

References

[1] A. Yilmaz and M. Gokmen, "Eigenhill vs. eigenface and eigenedge," in *15th International Conference on Pattern Recognition 2000*, Barcelona, Spain, 2000, pp. 827–830.

[2] Y. Gao and M. K. Leung, "Face recognition using line edge map," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 6, pp. 764–779, 2002.

[3] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 643–660, 2001.

[4] R. Basri and D. W. Jacobs, "Lambertian reflectance and linear subspaces," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 2, pp. 218–233, 2003.

[5] M. J. Black, D. J. Fleet, and Y. Yacoob, "Robustly estimating changes in image appearance," *Computer Vision and Image Understanding*, vol. 78, no. 1, pp. 8–31, 2000.

[6] X. Liu, T. Chen, and B. V. Kumar, "Face authentication for multiple subjects using eigenflow," *Pattern Recognition*, vol. 36, pp. 313–328, 2003.

[7] A. M. Martinez, "Recognizing expression variant faces from a single sample image per class," in *Computer Vision and Pattern Recognition*, Madison, Wisconsin, USA, 2003, pp. 353–358.

[8] ——, "Recognizing imprecisely localized, partially occluded and expression variant faces from a single sample per class," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 6, pp. 748–763, 2002.

[9] Y. Adinj, Y. Moses, and S. Ullman, "Face recognition: The problem of compensation for changes in illumination direction," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 721–732, 1997.

[10] 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.

[11] 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*, 2006, pp. 515–518.

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

[13] C. Liu and H. Wechsler, "Probabilistic reasoning models for face recognition," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Santa Barbara, California, USA, 1998.

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

[15] J. Daugman, "An information-theoretic view of analog representation in striate cortex," *Computational Neuroscience, MIT Press*, pp. 403–424, 1990.