Recognizing Faces by Weakly Orthogonalizing against Perturbations

Kenji NAGAO and Masaki SOHMA
nagao@mrit.mei.co.jp sohma@mrit.mei.co.jp

Matsushita Research Institute Tokyo, Higashimita, Tama-ku, Kawasaki, 214 JAPAN

Abstract. In this paper, we address the problem of face recognition under drastic changes of the imaging processes through which the facial images are acquired. A new method is proposed. Unlike the conventional algorithms that use only the face features, the present method exploits the statistical information of the variations between the face image sets being compared, in addition to the features of the faces themselves. To incorporate the face and perturbation features for recognition, a technique called *weak orthogonalization* of the two subspaces has been developed that transforms the two overlapped subspaces such that the volume of the intersection of the resulting two subspaces is minimized. Matching is performed in the transformed face space that has thus been weakly orthogonalized against perturbation space. Results using real pictures of the frontal faces from drivers' licenses demonstrate the effectiveness of the new algorithm.

1 Introduction

A considerable amount of literatures have been published in face recognition in recent years. Among those, some of the most successful schemes are based on the Karhunen-Loeve transform (KLT) (or principal component analysis PCA) of the gray level images [5, 6, 7, 3] (see also [2] for frequency domain representation). In this appoach, facial images are compared in a low dimensional subspace [3] called face subspace, included in the whole image vector space, that maximizes the scatters of the projected distribution of faces. This method works very well as long as the imaging conditions for both of the face image sets being compared are similar to each other. However, problems occur when it comes to data sets with large or complicated differences as addressed in the present work (see e.g. figure 1). This is because they do not explicitly take into account how the individual faces change in appearance between the image sets being compared, but such perturbation was handled only implicitly. For example, in the KLT approach, perturbations were excluded by somehow truncating the higher order eigenvectors of the face space in the PCA and matching was performed in the subspaces spanned by the remaining eigenvectors. However, we should note here that perturbations of faces happen independently of the face space configurations, as one would never be able to tell how the imaging conditions, such as lighting directions, and thus the resulting images could change from the canonical appearances. Hence, only using the static face space information is not sufficient for recognizing faces in images with severe deviations. This is also true for other non-subspace approaches such as correlation methods[8]. This is the point that was not considered in previous work and now motivates our research.

Suppose we try to find matches of given faces presented to the camera, let us call this set A, in the registered face images of the database, set B. An example of such pairs of frontal face images is given in figure 1, where the set A is the pictures of the subjects themselves while the set B is obtained through the pictures printed on their plastic drivers' license. Since the imaging processes of both face sets are thus totally different, drastic changes have been introduced between the corresponding face images as shown in the figure. Clearly, such type of face recognition is different from the traditional types that allow only small changes of the appearances. In this paper, we tackle such kind of difficulties of human face recognition, given the pair of frontal facial images with severe variation of the appearances due to the changes of imaging conditions.

To this end, our method exploits perturbation information between image sets explicitly, by introducing *perturbation subspace* and combining it with the face subspace that was used in previous work. A perturbation subspace is a feature space within which any variation between the given pairs of face images is restricted to exist. To incorporate this perturbation subspace with the face space, we develop a technique called *weak orthogonalization* of those two subspaces such that after a weak orthogonalization, the intersection of those two subspaces is minimized, thereby effectively excluding the major components of the perturbations from the face representations. Results on 181 pairs of pictures (362 facial images) of the individuals and their drivers' licenses show the effectiveness of the proposed method for face recogntion under drastic changes of the imaging conditions.

As a related work to our method, recently in [9](see also [1]) a method for face recognition was developed for dealing with image variations due to the changes of lighting, based on Fisher's discriminant analysis. In their work, training images acquired under a variety of lighting conditions were used to construct classes of individual faces, and the optimal projection matrix for classification was computed based on the Fisher ratio criterion.

2 Face and Perturbation Subspaces

The problem can be stated: Given a set of training facial images that are pairs of images of individuals in sets A^l and B^l, identify each face from the test set A^t in the test set B^t, where A^l, A^t and B^l, B^t are resepectively acquired under the same imaging conditions.

We emphasize here again that two kind of differences are included in the face recognition task: one is the difference due to face identity that is described by the face subspace and the other is the change which is brought by different imaging

type A type B

Fig. 1. An example of image pair

The left is the image of the subject himself, type A, while the right image is obtained through a picture printed on his plastic drivers' lisence, type B.

conditions. A perturbation subspace is a feature space for representing the latter change.

As mentioned, the face subspace is designed so that individual faces are discriminated most effectively [5, 6, 7, 3], by taking as its bases the principal components of the face distributions that are the eigenvectors of its covariance matrix: $C_F = \frac{1}{N}\sum_{i=1}^{N}(F_i - M)(F_i - M)^T$. where F_i is the representation of the face labeled i from the sets A^l and/or B^l, M is their mean over i, and N is the number of face images used for the training phase. Similarly, the perturbation subspace is spanned by the principal components obtained as the eigenvectors of the autocorrelation matrix defined below with the large eigenvalues (NOTE: for the sake of simplicity in this paper we may sometimes use the term covariance for the autocorrelation of perturbation ignoring the subtraction operation by its mean in covariance estimation): $C_P = \sum_{i=1}^{N} P_i {P_i}^T$ where P_i's are sample perturbation vectors that are simply computed by taking the differences of the sets A^l and B^l: $P_i = A_i^l - B_i^l$ where i is a label to individual person, and N is the number of faces used for learning the perturbation space. We assume here that the perturbation of the images happening to individual faces are statistically consistent over the different faces.

Here, now that we use the term subspace for perturbation, the perturbations should be enclosed in a low dimensional space like the face subspace in the high dimensional vector space of the images, which may not be guaranteed in general. However, suppose that for the image sets A and B the respective imaging conditions are somehow fixed, then, as analyzed later in the experiment, the corresponding perturbations constitute a subspace. Thus, such a perturbation subspace might vary depending on the changes of the imaging processes and may have to be computed for every face recognition task.

We also note that since the perturbation defined above is derived simply by the subtraction between face pictures in the set A^l and their corresponding pictures in B^l, if the two face sample spaces of A and B are compatible, from the mathematical definition of a subspace, the resulting perturbation space is simply a subspace of that common face subspace. Thus, at a glance it appears that any feature extracted from the perturbation space is nothing more than an element of that face space, implying that perturbation space may be useless. This is not true, however, because even if the perturbation space is thoroughly contained in the face space, the principal components, i.e., the major axes of those subspaces, might differ from each other. Then, if it is true, we might still have a chance to exclude or suppress the disturbance of perturbations in matching face images. This issue will also be examined empirically later.

3 Weakly Orthogonalizing Face and Perturbation Subspaces

In this section, a method for feature extraction is derived that is suited for recognition, incorporating face and perturbation information introduced so far. Particularly, we focus on the scheme to separate the perturbations from the descriptions of faces.

3.1 Orthogonal Subspace Method Revisited

In 1970's and early 80's, there was once a technical topic called Orthogonal Subspace Methods in statistical pattern classification. This aimed at feature extraction from multiclass patterns suited for classification and recognition. Motivated by the early work of Fukunaga&Koontz[10, 11], they sought ways of transforming patterns to be compared prior to classification such that resulting class subspaces are orthogonal with each other in a mathematical sense [17, 16, 13, 12]. The Fukunaga&Koontz method utilized the mathematical relationship between the two class autocorrelation matrices such that when one whitens the mixture autocorrelation of the two class distributions, one can obtain shared eigen vectors between the resulting two autocorrelations with respective eigenvalues in reversed order.

We reilluminate the orthogonal subspace method based on Fukunaga&Koontz by defining its relaxed concept called *Weakly Orthogonalized* subspaces and by deriving a generalized procedure for performing it.

[Definition: Weakly Orthogonalized Subspaces]
Suppose we are given two distributions $\{X_1\}$ and $\{X_2\}$ with covariance matrices Σ_1 and Σ_2. Let these covariances have SVD's such that $\Sigma_1 = \Phi_1 \Lambda^{(1)} \Phi_1$, $\Sigma_2 = \Phi_2 \Lambda^{(2)} \Phi_2$, where Φ_1, Φ_2 are orthonormal matrices and $\Lambda^{(1)}, \Lambda^{(2)}$ are diagonal matrices such that $\Lambda^{(1)} = diag[\lambda_1^{(1)}, \lambda_2^{(1)}, \lambda_3^{(1)}, \cdots]$, $\Lambda^{(2)} = diag[\lambda_1^{(2)}, \lambda_2^{(2)}, \lambda_3^{(2)}, \cdots]$. Here, if those covariances share the same eigenvectors, i.e., $\Phi_1 = \Phi_2 \equiv \Phi$, and at the same time if the orders of the the eigenvalues are reversed, i.e., if $\lambda_1^{(1)} \geq \lambda_2^{(1)} \geq \lambda_3^{(1)} \cdots$, then $\lambda_1^{(2)} \leq \lambda_2^{(2)} \leq \lambda_3^{(2)} \cdots$, the we say the two distributions $\{X_1\}$ and $\{X_2\}$ are *weakly orthogonalized.*

As described, this is exactly the state that was performed by Fukunaga&Koontz for two-class classification problem. Although, it has not been used so widely for multiclass classification, when the two classes are associated with the faces and their perturbations, it turns out to be a powerful tool for recognition. Given the second order statistics of face and perturbation, if we have a way to transform these two subspaces into mutually weakly orthogonalized ones, we could effectively reduce the effect of the perturbations contained in the face representations on matching by selecting a subspace that has only small overlaps with the perturbation subspace.

In the following, we show a generalized procedure for such weak orthogonalization by two steps: the first step transforms two covariances into ones sharing eivenvectors (let us call this operation TCSE for short), then the second step reorders the two series of eivenvalues so that their orders are reversed.

3.2 Simultaneous Diagonalization: Yielding Shared Eigenvectors

Suppose we transform two covariances Σ_1, Σ_2 by a non-singular affine transformation L to yield new covariances $L\Sigma_1 L^T, L\Sigma_2 L^T$. It is known that there exists a class of transformations L that can simultaneously diagonalize them. Apparently, the operation of TCSE is almost equivalent to simultaneous diagonalization because: (1) The operation of TCSE is an instance of simultaneous diagonalization. (2)Simultaneous diagonalization includes TCSE in the middle of the procedure, where the resulting diagonal matrix is exactly the eigenvalue matrix resulting from the TCSE operation.

A well known procedure for performing simultaneous diagonalization is the one that makes one of the two matrices, say Σ_1 an identity matrix, and the other, Σ_2, a diagonal matrix $\Lambda^{(2)}$. Specifically,

$$L\Sigma_1 L^T = I \tag{1}$$

$$L\Sigma_2 L^T = \Lambda^{(2)} \tag{2}$$

where matrix L^T is given as the eigenvector matrix of $\Sigma_1^{-1}\Sigma_2$(see e.g.[4]). Apparently, the class of affine transformations that can perform simultaneous diagonalization, and thus TCSE, includes an infinite number of elements. However, the following property regarding the description of this class is noteworthy:

[Proposition]

Suppose a non-singular matrix L simultaneously diagonalizes two covariance matrices Σ_1,Σ_2. Then, any matrix H that can simultaneously diagonalize those two convariances can be written as a product of L and some diagonal matrix D as $H = DL$

(proof of this proposition is found in the appendix). This property implies that the eigenvectors resulting from the TCSE operation are unique with respect to the distribution.

3.3 Diagonal Scaling for Reordering Eigenvalues

When simultaneous diagonalization has been performed on two covariances, we can weakly orthogonalize the two distributions as defined, by further applying a diagonal rescaling matrix to order corresponding eigenvalues in reverse in two covariances. Let affine transformation L diagonalize both Σ_1 and Σ_2:

$$L\Sigma_1 L^T = \Lambda^{(1)} \tag{3}$$

$$L\Sigma_2 L^T = \Lambda^{(2)} \tag{4}$$

where $\Lambda^{(1)}, \Lambda^{(2)}$ are diagonal matrices such that $\Lambda_1 = diag[\lambda_1^{(1)}, \lambda_2^{(1)}, \cdots]$, $\Lambda_2 = diag[\lambda_1^{(2)}, \lambda_2^{(2)}, \cdots]$. Let the rescaling diagonal matrix with positive diagonal components be Υ, premultiplying Υ and postmultiplying $\Upsilon^T (= \Upsilon)$ on both sides of eq's. (3)–(4), we can adjust the amplitude such that:

$$(\Upsilon L)\Sigma_1(\Upsilon L)^T = \Upsilon \Lambda^{(1)} \Upsilon^T$$

$$(\Upsilon L)\Sigma_2(\Upsilon L)^T = \Upsilon \Lambda^{(2)} \Upsilon^T.$$

To reorder eigenvalues in reverse in the two covariances, we have the following rules to follow:

$$rule1: \quad \lambda^{(1)}{}'_i + \lambda^{(2)}{}'_i = 1 \tag{5}$$

$$rule2: \quad \lambda^{(1)}{}'^2_i + \lambda^{(2)}{}'^2_i = 1 \tag{6}$$

$$rule3: \quad \lambda^{(1)}{}'^{\frac{1}{2}}_i + \lambda^{(2)}{}'^{\frac{1}{2}}_i = 1 \tag{7}$$

$$rule4: \quad \lambda^{(1)}{}'^n_i + \lambda^{(2)}{}'^n_i = 1 \tag{8}$$

$$rule5: \quad \lambda^{(1)}{}'_i \cdot \lambda^{(2)}{}'_i = 1 \tag{9}$$

where primes denote eigenvalues after rescaling, and n in rule4 is an arbitrary real number. Apparently, when the new eigenvalues satisfy any of the above rules, they are ordered in reverse, that is, when $\lambda^{(1)}{}'_1 \geq \lambda^{(1)}{}'_2 \geq \cdots$, then $\lambda^{(2)}{}'_1 \leq \lambda^{(2)}{}'_2 \leq \cdots$, and, thus, weak orthogonalization has been performed. To carry out the rescaling operation subject to these rules, we just need to apply an appropriate diagonal matrix Υ_i $(i = 1, 2 \cdots)$ as described in the following:

$$\Upsilon_1 = diag[(\lambda_1^{(1)} + \lambda_1^{(2)})^{-\frac{1}{2}}, (\lambda_2^{(1)} + \lambda_2^{(2)})^{-\frac{1}{2}}, \cdots] \tag{10}$$

$$\Upsilon_2 = diag[(\lambda^{(1)}{}_1^2 + \lambda^{(2)}{}_1^2)^{-\frac{1}{4}}, (\lambda^{(1)}{}_2^2 + \lambda^{(2)}{}_2^2)^{-\frac{1}{4}}, \cdots] \tag{11}$$

$$\Upsilon_3 = diag[(\lambda^{(1)}{}_1^{\frac{1}{2}} + \lambda^{(2)}{}_1^{\frac{1}{2}})^{-1}, (\lambda^{(1)}{}_2^{\frac{1}{2}} + \lambda^{(2)}{}_2^{\frac{1}{2}})^{-1}, \cdots] \tag{12}$$

$$\Upsilon_4 = diag[(\lambda^{(1)}{}_1^n + \lambda^{(2)}{}_1^n)^{-\frac{1}{2n}}, (\lambda^{(1)}{}_2^n + \lambda^{(2)}{}_2^n)^{-\frac{1}{2n}}, \cdots] \tag{13}$$

$$\Upsilon_5 = diag[(\lambda_1^{(1)}\lambda_1^{(2)})^{-\frac{1}{2}}, (\lambda_2^{(1)}\lambda_2^{(2)})^{-\frac{1}{2}}, \cdots] \tag{14}$$

Rescaling by rule1 is identical to Fukunaga&Koontz method, that is, to decorrelate the mixture of the face and perturbation distributions. Rule4 is the generalized version of rules 1,2,3, by which the sum of the nth power (n is a real number) of the correponding eigenvalues in both of the covariances are normalized to 1. If one applies the rule5, the transformed covariance matrices are inverses of each other. Thus, by controling the relationship between the corresponding eigenvalues as above, we can order them in reverse between the two covariances and can control the speed of the descent/ascent of the curves of the orderd eigenvalues. Thus, through a simultaneous diagonalization of face and perturbation covariances followed by a diagonal rescaling, we can obtain weakly orthogonalized subspaces, thereby minimizing the size the intersection of the two subspaces.

Summary

A notion called weakly orthogonalized two subspaces has been introduced. This aimed at suppressing the turbulance in recognition due to the perturbations contained in the representation of faces, by reducing the size of the intersection of face and perturbation subspaces. To perform this, a generalized version of Fukunaga&Koontz approach that was for orthogonal subspace method was derived. The generalized procedure consists of a simultaneous diagonalization for obtaining shared eigenvectors and a subsequent diagonal rescaling for ordering the eivenvalues of the face and perturbation covariances in reverse.

4 Recognition Algorithm

Now we can show the core of the algorithm for face recognition that can handle extreme changes of imaging processes, using the technique developed in the previous section. The algorithm is described by two parts: one for training phase for computing the transformation for weak orthogonalization, which is a off-line process, and the other for run time phase of recognition.

Training phase

- step 1-a: Given the face training sets A^l and B^l, estimate covariance C_F of the face distribution using both or either of the two sets.
- step 1-b: Compute the autocorrelation of the perturbations C_P between A^l and B^l as defined in the previous section.
- step 2: Simultaneously diagonalize two matrices C_F, C_P via some appropriate transformation L.
- step 3: Rescale the resulting diagonal elements of both of the transformed covariances using the diagonal transformations Υ described in (10)–(14).
- step 4: Compute the matrix Ψ as a product of Υ and L: $\Psi = \Upsilon L$, which weakly orthogonalizes the face and perturbation subspaces as derived in the previous section.
- step 5 : Transform each of the face vectors in the database B^t using the matrix Ψ and retain the result in the database, where the origin has been set to the mean of all the entry of the database.

Recognition phase

- step 5: Transform the input image of the set A^t by the same transformation Ψ and find the best match in the database.

Step 1-a and 1-b can be performed in parallel. As is true of the case with other subspace based face recognition algorithms, the run time process is computationally very efficient.

5 Experimental Results

In this section, we will present the empirical results of face recognition on two image sets with extreme variation due to the differences of their imaging processes. To examine the effectiveness of the proposed method, the performance has been compared with those of Turk&Pentland's Eigenface approach [5, 7] and another perturbation method based on ML (Maximum Likelihood) estimate (hereafter MLP). As for the density function for the MLP, when we assume that the distribution of the perturbation is normal and is, as described, independent of the face identity, the only parameter we need for describing the common perturbation distribution is the autocorrelation C_P, where the mean is assumed to be zero and the metric for the Gaussian kernel is C_P^{-1}.

5.1 Preliminaries

Three different algorithms were implemented on a Sparc Station SS-UA1. To strictly evaluate the effectiveness of our weak orthogonalization approach against changes of the imaging conditions, variations of the geometrical properties in the images such as the orientation and the size of the faces were normalized prior to the actual training and recognition process. This was done for each image from the sets A and B by first manually specifying the left and right eye centers, and then setting the interocular distance and the direction of eye-to-eye axis, performaing scale and rotation invariance of images. Although the above process includes manual operation, we understand that it can be replaced by an automated process using the techniques for face detection presented in, e.g., [5, 8]. The actual face parts used for the tests were masks for eyes (50x20 pixels), nose(25x25 pixels), and the whole face (45x60 pixels; the area from eyebrows downwards).

5.2 Recognition Experiments

The first experiment tests the recognition performance of Weak Orthogonalization method by a comparison with two other algorithms: Eigenface and MLP, demonstrating the superiority of the proposed method. In the second test of the present method, we check the effect of the selection of the principal components of the transformed face space on recognition performance.

[Test1: comparison with Eigenface and MLP]

Recognition performances of these different algorithms were compared by randomly choosing 80 different pairs of training (A^l, B^l) and test sets (A^t, B^t), where each pair consists of 91 pairs of images from types A and B. The training and test sets had no overlaps. In the following, explanations for using a single pair of training and test sets are given. Performance evaluations are simply by the results from the iterations on 80 different pairs of the learning and recognition image sets.

Training phase

Training operations were performed for each of the three different algorithms. For the Weak Orthogonalization method: first face covariance C_F was estimated by using the 91 type A images. Similarly, perturbation covariance C_P was estimated, where the differences between types A and B of each face were simply used for perturbation samples. Then, covariances C_F and C_P were simultaneously diagonalized using the affine transformation described in (1)–(2). From these transformed face covariances, 90 principal components were obtained and all of those were used, which means the dimension of the transformed face subspace was 90 (Note that this is exactly the rank of the face sample matrix of which each column face vector has been subtracted by their mean). For the diagonal rescaling, we selected rule 3. As will be described soon, as the number of images used for training was too small, the choice of diagonal rescaling matrix did not make any difference on recognition performance in this experiment. In estimating the covariance of face distribution for Eigenface, the mixture of sets A^l and B^l was used, instead of using C_F computed above. This is because using information from only one type of image set, e.g. set A, is not fair when comparing with weak orthogonalization method that uses information of both types, though original Eigenface used only one of the two comparing sets. For the MLP method, we simply used the same covariance C_P for computing the Mahalanobis distance between the two images as was used for weak orthogonalization method. For both the Eigenface and MLP algorithms, 90 principal components were used. Hence, the dimensions of the extracted feature vectors as well as the used face sample sets, each for computing feature extraction matrix by three different algorithms are the same. For each of the images in the set B^t feature extraction was performed using the three different algorithms and the results were stored in the database.

Recognition phase

Recognition tests were conducted using the remaining 91 pairs of potentially corresponding types A and B (A^t, B^t) images that were not used for the training phase. For each of the images from the set A^t feature extraction was done in the same way as for B^t, and the best match was found using the Euclidean distance between the input (A^t) and database (B^t) facial feature representations that had 90 dimensions.

In figure 2 the correct recognition percentage for the eye part using the three different algorithms are shown, where the horizontal axis simply denotes the different pairs of training and test sets: the top graph shows the results for our weak orthogonalization method, the middle the MLP method, and the bottom the Eigenface method. Here, correct recognition means that the correct pair was found to be the best match. For any of the training and test sets, the weak orthogonalization method performed best. The performance of MLP was always second. Eigenface performed very poorly on this experiment. In table 1,

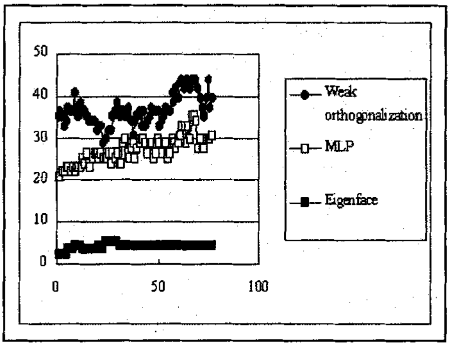

Fig. 2. Comparison with Eigenface and MLP

The correct recognition rates for the eye part are shown, where the horizontal axis denotes simply the different pairs of training and test sets: the top graph is our weak orthogonalization method, the middle the MLP method, and the bottom the Eigenface.

the cumulative percentage of the correct match being ranked within the top 10 ranks for eye images are given. It is seen that the weak orthogonalization method is very stable compared with the other methods.

In table 2, similar results are presented on regions of the nose and the whole face, where the average correct recognition rates are shown, which are also from the 60 different pairs of training and test image sets. From these results, the effectiveness of our weak orthogonalization method is confirmed.

Rank	Weak Orthogonalization	MLP	Eigenface
1	36.3	20.9	2.2
2	49.5	25.3	6.6
3	53.8	31.9	9.9
4	60.4	35.2	11.0
5	61.5	39.6	12.1
6	63.7	42.9	14.3
7	65.9	50.5	18.7
8	68.1	53.8	19.8
9	69.2	57.1	20.9
10	70.3	60.4	24.2

Table 1. Cumulative percentage of correct match within top 10 ranks

The weak orthogonalization method is very stable as compared with the other methods.

[Test2: The effect of the selection of the subspace]

	Weak Orthogonalization	MLP	Eigenface
Eye	36.6	27.0	4.2
Nose	18.6	12.6	3.2
Whole face	41.1	30.9	26.5

Table 2. Correct recognition rates: averages

The average correct recognition rates for nose and whole facial regions as well as eye part are presented, which are also from the 60 different pairs of training and test image sets. On any of the facial regions, weak orthogonalization method performed best.

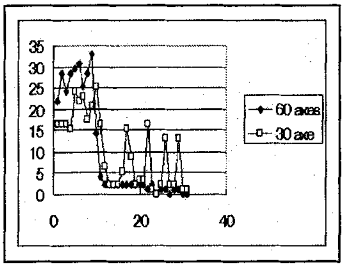

Fig. 3. Effect of the selection of the subspace

The horizontal axis is the order of the componet set (from the lowest to highest) and the vertical axis is the percentage of the correct recognition.

The dependency of the recognition performance of the weak orthogonalization method on the selection of principal components spanning the subspace of the faces was examined using the eye images. Tests were conducted by selecting 60 (and 30) components for spanning the subspace out of the 90 eigenvectors of the face distribution obtained by PCA after the weak orthogonalization operation. The selection was made in consequtive order from the lowest (having largest eigenvalues) 60 components to the highest. Figure 3 shows the results where the horizontal axis is the order of the componet set (from the lowest to the highest) and the vertical axis is the percentage correct recognition. Roughly, we can say that after the 10th order, the performance declined sharply for both of the 60 and 30 axes case. Thus, we tentatively conclude that there is a difference in the recognition power between the principal components after weak orthogonalization, and lowest components appear to have the maximum recognition ability.

[Configurations of weakly orthogonalized face and perturbation subspaces]

In the above tests, for constucting a face subspace that is weakly orthogonalized against perturbation space, we used 91 face image samples and 91 perturbation samples. Since this number was too small as compared with the dimension (1000)

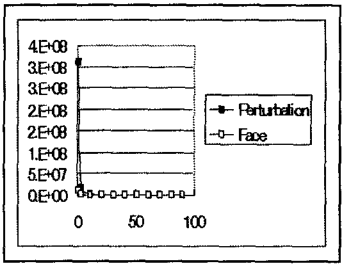

Fig. 4. Eigenvalues assciated with pricipal components

Eigenvalues along the major components of the face (set A^l) and the perturbation subspaces defined as $\{P_i|A_i - B_i\}$ are shown, where the horizontal axis is the order and the vertical axis is the eigenvalues. It can be observed that only a small set of principal components occupy a large part of the variation of the perturbation distribution.

of the eye images, when we performed simultaneous diagonalization, those two subspaces (exactly the spaces for the training sets) were completely orthogonalized in the mathematical sense. This implies that eigenvectors were not shared in practice: for the components of face subspace the eigenvalues for perturbation space were zeros, and vice versa. For this reason, in this paper we could not evaluate the variation of the rescaling methods as it assumes eigenvectors are truly shared.

6 Viability Tests

It may be of readers concern what kind of specific classes of the face recognition problem the present technique may be applied to, as was also addressed in section 2. At present, it is not easy for us to give exact answer to this question. However, we may give indications for this by defining the following two quantitative measures, both of which are computed using the PCA outputs of the training images: K: the degree of concentration of variations of perturbations to lower order components, and V: the degree of overlaps between the face and the perturbation subspaces.

$$K = \frac{1}{N} min[i]\{\frac{\sum_{j=1}^{i} \sigma_j}{\sum_{j=1}^{N} \sigma_j} > r\} \tag{15}$$

$$V = E\{log\left[\frac{min\{eval_1^i, eval_2^{f(i)}\} * cosine}{min\{eval_1^i, eval_2^i\}}\right]\} \tag{16}$$

where in the first equation, N is the number of obtained principal components, and $0 < r < 1$, and in the second equation, $eval_1^i$ is the square root of the ith eigenvalue for the first, let say perturbation, subspace, and $eval_2^i$ is the similar eigen property of the second, let say set A, subspace, $f(i)$ is the operation giving the order of component of the second subspace that has the maximum coincidence in direction with the ith component of the first subspace in terms of the consine. The notation $E\{\cdot\}$ denotes the averaging operation over i. Recall that our method assumes that perturbations consistute a subspace which might have some overlaps with face subspace, and also that it is the objective of the method to try to exclude or suppress the disturbance of such perturbations on matching faces. Therefore, we hypothesize that the smaller the values of K and V for the image sets the better our method should perform.

We have made an examination of these measures using the same image sets as those for the face recongition experiments. Figure 4 shows the eigenvalues along the 90 principal components of the eye images (set A) and their perturbation defined as $\{P_i|A_i - B_i\}$. When we set $r = 0.95$, the values of K were 0.23 for the perturbation and 0.22 for the set A. From this, it can be seen that the pertubation distribution can be represented as a small set of major componets, as their associated eigenvalues occupy the most part in the total variance, which allows us to define the perturbation subspace. We have also estimated the degree of overlap between the face and perturbation subspaces using the measure V. This was done for perturbation vs. face set A (P-FA) and perturbation vs. face set B (P-FB), and face set A vs. face set B (FA-FB) was also included for comparison. The results are -1.18 for P-FA, -1.58 for P-FB, and -0.48 for FA-FB. From this, it is observed that the degrees of overlap between perturbation and face subspaces (P-FA, P-FB) are lower than that between the two face subspaces (FA-FB). Thus, both of the measures K and V were low for those given specific image sets A and B used here, which supports the improved recognition performance of the proposed method over the existing techniques as demonstrated in the experiments.

7 Concluding remarks

We have proposed a face recogntion algorithm that can handle extreme changes of appearance due to drastic changes of imaging conditions. In the present method, the statistical information of the perturbation of the face images from the registered corresponding ones are exploited, in addition to the features of the faces themselves that were used in conventional methods. In order to incorporate both of the face and perturbation features for recognition, we developed a technique called weak orthogonalization of the two subspaces that transforms the given two overlapped subspaces such that the volume of the intersection of the resulting two subspaces is minimized. Empirical results using real pictures of the frontal faces from drivers' licenses demonstrated the effectiveness of the new algorithm. The viability of the proposed method for a given specific face recognition task may be evaluated using the measure K: the degree of concentration of the eigenvalues and the measure V: the degree of overlap between the perturbation and face subpaces.

Appendix

Proof of Proposition

Let non-singular matrix A simultaneously diagonalizes two covariances Σ_1, Σ_2:

$$A\Sigma_1 A^T = \Lambda_1 \quad (17)$$
$$A\Sigma_2 A^T = \Lambda_2 \quad (18)$$

Let matrix B be any one that performs similar diagonalization as:

$$B\Sigma_1 B^T = \Omega_1 \quad (19)$$
$$B\Sigma_2 B^T = \Omega_2 \quad (20)$$

Matrix B can be written as a product of A and some matrix C as $B = CA$. Substituting this into (19) and using (17), we have:

$$C\Lambda_1 C^T = \Omega_1 \quad (21)$$
$$C\Lambda_2 C^T = \Omega_2 \quad (22)$$

This can be rewritten as:

$$(C\Lambda_1^{\frac{1}{2}})(C\Lambda_1^{\frac{1}{2}})^T = (\Omega_1^{\frac{1}{2}})(\Omega_1^{\frac{1}{2}})^T \quad (23)$$
$$(C\Lambda_2^{\frac{1}{2}})(C\Lambda_2^{\frac{1}{2}})^T = (\Omega_2^{\frac{1}{2}})(\Omega_1^{\frac{1}{2}})^T \quad (24)$$

From (23), we have $C = \Omega_1^{\frac{1}{2}} U \Lambda_1^{-\frac{1}{2}}$, where U is an orthogonal matrix. Substituting this into (24) and rearranging, we have:

$$U\Lambda_{2/1} U^T = \Omega_{2/1} \quad (25)$$

where $\Lambda_{2/1} = \Lambda_2 \Lambda_1^{-1}$ and $\Omega_{2/1} = \Omega_2 \Omega_1^{-1}$.

Noting that (25) shows a similarity transformation between $\Lambda_{2/1}$ and $\Omega_{2/1}$ ($U^T = U^{-1}$), we have $\Lambda_{2/1} = \Omega_{2/1}$ and U is an identity matrix. Therefore, we obtain finally $C = \Omega_1^{\frac{1}{2}} \Lambda_1^{-\frac{1}{2}}$, which completes the proof.

References

1. T. Kurita, N. Ohtsu and T. Sato, "A Face Recognition Method Using Higher Order Local Autocorrelation and Multivariate Analysis", In Proc. IEEE ICPR92, pp. 213–216, 1992.
2. S. Akamatsu et. al., "An Accurate and Robust Face Identification Scheme", In Proc. IEEE ICPR92, pp. 217–220, 1992.
3. M. Kirby, L. Sirovich, "Application of the Karhunen-Loeve procedure for the characterization of human faces", *IEEE Trans. Patt. Anal. Machine Intell.*, vol. 12, pp. 103–108, 1990.
4. K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press 1972.

5. M. Turk, A. Pentland, "Face recognition using eigenfaces", In Proc. IEEE CVPR91, 1991.
6. M. Turk, A. Pentland, "Eigenfaces for recognition", Journal of Cognitive Neuroscience, vol. 3, No. 1, 1991.
7. A. Pentland, B. Moghaddam, T. Starner, "View-based and modular eigenspaces for face recognition", In Proc. IEEE CVPR94, 1994.
8. R. Brunelli, T. Poggio, "Face Recognition: Features versus Template", *IEEE Trans. Patt. Anal. Machine Intell.*, vol. PAMI-8, pp.34–43, 1993.
9. N. Belhumeur, P. Hespanha, J. Kriegman, "Eigenfaces vs. fisherfaces:recognition using class specific linear projection", In Proc. ECCV'96 vol1 pp. 45–58, 1996.
10. K. Fukunaga, W. C. G. Koontz, "Application of the Karhunen-Loeve expansion to feature extraction and ordering", *IEEE Trans. Computers*, Vol. C-19, pp. 311–318, 1970.
11. K. Fukunaga, *Introduction to Statistical Pattern Recognition*, Academic Press 1972.
12. D. H. Foley, J. W. Sammon, "An Optimal Set of Discriminant Vectors", IEEE Trans. Computers, Vol. C-24, NO. 3, pp. 281–289, March 1975.
13. E. Oja, J. Karhunen, "An Analysis of Convergence for a Learning Version of the Subspace Method", J. Math. Anal. Applications 91. pp. 102–111, 1983.
14. E. Oja, *Subspace Methods of Pattern Recognition*, Research Studies Press LTD and John Wiley & Sons Inc., 1983.
15. J. Kittler, "The subspace approach to pattern recognition", in *Progress in cybernetics and systems research*, p. 92, Hamisphere Publ. Co., Washington, 1978.
16. T. Kohonen, P. Lehtio, E. Oja, "Spectral classification of phonemes by learning subspacees", Helsinki University of Technology, Dept. of Technical Physics, Report TKK-F-A348. Also in Proc. IEEE ICASP, pp. 2–4, April 1979.
17. S. Watanabe, N. Pakvasa, "Subspace method of pattern recognition", In Proc. IJCPR pp. 25–32, 1973.