

Face Recognition based on Whitening Transformation of Distribution of Subspaces

Tomokazu Kawahara, Masashi Nishiyama, Tatsuo Kozakaya, Osamu Yamaguchi
 Corporate Research & Development Center Toshiba Corporation
 1, Komukai-Toshiba-cho, Saiwai-ku, Kawasaki-shi, Kanagawa 212-8582, Japan
 {tomokazu.kawahara, masashi.nishiyama, tatsuo.kozakaya, osamu1.yamaguchi}@toshiba.co.jp

Abstract

We propose a new face recognition method that separates subspaces representing individuals based on the mathematical analysis of angles between multiple subspaces. A low-dimensional subspace representation by principal component analysis is known to be an effective approach for describing variation of facial patterns. A similarity between individuals is defined by an angle between their subspaces. Since all facial patterns have the same structure of facial parts, it is significant to extract individual characteristics from each subspace by considering the cross-relationship between categories. Our method applies “whitening transformation of distribution of subspaces”, which can uniformize the distribution according to eigenvalues of the autocorrelation matrix of the subspaces. We derive the equation relating angles between subspaces to uniformity of the distribution of these subspaces. From this equation, the whitening transformation is effective for separation of the subspaces. Under the ideal condition, the whitening transformation orthogonalizes all subspaces. In other words, all similarities between each other are equal to 0. We show the proposed method works well even in a practical case through evaluation experiments on the FRGC 1.0 and the FERET databases and outperforms other methods.

1 Introduction

Many face identification methods have been proposed, which represent variation of patterns for an individual as a low-dimensional subspace generated from a set of patterns by principal component analysis (PCA) [2, 7, 11]. Since these methods are able to cope with variation in appearance, a robust face identification application can be built.

Yamaguchi et al. have proposed face recognition using the *Mutual Subspace Method* (MSM)[11]. They represent not only reference patterns as a reference subspace but also

input patterns as an input subspace. To compare an input subspace with the reference subspace representing an individual, a similarity of MSM is defined by an angle between the input subspace and the reference subspace. MSM has a problem in that reference subspaces crowd since all facial patterns have the same structure of facial parts and MSM does not have a function that separates the subspaces of individuals.

To improve the recognition accuracy by separating subspaces, Fukui et al. have extended MSM to the *Constrained Mutual Subspace Method* (CMSM)[2]. In CMSM, reference subspaces are projected onto a constraint subspace, which is designed to emphasize the difference between individuals. Fukui et al. confirmed empirically that the projection to the constraint subspace creates a larger angle between multiple reference subspaces and explained that subspaces are separated because the common components of subspaces are removed.

We propose a new method, the *Whitened Mutual Subspace Method* (WMSM), based on a more mathematical analysis of angles between subspaces, which uses the whitening transformation of the distribution of subspaces for separation of subspaces. Whitening is a process to make a distribution uniform. First, we derive the equation that relates angles between multiple subspaces to a standard deviation of eigenvalues of an autocorrelation matrix of these subspaces. This equation describes that uniformizing a distribution of multiple subspaces makes angles between these subspaces larger. In other words, the whitening transformation emphasize the difference between individuals. In particular, the whitening transformation of a distribution of subspaces orthogonalizes reference subspaces when the number of reference subspaces is small. We show the proposed method works well even in a practical case through evaluation experiments on the FRGC 1.0 and the FERET databases and outperforms other methods.

The remainder of this paper is organized as follows. First, to explain the reason for using the whitening transformation of a distribution of subspaces mathematically,

we analyze angles between multiple subspaces in section 2. Next, we describe the proposed method of face recognition in section 3. We demonstrate the effectiveness of our method by face recognition experiments in section 4.

2 Mathematical analysis of angles between subspaces

In this section, we explain the mathematical reason for using the whitening transformation of a distribution of subspaces for separation of these subspaces. For the purpose of the explanation, two mathematical objects that are calculated from multiple subspaces are prepared and the equation describing the relationship between these mathematical objects are derived. One of the mathematical objects is a measure of separability of multiple subspaces, that consists of canonical angles between these subspaces [1]. The other is an autocorrelation matrix of multiple subspaces[2]. We derived the equation that consists of a measure of separability of multiple subspaces and a standard deviation of eigenvalues of an autocorrelation matrix of these subspaces. This equation describes that a measure of separability of subspaces becomes large when a standard deviation of eigenvalues of this matrix becomes small. In other word, uniformizing distribution of subspaces separates these subspaces. Based on this mathematical analysis of angles between subspaces, we propose the method using whitening transformation for separation of multiple subspaces.

In MSM, a similarity between two subspaces is defined by an angle between these subspace. In this paper, therefore, we represent that subspaces separate when angles between these subspaces are large.

2.1 A measure of separability of subspaces

In this section, we define a measure of separability of two subspaces based on canonical angles between these subspaces and extend it to multiple subspaces.

To prepare the definition of a measure of separability of subspaces, we explain canonical angles between two subspaces, which are described in [1]. d canonical angles $\theta^{(1)}, \dots, \theta^{(d)}$ between the d -dimensional subspaces \mathbf{V}_1 and \mathbf{V}_2 in a vector space are defined as follows;

- $\mathbf{V}_1^{(1)} = \mathbf{V}_1$ and $\mathbf{V}_2^{(1)} = \mathbf{V}_2$.
- $\theta^{(i)}$ is the angle between $v_1^{(i)}$ and $v_2^{(i)}$, where $v_1^{(i)} \in \mathbf{V}_1^{(i)}$ and $v_2^{(i)} \in \mathbf{V}_2^{(i)}$ are the nearest vectors under the condition $|v_1^{(i)}| = |v_2^{(i)}| = 1$.
- $\mathbf{V}_1^{(i+1)} = \{v \in \mathbf{V}_1^{(i)} | v \perp v_1^{(i)}\}$ and $\mathbf{V}_2^{(i+1)} = \{v \in \mathbf{V}_2^{(i)} | v \perp v_2^{(i)}\}$.

where $i = 1, \dots, d$ and $|\cdot|$ denotes the norm. The subspaces $\mathbf{V}_j^{(1)}, \dots, \mathbf{V}_j^{(d)}$ have the following relation;

$$\mathbf{V}_j^{(1)} \supset \mathbf{V}_j^{(2)} \supset \dots \supset \mathbf{V}_j^{(d)} \quad (1)$$

where $j = 1, 2$. In particular, $\theta^{(1)}$ is equal to the angle between \mathbf{V}_1 and \mathbf{V}_2 . Therefore, we use canonical angles instead of a single angle because more detailed analysis of separability is possible. When two subspaces are identical and orthogonal, all canonical angles are equal to 0 and $\pi/2$, respectively. From the definition of canonical angles, we obtain the inequation $\theta^{(1)} \leq \dots \leq \theta^{(d)}$.

The canonical angles between these subspaces become large when two subspaces separate. Therefore, we define a measure of separability of two subspaces \mathbf{V}_1 and \mathbf{V}_2 as follows:

$$\text{Sep}(\mathbf{V}_1, \mathbf{V}_2) = 1 - \frac{1}{d} \sum_{i=1}^d \cos^2 \theta^{(i)}, \quad (2)$$

where $\theta^{(1)}, \dots, \theta^{(d)}$ are canonical angles between \mathbf{V}_1 and \mathbf{V}_2 . If two subspaces are identical and orthogonal, measures of separability of these subspaces are equal to 0 and 1, respectively. When this measure of two subspaces is large, these two subspaces separate.

For calculation of the measure of two subspaces (2) using orthonormal bases of these subspaces, we derive the equation between a measure of separability of two subspaces and projection matrices of these subspaces. The projection matrix \mathbf{P} of subspace \mathbf{V} is defined by equation (3) [8].

$$\mathbf{P} = \sum_{i=1}^d \psi_i \psi_i^T, \quad (3)$$

where $\{\psi_1, \dots, \psi_d\}$ is an orthonormal basis of \mathbf{V} . Generally, a projection matrix is defined by $d \times D$ matrix $(\psi_1, \dots, \psi_d)^T$ where D is dimension of the vector space. However, we use the former definition since the latter definition does not have the information of position of the subspace on the vector subspace. Let \mathbf{P}_j be the projection matrix of \mathbf{V}_j , where $j = 1, 2$. By calculation of the trace of $\mathbf{P}_1 \mathbf{P}_2$, the equation (4) is obtained.

$$\text{Sep}(\mathbf{V}_1, \mathbf{V}_2) = 1 - \frac{1}{d} \text{tr}(\mathbf{P}_1 \mathbf{P}_2), \quad (4)$$

where $\text{tr}(\cdot)$ is a trace of a matrix, which is a sum of diagonal components of the matrix. (See Appendix for a detailed calculation of (4)).

We extend a measure of two subspaces (2) to a measure of separability of multiple subspaces. Let $\mathbf{V}_1, \dots, \mathbf{V}_N$ be d -dimensional subspaces in a D -dimensional vector space. A measure of separability of subspaces $\mathbf{V}_1, \dots, \mathbf{V}_N$ is defined as an average of measures of \mathbf{V}_k and \mathbf{V}_l ($1 \leq k <$

$l \leq N$)

$$\text{Sep}(\mathbf{V}_1, \dots, \mathbf{V}_N) = \frac{2}{N(N-1)} \sum_{1 \leq k < l \leq N} \text{Sep}(\mathbf{V}_k, \mathbf{V}_l). \quad (5)$$

When all subspaces are identical and orthogonal, measures of separability of these subspaces are equal to 0 and 1, respectively. The more this measure of multiple subspaces is, the more these subspaces separate. We obtain the equation (6) from (5) and (4).

$$\text{Sep}(\mathbf{V}_1, \dots, \mathbf{V}_N) = 1 - \frac{2}{N(N-1)} \sum_{1 \leq k < l \leq N} \frac{1}{d} \text{tr}(\mathbf{P}_k \mathbf{P}_l), \quad (6)$$

where \mathbf{P}_k is the projection matrix of \mathbf{V}_k defined by (3). Therefore, we calculate a measure of separability of multiple subspaces using orthonormal bases of these subspaces.

2.2 An autocorrelation matrix of subspaces

To prepare calculation of a measure of multiple subspaces (5) we explain an autocorrelation matrix of distribution of subspaces, which is described in [2], and calculate an average and a standard deviation of its eigenvalues. An autocorrelation matrix of distribution of subspaces \mathbf{A} is defined as an average of all projection matrices, like an autocorrelation matrix of distribution of vectors [8], and its eigenvalue problem is solved as follows,

$$\mathbf{A} = \frac{1}{N} \sum_{k=1}^N \mathbf{P}_k = \mathbf{B} \mathbf{A} \mathbf{B}^T, \quad (7)$$

where \mathbf{B} is the matrix whose columns are the orthonormal eigenvectors of \mathbf{A} and $\mathbf{\Lambda}$ is the diagonal matrix of the corresponding eigenvalues $\lambda_1 \geq \dots \geq \lambda_D$.

We calculate an average and a standard deviation of eigenvalues of an autocorrelation matrix. Let m_λ and σ_λ be an average and a standard deviation of eigenvalues $\lambda_1, \dots, \lambda_D$, respectively. In the first step, we calculate an average of eigenvalues of an autocorrelation matrix. An average of eigenvalues of the autocorrelation matrix m_λ is equal to the constant value d/D regardless of arrangement of subspaces $\mathbf{V}_1, \dots, \mathbf{V}_N$ from the following calculation,

$$\begin{aligned} m_\lambda &= \frac{1}{D} \sum_{l=1}^D \lambda_l = \frac{1}{D} \text{tr} \mathbf{A} = \frac{1}{D} \text{tr} \left(\frac{1}{N} \sum_{k=1}^N \mathbf{P}_k \right), \\ &= \frac{1}{DN} \sum_{k=1}^N \text{tr}(\mathbf{P}_k) = \frac{1}{DN} \sum_{k=1}^N d = \frac{d}{D}. \end{aligned} \quad (8)$$

Next, we calculate a standard deviation of eigenvalues of an

autocorrelation matrix using (8) as follows,

$$\begin{aligned} \sigma_\lambda &= \frac{1}{D} \sum_{l=1}^D (\lambda_l - m_\lambda)^2, \\ &= \frac{1}{D} \sum_{l=1}^D \lambda_l^2 - m_\lambda^2 = \frac{1}{D} \text{tr} \mathbf{A}^2 - \left(\frac{d}{D} \right)^2. \end{aligned} \quad (9)$$

2.3 Equation between a measure of separability and an autocorrelation matrix

In this section, we show that a separability of multiple subspaces is decided from only a standard deviation of eigenvalues of an autocorrelation matrix from the equation that consists of a measure of separability of multiple subspaces, a standard deviation of eigenvalues of an autocorrelation matrix and a constant term.

Using (6), (8) and (9), a measure of separability of multiple subspaces $\mathbf{S} = \text{Sep}(\mathbf{V}_1, \dots, \mathbf{V}_N)$ is calculated as follows,

$$\begin{aligned} \mathbf{S} &= 1 - \frac{2}{N(N-1)} \sum_{1 \leq k < l \leq N} \frac{1}{d} \text{tr}(\mathbf{P}_k \mathbf{P}_l), \\ &= 1 - \frac{1}{dN(N-1)} \sum_{1 \leq k \neq l \leq N} \text{tr}(\mathbf{P}_k \mathbf{P}_l), \\ &= 1 - \frac{1}{dN(N-1)} \text{tr} \left(\sum_{k,l=1}^N \mathbf{P}_k \mathbf{P}_l - \sum_{k=1}^N \mathbf{P}_k \right), \\ &= 1 - \frac{1}{dN(N-1)} \text{tr}(N^2 \mathbf{A}^2 - N \mathbf{A}), \\ &= -\frac{DN}{d(N-1)} \sigma_\lambda^2 + \frac{N(D-d)}{(N-1)D}. \end{aligned} \quad (10)$$

From the equation (10), a transformation that decreases the standard deviation of eigenvalues σ_λ separates the subspaces $\mathbf{V}_1, \dots, \mathbf{V}_N$. In particular, all subspaces are separated most when all eigenvalues $\lambda_1, \dots, \lambda_D$ are the same values.

2.4 Whitening transformation of distribution of subspaces

We propose whitening transformation of distribution of subspaces for separation of multiple subspaces based on the analysis in the previous section. From the analysis in section 2.3, a transformation that decreases standard deviation of eigenvalues of autocorrelation matrix of subspaces separates these subspaces. In other words, whitening transformation of distribution of subspaces is effective to separate these subspaces (Fig. 1). ‘‘Whitening’’ is a process to make all eigenvalues of an autocorrelation matrix the same. The

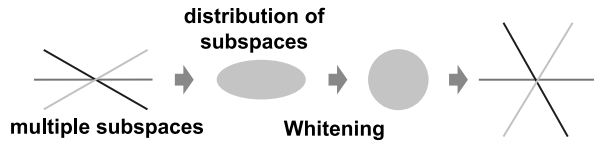


Figure 1. The ellipse and the circle in the center of the figure represent the distribution of subspaces. “Whitening” makes the distribution uniform.

matrix \mathbf{W} that represents whitening transformation of distribution of subspaces $\mathbf{V}_1, \dots, \mathbf{V}_N$ is defined and makes an autocorrelation matrix the identity matrix \mathbf{I} as follows:

$$\begin{aligned} \mathbf{W} &= \mathbf{\Lambda}^{-1/2} \mathbf{B}^T, & (11) \\ \mathbf{WAW}^T &= (\mathbf{\Lambda}^{-1/2} \mathbf{B}^T) \mathbf{B} \mathbf{A} \mathbf{B}^T (\mathbf{B} \mathbf{\Lambda}^{-1/2}) = \mathbf{I} & (12) \end{aligned}$$

where \mathbf{B} and $\mathbf{\Lambda}$ are defined in (7).

Another method, named the *Orthogonal Subspace Method* (OSM), in which whitening orthogonalizes subspaces, has been proposed by Fukunaga et al.[3] and Kittler[4]. In OSM, an autocorrelation matrix of each class is transformed by the whitening of the autocorrelation matrix generated from all samples in all classes before a subspace of each class is generated from the eigenvectors of the autocorrelation matrix of this class, the eigenvalues of which are large. In other words, a set of samples in each class is represented as a low-dimensional subspace after the distribution of all samples in all classes is made uniform. In this method, the eigenvector of an autocorrelation matrix of a class, the eigenvalue of which is 1, is orthogonal to all samples in other classes since all eigenvalues of the autocorrelation matrix generated from all samples in all classes are equal to 1.

In our method and OSM, subspaces are orthogonalized using whitening. The difference between our method and OSM is the order of the linearization and the transformation. In other words, an input subspace and a reference subspace are generated from a set of patterns before whitening in our method, but after whitening in OSM. Therefore, our method does not use eigenvectors of the autocorrelation matrix whose eigenvalues are small. Furthermore, when the number of subspaces is small, these subspaces are always orthogonalized in our method but not always orthogonalized in OSM (section 2.5).

2.5 Transformation under the ideal condition

We show that subspaces can be orthogonalized by the whitening transformation of distribution of these subspaces

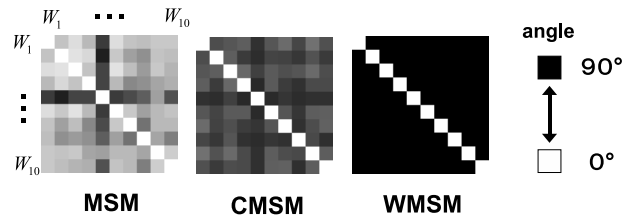


Figure 2. Similarity matrix: angles between ten reference subspaces in MSM, CMSM and WMSM in the case that the condition (14) is satisfied. The darker a pixel is, the larger the angle between subspaces is.

in the ideal case that the number of these subspaces is small.

In the first step, we prove the following proposition. Let u_1, \dots, u_N be bases of 1-dimensional subspaces in D -dimensional vector space. A matrix \mathbf{U} denotes (u_1, \dots, u_N) and $\mathbf{\Lambda}, \mathbf{\Lambda}^{-1/2}, \mathbf{B}, \mathbf{W}$ are defined as in (7) and (11). Let u'_i be $\mathbf{W}u_i$ for all i .

Proposition 1 if u_1, \dots, u_N is linearly independent, u'_1, \dots, u'_N is orthonormal.

Proof Let \mathbf{U}' be $\mathbf{W}\mathbf{U}$. Since $\mathbf{U}\mathbf{U}^T = \mathbf{A} = \mathbf{B}\mathbf{A}\mathbf{B}^T$ in the equation (7) and u_1, \dots, u_N is linearly independent,

$$\mathbf{U}'\mathbf{U}'^T = \mathbf{W}\mathbf{U}\mathbf{U}^T\mathbf{W}^T = \mathbf{\Lambda}^{-1/2}\mathbf{B}^T\mathbf{B}\mathbf{A}\mathbf{B}^T\mathbf{B}\mathbf{\Lambda}^{-1/2} = \tilde{\mathbf{I}}_N, & (13)$$

where $\tilde{\mathbf{I}}_N$ is a diagonal matrix in which the number of 1 on the diagonal is N and others are 0. The symmetric matrix $\mathbf{U}'^T\mathbf{U}'$ is an identity matrix because all eigenvalues of $\mathbf{U}'^T\mathbf{U}'$ are the same as those of $\mathbf{U}'\mathbf{U}'^T$ without 0 and the rank of $\mathbf{U}'^T\mathbf{U}'$ is N . Therefore, u'_1, \dots, u'_N is orthonormal because a component of $\mathbf{U}'^T\mathbf{U}'$ is an inner product of u'_k and u'_l . \square

Generally, we can orthogonalize all subspaces using the whitening transformation of the distribution of these subspaces if the following inequation is satisfied;

$$dN \leq D, & (14)$$

where D is the dimension of the vector space including these subspaces, because we apply Proposition 1 to bases of these subspaces. The equation (14) requires the dimension or the number of these subspaces to be small. In the case that the condition (14) is satisfied, the measure of separability of subspaces transformed by the whitening transformation is equal to 1 since the autocorrelation matrix of these subspaces is $\tilde{\mathbf{I}}_{dN}$ from the same calculation (13).

Fig. 2 shows similarity matrix images whose pixel values represent an angle between pairs of subspaces in MSM,

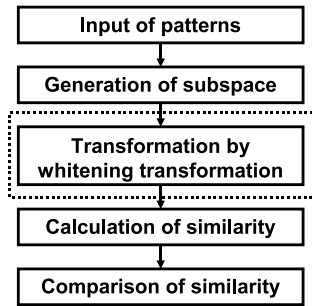


Figure 3. The flow chart of WMSM.

CMSM, and our method in the case that the condition (14) is satisfied. This figure shows that our method orthogonalizes all these subspaces.

3 Face Recognition using the whitening transformation of the distribution of subspaces

In this section, we describe the procedure of WMSM (Fig. 3).

3.1 Algorithm for face recognition

First, we located the face pattern from the positions of the feature points and cropped to 32×32 pixels using 3D normalization[5] and preprocessing[6]. In order to adapt localization error of feature points, we represent variation of face patterns due to the localization error as a subspace in the feature space by perturbation of the feature points and obtaining multiple face patterns from a single face image. We apply PCA to the vectors to generate an input subspace. Let $\{x_i\}_{i=1,\dots,n}$ be a set of vectors. The basis of the input subspace is the eigenvectors of the autocorrelation matrix $\mathbf{Z} = 1/n \sum_{i=1}^n x_i x_i^T$ [8].

The whitening transformation (11) is generated from an autocorrelation matrix of reference subspaces. To allow for the variation in appearance for each individual, it is effective to increase the dimension of the reference subspace by addition of other bases that are generated from reference patterns and not used for comparison with an input subspace.

To compare the input subspace with the reference subspace registered in a database for each individual, we calculate their similarities after transforming the input subspace and the reference subspaces by the whitening transformation of a distribution of reference subspaces. The person in the image is identified as the person who corresponds to the reference subspace with the highest similarity.

3.2 Transformation of a subspace and calculation of a similarity

In our proposed method, to transform the input subspace $\mathbf{V}_{\text{input}}$ and the reference subspace \mathbf{V}_{ref} by whitening of distribution of reference subspaces, we carry out the following steps:

1. Transform a basis of a subspace by the whitening transformation \mathbf{W} .
2. Apply Gram-Schmidt orthogonalization to them.

The orthonormal basis is a basis of the transformed subspace.

We define a similarity s between the d -dimensional subspaces $\mathbf{V}_{\text{input}}$ and \mathbf{V}_{ref} as $s = \cos^2 \theta$, where θ is the angle between $\mathbf{V}_{\text{input}}$ and \mathbf{V}_{ref} . The angle θ is equal to the 1-th canonical angle $\theta^{(1)}$ between $\mathbf{V}_{\text{input}}$ and \mathbf{V}_{ref} . If $\mathbf{V}_{\text{input}}$ and \mathbf{V}_{ref} are identical, the angle θ is equal to 0. The angle is calculated using the MSM[11]. The similarity s equals the largest eigenvalue λ_{max} of $\mathbf{X} = (x_{mn})$ using

$$x_{mn} = \sum_{l=1}^d (\psi_m, \phi_l)(\phi_l, \psi_n) \quad (m, n = 1 \dots d), \quad (15)$$

where $\{\psi_i\}_{i=1,\dots,d}$ and $\{\phi_j\}_{j=1,\dots,d}$ are the orthonormal bases of $\mathbf{V}_{\text{input}}$ and \mathbf{V}_{ref} , respectively; (ψ_m, ϕ_l) is the inner product of ψ_m and ϕ_l .

4 Evaluation with the FRGC 1.0 and FERET databases

We show the proposed method works well even in a practical case. We performed experiments using the controlled still images (exp1) in the FRGC 1.0 database [9] and the *fa* and the *fb* data sets in the FERET database[10]. The controlled still images in FRGC 1.0 consisted of 152 gallery images and 608 probe images. The *fa* and the *fb* in FERET consisted of images of 1196 people with one image per person and 1195 people with one image per person, respectively.

We compare five methods, namely, MSM, CMSM, *Multiple CMSM* (MCMSM) [7], WMSM and *Multiple WMSM* (MWMSM). MCMSM and MWMSM apply ensemble learning with bagging to CMSM and WMSM, respectively. In MCMSM, multiple constraint subspaces are generated from reference subspaces selected randomly in the same way of bagging. The input subspace and the reference subspaces are projected onto each constraint subspace and a similarity is determined with the similarities

Table 1. The methods and their parameters. d is the dimension of input and reference subspaces. L is the number of constraint subspaces and whitening transformations. d' is the dimension of reference subspaces that generate constraint subspaces and whitening transformations. C is the dimension of constraint subspaces.

	d	L	d'	C
MSM	7	—	—	—
CMSM	7	1	15	210
MCMSM	7	10	15	210
WMSM	7	1	15	—
MWMSM	7	10	15	—

Table 2. Experimental results using FRGC 1.0 in terms of Correct Match Rate (CMR) and Equal Error Rate (EER).

	CMR (%)	EER (%)
MSM	96.4	3.45
CMSM	96.5	2.47
MCMSM	97.2	2.28
WMSM	97.0	1.81
MWMSM	97.2	1.81

calculated on each constraint subspace. In MWMSM, multiple whitening transformations are generated from reference subspaces selected randomly and the similarity is determined with an average of the similarities calculated after transformation by each whitening transformation. Their parameters in the experiments are listed in Table 1.

Table 2 shows the evaluation results in FRGC 1.0 for each method in terms of Correct Match Rate (CMR) and Equal Error Rate (EER). Correct Match Rate is the probability that an input of the right person is correctly accepted. Equal Error Rate is the probability that false acceptance rate (FAR) equals the false rejection rate (FRR). It can be seen that the proposed method and the proposed method with ensemble learning are equivalent to MCMSM with regard to Correct Match Rate and superior to the other methods with regard to Equal Error Rate on FRGC 1.0.

To evaluate the generalization ability of our method, we performed experiments using another database. Fig. 4 shows the evaluation results for each method and the best result (UMD97) of the partially automatic algorithms reported in FERET'97 [10] in terms of Cumulative Match Rate. It can be seen that the proposed method and the proposed method with ensemble learning are superior to the other methods.

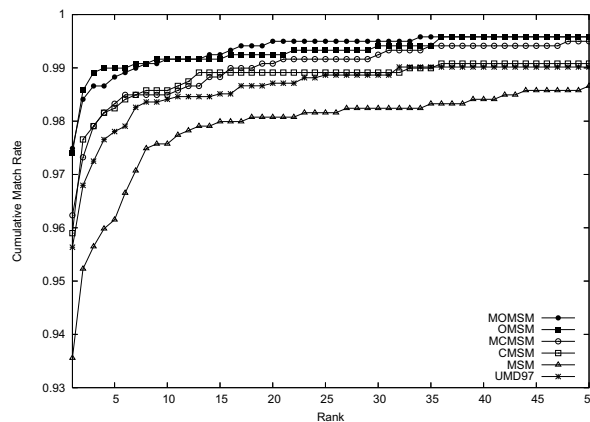


Figure 4. Experimental results using FERET database in terms of Cumulative Match Rate.

5 Conclusions

This paper presented a face recognition method based on mathematical analysis of angles between subspaces in which we apply whitening of a distribution of subspaces to emphasize the difference between individuals. We derived the equation (10) that relates angles between subspaces to a distribution of these subspace. This equation describes that the whitening transformation is effective for separation of these subspaces. In the experiment, we obtained high performance compared with other methods on the FRGC 1.0 and the FERET database.

References

- [1] F. Chaitin-Chatelin. *Eigenvalues of Matrices*. Johe Wiley & Sons Ltd, 1993. (Enlarged Translation of the French Publication with Masson).
- [2] K. Fukui and O. Yamaguchi. Face recognition using multi-viewpoint patterns for robot vision. In *11th International Symposium of Robotics Research*, pages 192–201, 2003.
- [3] K. Fukunaga and W. Koontz. Application of the Karhunen-Loeve expansion to feature extraction and ordering. *IEEE Transactions on Computers*, 19(5):311–318, 1970.
- [4] J. Kittler. The subspace approach to pattern recognition. In R. Trappl, G. J. Klir, and L. Ricciardi, editors, *Progress in cybernetics and systems research*, Washington, 1978. Hemisphere Publishing Corporation.
- [5] T. Kozakaya and O. Yamaguchi. Face recognition by projection-based 3d normalization and shading subspace orthogonalization. In *Proceedings IEEE 7th International Conference on Automatic Face and Gesture Recognition*, pages 163–168, 2006.
- [6] M. Nishiyama and O. Yamaguchi. Face recognition using the classified appearance-based quotient image. In *Proceed-*

ings *IEEE 7th International Conference on Automatic Face and Gesture Recognition*, pages 49–54, 2006.

- [7] M. Nishiyama, O. Yamaguchi, and K. Fukui. Face recognition with the multiple constrained mutual subspace method. In *Audio- and Video-based Biometric Person Authentication*, pages 71–80, 2005.
- [8] E. Oja. *Subspace Methods of Pattern Recognition*. Research Studies Press, 1983.
- [9] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek. Overview of the face recognition grand challenge. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2005.
- [10] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss. The FERET evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1090–1104, 2000.
- [11] O. Yamaguchi, K. Fukui, and K. Maeda. Face recognition using temporal image sequence. In *Proceedings of IEEE 3rd International Conference on Automatic Face and Gesture Recognition*, pages 318–323, 1998.

A Calculation of a measure of separability of subspaces

To obtain equation (4), we prove the following equation,

$$\frac{1}{d} \text{tr}(\mathbf{P}_1 \mathbf{P}_2) = \text{Sep}(\mathbf{V}_1, \mathbf{V}_2). \quad (16)$$

To calculate the trace of the product of projection matrices, we describe several facts about vectors $v_1^{(1)}, \dots, v_1^{(d)}$ and $v_2^{(1)}, \dots, v_2^{(d)}$ in section 2.1. A set of vectors $\{v_j^{(1)}, \dots, v_j^{(d)}\}$ is an orthonormal basis of \mathbf{V}_j since $v_j^{(i)}$ is orthogonal to $\mathbf{V}_j^{(i+1)}$, where $j = 1, 2$. Furthermore, $v_1^{(k)}$ is orthogonal to $v_2^{(l)}$ ($k \neq l$) since the equations (17) and (18) are derived from the definition of $v_1^{(i)}$ and $v_2^{(i)}$,

$$\mathbf{P}_2^{(i)} v_1^{(i)} = \cos \theta^{(i)} v_2^{(i)}, \quad (17)$$

$$\mathbf{P}_1^{(i)} v_2^{(i)} = \cos \theta^{(i)} v_1^{(i)}, \quad (18)$$

where $\mathbf{P}_j^{(i)}$ is projection matrix of $\mathbf{V}_j^{(i)}$ ($j = 1, 2$). From these facts, the following equations (19) and (20) is acquired.

$$(v_1^{(k)}, v_1^{(l)}) = (v_2^{(k)}, v_2^{(l)}) = \begin{cases} 1 & (k = l), \\ 0 & (k \neq l), \end{cases} \quad (19)$$

$$(v_1^{(k)}, v_2^{(l)}) = \begin{cases} \cos \theta^{(k)} & (k = l), \\ 0 & (k \neq l), \end{cases} \quad (20)$$

where (\cdot, \cdot) is the inner product of vectors.

We calculate the trace of the product of projection matrices in the equation (16). The projection matrix \mathbf{P}_j of \mathbf{V}_j

is defined by $\sum_{i=1}^d v_j^{(i)} v_j^{(i)T}$ since $\{v_j^{(1)}, \dots, v_j^{(d)}\}$ is an orthonormal basis, where $j = 1, 2$. By calculation of trace of $\mathbf{P}_1 \mathbf{P}_2$ using the equations (19) and (20) as follows, the equation (16) is obtained.

$$\begin{aligned} \frac{1}{d} \text{tr}(\mathbf{P}_1 \mathbf{P}_2) &= \frac{1}{d} \text{tr} \left(\left(\sum_{k=1}^d v_1^{(k)} v_1^{(k)T} \right) \left(\sum_{l=1}^d v_2^{(l)} v_2^{(l)T} \right) \right), \\ &= \frac{1}{d} \text{tr} \left(\sum_{k,l=1}^d v_2^{(l)T} v_1^{(k)} v_1^{(k)T} v_2^{(l)} \right), \\ &= \frac{1}{d} \sum_{k=1}^d \cos^2 \theta^{(k)} = \text{Sep}(\mathbf{V}_1, \mathbf{V}_2). \end{aligned}$$