Face Recognition Using Multi-viewpoint Patterns for Robot Vision — **Source link** ↗

Kazuhiro Fukui, Osamu Yamaguchi

Institutions: Toshiba

Related papers:

- Face recognition using temporal image sequence

- Discriminative Learning and Recognition of Image Set Classes Using Canonical Correlations

- Face Recognition from Long-Term Observations

- Learning over sets using kernel principal angles

- Robust Real-Time Face Detection

Face Recognition Using Multi-viewpoint Patterns for Robot Vision

Kazuhiro Fukui and Osamu Yamaguchi

Corporate Research and Development Center, TOSHIBA Corporation
1, KomukaiToshiba-cho, Saiwai-ku, Kawasaki 212-8582 Japan
kazuhiro.fukui@toshiba.co.jp / osamu1.yamaguchi@toshiba.co.jp

Abstract. This paper introduces a novel approach for face recognition using multiple face patterns obtained in various views for robot vision. A face pattern may change dramatically due to changes in the relation between the positions of a robot, a subject and light sources. As a robot is not generally able to ascertain such changes by itself, face recognition in robot vision must be robust against variations caused by the changes. Conventional methods using a single face pattern are not capable of dealing with the variations of face pattern. In order to overcome the problem, we have developed a face recognition method based on the *constrained mutual subspace method* (CMSM) using multi-viewpoint face patterns attributable to the movement of a robot or a subject. The effectiveness of our method for robot vision is demonstrated by means of a preliminary experiment.

1 Introduction

Person identification is a very important function for robots which work with humans in the real world. Face recognition is one of the essential methods identification since it is non-contact method and the subject can thus be unaware that recognition is being performed.

Many face recognition methods based on image appearance have been developed over the past few decades[6–8]. Most of these are based on a single face pattern and recognize a face using the model for the expected change. Although these methods have been in practical use for applications in which the lighting condition and face direction are stable, it is difficult to apply them in general. This is because the face pattern may change dramatically due to the changes of the relation between the positions of robot, subject and light sources. As a robot is not generally able to control such changes by itself, face recognition in robot vision must be robust against such variations.

In order to overcome the problem, we introduce a novel approach for face recognition using multiple face image patterns obtained in various views. Our approach exploits the following observations: (i) A robot can make a subject approach it or turn his/her face toward it by the means of visual or audio alerts. (ii) The robot can move to an advantageous position for capturing multi-viewpoint face patterns by itself. Face recognition by a robot is substantially different from that by a desktop computer in that a robot has the ability to actively capture face patterns.

Fig.1 shows a comparison between our method and a conventional method using a single face pattern. A face pattern obtained from one view can be represented

(a) Conventional method   (b) Our method
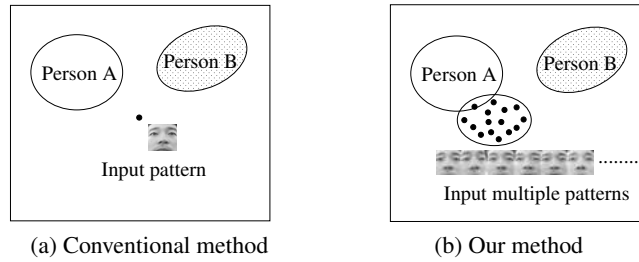
Fig. 1. Basic idea: (a) using a single static pattern, (b) using multiple patterns

as a point in a high-dimensional feature vector space where an n×n pixel pattern is treated as an $F(= n \times n)$ dimensional vector. In the conventional method the identification of a person is based on the minimum distance between an input pattern and a distribution of reference patterns as shown in Fig.1(a). The minimum distance is very unstable because the input face pattern varies easily due to the changes in face direction, expression and lighting. On the other hand, we can see that the similarity between the distributions of the input patterns and the reference patterns is more stable as shown in Fig.1(b). Consequently, our method based on the similarity between the two distributions is hardly affected by the changes mentioned above. Moreover, it should be noted that the similarity between the two distributions of face patterns implies implicitly the similarity between 3D shapes of faces[12]. This fact is one reason for the higher recognition rate of our method compared to single-view methods.

The distribution of face image patterns can be represented by a lower-dimensional linear subspace of the high-dimensional feature space. This subspace is generated using the Karhunen-Loève (KL) expansion, also known as principal component analysis (PCA). Moreover, the relationship between two subspaces is strictly defined by the multiple canonical angles[9], which are an extension of the angle between two vectors. Therefore, we can measure the structural similarity between the distributions of the face patterns by using the canonical angles between two subspaces. The canonical angles are calculated by the framework of the *mutual subspace method*(MSM)[3].

The MSM-based face recognition method using the multiple canonical angles can tolerate variations in the face patterns, considering the information due to the 3D face shape of each person and achieve a high recognition rate compared to the conventional methods. However, its classification ability still appears insufficient for face recognition because each subspace is created without considering the rival subspaces that are to be compared[2]. To overcome the problem, we consider employing the *constrained mutual subspace method* (CMSM). The essence of CMSM is to carry out the MSM framework in a constraint subspace $\mathcal{C}$ which satisfies the constraint condition: "it includes only the essential component for classification. The projection onto a constraint subspace enables CMSM to have a higher classification ability besides the ability to tolerate variations in the face patterns.
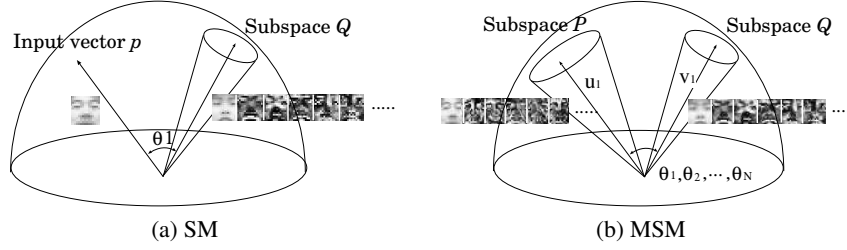
Fig. 2. Concept of SM and MSM

In the following sections, first, we explain the algorithm of MSM and CMSM. Then, we construct the face recognition method using CMSM. Finally, the effectiveness of our method for robot vision is demonstrated by experiments.

## 2 Basic identification by MSM

2.1 Algorithm of MSM

The *mutual subspace method* (MSM) is an extension of the subspace method (SM)[1,2] used widely for solving various pattern recognition problems. SM is for calculating the similarity as the minimum angle θ_1 between an input vector and a reference subspace which represents the variation of the learning set as shown in Fig.2(a). By contrast, in MSM, the similarity is defined by the minimum angle θ_1 between the input subspace and the reference subspace as shown in Fig.2(b). MSM utilizes only the minimum canonical angle. However, given an M-dimensional subspace $\mathcal{P}$ and an N-dimensional subspace $\mathcal{Q}$ in the F-dimensional feature space, we can obtain N canonical angles (for convenience $N \leq M$) between $\mathcal{P}$ and $\mathcal{Q}$[9]. Therefore, we use these canonical angles to define the similarity between these subspaces. The canonical angle θ_i between $\mathcal{P}$ and $\mathcal{Q}$ is defined as

$$cos^2\theta_i = \max_{\substack{\mathbf{u}_i \perp \mathbf{u}_{j(=1,\ldots,i-1)} \\ \mathbf{v}_i \perp \mathbf{v}_{j(=1,\ldots,i-1)}}} \frac{|(\mathbf{u}_i, \mathbf{v}_i)|^2}{||\mathbf{u}_i||^2||\mathbf{v}_i||^2} \quad (i = 1, \ldots, N) \tag{1}$$

where $\mathbf{u}_i \in \mathcal{P}, \mathbf{v}_i \in \mathcal{Q}, ||\mathbf{u}_i|| \neq 0$, and $||\mathbf{v}_i|| \neq 0$. Let the $F \times F$ dimensinal projection matrix corresponding to projection of a vector on the M-dimensional subspace $\mathcal{P}$ be $\mathbf{P}$, and the $F \times F$ dimensional projection matrix corresponding to the N-dimensional subspace $\mathcal{Q}$ be $\mathbf{Q}$. $cos^2\theta$ of the angle θ between $\mathcal{P}$ and $\mathcal{Q}$ is equal to the eigenvalue of $\mathbf{PQP}$ or $\mathbf{QPQ}$[3]. The largest eigenvalue of these matrices represents $cos^2\theta_1$ of the smallest canonical angle θ_1, whereas the second largest eigenvalue represented $cos^2\theta_2$ of the smallest angle θ_2 in the direction perpendicular to that of θ_1. $cos^2\theta_i$ for $i = 3, \ldots, N$ are calculated similarly.

The eigenvalue problem of $\mathbf{PQP}$ can be transformed to that of a matrix with smaller dimensions[3]. Let $\mathbf{\Phi}_i$ and $\mathbf{\Psi}_i$ denote the i-th F-dimensional orthogonal
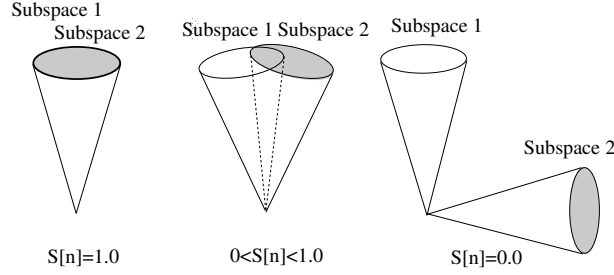
Fig. 3. Similarity $S[t]$ between two subspaces: it changes from 0.0 to 1.0 depending on the relation between two subspaces

basis vectors of the subspace $\mathcal{P}$ and $\mathcal{Q}$, respectively. The eigenvalue problem of **PQP** is then attributed to that of the $N \times N$ matrix $\mathbf{X}$[1] as:

$$\mathbf{X}\mathbf{c} = \lambda\mathbf{c} \tag{2}$$

where $\mathbf{X} = (x_{ij})$, $x_{ij} = \sum_{k=1}^{M}(\mathbf{\Psi}_i \cdot \mathbf{\Phi}_k)(\mathbf{\Phi}_k \cdot \mathbf{\Psi}_j)$.

2.2   Definition of similarity using multiple canonical angles

We consider the value of the mean of the canonical angles, $S[t] = \frac{1}{t}\sum_{i=1}^{t}\cos^2\theta_i$, as the similarity between two subspaces. The similarity $S[t]$ has the characteristic shown in Fig.3. In the case that two subspaces coincide completely with each other, $S[t]$ is 1.0, since all canonical angles are zero. The similarity $S[t]$ becomes smaller as the two subspaces separate. Finally, the similarity $S[t]$ is zero when the two subspaces are orthogonal to each other.

For practical use of the similarity $S[t]$, we should consider the situation in which both subject and robot stand still. In this case, $S[t]$ ($t \geq 2$) can not be used, since the distribution of input patterns degenerates and the canonical angles excepting the minimum canonical angle are unreliable[2]

3   High-performance identification by CMSM

3.1   Algorithm of CMSM

MSM does not have the ability to reject the influence of undesirable changes in a face pattern, such as changes due to lighting condition, a face direction or an expression. In Fig.4 the difference vector, $\mathbf{d}(|\mathbf{u}| = |\mathbf{v}| \neq \mathbf{0})$, of the two vectors, $\mathbf{u}$ and $\mathbf{v}$, composing the minimum angle, θ, includes some components derived from such changes and the angles between the subspaces will accordingly deviate from what should be observed solely between individuals.

---

[1] This matrix can be also derived by several other methods[9].

[2] In such situation, a multi-camera system is valid, since the system can obtain various face patterns at the same time without depending on the movement of the subject and the robot.
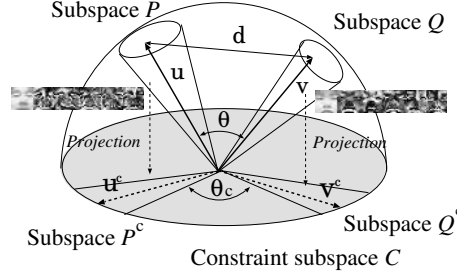
Fig. 4. Concept of CMSM

In order to attenuate the influence of such undesirable variation, we consider employing the *constrained mutual subspace method* (CMSM). In CMSM, we introduce a constraint subspace $\mathcal{C}$ which satisfies the constraint condition: "it includes the effective component for recognition, that is the difference between people, but does not include any unnecessary component for recognition, namely, undesirable variation.". Then, we calculate the canonical angles between the projected input subspace $\mathcal{P}^c$ and the projected reference subspace $\mathcal{Q}^c$ using MSM. The canonical angle θ_c between the M-dimensional projected input subspace $\mathcal{P}^c$ and the N-dimensional projected reference subspace $\mathcal{Q}^c$ is defined as:

$$cos^2\theta_i = \max_{\substack{\mathbf{u}_i^c \perp \mathbf{u}_{j(=1,\ldots,i-1)}^c \\ \mathbf{v}_i^c \perp \mathbf{v}_{j(=1,\ldots,i-1)}^c}} \frac{|(\mathbf{u}_i^c, \mathbf{v}_i^c)|^2}{||\mathbf{u}_i^c||^2||\mathbf{v}_i^c||^2} \quad (i = 1,\ldots,N) \tag{3}$$

where $\mathbf{u}_i^c \in \mathcal{P}^c, \mathbf{v}_i^c \in \mathcal{Q}^c, ||\mathbf{u}_i^c|| \neq 0, ||\mathbf{v}_i^c|| \neq 0$. The subspace $\mathcal{P}^c$ is calculated by the following steps:

1. M orthogonal basis vectors of the subspace $\mathcal{P}$ are projected onto the constraint subspace $\mathcal{C}$.
2. The length of each projected vector is normalized.
3. Gram-Schmidt orthogonalization is applied to the normalized vectors to obtain M orthogonal basis vectors of the subspace $\mathcal{P}^c$.

Similarly $\mathcal{Q}^c$ is calculated.

3.2   Generation of constraint subspace

How to generate the constraint subspace is an important problem. Here, we derive the constraint subspace based on the concept of *difference subspace* that we have proposed in [5]. First, the *difference subspace* is defined geometrically with the canonical angles. Then, we redefine the *difference subspace* using the projection matrices analytically. Finally, we generalize the concept of the *difference subspace* for multiple subspaces and show that the *generalized difference subspace* is available as the constraint subspace.
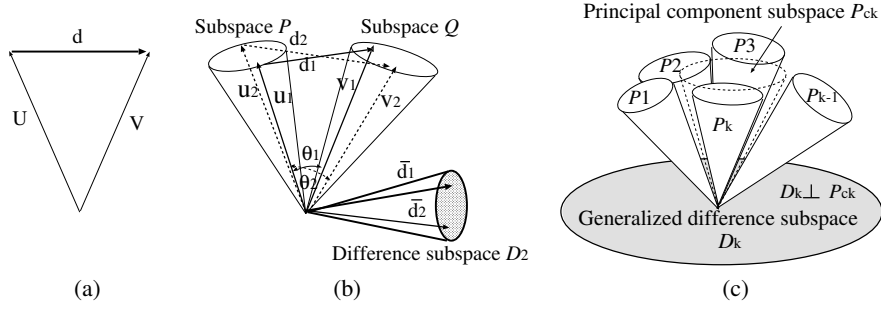
Fig. 5. Concept of difference subspace: (a) difference vector, (b) difference subspace, (c) generalized difference subspace for k subspaces

Definition of difference subspace Here, we formulate the *difference subspace* between two subspaces. The *difference subspace* is an extension of the difference vector between two vectors in multidimensional space as shown in Fig.5(a)(b). The *difference subspace* is an effective component for discriminating two subspaces.

When MSM is applied to M-dimensional subspace $\mathcal{P}$ and N-dimensional subspace $\mathcal{Q}$, N canonical angle $\theta_i, i = 1, \ldots, N$ (for convenience $N \leq M$) are obtained. Let $\mathbf{d}_i$ be the difference vector $\mathbf{u}_i$-$\mathbf{v}_i$ between vector $\mathbf{u}_i$ and vector $\mathbf{v}_i$ forming the i-th canonical angle θ_i. All $\mathbf{d}_i$ are then orthogonal to each other. By normalizing the length of each difference vector $\mathbf{d}_i$ to 1.0, we regard these normalized difference vectors $\bar{\mathbf{d}}_i$ as the basis vectors of the *difference subspace* $\mathcal{D}_2$.

We can redefine the *difference subspace*, defined geometrically, analytically using the projection matrices. For this purpose, we show that half of all the eigenvectors of matrix $\mathbf{P} + \mathbf{Q}$ corresponding to eigenvalues smaller than 1.0 span the *difference subspace* between the two subspaces. Let the i-th largest eigenvalue of matrix $\mathbf{P} + \mathbf{Q}$ be $\lambda_i(\mathbf{P} + \mathbf{Q})$ and the i-th largest eigenvalue of matrix $\mathbf{PQ}$ be $\lambda_i(\mathbf{PQ})$. Starting from $\lambda_i(\mathbf{PQ})=(\lambda_i(\mathbf{P} + \mathbf{Q}) - 1)^2$, proved in [10], we can obtain the relation between the eigenvalues of matrix $\mathbf{P} + \mathbf{Q}$, the *difference subspace* $\mathcal{D}_2$, and principal component subspace $\mathcal{P}_{c2}$ which represents the "principal component" of multiple subspaces as follows

1. N eigenvectors of matrix $\mathbf{P} + \mathbf{Q}$ corresponding to eigenvalues smaller than 1.0 span the *difference subspace* $\mathcal{D}$.
2. N eigenvectors of matrix $\mathbf{P} + \mathbf{Q}$ corresponding to eigenvalues larger than 1.0 span the *principal component subspace* $\mathcal{P}_c$.

The relations lead us to the conclusion that the sum subspace $\mathcal{S}_2$ spanned by all the eigenvectors of matrix $\mathbf{P} + \mathbf{Q}$ is represented by the orthogonal direct sum of the principal component subspace $\mathcal{P}_{c2}$ and the *difference subspace* $\mathcal{D}_2$. In other words, the *difference subspace* $\mathcal{D}_2$ can be defined as the subspace which is produced by removing the principal component subspace $\mathcal{P}_{c2}$ from the sum subspace $\mathcal{S}_2$.

From the above discussion, we see that the *difference subspace*, defined as the subspace which represents "difference" between two subspaces becomes at the same time the subspace, which does not include the principal components of the two subspaces. Thus, the *difference subspace* satisfies the constraint condition mentioned in Section 3.1.

Generalization of difference subspace Now, we generalize the concept of the difference subspace for multiple subspaces on the basis of the definition using projection matrices. Given $k(\geq 2)$ N-dimensional subspaces, a generalized difference subspace $\mathcal{D}_k$ can be defined as the subspace which is produced by removing the principal component subspace $\mathcal{P}_{ck}$ of all the subspaces from the sum subspace $\mathcal{S}_k$ of that. According to this definition, the generalized difference subspace $\mathcal{D}_k$ is actually spanned by N_c eigenvectors of the sum matrix $\mathbf{G} = \sum_{i=1}^{k} \mathbf{P}_i$ of projection matrices $\mathbf{P}_i$, $\mathbf{d}_i(i = N \times k - N_c, \dots, N \times k)$.

$$\mathbf{Gd} = \lambda \mathbf{d}, \tag{4}$$

where eigenvectors $\mathbf{d}_i$ correspond to the i-th eigenvalue λ_i in descending order and an optimal dimension N_c of the difference subspace is set experimentally[3].

Fig.5(c) shows the concept of the generalized difference subspace $\mathcal{D}_k$ for k subspaces. We can see that the generalized difference subspace $\mathcal{D}_k$ includes only the essential component for face recognition since it is orthogonal to the principal component subspace $\mathcal{P}_{ck}$, which represents the intersections of reference subspaces of each person.

3.3   Framework of face recognition using the CMSM

Fig.6 shows the framework of our face recognition based on CMSM. First, the pupils and nostrils are directly detected from an input image using a feature detection method based on a combination of separability filter and pattern matching[11]. The input subspace is updated, in real time, whenever a normalized face pattern is extracted on the basis of the position of the four facial feature points, then projected onto the constraint subspace. To update the input subspace we have adopted the simultaneous iteration method[2]. Then, we compute the similarity between the projected input subspace and the projected reference subspace on a database, and the reference subspace that has the highest similarity is determined to be that of the identified person given the similarity is above a threshold. The whole process from inputting an image to identification can be executed at the speed of 25 frames/sec using a PC (Pentium-III 800MHz).

---

[3] Eq.(4) appears to be an equation of principal component analysis (PCA) in that basis vectors of each subspace are considered as a sample vector. However, it is completely different from PCA in that m subspaces $\mathcal{P}_{ij}$ $(j = 1 \sim m)$ which belong to i-th class should be integrated into the same subspace $\mathcal{P}_{i*}$. Actually, $\mathcal{P}_{i*}$ can be generated as the subspace which is spanned by all the eigenvectors of the sum matrix $\sum_{j=1}^{m} \mathbf{P}_{ij}$.
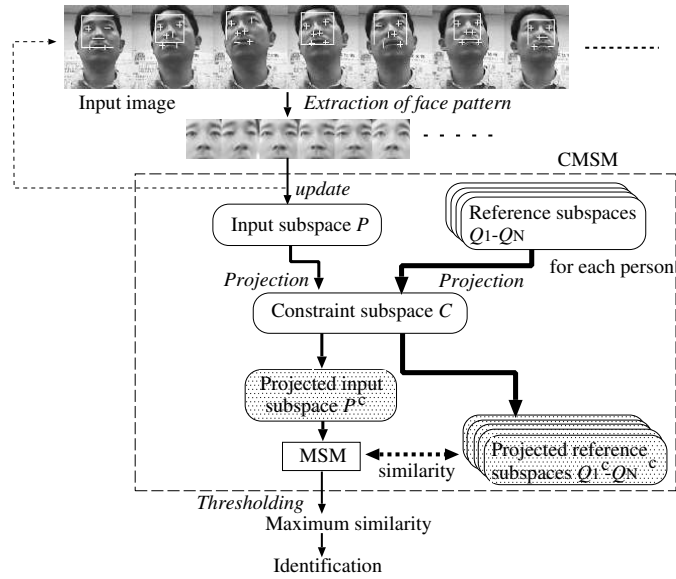
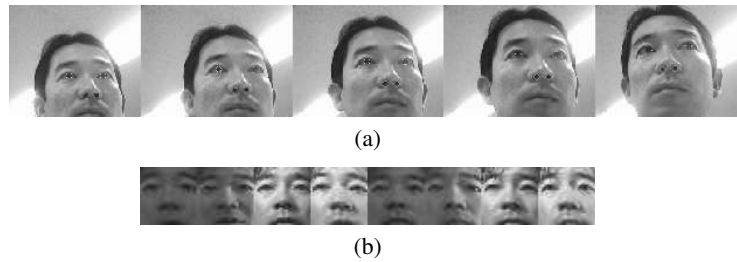**Fig. 6.** Flow of face recognition based on CMSM

(a)

(b)

Fig. 7. Data for evaluation: (a) the collected sequential face images, (b) the normalized face patterns of subject1 in lighting condition L1-L8 (from left)

4 Experiments

We have compared our method with conventional methods using a single face pattern: the subspace method (SM) and the eigenface method (EFM)[6][4], in addition, MSM in terms of the recognition rate and the separability[5] which is a normalized index of classification ability between subjects.

We have considered the situation in which a subject moves and a robot stands still, for example, a subject approaches a robot or turns his/her face toward the

---

[4] EFM is widely used as the base of various commercial products.

[5] The higher the separability is, the higher is the classification ability allowing larger for possible choice of rejection threshold.

Table 1. Recognition performance of each method, similarity $S[1]$ and $S[3]$ are defined in Section 2.2

Methods	SM	EFM	MSM$S[1]$	MSM/$S[3]$	CMSM/$S[1]$	CMSM/$S[3]$
Recognition rate	75%	75%	80%	89%	99%	99%
Separability	0.10	0.12	0.12	0.39	0.57	0.65

robot. A situation in which conversely the subject stands still and the robot moves is fundamentally the same, although how to move to a position where the robot can obtain various face patterns useful for face recognition is a difficult problem.

To simulate the first situation we collected sets of 320×240 pixels face images while making the subject change position and direction of their face as shown in Fig.7(a). The system for data collection consists of a PC (Pentium-III 800MHz), an image capture board, and a CCD camera with a lens with focal length=7.5mm. The camera is installed below the monitor so that it looks up at the face of the subject. We have carried out the collection to obtain 80 face images for each subject in each lighting condition and extracted the 15×15 pixels normalized face patterns from these images as shown in Fig.7. The number of subjects was 25. The lighting condition was changed by switching three lamps off/on: making eight combinations (L1-L8).

The normalized face patterns of subjecters 1-12 in lighting conditions L1-L4 were used for generating the constraint subspace $\mathcal{C}$. The face patterns extracted from the images of the other subjects, 13-25, in lighting conditions L5-L8 were used for evaluation. Using the normalized face patterns, we generated the 7-dimensional subspace $\mathcal{P}_{ij}$ for each subject i in lighting condition j by the KL expansion.

The constraint subspace $\mathcal{C}$ was then generated as follows. First, for subject i, four 10-dimensional subspaces $\mathcal{P}_{ij}$, (j=1~4) are integrated into the 40-dimensional $\mathcal{P}_{i*}$. Then, we calculated the sum matrix $\mathbf{G}$ from the projection matrices of the twelve 40-dimensional subspaces $\mathcal{P}_{i*}$ of subject1-12. Finally, the eigenvectors of the matrix $\mathbf{G}$ corresponding to the N_c smallest eigenvalues become the basis vectors of the N_c dimensional constraint subspace $\mathcal{C}$. N_c is set, experimentally, to 170.

In the evaluation of each methods, assuming that each subspace under lighting condition j is an input subspace and all the subspaces under lighting condition j^* are reference subspaces, we calculated the similarity between the subspaces while changing the combination of lighting conditions. The dimensions of an input subspace and the reference subspaces are set at 5 based on a preliminary experiment. The dimension of a reference subspace of SM is also set at 5. SM utilizes the mean of the 80 face patterns as an input vector so that the number of face patterns used for calculating the similarity is the same in all the methods. FEM also utilizes the mean of the 80 face patterns as an input vector and a reference vector.

Table 1 shows the evaluation result of each method in terms of the recognition rate and the separability. We can see that the method using CMSM is superior to the others with regard to both indices. This result also shows that the performance of face recognition has been further improved by using the multiple canonical angles. This

is results from information on 3D face shape, from the multiple canonical angles, being reflected in the recognition process.

5   Conclusion

This paper has presented a face recognition method on the basis of CMSM using multiple face patterns. The effectiveness of our method for robot vision is demonstrated by means of a preliminary experiment, which shows improvement based on the fact that our method is able to utilize various multiple face patterns. In future work, we intend to consider in particular of planning function for data collection and the actual implementation of our method in a robot.

References

1. S. Watanabe, N. Pakvasa, "Subspace method of pattern recognition," Proc. 1st Int. J. Conf. on Pattern Recognition, 1973.
2. E. Oja, "Subspace methods of pattern recognition," Research Studies Press, England, 1983.
3. K. Maeda, S. Watanabe, "A pattern matching method with local structure," Trans. IEICE, vol.J68-D, no.3, pp.345-352, 1985 (in Japanese).
4. O. Yamaguchi, K. Fukui, K. Maeda, "Face recognition using temporal image sequence," Proc. Third International Conference on Automatic Face and Gesture Recognition, pp.318-323, 1998.
5. K. Fukui, O.Yamaguchi, K.Suzuki, K.Maeda, "Face recognition under variable lighting condition with constrained mutual subspace method," Trans. IEICE (D-II), vol.J82-D-II, no.4,pp.613-620, 1999 (In Japanese).
6. M. Turk, A. Pentland, "Face recognition using eigenfaces," Proc. CVPR, pp.453-458, 1993.
7. P. N. Belhumeur, J. P. Hespanha, D. J. Kriegman, "Eigenface vs. Fisherfaces: Recognition using class specific linear projection," IEEE Trans. Pattern Analysis and Machine Intelligence, vol.19, no.7, pp.711-720, 1997.
8. P. S. Penev, J. J. Atick, "Local feature analysis: A general statistical theory for object representation," Network: Computation in Neural Systems, 7(3), pp.477-500, 1996.
9. F. Chatelin, "Eigenvalues of Matrices," Wiley, Chichester, 1993. (Enlarged Translation of the French Publication with Masson)
10. A. Gif, "Nonlinear multivariate analysis," John Wiley & Sons, 1989.
11. K. Fukui, O. Yamaguchi, "Facial feature point extraction method based on combination of shape extraction and pattern matching," Systems and Computers in Japan, vol.29, no.6, pp.49-58, 1998.
12. A. Shashua, "On photometric issues in 3D visual recognition from a single 2D image," IJCV, vol.21, pp.99-122, 1997.