Face recognition using temporal image sequence

Osamu Yamaguchi, Kazuhiro Fukui, and Ken-ichi Maeda
Kansai Research Laboratories, Toshiba Corporation
8-6-26 Motoyama-Minami-Machi, Higashinada-Ku,
Kobe 658, Japan
E-mail:{osamu, fki, maeda}@krl.toshiba.co.jp

Abstract

We present a face recognition method using image sequence. As input we utilize plural face images rather than a 'single-shot', so that the input reflects variation of facial expression and face direction. For the identification of the face, we essentially form a subspace with the image sequence and apply the Mutual Subspace Method in which the similarity is defined by the angle between the subspace of input and those of references. We demonstrate the effectiveness of the proposed method through several experimental results.

1. Introduction

Face recognition is a promising means for man-machine interface and security. For practical use of face recognition, it is necessary to overcome the variation of the face, such as facial expression, face direction, lighting condition and aging. In this paper, we present a method that is especially tolerant of the changes of facial expressions and face directions.

To restrain the influence of the variation, Pentland used modular eigenspace representation and the multiple 'Eigenface' for variable facial poses[4]. Belhumeur developed the 'Fisherface' method to separate classes in a low-dimensional subspace under variation of lighting and facial expressions[5]. Ariki proposed a modified CLAFIC method for face recognition[6], which is one of the modified subspace methods that can efficiently represent the variation of patterns. Those previous works are based on a 'single-shot' face recognition. If the real-time face recognition system fails in capturing or processing the image, the recognition result will not be reliable due to the inclusion of imperfect data.

We propose a face recognition method using temporal image sequence. To recognize a face, a sequence of input images is accumulated and a subspace is formed from the sequence. Then *the Mutual Subspace Method* (MSM) is applied in which the similarity is defined by the angle between the input and the reference subspaces. This method is stable against the influence of facial expressions and face directions because the subspace can efficiently represent the distribution of the changes.

2. Face recognition using image sequence

Our approach is based on the following observation.
1. The changes of facial expressions and face directions are time-varying information, so a face recognition method should employ a time-stable measure.
2. Conventional methods recognize a face using only a single-shot image input. If the input image is noisy or image processing accidentally fails, the recognition result will not be reliable.

Figure 1 shows the similarity in a 'single-shot' method (using the Conventional Subspace Method (CSM)[1][6]) for a face image sequence. This sequence was captured while a person was changing the direction of his face and speaking at the same time. The similarity value "CSM" randomly varies, as it is influenced by facial expression and face direction. If the system happens to take an instance when the line dips, the recognition may be a failure.

We have developed a new method using temporal image sequence. The use of plural images stabilizes the recognition against the noise and variation of the input. And the MSM plays an important role by using plural input images efficiently.

2.1. The Mutual Subspace Method

The subspace method[1] is a statistical pattern recognition method in which each class is represented by its own subspace that is spanned with a rather small set of basis.

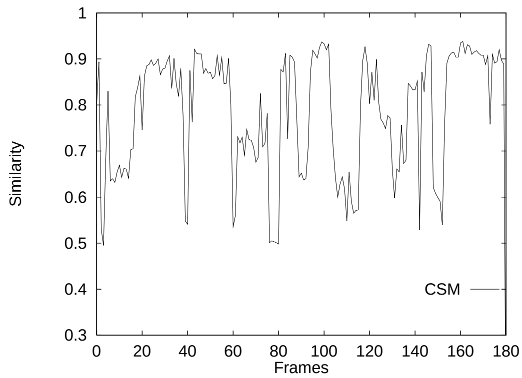


Figure 1. Similarity in a "single-shot" face recognition.

In the CSM, a similarity measure of input pattern g is defined as

$$S_{CSM}(g) = \frac{1}{||g||^2} \sum_{n=1}^{N} (g, \phi_n)^2, \tag{1}$$

where ϕ_n denotes a basis of the subspace, N is the dimensions of the subspace, $||.||$ is Euclidean norm, and $(\cdot, \cdot)$ is inner product.

Figure 2 shows a simple model of the CSM. The similarity $S_{CSM}(g)$ is equal to $cos^2\theta$, where θ is the angle between the vector g and the subspace D. The basis of the subspace is calculated by the principle component analysis (or the Karhunen-Lòeve expansion) of sample patterns.

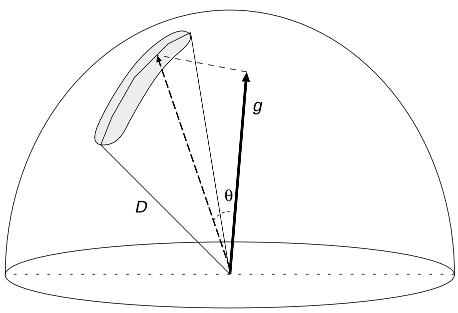


Figure 2. Subspace Method.

The MSM, on the other hand, defines the similarity by the angle between two subspaces. Each subspace is computed from input patterns or pre-registered patterns. Figure 3 shows a model of the angle between two subspaces, D and G. The angle θ is defined as

$$cos^2\theta = \sup_{d \in D, g \in G, ||d|| \neq 0, ||g|| \neq 0} \frac{|(d, g)|^2}{||d||^2 ||g||^2}. \tag{2}$$

While Equation (2) has local maxima, the largest eigenvalue λ_{max} of the matrix PQP gives $cos^2\theta$ as shown in Equation (3), where P and Q are respectively the orthogonal projection matrices onto the subspaces D and G[2][3].

$$cos^2\theta = \lambda_{max}. \tag{3}$$

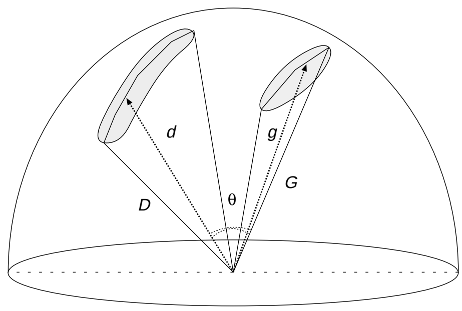


Figure 3. Mutual Subspace Method.

Since the matrix PQP is large, it is expensive to compute its eigenvalue. The dimensions of the subspaces D and G are smaller than those of the whole space. Hence the rank of PQP is generally much smaller than the dimension of the whole space. Therefore, the eigenvalue problem of PQP is reduced to that of a smaller matrix[2].

Let matrix X be

$$X = (x_{ij}) = \sum_{m=1}^{M} (\psi_i, \phi_m)(\phi_m, \psi_j), \tag{4}$$

where ϕ_i and ψ_j are the eigenvectors of subspaces D and G, and M and L are the number of dimensions of the subspace D and G (assuming $L \leq M$) respectively. The matrix X is then decomposed as,

$$W^T X W = \Lambda, \tag{5}$$

where W is the diagonalization matrix. The $diag(\Lambda)$ is the eigenvalues of X, and λ_{max} is the largest one. Having G and D as input and reference subspaces respectively, *the inter-subspace similarity* is defined as,

$$S_{MSM}(G, D) = \lambda_{max}. \tag{6}$$

Figure 4 shows the comparison between the CSM(dashed line) and the MSM(solid line) where each

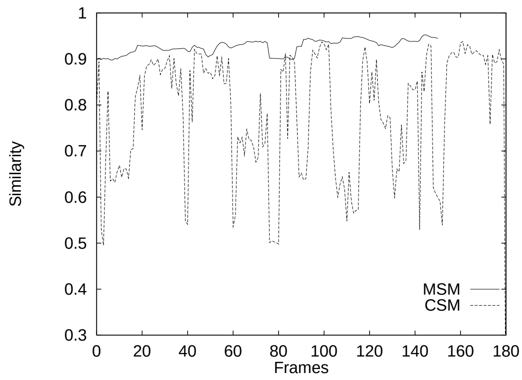


Figure 4. Comparison between the conventional subspace method and the proposed method.

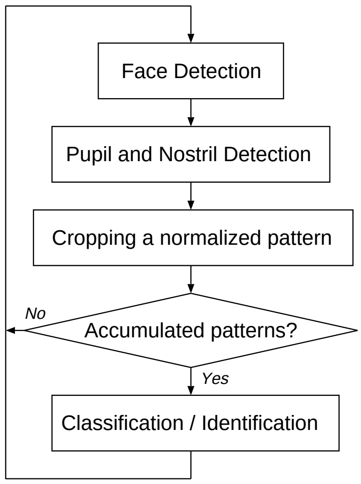


Figure 5. Flow of the processing.

method uses the reference subspace that is formed with the same reference sample patterns.

In the MSM, the similarity of the time t is the inter-subspace similarity between input subspace and reference subspace. The input subspace is formed with 30 face images from t to $t + 29$. As shown in Fig.4, the MSM is stable and keeps high similarity in comparison with the CSM.

3. Preprocessing

An experimental system is implemented on an SGI Indy with a 180MHz R5000 processor, whose average frame rate for preprocessing is 7Hz. Preprocessing flow is shown in Fig.5.

3.1. Face detection

Face detection is also based on a subspace method where two subspaces, the face subspace and the non-face subspace, are stored for distinguishing the face class from the non-face class. The similarities of the subspaces are calculated for each 15 $\times$ 15 sliding window. This calculation is done for plural scaled input images. The location of the face is determined as the position which gives the global maximum of similarity of the face class while the similarity of the face class is greater than that of the non-face class.

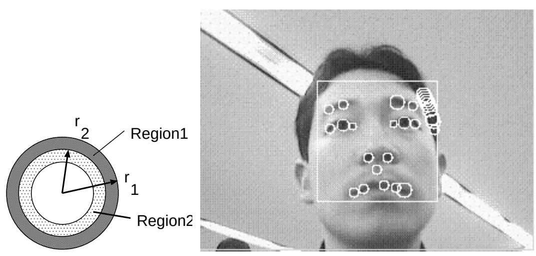
3.2. Pupil and nostril detection

The system uses the positions of pupils and nostrils to normalize the region for recognition. The detection method for these facial parts is based on a combination of shape extraction and pattern matching[7]. First, the candidates of features are extracted by a circular separability filter. Separability η $(0.0 \leq \eta \leq 1.0)$ is defined in[8], as

$$\eta = \frac{n_1(M_1 - M_m)^2 + n_2(M_2 - M_m)^2}{\sum_{i=1}^{n_1+n_2}(P_i - M_m)^2}, \tag{7}$$

where P_i is the image intensity at a pixel i, n_1 and n_2 are the number of pixels in regions 1 and 2, M_m is the mean of the image intensity for the combined local region, and M_1 and M_2 are the means of the image intensity in regions 1 and 2, respectively. Figure 6 shows the separability filter with circular mask(a) and local maximum points of separability(b).

The features are then verified using the subspace method and only four candidates are selected.



(a) **(b)**

Figure 6. (a)Separability filter with circular mask. (b)Candidates of pupils and nostrils by local maximum points of separability.

3.3. Cropping a normalized pattern

Normalization is based on the extracted four feature points (pupils and nostrils). A normalized pattern image is cropped as 30×30 image by subsampling the input image.

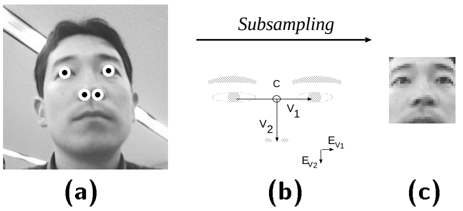


Figure 7. Cropping a normalized pattern. (a)Selected four feature points (b) the vectors for subsampling (c)a normalized pattern

Figure 7 illustrates the normalization process. The vectors V_1 and V_2 are determined from the four points, and $E_{V_1} = V_1/m_1$ and $E_{V_2} = V_2/m_2$ are calculated using m_1 and m_2 as constant weights. Point C is the center of the two pupils. The normalized pattern is generated by subsampling the pixels of the locations as $C + aE_{V_1} + bE_{V_2} (-15 < a, b \leq 15)$. Since the motion of the mouth is an unstable feature, we determine the parameters m_1 and m_2 to discard the mouth area as shown in Fig.7(c). Figure 8 shows examples of normalized patterns, which were gathered automatically from an image sequence.

Figure 8. Example of normalized patterns.

3.4. Accumulation of patterns to form the input subspace

The preprocessed data are accumulated to form a subspace. To avoid changes of the illumination brightness, the histogram equalization processing is applied to the normalized patterns. Also, in order to reduce the calculation cost, we obtain a feature vector of 225-dimension(15×15) by subsampling to a quarter size of the normalized pattern.

When the number of accumulated patterns is equal to r(constant), the input subspace is computed from the correlation matrix of the feature vectors. Let the correlation matrix C be

$$C = \frac{1}{r} \sum_{k=1}^{r} N_k N_k^T, \tag{8}$$

where $N_i (1 \leq i \leq r)$ represents a feature vector. Then the eigenvalue method is performed to find the eigenvectors $\phi_t (1 \leq t \leq L)$ of C.

3.5. Recognition by the MSM

Having generated the input subspace as shown above, we compute the inter-subspace similarity between this and the reference subspaces on a database. For the identification task, the input is identified if the similarity exceeds a threshold. For the classification task, the input is classified as the class that gives the highest similarity.

4. Experimental results

We have collected faces of 101 individuals. Figure 9 shows several of those captured sample images.

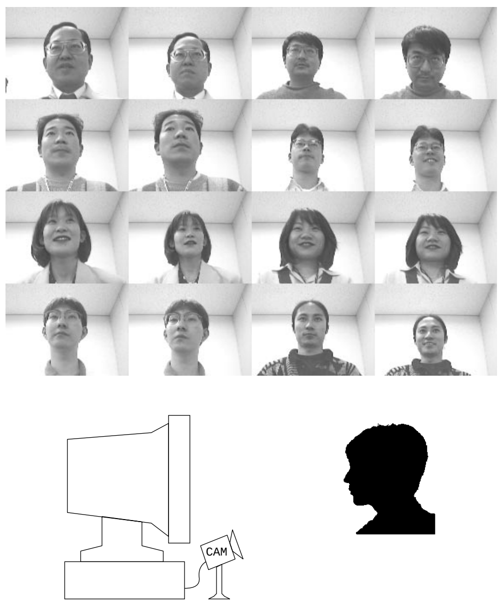


Figure 9. Samples of face database.(The arrangement of CCD camera is shown below.)

The camera is pointing upwards in front of the monitor of an SGI as shown in Fig.9, which makes it easy to take clear images of pupils and nostrils. We have separated the data into two subsets; one for registration and another for recognition test. Both sets consist of 180×101 face patterns. To acquire various face images, each person turned his or her head in several directions. The test set includes various facial expressions and movements derived for example by speaking. The databases were, however, acquired under a constant lighting condition in order to focus on the problems caused by facial expressions and face directions.

The experiments were carried out off-line in order to keep the condition for the reexperiment. The results are evaluated in terms of two criteria, i.e. False Rejection Rate(FRR) and False Acceptance Rate(FAR).

Identification experiments were performed to compare:

1. the CSM(a 'single-shot' recognition method),
2. the MSM(the proposed method),
3. the multiple CSM, and
4. the CSM using mean vector

in the identification task. The reference subspaces are the same in all experiments and are formed with 180 face patterns of a database for the registration.

The first experiment is for comparing the proposed method with the CSM. Figure 10 shows the result of identification experiments for the CSM. The vertical axis corresponds to the error rate and the horizontal axis corresponds to the identification threshold of the similarity. The dimensions of reference subspaces of the CSM are all twenty-five.

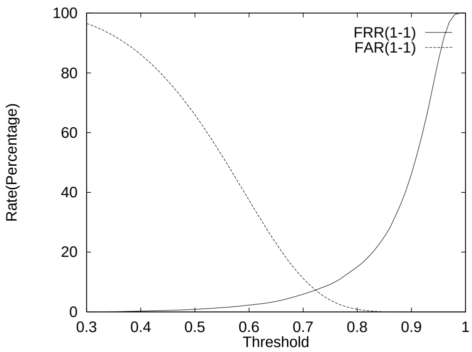


Figure 10. Performance of the CSM.

Figure 11 shows the performance of the MSM. The number of accumulated patterns is thirty, and the dimensions of subspace to calculate the inter-subspace similarity are five both for input and reference subspaces.

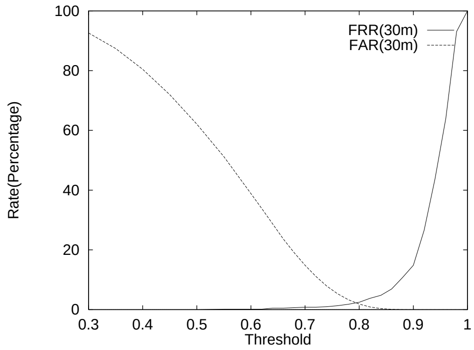


Figure 11. Performance of the MSM.

The crosspoint of FRR and FAR in Fig.11 is much smaller than that in Fig.10 reflecting better identification of the MSM. Another evaluation is to compare the minimum FARs at the points where FRR becomes 0%. The FAR in Fig.11 is 30% whereas that in Fig.10 is 80%, showing that the performance of the MSM greatly surpasses that of the CSM also respecting false acceptance.

We have also compared the proposed method with two other possible methods using plural patterns as input. The first method is the multiple CSM, where the CSM is continuously applied 30 times and the number of identification is counted. The input is identified only in the case that the number is greater than a threshold (ten in this case). The threshold is determined so that the crosspoint of FAR and FRR becomes equivalent to that in Fig.11. Figure 12 shows the result of performance of the multiple CSM. FAR of this method is not so low as that of the proposed method.

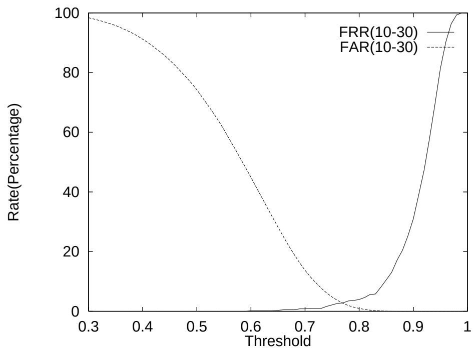


Figure 12. Performance of the multiple CSM.

The second method uses the mean vector as the input of the CSM. Figure 13 shows the result. The crosspoint of FRR and FAR is higher than that of the proposed method. Since the input of the MSM is represented with subspace using eigenvectors corresponding to plural largest eigenvalues, it reflects more information than that of the CSM whose input is the mean vector which is almost equal to the eigenvector corresponding to the largest eigenvalue. The use of more information is effective for achieving low error rate.

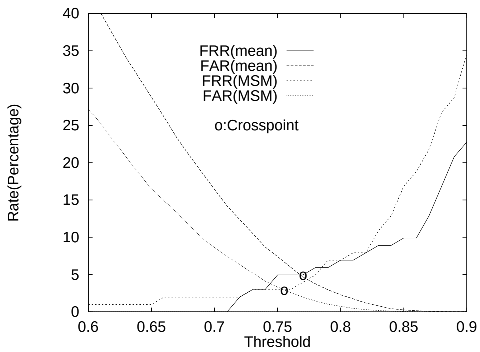


Figure 13. Performance of the CSM using mean vector.

5. Conclusion

In this paper, we have proposed a new method for face recognition using temporal image sequence. The proposed method uses an image sequence not only in registration process but also in recognition process, and applies the MSM for recognition. We have shown the effectiveness of the proposed method through the experimental results, and that the method is robust against varying face expression or face direction changes in comparison with other methods using plural images.

Subjects for future work include the solution of lighting condition and aging problems, and evaluation with a much larger number of persons.

References

[1] E. Oja, “Subspace Methods of Pattern Recognition”, Research Studies Press, England(1983)

[2] K. Maeda and S. Watanabe, “A Pattern Matching Method with Local Structure”, Trans. IEICE(D), vol.J68-D, No.3, pp. 345–352(1985)(In Japanese)

[3] Å. Björck and G.H. Golub, “Numerical Methods for computing Angles Between Linear Subspaces”, Mathematics of Computation, Vol.27, pp. 579–594(1973)

[4] A. Pentland, B. Moghaddam, and T. Starner, “View-based and modular eigenspaces for face recognition” , CVPR '94, pp.84-91 (1994)

[5] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman, “Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection”, ECCV '96, vol.1, pp.45-58 (1996)

[6] Y. Ariki and N. Ishikawa, “Integration of face and speaker recognition by subspace method”, 13th ICPR '96,vol. 3, pp. 456–460 (1996)

[7] K. Fukui and O. Yamaguchi, “Facial Feature Point Extraction Method Based on Combination of Shape Extraction and Pattern Matching”, Trans.IEICE (D-II), vol.J80-D-II, No.8, pp.2170-2177 (1997) (In Japanese)

[8] K. Fukui, “Edge Extraction Method Based on Separability of Image Features” , IEICE Trans. Inf.& Syst., Vol E78-D, No.12, pp.1533-1538(1995)

[9] K. Sung and T. Poggio, “Example-based Learning for View-based Human Face Detection”, Technical Report A.I.Memo No.1521,A.I. Lab., MIT(1994)