Face Recognition Based on Mutual Projection of Feature Distributions

Akira INOUE\(^{(a)}\) and Atsushi SATO\(^{1}\), Members

SUMMARY This paper proposes a new face recognition method based on mutual projection of feature distributions. The proposed method introduces a new robust measurement between two feature distributions. This measurement is computed by a harmonic mean of two distance values obtained by projection of each mean value into the opposite feature distribution. The proposed method does not require eigenvalue analysis of the two subspaces. This method was applied to face recognition task of temporal image sequence. Experimental results demonstrate that the computational cost was improved without degradation of identification performance in comparison with the conventional method.

key words: face recognition, image sequence, subspace, mutual projection

1. Introduction

Person identification by face recognition has an advantage of lower psychological stress for uses than other biometrics technologies because it does not use a contact sensor. Therefore, face recognition technologies\(^{[1]}\) have gained attention in several applications such as an entrance control system, human machine interfaces and personal robots. However, face recognition technologies in general have a problem of robustness under environment with illumination and pose variations. In recent years, recognition methods by using a temporal image sequence instead of a single image have been suggested to cope with the problem and improve identification performance\(^{[2]}-[5]\).

In the field of face recognition by using temporal image sequence, Mutual Subspace Method (MSM)\(^{[2]}-[3]\) has been proposed, and it was reported a better recognition performance in illuminant varying environment in comparison with a single image recognition\(^{[3]}\). In MSM, the minimum angle (square of cosine) between two subspaces uses as a similarity measurement between query and enrollment feature distributions. In\(^{[4]}\), it has proposed that kernel function was applied to MSM, and expected to improve in the case of nonlinear distributions. MSM and the expansion method have better characteristics for robust identification because they only use a few eigenvectors and decrease noise influences. However, they require eigenvalue analysis to compute the minimum angle between the two subspaces. Namely, MSM requires maximizing of the following matrix \(X\) for each test:

\[
X = U^TV
\]

where, \(U\) is formed by the eigenvectors for query feature subspace, and \(V\) is for enrollment feature subspace.

Recently, Inter-subspace distance (ISD) was proposed for face recognition\(^{[5]}\). This method uses the minimum distance between two subspaces. The method has reported a similar identification performance to MSM, and it also needs an eigenvalue analysis to find the distance. In aspect of practical application, a processing time is an important issue and some applications like robot systems need lower computational cost. However, these conventional face recognition methods\(^{[2]}-[5]\) for image sequences require eigenvalue analysis of the two subspaces, and this causes an increase of computational cost.

This paper presents a new face recognition method using mutual projection of feature distributions. This method is referred to as, Mutual Projection Method (MPM) in this paper. The proposed method introduces a new robust measurement between two feature distributions. This measurement is computed by a harmonic mean of two distance values gotten by using projection each mean vector into the opposite feature distribution.

In Sect. 2, we describe the algorithm of our proposed method. Experimental results to evaluate performances are demonstrated in Sect. 3.

2. Mutual Projection Method

A new face recognition method, Mutual Projection Method (MPM) is described in this section. Processing flow of MPM is represented by following steps.

(1) Query facial image sequences are entered.
(2) Distance between a query and an enrollment feature distribution is obtained for each person. The inter-distribution distance is referred to as Mutual Projection Distance (MPD). MPD is calculated using two distance values gotten by projecting each mean vector into the opposite feature distribution.
(3) The face recognition is done by choosing the person obtaining smallest MPD.

2.1 Definition of Mutual Projection Distance

Recognition of image sequence is considered to evaluate distance between a query and an enrollment feature distribution formed by the image sequence.
We consider an input feature distribution $C_1$ and an enrolment feature distribution $C_2$. It is assumed that the distances between a feature vector $x$ and the distributions $C_1$, $C_2$ are defined as $d_1(x)$, $d_2(x)$ respectively. If $m_1$ and $m_2$ represent the centers of each distribution, the distance between $m_1$ and $C_2$ is shown as $d_2(m_1)$. The distance between $m_2$ and $C_1$ is also shown as $d_1(m_2)$.

Here, we define a new measurement for inter distribution which is referred to as Mutual Projection Distance (MPD). Desired distance value $D$ between distributions is considered to be represented as the function $d_1(x) + d_2(x)$ for a certain vector $x$, as shown in Fig. 1. The vector $x$ is defined on the line between $m_1$ and $m_2$. $d_1(x)$ and $d_2(x)$ have the minimum value 0 when $x = m_1, m_2$, and these are monotonic increasing functions according to $\|x - m_1\|$ and $\|x - m_2\|$ respectively. In this condition, there exists a feature vector $a \{a \in x\}$ satisfying $d_1(x) = d_2(x)$. The vector $a$ is considered the equal distance point from both distributions’ means. Therefore, the distance between the two distributions can be defined the 2 times of the distance value at the point $a$. In other words, the distance value $D$ is obtained by sum of $d_1(a)$ and $d_2(a)$ (as shown in Eq. (2)). The formula of $D$ is the definition of the inter distribution distance, MPD. Figure 1 shows the distance $D$ for examples of $d_1(x)$ and $d_2(x)$.

$$D = d_1(a) + d_2(a)$$  

By combining both distance values $d_1(a)$ and $d_2(a)$, we can take both feature distributions into account for the inter distribution measurement.

2.2 Computation of Mutual Projection Distance

In order to obtain MPD, it is required the following two steps. At first it needs to compute the distance value between a vector and a distribution, $d_1(x)$ or $d_2(x)$. Then, we combine the two distances by considering the equal distance point from both distributions. These computations are presented in the following subsections.

2.2.1 Distance between a Vector and a Distribution

The distance a vector and a distribution, $d_1(x)$ or $d_2(x)$ can be calculated by using simple subspace projection. However, we have defined a formula for $d_1(x)$ and $d_2(x)$ based on Mahalanobis distance, in order to consider variances for axes of the subspace.

The Mahalanobis distance is widely used for a normalized distance in the field of pattern recognition. The distance $d_m$ is defined as Eq. (3):

$$d_m = (x - m)^T \Sigma^{-1} (x - m)$$  

where $x$ is a $n$-dimensional feature vector, $m$ is a mean vector of a distribution and $\Sigma$ shows a covariance matrix. However, Mahalanobis distance becomes unstable when the $\Sigma$ is singular. This situation often occurs when the number of training samples of recognition target is small. To avoid this instability, several distance measurements are proposed in this field [6], [7]. In this paper, the covariance matrix is estimated as following equation.

$$\Sigma = \hat{\Sigma} + \sigma^2 I$$  

In Eq. (4), $\hat{\Sigma}$ is a covariance matrix calculated using training samples, and the second term shows an initial estimation ($\sigma^2$ is a constant and $I$ is an identity matrix.). Eigenvalues and eigenvectors obtained are represented as $\lambda_i$ and $\Phi_i$. Then, Mahalanobis distance formula is transformed into Eq. (5).

$$d_{\Sigma}^2 = \sum_{i=1}^{n} \frac{1}{\lambda_i + \sigma^2} |\Phi_i^T (x - m)|^2$$  

Components of subspace spanned by eigenvectors with small eigenvalues are dominated by noise. Therefore, we assumed $\lambda_i \ll \sigma^2$ when $i > k$, and obtained the formula Eq. (6).

$$d_m^2(x) = ||x - m||^2 - \sum_{i=1}^{k} \frac{\lambda_i}{\lambda_i + \sigma^2} |\Phi_i^T (x - m)|^2$$  

In this paper, $d_m(x)$ which is the root of Eq. (6), are called Pseudo Mahalanobis Distance (PMD). In our method, PMD is used for the distance function $d_1(x)$ and $d_2(x)$. It is obvious that PMD coincide with the projective distance to the subspace in the case of $\sigma^2 = 0$. The transformations of the equation from Eq. (4) to Eq. (6) are based on the literature [6].

2.2.2 Combining Two Projective Distances

Figure 2 shows a distance value when PMD is applied. It is assumed that the statistical properties (mean, eigenvalues and eigenvectors) of one distribution are represented
as \{m_1, \lambda, \Phi\} and the other distribution is represented as \{m_2, \mu, \Psi\}. \(d_1^2(x)\) and \(d_2^2(x)\) using PMD can be described in Eq. (7).

\[
\begin{align*}
d_1^2(x) &= \|x - m_1\|^2 - \sum_{i=1}^{k} \frac{\lambda_i}{\lambda_i + \sigma^2}(\Phi^T(x - m_1))^2 \\
d_2^2(x) &= \|x - m_2\|^2 - \sum_{i=1}^{k} \frac{\mu_i}{\mu_i + \sigma^2}(\Psi^T(x - m_2))^2
\end{align*}
\]  

(7)

Because \(d_2^2(x)\) is proportional to \(\|x - m\|^2\) shown in Eq. (6), \(d_2(x)\) must be proportional to \(\|x - m\|\). The \(d_1(x)\), \(d_2(x)\) using PMD are defined on the line between \(m_1\) and \(m_2\) in Eq. (8):

\[
\begin{align*}
d_1(x) &= \frac{d_1(m_2)}{L}(m_2 - x) \\
d_2(x) &= \frac{d_2(m_1)}{L}(x - m_1)
\end{align*}
\]  

(8)

where \(L = |m_2 - m_1|\). Equal distance point \(a\) is obtained by solving \(d_1(x) = d_2(x)\).

\[
a = \frac{d_1(m_2)m_1 + d_2(m_1)m_2}{d_1(m_2) + d_2(m_1)}
\]  

(9)

Therefore, Mutual Projection Distance (MPD) is represented as \(D\) of Eq. (10) in this condition (derived from Eq. (3) and Eq. (9)).

\[
D = 2 \cdot \frac{d_1(m_2)m_1 + d_2(m_1)m_2}{d_1(m_2) + d_2(m_1)}
\]  

(10)

The \(D\) of Eq. (10) is the MPD formula based on PMD. The MPD is computed as a harmonic mean of two PMDs, which obtained by projecting each mean vector into the opposite subspace.

Mutual Projection Method (MPM) is a recognition method by using MPD for measurement of inter-distribution.

3. Face Recognition Experiments

Several applications such as gate control system and human machine interface are considered for the proposed method. These applications are used under various illuminations, and target face images include various poses or expressions. They require robustness in these conditions.

We have applied MPM to several face recognition tasks for image sequences. Two experiments were performed to investigate face recognition performance of MPM. The first experiment used image sequences taken under various illuminations, and the second experiment applied pose variations.

The face recognition experiments have used temporal image sequences for both query and enrollment samples. The following steps were used for the two experiments.

1. Beforehand, statistical properties of enrollment samples \{\(m_1, \lambda, \Phi\)\} are calculated for each person.
2. Sequential \(N\) frame images are obtained from a query image sequence. (Frame No.: \(f = 1\) to \(N\))
3. Statistical properties of query samples \{\(m_2, \mu, \Psi\)\} are calculated for \(N\) frame images.
4. MPD is computed for each person and the identification is performed by using the distance.
5. The next \(N\) frame images (Frame No.: \(f = 2\) to \(N + 1\)) are obtained. The identification process continues until the sequence ends.

3.1 Pseudo Divergence

Distance between two distributions has been studied for years in statistic research field, such as Bhattacharyya distance and Divergence\[8\]. In assumption that two distributions are normal, the divergence is represent as Eq. (11).

\[
div = \frac{(M_1 - M_2)^T(Sigma_1^{-1} + Sigma_2^{-1})(M_1 - M_2)}{2} + tr\left[Sigma_1^{-1}Sigma_2 + Sigma_2^{-1}Sigma_1 + 2 \cdot I\right]
\]  

(11)

Divergence value is thought to be useless for our applications because it needs the inverse of covariance matrices, which is empirically unstable to noise. However, we found the divergence formula of Eq. (11) becomes arithmetic mean of Maharanobis distances by eliminating the second term. Therefore, we applied PMD to the divergence formula, and defined Pseudo Divergence (PD) shown in Eq. (12). This measurement is used for comparison with MPM in following experiments.

\[
D_P = \frac{1}{2}(d_1(m_2) + d_2(m_1))
\]  

(12)

3.2 Experiment I (Under Various Illuminations)

In order to investigate performance under various illuminations, following experiment was performed. Facial images used in the experiments were captured by Digital Video Camera in home environment. They have taken under 12 different illumination conditions.

Preprocessing of face images are described as follows.

1. Images were converted to gray scale images beforehand. Eyes locations (pupil’s center) for face images were given by hand.
(2) The images were geometrically transformed by using both eyes locations as shown in Fig. 3. The $L$ shows a distance between eyes. A face region that is a rectangle region around face center was defined by $L$ and parameters $\alpha, \beta, \gamma$. The face region was transformed, rotated and cropped into $w \times h$ image (it is called, transformed face image).

(3) Pixel values of the transformed face images were normalized by using mean and variance of each image. Normalization is done by Eq. (13) and Eq. (14), where $I(x, y)$ is a transformed $w \times h$ image. $I'(x, y)$ represents a normalized face image. $\bar{I}$ is an average of the pixel values and $\sigma_I$ is a standard deviation for them. The $a$, $b$ in Eq. (13) are constant values for adjusting pixel value’s range. The normalized face images were used for feature vectors in our experiments.

$$I'(x, y) = \frac{I(x, y) - \bar{I}}{3\sigma_I} \cdot a + b \quad (13)$$

$$\bar{I} = \frac{1}{w \cdot h} \sum_{x=1}^{w} \sum_{y=1}^{h} I(x, y)$$

$$\sigma_I = \sqrt{\frac{1}{w \cdot h} \sum_{x=1}^{w} \sum_{y=1}^{h} (I(x, y) - m)^2} \quad (14)$$

In experiment I, images were captured by Digital Video Camera, and the original size was VGA (640 × 480). The transformed face image size was ($w = 12, h = 18$). Transform parameters we used were $\alpha = 1.6, \beta = 1.0, \gamma = 1.4$. Figure 4 shows normalized examples of face images. Table 1 represents the image dataset details.

In the experiment, a person identification performance was evaluated. The evaluation was done by ROC curve (plotting FRR and FAR). Figure 5 shows the experimental result. The number of the projective dimension $k$ was 20 for query and 20 for enrollment, the number of input frames $N$ was 30, and $\sigma^2$ was 0.001. The performance of the proposed method was compared with several other methods: Subspace Method (SM), Mutual Subspace Method (MSM) and Pseudo Divergence (PD). Subspace Method represents the identification method by using a distance of subspace projection. Note that average score of $N$ frames was used in identification with SM.

Experimental results in Fig. 5 show that the identification performance (ROC curve) of MPM was better than SM and PD. The results also represent the error rate of MPM was similar to MSM.

Figure 6 presents EER (the value when FAR = FRR) transition when the projective dimension $k$ varies. In the condition of $k = 20$, EER values of SM, MSM, MPM
When \( k < 20 \), MPM (\( \sigma^2 = 0.001 \)) performance was better than MSM. However, MSM results were better when \( k \) was large. In practice, it is thought that the difference of identification performance between MPM and MSM is quite small. The results indicate that \( k \) was important factors for identification performance of MPM.

Figure 7 shows EER results in several \( \sigma^2 \) values, 0.01, 0.001, 0.0001 and 0.0. EER in \( \sigma^2 = 0.01 \) was worse than the other values. When \( \sigma^2 \) is 0.001, the best performance was shown in \( k = 30 \) (EER = 0.0073).

### 3.3 Experiment II (Under Small Pose Variations)

We investigated performance of MPM under pose variations. Face image dataset we used was consisted of images with pose variation within around 20 degrees and small expression changes. Query image sequences were captured on a few days after we got enrollment image sequences. Illumination conditions for query and enrollment images were almost same, but they were not strictly controlled. Subjects have slightly changed their expressions when the query images captured. Table 2 shows the image dataset details.

Images were captured by a small CCD camera, and capturing process is shown in Fig. 8. Subjects faced the CCD camera and moved their heads among recording. The original image size was QVGA (320 × 240). Preprocessing method was same as experiment I (applying geometrically transformation and normalization). Figure 9 represents normalized face images.

Figure 10 shows the experimental result. The number of the projective dimension \( k \) was 20 for query and 20 for enrollment, the number of input frames \( N \) was 30, and \( \sigma^2 \) was 0.001. In Fig. 10, MPM showed the best performance among the compared method. SM represents the worst result as we expected. MSM performance was worse than PD in experiment II (different from experiment I).

Figure 11 presents EER (the value when \( \text{FAR} = \text{FRR} \)) transition when the projection dimension \( k \) varies. In the condition of \( k = 20 \), EER values of SM, MSM, MPM (\( \sigma^2 = 0.001 \)) and PD are 0.18, 0.054, 0.041 and 0.044 respectively.
The results also presented EER values did not depend on the k value in experiment II.

Figure 12 shows EER results in several $\sigma^2$ values, 0.001, 0.0001 and 0.0. In this experiment, when $\sigma^2$ is 0.0, the best performance was shown in k=30 (EER=0.036). From both experiments, it is thought that the $\sigma^2$ value should be smaller than 0.001, and the best value is depend on applications.

3.4 Discussion for Identification Performance

Proposed inter distribution measurement, Mutual Projection Distance (MPD) used in MPM, was derived by using equal distance point from both distributions’ means. When PMD is applied to vector-distribution measurement, MPD formula becomes the harmonic mean of PMD.

The experimental results using real face images indicated that identification performance of MPM was similar to MSM. The results also presented that MPM (the harmonic mean of PMD) was slightly better than PD (the arithmetic mean of PMD). However, identification performance for the harmonic mean of PMD is not theoretically guaranteed to be better than the arithmetic mean. Theoretical analysis for MPM behavior for various distributions is a topic of further study.

3.5 Comparison of Computational Performance

Table 3 represents an average processing time for each method. The processing time was calculated by taking an average for each person and each query in experiment I. From a series of MPM results for identification performance, $k = 20$ was thought to be reasonable under our experimental condition. The results have shown that the processing time of MPM was about 50% of MSM.

From the result, MPM has an advantage of computational cost in comparison with MSM. The reason is thought that while MSM needs to compute the maximum eigenvalue of the matrix $X = U^TV$ for each pattern matching, MPM does not require such an eigenvalue analysis. (PD has also shown a good computational performance by the same reason.)

Computational cost detail for MSM and MPM was analyzed on another CPU in condition of experiment I (8 persons enrolled). Figure 13 shows processing time measurement results in $k = 10$, 20 and 30. Processing of both methods (MSM and MPM) can be divided into three parts, 1) Preprocessing, 2) Computing eigenvector for query images and 3) Pattern matching to enrolled data.

MSM have to compute the maximum eigenvalue of matrix $X$ in the pattern matching stage. The result of MSM in $k = 20$ was 0.10, 3.9 and 8.4 milliseconds respectively. On the other hand, MPM result was 0.10, 4.2 and 0.32 milliseconds respectively. The results represent matching cost of MPM was obviously smaller than MSM. Furthermore, the improvement in $k = 30$ was much better. The projective dimension $k$ has a large impact for eigenvalue computation. The pattern matching time is also proportional to the number of enrolled persons.

4. Conclusion

This paper has proposed a new face recognition method using mutual projection of feature distributions. The proposed method, referred to as Mutual Projection Method
(MPM), introduced a new measurement between two feature distributions. This measurement was computed by a harmonic mean of two distance values obtained by projection of each mean value into the opposite feature distribution. The MPM does not require eigenvalue analysis of the two subspaces. The MPM was applied to several face recognition tasks of temporal image sequence. The experimental results have demonstrated that the computational cost was improved compared with the Mutual Subspace Method. The identification performance was much better than Subspace Method (a single image based method), and represents similar results to the Mutual Subspace Method for various conditions of illumination and facial pose. Therefore, MPM is promising for various applications using image sequence recognition.

Acknowledgments

The authors would like to thank Dr. Akihiko Sugiyama and Dr. Masahiro Serizawa of Common Platform Software Research Laboratories, NEC Corporation for their useful comments on this paper.

References


Akira Inoue received the BS degree in mechanical engineering from Osaka University, Japan in 1988. He joined NEC Corporation in 1988, and he has been engaged in research on image processing and pattern recognition. He was a guest researcher at the Department of Engineering of Cambridge University, UK, from 1999 to 2000. He is now a principle researcher in Common Platform Software Research Laboratories, NEC Corporation.

Atsushi Sato received the M.S. and Ph.D. degrees in physics from Tohoku University in 1986 and 1989, respectively. He joined NEC Corporation in 1989, and he is now a senior principal researcher at the Common Platform Software Research Labs. He has been engaged in the research on pattern recognition and artificial neural networks. During 1994–1995, he was a visiting researcher at the Intelligent Systems Laboratory, University of Washington. His current research interests include studies on face recognition and machine learning. Dr. Sato is a member of the Physical Society of Japan and the IEEE Computer Society. He received the Convention Award from Information Processing Society of Japan (IPSJ) and from IEICE in 1991 and 1994, respectively.