2004 8th International Conference on
Control, Automation, Robotics and Vision
Kunming, China, 6-9th December 2004

Feature fusion method based on Canonical Correlation Analysis and Handwritten character recognition

Quan-Sen Sun[a, b], Sheng-Gen Zeng[a], Pheng-Ann Heng[c], De-Sen Xia[a]

[a]Department of Computer Science, Nanjing University of Science &Technology, Nanjing 210094, People' Republic of China
[b]Department of Mathematics, Jinan University, Jinan 250022, People' Republic of China
[c]Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong
Email: qssun@beelink.com

Abstract

A new feature extraction method, based on feature fusion, according to the idea of canonical correlation analysis (CCA) , is proposed in this paper. A framework of CCA used in pattern recognition is described. The overall process comprises: extracting two groups of feature vectors with the same pattern; establishing the correlation criterion function between the two groups of feature vectors, and extract their canonical correlation features in order to form effective discriminant vectors for recognition. The inherent essence of this method used in recognition is theoretically analyzed. This method uses correlation features between two groups of feature vectors as effective discriminant information, so it not only is suitable for information fusion, but also eliminates redundant information within features, a new way for classification is proposed. Experimental results of our method applying on Concordia University CENPARMI handwritten numeral database has shown that our recognition rate is higher than that of the algorithm adopting single feature or the existing fusion algorithm.

1 Introduction

In recent years, feature level fusion plays an important role in the process date fusion, which has achieved delightful development[1]. The advantage of feature level fusion is obvious. As a matter of fact, the different feature vector extracted from the same pattern always reflects the different feature of patterns. By optimizing and combining these different features, it not only keeps the effective discriminant information of multi-feature, but also eliminates redundant information to certain degree. This is especially important to the problem of classification and recognition.

There exist two feature fusion methods. One is to group two sets of feature vectors into one union-vector [2], and then to extract features in a higher-dimension real vector space. Another one is to combine two sets of feature vectors by a complex vector[1], and then to extract features in the complex vector space. Both feature fusion methods aim at increasing the recognition rate, the feature fusion method based on the union-vector is referred as serial feature fusion while the method based on the complex vector is called parallel feature fusion [1].

Canonical Correlation Analysis (CCA) is one of the statistical methods dealing with the mutual relationships between two random vectors, its importance is comparable to Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) in Multivariate Statistical Analysis. CCA is one of the valuable multi-data processing methods[3]. In recent years, CCA has been used in various fields such as signal processing, computer vision, and sound recognition [4-7], and has induced great attention.

In this paper, a new method of feature fusion is proposed by adopting the idea of CCA. A framework of CCA used in pattern recognition is discussed. We extract two groups of feature vectors with the same sample, then establish a correlation criterion function between two groups of feature vectors, extract their canonical correlation features according to this criterion, and finally form effective discriminant vectors for recognition. This method uses correlation features between two groups of feature vectors as effective discriminant information, so it not only suits for information fusion, but also eliminates the redundant information within the features. This method can be applied to classification and recognition.

Experimental results based on the Concordia University CENPARMI database of handwritten Arabic numerals show that recognition rate of our method is higher than that of the method adopting a single feature or the existing fusion algorithm. Our simulation results have shown that this algorithm not only realize the compression of primitive feature dimensions, but also performs well in classification by reflecting the essential feature of images.

The rest of this paper is organized as follows. In Section 2, the theory and method of CCA used in feature fusion are presented. In Section 3, we discuss the inherent essence of this method in theory and explain why it is effective in recognition. The proposed feature fusion method has been tested on Concordia University CENPARMI handwritten numeral database and compared to other methods in Section 4. Finally, conclusions are drawn in Section 5.

2 The theory and method of feature fusion

2.1 The basic idea of CCA

In Multivariate Statistical Analysis, the correlation problem of two random vectors are often studied, that is to convert the correlation research of two random vectors into that of a few pairs of variables, which are uncorrelated. H.Hotelling developed this idea in 1936 [8].

Considering two zero-mean random vectors X and Y, CCA finds a pair of directions α and β that maximize the correlation between the projections $a_1 = \alpha^{\mathrm{T}} X$ and $b_1 = \beta^{\mathrm{T}} Y$ .This correlation is called the canonical correlation, projections a_1 and b_1 are called the first pair of canonical variates. Then we find the second pair of canonical variates a_2 and b_2, which are uncorrelated with canonical variates a_1 and b_1 respectively and maximize the correlation between them. This process is repeated until all correlation features of X and Y are extracted. In order to study the correlation of X and Y, we only need to analyze the correlation of a few pairs of canonical variates.

2.2 The theory and algorithm of combine feature extraction

Suppose $\omega_1, \omega_2, \cdots, \omega_c$ are c known pattern classes. Let $\Omega = \{\xi \mid \xi \in \mathbf{R}^n\}$ is a training sample space. Given $A = \{x \mid x \in \mathbf{R}^p\}, B = \{y \mid y \in \mathbf{R}^q\}$, where x and y are two feature vectors of the same sample ξ extracted by different means respectively. We will discuss the feature fusion in the transformed training sample feature space A and B.

Our idea is to extract the canonical correlation features between x and y based on the idea of CCA proposed in Section 2.1, denoted them as $\alpha_1^{\mathrm{T}} x$ and $\beta_1^{\mathrm{T}} y$ (the first pair), $\alpha_2^{\mathrm{T}} x$ and $\beta_2^{\mathrm{T}} y$ (the second pair), $\cdots$, $\alpha_d^{\mathrm{T}} x$ and $\beta_d^{\mathrm{T}} y$ (the d th pair). Given the following:

$$\begin{aligned} X^* &= (\alpha_1^{\mathrm{T}} x, \alpha_2^{\mathrm{T}} x, \cdots, \alpha_d^{\mathrm{T}} x)^{\mathrm{T}} = (\alpha_1, \alpha_2, \cdots, \alpha_d)^{\mathrm{T}} x \\ &= W_x^{\mathrm{T}} x; \end{aligned} \tag{1}$$

$$\begin{aligned} Y^* &= (\beta_1^{\mathrm{T}} y, \beta_2^{\mathrm{T}} y, \cdots, \beta_d^{\mathrm{T}} y)^{\mathrm{T}} = (\beta_1, \beta_2, \cdots, \beta_d)^{\mathrm{T}} y \\ &= W_y^{\mathrm{T}} y. \end{aligned} \tag{2}$$

Following linear transformation

$$Z = \begin{pmatrix} X^* \\ Y^* \end{pmatrix} = \begin{pmatrix} W_x^{\mathrm{T}} x \\ W_y^{\mathrm{T}} y \end{pmatrix} = \begin{pmatrix} W_x^{\mathrm{T}} & 0 \\ 0 & W_y^{\mathrm{T}} \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \tag{3}$$

as the combined feature projected, used for classification and recognition, while the transformation matrix is

$$W = \begin{pmatrix} W_x & 0 \\ 0 & W_y \end{pmatrix}.$$

where $W_x = (\alpha_1, \alpha_2, \cdots, \alpha_d)$, $W_y = (\beta_1, \beta_2, \cdots, \beta_d)$.

Definition 1. We call α_i and β_i as the i th pair of canonical projective vectors (CPV) of x and y, and $\alpha_i^{\mathrm{T}} x$ and $\beta_i^{\mathrm{T}} y$ as the i th canonical correlation features of x and y. We also call W as canonical projective transformation matrix (CPTM), Z as canonical correlation discriminant features (CCDF).

Next, we will discuss how to get the solution and the quality of CPV and CCDF. Supposed that S_{xx} and S_{yy} are the covariance matrixes of A and B respectively, while S_{xy} denotes their between-set covariance matrix, then the covariance matrix of $\begin{bmatrix} x \\ y \end{bmatrix}$ can be denoted as follows:

$$S = \begin{pmatrix} Var(x) & Cov(x,y) \\ Cov(y,x) & Var(y) \end{pmatrix} = \begin{pmatrix} S_{xx} & S_{xy} \\ S_{yx} & S_{yy} \end{pmatrix}.$$

where S_{xx} and S_{yy} are positive definite, and $S_{xy}^{\mathrm{T}} = S_{yx}$, $r = rank(S_{xy})$. Considering the linear combination of x and y: $\alpha^{\mathrm{T}} x$ and $\beta^{\mathrm{T}} y$, where $\alpha \in \mathrm{R}^p$ and $\beta \in \mathrm{R}^q$ are two arbitrary nonzero vectors. CCA finds pairs of directions α and β that maximize correlation between the canonical correlation features $\alpha^{\mathrm{T}} x$ and $\beta^{\mathrm{T}} y$. Since

$$\begin{aligned} Var(\alpha^{\mathrm{T}} x) &= \alpha^{\mathrm{T}} Var(x) \alpha = \alpha^{\mathrm{T}} S_{xx} \alpha \\ Var(\beta^{\mathrm{T}} y) &= \beta^{\mathrm{T}} Var(y) \beta = \beta^{\mathrm{T}} S_{yy} \beta \\ Cov(\alpha^{\mathrm{T}} x, \beta^{\mathrm{T}} y) &= \alpha^{\mathrm{T}} Cov(x,y) \beta = \alpha^{\mathrm{T}} S_{xy} \beta \end{aligned}$$

We can give the criterion function as the following:

$$J(\alpha,\beta)=\frac{\alpha^{\mathrm{T}}S_{xy}\beta}{(\alpha^{\mathrm{T}}S_{xx}\alpha\cdot\beta^{\mathrm{T}}S_{yy}\beta)^{1/2}} \tag{4}$$

Obviously, the criterion (4) has the character as follows:

- $J(k\alpha,l\beta)=J(\alpha,\beta).\ \forall k,l\in\mathbf{R}$;
- The extremum of $J(\alpha,\beta)$ is nothing to do with length of α and β, but has something to do with their direction.

According the above character, we can think that

$$\alpha^{\mathrm{T}}S_{xx}\alpha=\beta^{\mathrm{T}}S_{yy}\beta=1 \tag{5}$$

Now the question is changed to the solution of CPV α and β in Eq.(5), with the extremum of criterion (4). Supposed the first pair of CPV(α_1,β_1) has been done, after the first $(k-1)$th $(\alpha_2,\beta_2),\cdots,(\alpha_{k-1},\beta_{k-1})$ have been chosen, the kth can be determined by solving the following optimization problem:

$$Model\,I\begin{cases}\max J(\alpha,\beta)\\ \alpha^{\mathrm{T}}S_{xx}\alpha=\beta^{\mathrm{T}}S_{yy}\beta=1\\ \alpha_i^{\mathrm{T}}S_{xx}\alpha=\beta_i^{\mathrm{T}}S_{yy}\beta=0\ (i=1,2,\cdots,k-1)\\ \alpha\in\mathrm{R}^p,\beta\in\mathrm{R}^q\end{cases} \tag{6}$$

We'll discuss the solution of *model I*. According to the method of Lagrange multipliers, choose function:

$$L(\alpha,\beta)=\alpha^{\mathrm{T}}S_{xy}\beta-\frac{\lambda_1}{2}(\alpha^{\mathrm{T}}S_{xx}\alpha-1)-\frac{\lambda_2}{2}(\beta^{\mathrm{T}}S_{yy}\beta-1)$$

where λ_1 and λ_2 are Lagrange multipliers. Setting the partial derivatives of $L(\alpha,\beta)$ with respect to α and β equal to zero:

$$\begin{cases}\dfrac{\partial L}{\partial\alpha}=S_{xy}\beta-\lambda_1S_{xx}\alpha=0 & (7)\\[2ex] \dfrac{\partial L}{\partial\beta}=S_{yx}\alpha-\lambda_2S_{yy}\beta=0 & (8)\end{cases}$$

Multiplying both side of Eq.(7) and (8) by α^{T} and β^{T} respectively, and under the restricted condition (5), we obtain:

$$\alpha^{\mathrm{T}}S_{xy}\beta=\lambda_1\alpha^{\mathrm{T}}S_{xx}\alpha=\lambda_1$$
$$\beta^{\mathrm{T}}S_{yx}\alpha=\lambda_2\beta^{\mathrm{T}}S_{yy}\beta=\lambda_2$$

Since $S_{yx}^{\mathrm{T}}=S_{xy}$, so

$$\lambda_1=\lambda_1^{\mathrm{T}}=(\alpha^{\mathrm{T}}S_{xy}\beta)^{\mathrm{T}}=\beta^{\mathrm{T}}S_{yx}\alpha=\lambda_2.$$

Let $\lambda_1=\lambda_2=\lambda$. then

$$J(\alpha,\beta)=\alpha^{\mathrm{T}}S_{xy}\beta=\beta^{\mathrm{T}}S_{yx}\alpha=\lambda \tag{9}$$

This shows that the Lagrange multipliers λ_1 and λ_2 are equal to the correlation coefficient of $\alpha^{\mathrm{T}}x$ and $\beta^{\mathrm{T}}y$. So Eq. (7) and (8) can also be written as:

$$\begin{cases}S_{xy}\beta-\lambda S_{xx}\alpha=0 & (10)\\ S_{yx}\alpha-\lambda S_{yy}\beta=0 & (11)\end{cases}$$

Since S_{xx} and S_{yy} are both positive definite, from Eq. (10) and (11), we obtain:

$$\begin{cases}S_{xy}S_{yy}^{-1}S_{yx}\alpha=\lambda^2S_{xx}\alpha & (12)\\ S_{yx}S_{xx}^{-1}S_{xy}\beta=\lambda^2S_{yy}\beta & (13)\end{cases}$$

Now the question has been converted to a problem of solving two generalized eigenproblem. Given $M_{xy}=S_{xx}^{-1}S_{xy}S_{yy}^{-1}S_{yx}$ and $M_{yx}=S_{yy}^{-1}S_{yx}S_{xx}^{-1}S_{xy}$, Eq.(12) and (13) can be changed to:

$$\begin{cases}M_{xy}\alpha=\lambda^2\alpha & (14)\\ M_{yx}\beta=\lambda^2\beta & (15)\end{cases}$$

The theorem about the eigenvalue and eigenvector of M_{xy} and M_{yx}, is as follows [3].

Theorem 1. M_{xy} and M_{yx} have the same nonzero eigenvalues, which satisfies $1>\lambda_1^2\geq\lambda_2^2\geq\cdots\geq\lambda_r^2>0$, where $r=rank(S_{xy})$.

In order to obtain the solution under the restricted condition (6), suppose that:

$$G_1=S_{xx}^{-1/2}S_{xy}S_{yy}^{-1}S_{yx}S_{xx}^{-1/2},G_2=S_{yy}^{-1/2}S_{yx}S_{xx}^{-1}S_{xy}S_{yy}^{-1/2} \tag{16}$$

Then we can get that M_{xy} and G_1 have the same nonzero eigenvalues according to the related theorem of matrix, so do M_{yx} and G_2. Such that the nonzero eigenvalues are all equal to $\lambda_1^2,\cdots,\lambda_r^2$. Again given $H=S_{xx}^{-1/2}S_{xy}S_{yy}^{-1/2}$, then $G_1=HH^{\mathrm{T}}$, $G_2=H^{\mathrm{T}}H$. Applying Singular Value Decompose (SVD) theorem on the matrix H, we obtain: $H=\sum_{i=1}^{r}\lambda_iu_iv_i^{\mathrm{T}}$, where $\lambda_1^2,\cdots,\lambda_r^2$ are the entirely nonzero eigenvalues of G_1 and G_2, u_i and $v_i(i=1,2,\cdots,r)$ are the orthogonal eigenvectors of G_1 and G_2 corresponding to the nonzero eigenvalue λ_i^2, respectively.

From above, we can infer another important theorem as follows[3]:

Theorem 2. Given $\alpha_i = S_{xx}^{-1/2} u_i$, $\beta_i = S_{yy}^{-1/2} v_i (i = 1, \cdots, r)$, then

(1) α_i and β_i are eigenvectors of M_{xy} and M_{yx} corresponded to λ_i^2;

(2) $$\begin{cases} \alpha_i^{\mathrm{T}} S_{xx} \alpha_j = \beta_i^{\mathrm{T}} S_{yy} \beta_j = \delta_{ij} \\ \alpha_i^{\mathrm{T}} S_{xy} \beta_j = \lambda_i \delta_{ij} \qquad (i, j = 1, 2, \cdots, r). \end{cases} \tag{17}$$

As the matter of fact, the theorem 2 gives the solution under the restricted condition (5) and (6), and the extremum of criterion (4), that is also intended to the optimized solution of *model 1*. According to Eq. (9) and theorem 2, we can get the corollary as follows:

Corollary. Given all eigenvalues of G_1 or M_{xy} as $\lambda_1^2 \geq \cdots \geq \lambda_r^2 > \lambda_{r+1}^2 = \cdots = \lambda_p^2 = 0$. The criterion function (4) has $J(\alpha_i, \beta_i) = \lambda_i \ (i = 1, \cdots, p)$.

From the above discussion, we can draw the following conclusion.

Theorem 3. Under criterion (4), the number of the efficient CPV, satisfying the restricted condition (5) and (6), is r at most, where $r = rank(S_{xy})$, and getting $d (\leq r)$ CPV are compose of the eigenvectors corresponding to the first d maximum eigenvalues of two eigenequation (14) and (15) that satisfy Eq.(17).

Prove. From theorem 1 and Corollary, we know

$$\lambda_i > 0, i = 1, \cdots, r \,;\, \lambda_{r+1}^2 = \cdots = \lambda_p^2 = 0 .$$

$$J(\alpha_i, \beta_i) = \lambda_i, \ i = 1, \cdots, p .$$

Hence $J(\alpha_i, \beta_i) = 0$ ($j = r+1, \cdots, p$). Here the efficient CPV can't be extracted, which means that the number of the efficient CPV is r at most. From theorem 2, we know that, $d (\leq r)$ CPV can be composed of the eigenvectors corresponding to first d maximum eigenvalues of two eigenequation $M_{xy} \alpha = \lambda^2 \alpha$ and $M_{yx} \beta = \lambda^2 \beta$ respectively, and satisfying condition (17). □

2.3 The steps of overall algorithm

Step1. Extracting sets of two different feature vectors with the same pattern sample in order to form training sample spaces A and B, which are transformed from the original pattern sample space Ω.

Step2. Computing the total scatter matrix S_{xx} and S_{yy} of the samples in A and B, and computing their between-set covariance matrix S_{xy}.

Step3. Computing G_1 and G_2 according to Eq.(16), then finding their nonzero eigenvalues $\lambda_1^2 \geq \lambda_2^2 \geq \cdots \geq \lambda_r^2$, and the corresponding orthonormal eigenvectors u_i and $v_i (i = 1, 2, \cdots, r)$.

Step4. Getting the CPV α_i and β_i ($i = 1, 2, \cdots, r$), according to theorem 2, choose the canonical projective transformed sub-matrices W_x and W_y composed by first d vectors.

Step5. Projecting feature vectors x and y in d CPVs to get d-dimensional discriminant feature X^* and Y^* respectively, according to:

$$X^* = W_x^{\mathrm{T}} x = (\alpha_1, \alpha_2, \cdots, \alpha_d)^{\mathrm{T}} x$$

$$Y^* = W_y^{\mathrm{T}} y = (\beta_1, \beta_2, \cdots, \beta_d)^{\mathrm{T}} y$$

where $x \in A$, $y \in B$.

Step6. Using Eq.(3) $Z = \begin{pmatrix} X^* \\ Y^* \end{pmatrix}$ for classification.

From Eq.(10) and (11), we know that we only need to solve one α_i among each pair of canonical projective vectors, the other one can be solved by $\beta_i = \frac{1}{\lambda_i} S_{yy}^{-1} S_{yx} \alpha_i$. In general, we can choose the low-dimensional matrix G_1 ($p = \min(p, q)$) for determining its eigenvalues and eigenvectors, so as to lower the computation complexity.

3 The analysis of the algorithm validity

In the theory of pattern recognition, a common principle of feature extraction is the smaller statistical correlation of selected features the better. In other words, it would be better if the extracted features are uncorrelated. A method of the uncorrelated optimal discriminant vectors is proposed by Jin and Yang[9, 10], which has been applied to the realms of face recognition and character recognition, obtains good results. The essence of the theory is that the components of discriminant vectors are uncorrelated, and the projective vectors are orthonormal to each other about the total scatter matrix S_t.

From Theorem 2, CPV α_i and α_j , β_i and β_j are orthonormal about S_{xx} and S_{yy} respectively, i.e. the components of CCDF:

$$E\,[(\alpha_i^{\mathrm{T}} x - E\alpha_i^{\mathrm{T}} x)(\alpha_j^{\mathrm{T}} x - E\alpha_j^{\mathrm{T}} x)] = \alpha_i^{\mathrm{T}} S_{xx} \alpha_j = \delta_{ij}$$

$$E\,[(\beta_i^{\mathrm{T}} y - E\beta_i^{\mathrm{T}} y)(\beta_j^{\mathrm{T}} y - E\beta_j^{\mathrm{T}} y)] = \beta_i^{\mathrm{T}} S_{yy} \beta_j = \delta_{ij}$$

are uncorrelated. In this sense, this projective transform is optimal.

Commonly, the time taken greatly depends on the computational process of projective vectors, which uses algebraic method to extract the discriminant feature. When the rank of the matrix is very large, the computation of eigenvalues and eigenvectors would be very time-consuming. In the same pattern, supposing that the dimension of feature vectors x and y are p and q respectively, projective vectors described in [2] has been computed in real vector space of $p+q$ dimension, while that described in [1] has been done in the complex vector space of $\max(p,q)$ dimension. In one of our method these vectors can be done in real vector space of $\min(p,q)$ dimension. When the dimension of p and q are large, the advantage of our method in computational speed is significant.

The linear discriminant analysis based on Fisher criterion (FLDA) is one of the most effective ways in feature extraction and recognition. FLDA can be treated as a special situation of CCA and may be solved in theory, some algorithms based on Fisher criterion can be translated into the method presented in this paper (we should discuss in other paper). Applying CCA in pattern recognition is more general and has more potential development.

3 Experiment and analysis

We have tested our method using the Concordia University CENPARMI handwritten numeral database, which is popular in the world. In this database, there are 10 class, i.e. 10 digits (from 0 to 9), and 600 sample for each. The training samples and testing samples are 4000 and 2000, respectively. Some images of original samples are shown in Fig.1.

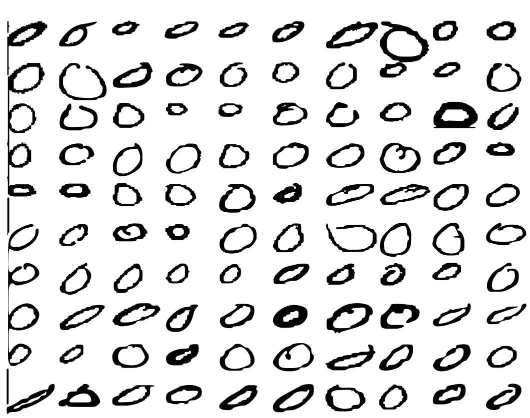

Fig.1. Some images of digits in CENPARMI handwritten numeral database

In [9], Hu et al. had performed some preprocessing work and extracted four kinds of features as follows:

X^G : 256-dimensional Gabor transformation feature ,
X^L: 121-dimensional Legendre moment feature,
X^P: 36-dimensional Seudo-Zernike moment feature,
X^Z : 30-dimensional Zernike moment feature .

Combining any two features of the above four features in the original feature space, and using the algorithm described in this paper (Section 2.3), we obtain the canonical projected sub-matrixes W_x and W_y, using the linear transformation (3) to extract the CCDF. The minimum-distance classifier is used for the classification, and the classification error rate is shown in table 1.

Notice that in getting the CPV, we should get the low-dimensional CPV first, and then high-dimensional according to Eq.(10) or (11). For example, when we want to combine two features X^G and X^P, for getting the CPV with 36-dimension and 256-dimension, we only need to get the eigenvalue and the corresponding eigenvector from the 36-dimensional matrix G_{X^P} .

To compare this algorithm with two existing feature level fusion method, experimental results are given here. Firstly, each sample should be normalized. Then, we group two sets of feature vectors into one union-vector according to the serial feature fusion in [2], and we combine two sets of feature vectors by a complex vector according to the parallel feature fusion in paper [1]. At last, we extract the combined features in the combined space, namely the total scatter matrix of their training sample of K-L transformation. Also the minimum-distance classifier is used for the classification, and the classification error rate is shown in table 1.

Table 1. The classification error rates based on the different feature fusion methods

Error rate	Serial fusion	Parallel fusion	Our method
X^G—X^L	0.1920	0.1925	0.1170
X^G—X^P	0.2280	0.2285	0.2050
X^G—X^Z	0.2295	0.2290	0.2230
X^L—X^P	0.2400	0.2415	0.1810
X^L—X^Z	0.2510	0.2505	0.2110
X^P—X^Z	0.4760	0.4785	0.3230

To explain the advantage of our algorithm, we list the experimental results on the same database, based on the single feature for recognition in recent years, in table 2. Such as the UODV and FSODV algorithm proposed by Jin-Yang in [11], the ULDA and FSLDA algorithm proposed by Xu-Yang in [12], and so on. In addition, we give the recognition results based on primitive feature. And the above experiments are all done using the minimal-distance classifier.

Table 2. The classification error rates based on the different single feature methods

Error rate	Paper [11] UODV	Paper [11] FSODV	Paper [12] ULDA	Paper [12] FSLD	Xu-Yong	Primal feature
X^G	0.183	0.260	0.199	0.274	0.198	0.269
X^L	0.106	0.268	0.141	0.270	0.150	0.479
X^P	0.285	0.430				0.429
X^Z	0.299	0.436				0.449

From table 1, we can observe that, the recognition error rate of our method is lower than that of the serial and parallel fusion method. From table 1 and table2, we can also see that, when combining the Gabor feature and Legendre feature, our method performs better than the other single feature methods based on (the recognition error rate is only corresponding to the method adopting UODV of Jin-Yang in Legendre feature). It shows that, the algorithm of combined feature extraction proposed in this paper increases the recognition rate to some extent.

4. Conclusion

In this paper, CCA is applied to feature fusion and image recognition. A new feature fusion method is proposed, which uses correlation feature of two groups of feature as effective discriminant information, so it not only suits for information fusion, but also eliminates the redundant information within the features. This offers a new way of applying two groups of feature fusion to discrimination.

Moreover, comparison between our method and two existing feature fusion methods is conducted. The inherent essence of using our method in recognition is put forward. From experimental results, the extraction of CCDF not only realizes the reduced primitive feature dimension, but also performs well in classification so that images' essential feature can be reflected.

The method in this paper is a good approach for information fusion on feature level. We should improve this method, and the relation between CCA and FLDA be researched.

Acknowledgements

We wish to thank the CUHK fund from HKSAR Government under Grant No. 4185 / 00E for supporting.

References

[1] Yang Jian, Yang Jing-yu, Zhang David, Lu Jian-feng, Feature fusion: parallel strategy vs. serial strategy, Pattern Recognition, 36 (2003) 1961-1971.

[2] Liu C J, Wechsler H, A shape-and texture-based enhanced Fisher classifier for face recognition, IEEE Transactions on Image Processing, 10 (4) (2001)598~608.

[3] Wen-Shuang Sun, Lan-Xiang Chen, Multivariate Statistical Analysis(in Chinese), Beijing: Higher Education Press, 1994.

[4] M. Borga, Learning Multidimensional Signal Processing, Linköping Studies in Science and Technology, issertations, No.531, Department of Electrical Engineering, Linköping University, Linköping , Sweden, 1998.

[5] Yu, Shiann-Jeng, Direct blind channel equalization via the programmable canonical correlation analysis, Signal Processing, 81 (2001)1715-1724.

[6] Thomas Melzer, Michael Reiter, Horst Bischof, Appearance models based on kernel canonical correlation Analysis, Pattern Recognition , 36 (2003) 1961-1971.

[7] H. C. Choi, R. W. King, Speaker adaptation through spectral transformation for HMM based speech recognition, IEEE International Symposium on Speech, Image Proce- ssing and Neural Networks, 2 (1994) 686-689.

[8] H. Hotelling, Relations between two sets of variates, Biometrika, 8(1936)321-377.

[9] Z. S. Hu, Z. Lou, J-Y. Yang, K. Liu, C. Y. Suen, Handwritten digit recognition basis on multi-classifier combination, Chinese J. Computer(in Chinese), 22 (4) (1999) 369-374.

[10] Jin Z, Yang J Y, Hu Z S, Face recognition based on uncorrelated discriminant transformation, Pattern Recognition, 34 (2001)1405-1416.

[11] Zhong Jin, Jing-Yu Yang, Zhen-Min Tang, Zhong-Shan Hu, A theorem on the uncorrelated optimal discriminant vectors. Pattern Recognition, 34 (2001)2041-2047.

[12] Yong Xu., Jing-Yu Yang, Zhong Jin, A novel method for Fisher discriminant analysis. Pattern Recognition, 37 (2004) 381-384.