Enhanced Linear Discriminant Canonical Correlation Analysis for Cross-modal Fusion Recognition

Chengnian Yu, Huabin Wang$^{(\boxtimes)}$ [ORCID], Xin Liu, and Liang Tao

School of Computer Science and Technology, Anhui University,
230601 Hefei, China
wanghuabin@ahu.edu.cn

Abstract. Based on discriminant canonical correlation analysis of LDA, a new method of multimodal information analysis and fusion is proposed in this paper. We process data from two perspectives, single modality and cross-modal. More specifically, firstly, LDA is utilised to obtain the best projection matrix, this way, the data in each within-modal can be as centralized as possible. Secondly, the improved DCCA is used to process the output of first step in order to maximize within-class correlation and minimize between-class correlation. The above two steps prove beneficial to obtain the feature with higher discriminating ability which is essential for the average fusion recognition accuracy improvement. We show state-of-art results or better than state-of-art on widely used USM benchmarks against all existing results include CCA, LDA, DCCA, GCCA and KCCA.

Keywords: Feature level fusion · Multimodal analysis
Canonical correlation analysis · Linear discriminant analysis

1 Introduction

Single modality biometric identification technology, such as face recognition [1], the gesture recognition [2], the iris recognition [3], has plagued researchers with a variety of formidable challenges over the years. However, multimodal fusion recognition [4, 5] is a technology that combines multiple biological features and uses fusion algorithms to recognition. Most recently, the concept has attracted increasing attention for feature set matching in [6, 7]. Wang et al. [8] proposed dynamic fusion methods that corresponding weights were assigned to each modality. It has strong robustness. Liu et al. [9] proposed an adaptive multi-feature fusion algorithm. This refer to as infrared object tracking method based on adaptive multi-feature fusion and Mean Shift (MS), which can achieve target tracking in complex scene. Liu et al. [10] proposed a new learning framework for projection dictionary was established to solve the problem of weak

matching and multimodal information fusion. Zeng et al. [11], from the viewpoint of fusing appearance statistical features, proposed human target recognition algorithm based on appearance statistics feature fusion.

According to the fusion of different information, the multimodal biometric fusion technology can be divided into sensor level fusion [12], feature level fusion [13], matching-score level fusion [14], and decision level fusion [15]. Among them, feature level fusion has great advantages. It not only retains more information of original samples, but also eliminates redundant information between different features. Theoretically, the superior fusion performance can be obtained. However, feature level fusion currently met with many challenges, such as the Curse of dimensionality, the incompatibility of the feature space, poor correlation between features, and how to design effective fusion strategy. We hope to solve above problems, so the goal for mutual utilization and supplement between different features can be achieved. At present, two well-known and typical feature fusion methods are: serial feature fusion [16] and parallel feature fusion [17, 18].

Recently, Sun et al. [19] used canonical correlation analysis (CCA) to achieve feature fusion by maximizing the correlation of two sets of variables in the projection space. A discriminative canonical correlation analysis (DCCA) algorithm, proposed by Kim et al. [20], this method can simultaneously maximize the within-class correlation and minimize the between-class correlation. Generalized canonical correlation analysis (GCCA) [21] made full use of class label information by minimizing and Constraint the within-class scatter matrix, so as to improve the discriminating ability of features. Kernel canonical correlation analysis (KCCA) [22], an extension method of CCA, cast a light on nonlinear problem, in this way, samples were mapped to kernel space through kernel functions and extracted features in kernel space. Haghighat et al. [23] proposed the discriminant correlation analysis (DCA), this work removed the correlation between feature sets of different classes through between-class scatter matrix, thus realizing the purpose of using class information.

CCA and derivative algorithm had solved some specific problems and gained exceeding recognition performance. The above performance, however, is achieved at the cost of not taking into account solving the problem of high dimension and similarity between different features at the same time. Additionally, it ignored the between-class relationship of the same modality feature sample. In this paper, we propose an enhanced fusion algorithm for linear discriminant canonical correlation analysis. The algorithm is described as follows: Firstly, each modality is treated widthwise, and LDA is applied to process each modality dataset respectively. In this way, data in each modality will be more centralized as much as possible. Secondly, the improved DCCA is used to process the outputs of first step, for the purpose of simultaneously maximize the within-class correlation and minimize the between-class correlation. The establishment and solution of the objective function is based on (DMCCA) [24]. The benefits of the proposed paper are multi-fold: 1, dimension curse problem of feature fusion is successfully solved: 2, Feature sets of each modality and different modalities have been well processed. The proposed algorithm simultaneously maximizes the within-class correlation and minimizes the between-class correlation. More importantly, the within-class feature is more centralized in each modality.

2   Related Knowledge-Canonical Correlation Analysis

The canonical correlation analysis was proposed by Hotelling et al. [25] in 1936. The object of CCA is to capture the correlations between two sets of variables. The methods are based on singular value decomposition for both representations. The details are as follows: let $x \in R^{p \times 1}$, $y \in R^{q \times 1}$ be two sets of zero-mean random variables, the CCA is to find a pair of projection directions w_x and w_y, such that the linear correlations between the projections onto these basis vectors are mutually maximized: $X_1 = w_x^T x, Y1 = w_y^T y$, then X_1, Y_1 is the first pair of canonical variables. It needs to satisfy the canonical property that the first projection is uncorrelated with the second projection, etc. All the correlation features of x and y are extracted. The criterion function is defined as follows:

$$\arg \max_{w_x, w_y} w_x^T * C_{xy} * w_y \quad s.t. w_x^T C_{xx} w_x = 1, w_y^T C_{yy} w_y = 1 \tag{1}$$

Where C_{xx}, C_{yy} are the within-sets covariance matrices of X and Y respectively, C_{xy} represents the between-set covariance matrix (note that $C_{yx} = C_{xy}^T$). The way to solve these problems can be referred to [26] for detail. In the aspect of fusion, according to the theory of canonical correlation analysis of Sun et al. [27], the method of concatenation or summation is put forward to fuse the feature vectors. The method is called FFS-1(F1) and FFS-2(F2), respectively.

Feature layer fusion strategy FFS-1: $Z_1 = \begin{pmatrix} X^* \\ Y^* \end{pmatrix} = \begin{pmatrix} W_x^T X \\ W_y^T Y \end{pmatrix}$

Feature layer fusion strategy FFS-2: $Z_2 = X^* + Y^* = W_x^T X + W_y^T Y$

3   Fusion Algorithm

In order to tackle the problem that covariance irreversibility caused by the number of samples is smaller than the number of dimensions. The traditional feature fusion method such as proposed in [28], which adopt method in two stages way, PCA + CCA. However, this method ignores that PCA may cause the loss of feature information, and ignores the discriminative information between samples as well. Motivated by the ideas of this combination. By this, algorithm is based on the dimension reduction method of LDA [29] to find the projection of the best separation class. In this paper, not only the problem of covariance irreversible is solved, but also achieves better results for each modality and cross-modal data sets.

3.1   Linear Discriminate Analysis

The main contribution of LDA is the idea that by finding a better projection vector space for high-dimensional data sets, so the distances within-class become smaller and the distances between-classes become larger in the projected space. Let's assume that the samples can be divided into C separate groups, where n_i columns belong to the ith

class $\left(N = \sum_{i=1}^{C} n_i\right)$. Suppose that $X = \{X_1, X_2 \cdots X_N\}$ denote a matrix, contains N training feature vectors. Let $X_i = R^d$ denote a feature vector. $\overline{u_i}$ is the means of the X_i vectors in the ith class, $\overline{u}$ is the global mean of the entire sample set. The within-class scatter matrix is defined as

$$S_W = \sum_{i=1}^{C} S_i = \sum_{i=1}^{C} \sum_{X_i \in W_i} (X_i - \overline{u_i})(X_i - \overline{u_i})^T \tag{2}$$

The between-class scatter matrix is defined as

$$S_B = \sum_{i=1}^{C} n_i(\overline{u_i} - \overline{u})(\overline{u_i} - \overline{u})^T \tag{3}$$

Therefore, a criterion function can be established:

$$J(W) = \frac{W^T S_B W}{W^T S_W W} \tag{4}$$

By constraining the denominator $W^T S_W W = 1$, the objective function is maximized to find the optimal value. The Lagrangian multiplier method is employed and then Eq. (4) is converted to find the eigenvalue of $S_W^{-1} S_B W = \lambda_i W$ to solve the problem.

3.2   Linear Discrimination Canonical Correlation Analysis

By the 3.1 algorithm, thereby achieving between-class is centralized and the effects of within-class are decentralized. Due to the difference in the dimension of the two modal samples, the two Training sets are normalized. Next, the algorithm of 3.2 is to establish correlations between the samples of the two modalities. The purpose is to maximize the similarities of a pairs of sets of within-class while minimizing the correlations between-class. $X = (X_1, X_2, \cdots X_C), Y = (Y_1, Y_2 \cdots Y_C)$ are training sets of two modalities, where $X_i \in R^{p \times ni}, Y_i \in R^{q \times ni}$, each class has ni samples, N represents the total number of training samples, X and Y have C = $(W_1, W_2, \ldots, W_c)$ class samples, and subsamples of each class can be represented as $Y_i = (y_1, y_2, \cdots, y_{ni})X_i = (x_1, x_2, \cdots, x_{ni})$, $x_i \in R^{p \times 1}, y_i \in R^{q \times 1}$. The correlation of category W_i in sample space X and sample space Y can be expressed as:

$$C_{WXY,i} = X_i(Y_i)^T \tag{5}$$

$$C_{BXY,i,j} = X_i(Y_j)^T \tag{6}$$

The within-class correlation matrix is expressed as:

$$C_{WXY} = \sum_{i=1}^{C} C_{WXY,i} = \sum_{i=1}^{C} X_i (Y_i)^T \tag{7}$$

The between-class correlation matrix is expressed as:

$$C_{BXY} = \sum_{i=1}^{C} \sum_{j=1,j\neq i}^{C} C_{BXY,i,j} = \sum_{i=1}^{C} \sum_{j=1,j\neq i}^{C} X_i (Y_j)^T$$

$$= \sum_{i=1}^{C} \sum_{j=1}^{C} X_i (Y_j)^T - \sum_{i=1}^{C} X_i (Y_i)^T \tag{8}$$

$$= \left(\sum_{i=1}^{C} X_i \right) \left(\sum_{j=1}^{C} Y_i \right)^T - \sum_{i=1}^{C} X_i (Y_i)^T$$

The criterion function model can be established as:

$$J(W,V) = \max_{W,V} \frac{W^T \tilde{C}_{XY} V}{\sqrt{W^T C_{XX} W} \sqrt{V^T C_{YY} V}} \quad s.t. W^T C_{XX} W = 1, V^T C_{YY} V = 1 \tag{9}$$

Among them, C_{xx}, C_{yy} are the within-sets covariance matrixes of X and Y, respectively, where $\tilde{C}_{XY} = C_{Wxy} - \eta C_{Bxy}, (\eta > 0)$. From the above equation, by adjusting the size of η, so that $\tilde{C}_{XY}$ makes certain trade-offs between within-class and between-class. The above problem becomes that under the constraint condition, the maximum projection vector W, V of the criterion function is obtained.

$$L(W,V) = W^T \tilde{C}_{XY} V - \lambda_1 (W^T C_{XX} W - 1) - \lambda_2 (V^T C_{YY} V - 1) \tag{10}$$

Where λ_1, λ_2 is a Lagrangian multiplier and the derivation of W and V respectively, and then obtained:

$$\frac{\partial L}{\partial W} = \tilde{C}_{XY} V - \lambda_1 C_{XX} W = 0 \tag{11}$$

$$\frac{\partial L}{\partial V} = \tilde{C}_{XY}^T W - \lambda_2 C_{YY} V = 0 \tag{12}$$

Multiply W^T, V^T by (11), (12) above, and then convert (11), (12) to (13):

$$\left. \begin{array}{l} W^T \tilde{C}_{XY} V = \lambda_1 W^T C_{XX} W = \lambda_1 \\ \left(W^T \tilde{C}_{XY} V \right)^T = \lambda_2 V^T C_{YY} V = \lambda_2 \end{array} \right\} \tag{13}$$

From the constraints,$\lambda_1 = \lambda_2 = \lambda, \tilde{C}_{YX} = \tilde{C}_{XY}^T$ can be obtained, C_{XX}, C_{YY} is reversible, then (13) can be converted to:

$$\left.\begin{array}{c}\tilde{C}_{XY}V = \lambda C_{XX}W \\ \tilde{C}_{XY}^T W = \lambda C_{YY}V\end{array}\right\} \Rightarrow \left.\begin{array}{c}\left(C_{XX}^{-1/2}\tilde{C}_{XY}C_{YY}^{-1}\tilde{C}_{YX}C_{XX}^{-1/2}\right)\left(C_{XX}^{1/2}W\right) = \lambda^2\left(C_{XX}^{1/2}W\right) \\ \left(C_{YY}^{-1/2}\tilde{C}_{YX}C_{XX}^{-1}\tilde{C}_{XY}C_{YY}^{-1/2}\right)\left(C_{YY}^{1/2}W\right) = \lambda^2\left(C_{YY}^{1/2}W\right)\end{array}\right\} \quad (14)$$

At this point, finding a maximizing projection vector for a criterion function is converted to solve generalized characteristic equations of formula (14). The matrix required to solve the eigenvalue decomposition (EVD) is a square matrix and singular value decomposition (SVD) [30] is a decomposition method applicable to any matrix, so this paper uses SVD decomposition for Eq. (15). Let $H = C_{XX}^{-1/2}\tilde{C}_{XY}C_{YY}^{-1/2}, \bar{W} = C_{XX}^{1/2}W, \bar{V} = C_{YY}^{1/2}V$, then (14) becomes the following form:

$$\begin{array}{c}HH^T\bar{W} = \lambda^2\bar{W} \\ H^TH\bar{V} = \lambda^2\bar{V}\end{array} \quad (15)$$

It can be seen that $\bar{W} = C_{XX}^{1/2}W$ and $\bar{V} = C_{YY}^{1/2}V$ are the feature vectors of HH^T and H^TH, respectively, and singular value decomposition $H = P\Lambda Q^T$ is performed on H, where $\Lambda = diag(\lambda_1, \lambda_2, \cdots, \lambda_r)$, $P = (p_1, p_2, \cdots, p_r)$, $Q = (q_1, q_2, \cdots q_r)$, $r = $ rank $(\tilde{C}_{XY})$. HH^T and H^TH have common non-zero eigenvalues, the column vector of P is the eigenvector corresponding to the eigenvalue λ^2 of HH^T, and the column vector of Q is the eigenvector corresponding to the eigenvalue λ^2 of H^TH.Then, the maximum solution of the criterion function model is obtained:

$$\left.\begin{array}{c}W = C_{XX}^{-1/2}\bar{W} \\ V = C_{YY}^{-1/2}\bar{V}\end{array}\right\} \quad (16)$$

The resulting W, V is the standard function projection vector. Therefore, the feature set X^*, Y^* is obtained, where X', Y' are the feature sets obtained in the first time, and the projection matrix W_x, W_y is as follows:

$$\left.\begin{array}{c}X^* = W^TX' = W^TW_x^TX = W_XX \\ Y^* = V^TY' = V^TW_y^TY = W_YY\end{array}\right\} \quad (17)$$

3.3   The Flow of the Algorithm

(1) Enter the two sets of centered training feature sets X and Y, and the category label information.
(2) According to the algorithm in Sect. 3.1, the LDA algorithm is used for the two feature sets, the projection matrix $W_{x'}, W_{y'}$ is obtained.

(3) According to the algorithm in Sect. 3.2, similarity processing is performed on the processed data sets of the two modalities. The projection matrix W, V is obtained from the criterion function in 3.2.

(4) According to the fusion strategy of Part 2, the test samples are classified and identified by using the nearest neighbor classifier.

4 Simulation Experiment

To verify the validity of algorithm in this paper, the database used for the USM database is the finger vein database, the open palm print database and the finger database of Hong Kong Polytechnic University, and the multi-feature handwritten data sets. Among them, the mean LBP method was used to extract finger veins and palm veins features. The finger feature is calculated by histogram of gradients (HOG) for detecting the contour image of finger vein. In this paper, we utilized the mean classification accuracy (MCA) [31] to evaluate the performance of the fusion algorithm, which is defined as follows:

Let's assume that the N samples as the testing subsets, while truth class labels are $\{\bar{y}_1, \bar{y}_2, \ldots, y_N\}$ and the predicted class labels are $\{\bar{y}_1, \bar{y}_2, \ldots, y_N\}$.

$$MCA = \frac{N_1}{N} \tag{18}$$

Where N_i denotes the number of samples with an error not greater than k between the predicted category label and the real category label, parameter k is set to 1.

4.1 Experiment 1

Multi-feature hand-written data sets are multimodal datasets in the UCI dataset, including 0 to 9 total 10 hand-written digital features, 200 samples in each category, 2000 samples in total. Each sample contains 6 features, which are morphological features (mfeat_mor), Zernike moment features (mfeat_zer), KL expansion coefficients (mfeat_kar), Fourier coefficients (mfeat_fou), contour correlation features (mfeat_fac), and pixel averaging (mfeat_pix). If you select any two features as fusion feature sets, there are 15 combinations. Randomly select 100 samples for each class as the training sets. The remaining samples are used as the testing sets. The nearest neighbor method is used for classification. The testing process repeats 20 cycles independently, and the average of 20 results is used as the final performance measure. Table 1 provides the recognition rate in single modality. Table 2 is based on the FFS-1(denote F1) and FFS-2 (denote F2) feature fusion strategies.

Table 1. Average classification accuracy in single modality

Single modality	fac	fou	kar	mor	pix	zer
MCA	0.944	0.823	0.967	0.422	0.967	0.809

Table 2. Average classification accuracy of the F1 and F2 under cross- modal

Multi-feature fusion	MCA											
	CCA		DCCA		GCCA		DCA		KCCA		ELDCCA	
	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2
fac + fou	0.85	0.89	0.84	0.91	0.97	0.97	0.94	0.90	0.70	0.69	0.98	0.98
fac + kar	0.91	0.93	0.92	0.93	0.97	0.95	0.96	0.96	0.48	0.47	0.98	0.98
fac + mor	0.89	0.93	0.91	0.94	0.95	0.96	0.92	0.86	0.41	0.49	0.98	0.98
fac + pix	0.83	0.84	0.81	0.85	0.96	0.96	0.96	0.94	0.65	0.67	0.98	0.98
fac + zer	0.92	0.94	0.92	0.93	0.97	0.95	0.90	0.88	0.60	0.62	0.98	0.97
fou + kar	0.88	0.87	0.87	0.88	0.96	0.96	0.94	0.90	0.72	0.78	0.97	0.96
fou + mor	0.75	0.77	0.74	0.75	0.51	0.49	0.76	0.56	0.52	0.47	0.85	0.84
fou + pix	0.70	0.75	0.69	0.79	0.93	0.93	0.95	0.89	0.81	0.86	0.97	0.96
fou + zer	0.79	0.81	0.80	0.80	0.84	0.84	0.84	0.81	0.69	0.64	0.86	0.85
kar + mor	0.93	0.95	0.93	0.95	0.92	0.90	0.81	0.48	0.68	0.64	0.97	0.97
kar + pix	0.77	0.86	0.79	0.85	0.96	0.90	0.95	0.94	0.70	0.68	0.94	0.94
kar + zer	0.94	0.93	0.94	0.95	0.96	0.93	0.76	0.83	0.67	0.66	0.96	0.95
mor + pix	0.69	0.77	0.67	0.77	0.91	0.90	0.93	0.93	0.27	0.26	0.97	0.97
mor + zer	0.81	0.81	0.80	0.81	0.74	0.74	0.73	0.66	0.36	0.30	0.84	0.82
pix + zer	0.79	0.87	0.76	0.84	0.94	0.87	0.88	0.96	0.80	0.73	0.96	0.94

From Tables 1 and 2, it can be seen that under the double modality, the MCA of the proposed method is higher than that in the single modality. Under the combinations of 13 features situation, we show that our approach performs better than single modality. However, the fusion recognition accuracy is only slightly lower than single modality in combinations of two features situation. Firstly, the processing of features makes the two modal data more centralized. Secondly, cross modal maximizes the within-class sample correlation between two modalities. The above two steps prove beneficial to the improvement of recognition accuracy. From Table 2, it can be seen that the recognition performance is better than the existing algorithms whether in series or parallel way.

From the analysis of the algorithm in Table 3 in Sects. 3.1 and 3.2, this illustrates the reasonableness of the 3.1 algorithm that make the data more centralized in advance. A large of experiments show that put the 3.1 algorithm into the standard function of algorithm 3.2 directly can't achieve double standard effect at the same time. Therefore, in Table 3, the recognition rate of the combination algorithm in the above manner is also better than the recognition rate of single algorithm. The Sect. 3.1 algorithm improves the performance of the Sect. 3.2. The combination of the two algorithms shows the rationality and effectiveness of the proposed algorithm.

4.2 Experiment 2

The datasets we use are finger vein library provided by the USM database, palm print database and finger database of Hong Kong Polytech University. However, there are no multimodal databases with multiple biometric features from the same user. Taking into

Table 3. Uses the F1 and F2 to combine the average classification accuracy of the proposed algorithm

Multi-feature fusion	MCA					
	Algorithm 3.1		Algorithm 3.2		ELDCCA	
	F1	F2	F1	F2	F1	F2
fac + fou	0.973	0.973	0.924	0.944	0.983	0.984
fac + kar	0.974	0.971	0.921	0.931	0.983	0.980
fac + mor	0.975	0.975	0.915	0.952	0.986	0.986
fac + pix	0.979	0.961	0.834	0.852	0.982	0.981
fac + zer	0.978	0.974	0.930	0.956	0.983	0.979
fou + kar	0.957	0.956	0.946	0.957	0.975	0.968
fou + mor	0.932	0.814	0.829	0.829	0.851	0.842
fou + pix	0.933	0.937	0.700	0.805	0.979	0.965
fou + zer	0.814	0.818	0.849	0.854	0.869	0.850
kar + mor	0.960	0.955	0.937	0.952	0.979	0.970
kar + pix	0.949	0.916	0.782	0.855	0.948	0.947
kar + zer	0.958	0.924	0.949	0.949	0.968	0.957
mor + pix	0.941	0.936	0.715	0.776	0.970	0.970
mor + zer	0.789	0.780	0.808	0.806	0.841	0.829
pix + zer	0.958	0.935	0.806	0.861	0.965	0.945

account the independence of biometric feature, so we choose the single modal feature of each user in compromise.64 users were selected in total, each user's single modality image consist of 6 samples, 3 training samples and the remaining 3 as testing samples. The nearest neighbor method is leveraged to feature classification. The test process was repeated 20 times independently and the average results were taken as the final performance metric.

Figures 1 and 2 refers to the combination of vein and finger shape. Serial and parallel strategies are adopted respectively. The recognition rate of each algorithm is obtained by setting different feature dimensions, and then the recognition results are compared with the existing algorithms. Figures 3 and 4 is the combination of finger veins and palmprint. In the same spirit, do as operations of Figs. 1 and 2 did. The conclusions can be drawn from Figs. 1 to 4: the algorithm proposed in this paper is more robust than the existing method. According to the following figure, the recognition rate of some algorithms improved with the increase of the feature dimension, while others fluctuate greatly. The reason for the decline is that redundant information is added to the feature dimension, which does harm to the acquisition of information, worst of all, and the recognition accuracy. In this paper, the recognition accuracy improves steadily.

Tables 4 and 5 shows the comparison between our algorithm and CCA and their improved methods in hand features. It can be seen from the above table that our algorithm has shown outstanding results on hand biometrics. This also fully demonstrates that a projection matrix can't simultaneously achieve data within-modal more
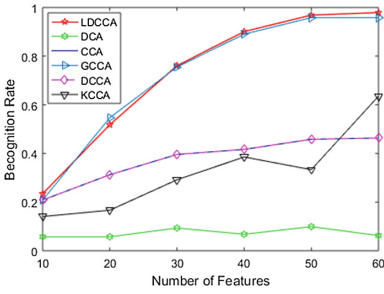
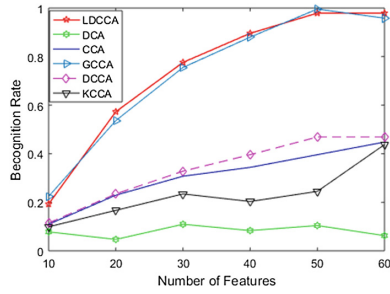**Fig. 1.** Average classification accuracy of finger veins and finger outline using the FFST-1

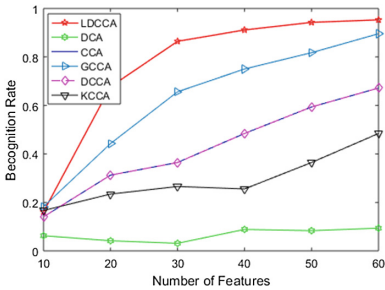**Fig. 2.** Average classification accuracy of finger veins and finger outline using the FFST-2

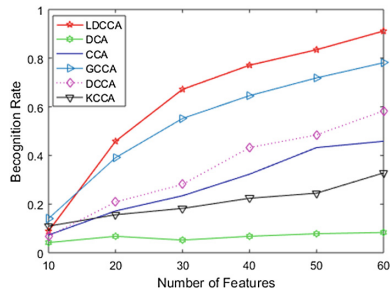**Fig. 3.** Average classification accuracy of finger veins and palm prints using the FFST-1

Fig. 4. Average classification accuracy of finger veins and palm prints using the FFST-2

Table 4. The average classification accuracy of the F1 and F2 under cross-modal

Multi-feature fusion	MCA											
	CCA		DCCA		GCCA		DCA		KCCA		ELDCCA	
	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2	F1	F2
Finger Vein + Contour	0.77	0.65	0.77	0.77	0.97	0.97	0.27	0.29	0.76	0.61	0.99	0.99
Finger Vein + Palm	0.93	0.90	0.93	0.91	0.26	0.25	0.93	0.93	0.94	0.91	0.97	0.98

centralized and maximum correlation between cross-modal. Table 5 show the results obtained from different feature combinations, using serial and parallel feature fusion strategies respectively. The original feature and the double standard effect are preserved. We can see that the recognition performance has been greatly improved. More importantly, the connection between single modal data has not been destroyed. This combination has achieved outstanding classification performance.

Table 5. Uses the F1 and F2 to combine the average classification accuracy of the proposed algorithm

Multi-feature fusion	MCA					
	Algorithm 3.1		Algorithm 3.2		ELDCCA	
	F1	F2	F1	F2	F1	F2
Finger Vein + Contour	0.994	0.994	0.993	0.994	0.997	0.998
Finger Vein + Palm	0.947	0.963	0.848	0.947	0.973	0.984

5   Conclusions

In this paper, a canonical correlation analysis algorithm based on linear discriminant analysis is proposed. In order to improve the effect of feature layer fusion, firstly, this article replaces the correlation analysis of image set classes in DCCA with the correlation analysis between samples. It can simultaneously maximize the within-class correlation and minimize the between-class correlation. At the same time, in order to extract feature sets with higher discriminative ability, LDA is adopted to make Samples in the same class for each modality as concentrat as possible. Results demonstrated, that in the USM database, the palmprint open database, the finger database and the multi feature handwritten data set of Hong Kong Polytech University, the algorithm is superior to other internationally popular comparison methods.

References

1. Li, L., Jun, Z., Fei, J., et al.: An incremental face recognition system based on deep learning. In: 2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA), pp. 238–241. IEEE (2017). https://doi.org/10.23919/mva.2017.7986845
2. Sun, Z., Zhang, H., Tan, T., et al.: Iris image classification based on hierarchical visual codebook. IEEE Trans. Pattern Anal. Mach. Intell. **36**(6), 1120–1133 (2014). https://doi.org/10.1109/tpami.2013.234
3. Liu, S., Liu, Y., Yu, J., et al.: Hierarchical static hand gesture recognition by combining finger detection and HOG features. J. Image Graph (2015)
4. Seng, K., Ang, L.M., Ooi, C.: A combined rule-based and machine learning audio-visual emotion recognition approach. IEEE Trans. Affect. Comput. (2016). https://doi.org/10.1109/TAFFC.2016.2588488
5. Ahlawat, M., Kant, C.: An introduction to multimodal biometric system: an overview. Int. J. Sci. Res. Dev. **3**(02), 2321–0613 (2015)
6. Zheng, H., Geng, X.: A multi-task model for simultaneous face identification and facial expression recognition. Neurocomputing **171**, 515–523 (2016). https://doi.org/10.1016/j.neucom.20-15.06.079
7. Wen, H., Liu, Y., Rekik, I., et al.: Multi-modal multiple kernel learning for accurate identification of tourette syndrome children. Pattern Recogn. **63**, 601–611 (2017). https://doi.org/10.1016/j.patcog.2016.09.039
8. Wang, S., Zhang, J., Zong, C.: Learning multimodal word representation via dynamic fusion methods. arXiv preprint arXiv:1801.00532 (2018)

9. Liu, Q., Tang, L., Zhao, B.-J., et al.: Infrared target tracking based on adaptive multiple features fusion and mean shift. J. Electron. Inf. Technol. **34**(5), 1137–1141 (2012). https://doi.org/10.3724/SP.J.1146.2011.01077

10. Liu, H., Wu, Y., Sun, F., et al.: Weakly paired multimodal fusion for object recognition. IEEE Trans. Autom. Sci. Eng. **15**(2), 784–795 (2018). https://doi.org/10.1109/TASE.2017.2692271

11. Zeng, M., Wu, Z., Tian, C., et al.: Fusing appearance statistical features for person re-identification. J. Electron. Inf. Technol. **36**(8), 1845–1851 (2014). https://doi.org/10.3724/SP.J.1146.2013.01389

12. Meng, W., Wong, D.S., Furnell, S., et al.: Surveying the development of biometric user authentication on mobile phones. IEEE Commun. Surv. Tutorials **17**(3), 1268–1293 (2015). https://doi.org/10.1109/COMST.2014.2386915

13. Haghighat, M., Abdel-Mottaleb, M., Alhalabi, W.: Discriminant correlation analysis for feature level fusion with application to multimodal biometrics. In: IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 1866–1870 (2016). https://doi.org/10.1109/icassp.2016.7472000

14. He, M., Horng, S.J., Fan, P., et al.: Performance evaluation of score level fusion in multimodal biometric systems. Pattern Recogn. **43**(5), 1789–1800 (2010). https://doi.org/10.1016/j.patcog.2009.11.018

15. Liu, H., Li, S.: Decision fusion of sparse representation and support vector machine for SAR image target recognition. Neurocomputing **113**, 97–104 (2013). https://doi.org/10.1016/j.neucom.2013.01.033

16. Liu, C., Wechsler, H.: A shape-and texture-based enhanced fisher classifier for face recognition. IEEE Trans. Image Process. **10**(4), 598–608 (2001). https://doi.org/10.1109/83.913594

17. Yang, J., Yang, J., Zhang, D., et al.: Feature fusion: parallel strategy vs. serial strategy. Pattern Recogn. **36**(6), 1369–1381 (2003). https://doi.org/10.1016/S0031-3203(02)00262-5

18. Sun, Q., Zeng, S., Yang, M., et al.: A new method of feature fusion and its application inimage recognition. Pattern Recogn. **38**(12), 2437–2448 (2005). https://doi.org/10.1016/j.patcog.2004.12.013

19. Sun, Q., Zeng, S., Yang, M., et al.: Combined feature extraction based on canonical correlation analysis and face recognition. J. Comput. Res. Dev. **42**(4), 614–621 (2005). https://doi.org/10.1360/crad20050413

20. Kim, T.K., Kittler, J., Cipolla, R.: Discriminative learning and recognition of image set classes using canonical correlations. IEEE Trans. Pattern Anal. Mach. Intell. **29**(6), 1005–1018 (2007). https://doi.org/10.1109/TPAMI.2007.1037

21. Tenenhaus, A., Philippe, C., Guillemot, V., et al.: Variable selection for generalized canonical correlation analysis. Biostatistics. **15**(3), 569–583 (2014). https://doi.org/10.1093/biostatistics/kxu001

22. Jia, Z.: Multi-feature combination face recognition based on kernel canonical correlation analysis. Int. J. Signal Process. Image Process. Pattern Recogn. **9**(7), 221–230 (2016)

23. Haghighat, M., Abdel-Mottaleb, M., Alhalabi, W.: Discriminant correlation analysis: real-time feature level fusion for multimodal biometric recognition. IEEE Trans. Inf. Forensics Secur. **11**(9), 1984–1996 (2016). https://doi.org/10.1109/TIFS.2016.2569061

24. Gao, L., Qi, L., Chen, E., et al.: Discriminative multiple canonical correlation analysis for information fusion. IEEE Trans. Image Process. **27**(4), 1951–1965 (2018). https://doi.org/10.1109/TIP.2017.2765820

25. Hotelling, H.: Relations between two sets of variates. Biometrika **28**(3–4), 321–377 (1992). https://doi.org/10.2307/2333955

26. Weenink, D.: Canonical correlation analysis. In: Proceedings of the Institute of Phonetic Sciences of the University of Amsterdam, vol. 25, pp. 81–99. University of Amsterdam (2003)
27. Sun, Q., Zeng, S., Yang, M., et al.: A new method of feature fusion and its application in image recognition. Pattern Recogn. **38**(12), 2437–2448 (2005). https://doi.org/10.1016/j. patcog.2004.12.013
28. Correa, N.M., Adali, T., Li, Y.-O., Calhoun, V.D.: Canonicalcorrelation analysis for data fusion and group inferences. IEEE Signal Process. Mag. **27**(4), 39–50 (2010). https://doi.org/ 10.1109/34.598228
29. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.: Eigenfaces vs. fisherfaces: recognition using class specific linear projection. IEEE Trans. Pattern Anal. Mach. Intell. **19**(7), 711–720 (1997). https://doi.org/10.1109/34.598228
30. Dan, K.: A singularly valuable decomposition: the SVD of a matrix. Coll. Math. J. **27**(1), 2–23 (1996). https://doi.org/10.2307/2687269
31. Zhou, H., Chen, S.: Ordinal discriminative canonical correlation analysis. J. Softw. **25**(9), 2018–2025 (2014). https://doi.org/10.13328/j.cnki.jos.004649