Cross-modal correlation feature extraction with orthogonality redundancy reduce and discriminant structure constraint

Qianjin Zhao and Xinrui Ping

School of Mathematics and Big Data,
Anhui University of Science and Technology,
Huainan 232001, China
Email: qjzhao@aust.edu.cn
Email: 1845112028@qq.com

Shuzhi Su*

School of Computer Science and Engineering,
Anhui University of Science and Technology,
Huainan 232001, China
Email: sushuzhi@foxmail.com
*Corresponding author

Abstract: Canonical correlation analysis (CCA) is a classic feature extraction method that is widely used in the field of pattern recognition. Its goal is to learn correlation projection directions to maximise the correlation between the two sets of variables, but it does not take into consideration the class label information among samples and the within-modal redundancy information from the correlation projection directions. To this end, a novel method of class label embedding orthogonal correlation feature extraction method is proposed in this paper. The label-guide discriminant structure information is deeply embedded in correlative analytical theories for improving discrimination of correlation features, and within-modal orthogonality constraints are added to further reduce the projection redundancy of correlation features. Several validation experiments on the GT, Umist and YALE datasets are designed, which demonstrates that DOCCA method is superior to other feature extraction methods and achieves state-of-the-art performance. This method provides a new solution to pattern recognition.

Keywords: feature extraction; correlation analysis theory; discriminative subspace learning; orthogonality redundancy reduce.

Reference to this paper should be made as follows: Zhao, Q., Ping, X. and Su, S. (2023) 'Cross-modal correlation feature extraction with orthogonality redundancy reduce and discriminant structure constraint', *Int. J. Computational Science and Engineering*, Vol. 26, No. 1, pp.101–109.

Biographical notes: Qianjin Zhao is a Professor with the School of Mathematics and Big Data, Anhui University of Science and Technology, Huainan, China. He received his PhD in the School of Computer Science, Hefei University of technology, Hefei, China. His research interests include rational interpolation and approximation, computer aided geometric design and digital image processing.

Xinrui Ping will receive his Master's in Mathematics in 2022 from the Anhui University of Science and Technology, Huainan, China. His research interests include multimodal information fusion, feature extraction and data processing.

Shuzhi Su is an Associate Professor with the School of Computer Science and Engineering, Anhui University of Science and Technology, Huainan, China. He received his PhD in the School of Internet of Things Engineering, Jiangnan University, Wuxi, China. His research interests include multimodal pattern recognition, information fusion, feature learning, and image processing.

1 Introduction

In real-world applications, multimodal data is abundant and widespread. Multimodal data is description of multiple types of data for the same target. For example, in video analysis, video can be decomposed into audio, images, captions and so on. These different data can describe video

information from different angles, and they are complementary. We named them multimodal data. Compared with the single modal data, multimodal data contains more complete internal structural information. Feature information contained in multimodal data can be used in image classification (Ying et al., 2021b), image retrieval (Ying et al., 2021a), semantic segmentation (Wu et al., 2019), information fusion and other fields. How to extract effective features from multimodal data is another important and challenging problem in pattern recognition (Xu et al., 2018).

Canonical correlation analysis (CCA) (Yang et al., 2021c; Shu et al., 2020) is a powerful tool for feature extraction (Qian et al., 2020; Srivastava and Biswas, 2020) which finds a pair of projection directions that maximise the correlation between the projections to handle with the relationships between two random vectors. It is widely used in neuroscience (Wang et al., 2020), emotion recognition (Zheng, 2016), fault detection (Bhowmik et al., 2020), process monitoring (Bao et al., 2019), remote sensing image (Samat et al., 2017) and so on. CCA can commendably capture the correlation between different modalities, but the correlation does not reflect the similarity between different classes of the same sample. In other words, the class information of the samples is ignored. Through introducing discrimination information, a new feature extraction method called discriminative CCA (DCCA) (Sun et al., 2008) is proposed. DCCA simultaneously maximises within-class correlation and minimises inter-class correlation in correlative analysis theories, which can learn correlative feature with well class separability.

Several feature extraction methods mentioned above based on correlation analysis theories do not exploit the graph structure information. Graph structure information is invoked by graph regularisation, graph CCA (Chen et al., 2018) method was proposed accordingly. A neat link was introduced between graph embedding and canonical correlations. Graph CCA pursues maximally correlated linear projections, and can capture structural information about data vectors. Therefore, the linearity of CCA is difficult to reveal the local geometric structure in the original high-dimensional data. To address this problem, Sun and Chen (2007) proposed locality preserving CCA (LPCCA) method, which explored the local relations among the original high-dimensional data with the use of Euclidian distance. In addition, the local nonlinear structure is retained in the learning of low-dimensional features. This method is successfully used in attitude estimation (Al-Sharman et al., 2019). Simultaneously, we consider the correlation between the sample pairs, the correlation between the sample and the domain area. A CCA model called local discrimination CCA (Peng et al., 2010) was proposed.

As a linear subspace method, CCA has not found a nonlinear correlation between the two sets of features. Kernel technology is commonly used nonlinear auxiliary technology in feature extraction. With the help of kernel technology, kernel CCA (Lisanti et al., 2014) maps data to a higher dimensional kernel space, and then executes the linear CCA method in the kernel space to obtain nonlinear low-dimensional features. Therefore, kernel CCA can solve the nonlinear related problems to a certain extent. It is a promising research method, which is commonly used for optical blind separation of sources. Meanwhile, there is another alternative method that can capture the nonlinear correlation between different modalities called deep CCA (Andrew et al., 2013). Deep CCA can be seen as a nonlinear extension of the linear method of CCA. Deep CCA is a method of converting cross-modal nonlinear data onto highly linear correlation, which is commonly used in feed-forward neural networks. Combining deep CCA with within-class and inter-class information, deep DCCA (Elmadany et al., 2016) is proposed. It simultaneously learns two deep mapping networks of two sets to maximise the within-class correlation and minimise the inter-correlation, and improves the classification accuracy of handwritten digits and emotional data.

From different aspects, the above methods improve and optimise correlation analysis theories for enhancing the class separability of low-dimensional correlation feature. Orthogonality is also a popular constraint criterion in feature extraction. Orthogonal projection system is less susceptible to the influence of data distribution and noise, and can keep Euclidean distances between samples in the process of dimension reduction. Cai et al. (2006) proposed an orthogonal locality preservation analysis method based on orthogonal constraints and successfully applied to face recognition. Coupled with the idea of locally linear embedding (Roweis and Saul, 2000), orthogonal neighbourhood preserving projection method (Kokiopoulou and Saad, 2007) was proposed, which was successfully applied to data visualisation. In the way of adding orthogonality to correlation analysis theories, Shen et al. (2013) propose an orthogonal correlation feature extraction method. This method obtains orthogonal canonical projection vectors through dual feature decomposition, which will have higher recognition rate and the noise robustness when facing small sample size (SSS) (Sun et al., 2005) problems.

In this paper, the idea of orthogonality and class information was combined with correlation analysis. Then a novel cross-modal correlation feature extraction method with orthogonality redundancy reduction and discriminant structure constraint was proposed, called discriminative orthogonal canonical correlation analysis (DOCCA). Label-guide discriminant structure information is embedded into correlation analysis theories for improving discrimination of correlation features in this method. To further reduce the projection redundancy of correlation features, a novel orthogonality redundancy reduction model is constructed, and the analytical solutions of the model can be directly derived by theoretical derivation. This feature extraction method can be employed in image pre-processing to improve the accuracy of tumour detection. In order to verify the effectiveness of the proposed method, the experimental results from various image datasets show superiority of our method in image recognition tasks.

The rest is distributed as follows. Section 2 is a review of the CCA method. Section 3 is a detailed description of the model construction and optimisation solution of the DOCCA method. The experimental part and the thesis summary are in Sections 4 and 5 respectively.

2 Review of CCA

Given N pair-wise samples $\{x_i, y_i\}$ ($i = 1, 2, \ldots, N$). Suppose that $X = [x_1, x_2, \ldots, x_N] \in R^{q\times N}$, $Y = [y_1, y_2, \ldots, y_N] \in R^{q\times N}$ are sample sets of two modalities, where p and q are the dimensions of X and Y. CCA aims to find pairs of correlation projection directions ω_x and ω_y that maximise the correlation between the projections $\omega_x^T X$ and $\omega_y^T Y$. CCA can be described as follows:

$$J(\omega_x, \omega_y) = \frac{\omega_x^T S_{xy} \omega_y}{\sqrt{\omega_x^T S_{xx} \omega_x}\sqrt{\omega_y^T S_{yy} \omega_y}} \tag{1}$$

where $S_{xy} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})^T$ is the covariance matrix of X and Y, $S_{xx} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(x_i - \bar{x})^T$ and $S_{yy} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y})(y_i - \bar{y})^T$ are the variances of X and Y. Besides, $\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$, $\bar{y} = \frac{1}{N}\sum_{i=1}^{N} y_i$ are respectively the sample means of X and Y. S_{xy} reveals the correlation between views, while S_{xx} and S_{yy} reflect the overall distribution of data within views.

3 Discriminative orthogonal canonical correlation analysis

CCA method is an unsupervised linear feature extraction method. This method can extract two sets of features and perform linear fusion. However, CCA ignores the class information of samples, which could lead to the limitation of recognition performance. In order to make extractive features having better discriminative performance, in the proposed DOCCA method, the orthogonality redundancy reduction and discriminant structure constraint can be exploited into the cross-modal correlation feature extraction framework to extract more powerful features and reduce the influence of data distribution and noise. Advantages of DOCCA can be summarised as the following points:

1 Effective discriminative structure in the discriminative subspace is kept and is insensitive to the influence of data distribution and noise

2 This method constructs the novel orthogonality redundancy reduce model to further reduce the projection redundancy of correlative features, and the analytical solutions of the model can be directly derived by theoretical derivation.

Our method is described as follows.

The optimisation object of equation (1) is difficult to compute the optimal solution of correlation projection directions. Thus, we hope to further optimise and ameliorate it.

Theorem 1: $J(k\omega_x, l\omega_y) = J(\omega_x, \omega_y)$, $\forall k, l \in \mathbb{R}$

Proof: For $\forall k, l \in \mathbb{R}$

$$\begin{aligned} J(k\omega_x, l\omega_y) &= \frac{k\omega_x^T S_{xy} l\omega_y}{\sqrt{k\omega_x^T S_{xx} k\omega_x}\sqrt{l\omega_y^T S_{yy} l\omega_y}} \\ &= \frac{kl\omega_x^T S_{xy} \omega_y}{\sqrt{k^2\omega_x^T S_{xx} \omega_x}\sqrt{l^2\omega_y^T S_{yy} \omega_y}} \\ &= \frac{kl\omega_x^T S_{xy} \omega_y}{kl\sqrt{\omega_x^T S_{xx} \omega_x}\sqrt{\omega_y^T S_{yy} \omega_y}} \\ &= \frac{\omega_x^T S_{xy} \omega_y}{\sqrt{\omega_x^T S_{xx} \omega_x}\sqrt{\omega_y^T S_{yy} \omega_y}} \\ &= J(\omega_x, \omega_y) \end{aligned}$$

According to Theorem 1, if the numerator and denominator of the target function proportional increase through the same multiple, the optimisation objective results are constants. We fix the numerator, optimise the denominator and rewrite the CCA optimisation formula as follows:

$$\begin{aligned} &\max \ \omega_x^T S_{xy} \omega_y \\ &\quad \omega_x, \omega_y \\ &\text{s.t.} \quad \omega_x^T S_{xx} \omega_x = 1, \omega_y^T S_{yy} \omega_y = 1. \end{aligned} \tag{2}$$

In the above equation (2), adding and subtracting constants to the optimisation target does not affect the final optimisation result the optimisation objective. Thus, it can be modified in following form:

$$\begin{aligned} &\max \ \omega_x^T S_{xy} \omega_y - \omega_x^T S_{xx} \omega_x - \omega_y^T S_{yy} \omega_y \\ &\quad \omega_x, \omega_y \\ &\text{s.t.} \quad \omega_x^T S_{xx} \omega_x = 1, \omega_y^T S_{yy} \omega_y = 1. \end{aligned} \tag{3}$$

The above model maximises the differences between the two samples. To gain more discriminative power, it is desirable to minimise the differences in within-class samples.

The orthogonality property is pivotal for redundancy reduction. To further improve the discriminability of the projection vectors and eliminate redundant information, orthogonal constraint is imposed on the projection matrix. Therefore, for purpose of obtaining a set of orthogonal projection vectors, a constraint is added to the objective function in equation (3). The orthogonal direction matrices ω_x and ω_y can be computed in an iterative way as follows:

$$\begin{aligned} &\max\ \omega_x^T S_{xy}\omega_y \\ &\quad \omega_x, \omega_y \\ &\begin{cases} \omega_{x_1}^T\omega_{x_k} = \omega_{x_2}^T\omega_{x_k} = \cdots = \omega_{x_{k-1}}^T\omega_{x_k} = 0, \\ \omega_{y_1}^T\omega_{y_k} = \omega_{y_2}^T\omega_{y_k} = \cdots = \omega_{y_{k-1}}^T\omega_{y_k} = 0, \\ \omega_{x_k}^T S_{xx}\omega_{x_k} = 1, \\ \omega_{y_k}^T S_{yy}\omega_{y_k} = 1. \end{cases} \end{aligned} \tag{4}$$

By means of orthogonal constraints, we can further normalise the optimal projection directions. The new CCA model can preserve the intrinsic correlation, that means orthogonal transformation can preserve structure information between modals.

Simplifying the orthogonal constraints in equation (4) relies on the deflation scheme strategy (Zhang et al., 2020), then we obtain

$$\omega_x^T\omega_x = 1,\ \omega_y^T\omega_y = 1. \tag{5}$$

We change the constraint condition and a new optimisation problem show itself upon the water. Adding orthogonal constraints on the above model equation (3) to reconstruct the model

$$\begin{aligned} &\max\ \omega_x^T S_{xy}\omega_y - \omega_x^T S_{xx}\omega_x - \omega_y^T S_{yy}\omega_y \\ &\quad \omega_x, \omega_y \\ &\text{s.t.}\quad \omega_x^T\omega_x = 1,\ \omega_y^T\omega_y = 1 \end{aligned} \tag{6}$$

which orthogonality can make sure that features exacted by our method are uncorrelated as much as possible. Sample sets represent two modalities of the object, so the irrelevance of features are supposed to reduce the redundancy that exists in each modal features.

However, the multivariate eigenvalue problem equation (6) not only has large time complexity, but also has no analytical solutions. Thus, we use the constraint relaxation strategy (Yuan and Sun, 2020) to simplify the model in equation (6) to

$$\begin{aligned} &\max\ \omega_x^T S_{xy}\omega_y - \omega_x^T S_{xx}\omega_x - \omega_y^T S_{yy}\omega_y \\ &\quad \omega_x, \omega_y \\ &\text{s.t.}\quad \omega_x^T\omega_x + \omega_y^T\omega_y = 1. \end{aligned} \tag{7}$$

Because of CCA method does not utilise class label information, the obtained projection vectors may not have good discriminative performance, so it is difficult to guarantee well classification effect. At this point, we embed class label information into the scatter matrix S_{xx} and S_{yy} to extract more discriminative cross-modal features, and we get the within-class scatter matrix $S_{wx} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x}_i)(x_i - \bar{x}_i)^T$ and $S_{wy} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \bar{y}_i)(y_i - \bar{y}_i)^T$, $\bar{x}_i$ is the mean of the class where x_i is, and $\bar{y}_i$ is the mean of the class where y_i is. S_{wx} and S_{wy} measure the within-class cohesion. A new optimisation model for identifying orthogonal CCA is obtained by rearrangement

$$\begin{aligned} &\max\ \omega_x^T S_{xy}\omega_y - \omega_x^T S_{wx}\omega_x - \omega_y^T S_{wy}\omega_y \\ &\quad \omega_x, \omega_y \\ &\text{s.t.}\quad \omega_x^T\omega_x + \omega_y^T\omega_y = 1. \end{aligned} \tag{8}$$

S_{xy} reflects the correlation between modalities, S_{wx} and S_{wy} constrain the correlation within modalities. Therefore, the optimisation criterion of DOCCA can be regarded as maximising the correlation between modes and minimising the intra-class information within modalities.

We can apply the Lagrange multiplier (Kheyrinataj and Nazemi, 2020) method to solve equation (8). Firstly, the Lagrange function can be constructed as follows:

$$\begin{aligned} L(\omega_x, \omega_y) = {} & \omega_x^T S_{xy}\omega_y - \omega_x^T S_{wx}\omega_x - \omega_y^T S_{wy}\omega_y \\ & -\lambda(\omega_x^T\omega_x + \omega_y^T\omega_y - 1) \end{aligned} \tag{9}$$

where λ is the Lagrange multiplier.

Let the partial derivatives of $L(\omega_x, \omega_y)$ with respect to ω_x and ω_y to zero

$$\frac{\partial L}{\partial \omega_x} = S_{xy}\omega_y - 2S_{wx}\omega_x - 2\lambda\omega_x = 0 \tag{10}$$

$$\frac{\partial L}{\partial \omega_y} = S_{yx}\omega_x - 2S_{wy}\omega_y - 2\lambda\omega_y = 0 \tag{11}$$

in equations (10) and (11), S_{wx} and S_{wy} are positive semi-definite matrices. Through simple algebraic operations, the above equations (10) and (11) can be further written as:

$$S_{xy}\omega_y - 2S_{wx}\omega_x = 2\lambda\omega_x \tag{12}$$

$$S_{yx}\omega_x - 2S_{wy}\omega_y = 2\lambda\omega_y. \tag{13}$$

Equations (12) and (13) can be equivalently transformed into the following generalised eigenvalue problem

$$\begin{bmatrix} -2S_{wx} & S_{xy} \\ S_{yx} & -2S_{wy} \end{bmatrix}\begin{bmatrix} \omega_x \\ \omega_y \end{bmatrix} = 2\lambda\begin{bmatrix} \omega_x \\ \omega_y \end{bmatrix} \tag{14}$$

that is

$$\begin{bmatrix} -S_{wx} & \frac{1}{2}S_{xy} \\ \frac{1}{2}S_{yx} & -S_{wy} \end{bmatrix}\begin{bmatrix} \omega_x \\ \omega_y \end{bmatrix} = \lambda\begin{bmatrix} \omega_x \\ \omega_y \end{bmatrix}. \tag{15}$$

By solving equation (15), the optimal projection matrix $\omega_x = [\omega_{x1}, \ldots, \omega_{xd}]^T R^{p\times d}$ and $\omega_y = [\omega_{y1}, \ldots, \omega_{yd}]^T R^{q\times d}$ can be formed by the first d sets of eigenvectors $[\omega_{x1}, \ldots, \omega_{xd}]$ and $[\omega_{y1}, \ldots, \omega_{yd}]$, then the relevant features of X and Y are extracted, expressed as $\omega_x^T X$ and $\omega_y^T Y$.

The specific steps of DOCCA method are as follows:

Step 1 For given N pair-wise samples $\{x_i, y_i\}$ ($i = 1, 2, \ldots, N$), calculate within-set S_{xx}, S_{yy} and between-set covariance matrices S_{xy}.

Step 2 Calculate within-class scatter matrix S_{wx} and S_{wy}. According to orthogonal constraints condition, get optimisation model (8).

Step 3 According to generalised eigenequation in equation (15), obtain eigenvectors $[\omega_{x1}, \ldots, \omega_{xd}]$ and $[\omega_{y1}, \ldots, \omega_{yd}]$.

Step 4 Take eigenvectors as the projection vector matrix of the sample sets and use the fusion strategy in the database recognition experiment.

4 Experiment and analysis

In order to verify the effectiveness of the proposed method, we designed several experiments on real GT image database (Zhang et al., 2018), Umist image database (Hu and Zhang, 2020) and YALE image database (Wang and Zhang, 2013). Different modal data onto each image is obtained by modal strategies. We use Coiflets (Yu et al., 2018) and Daubechies (Waqas et al., 2020) wavelet transform to extract feature vectors from the low-frequency sub-images. Wavelet transforms (Chellappan et al., 2021) can not only retain the original information on digital images, but also holds the unique function of decomposition and de-correlation. It is a reversible transformation of energy conservation. Meanwhile, in order to decrease the impact of the SSS problems, the principal component analysis (PCA) (Jiang et al., 2018, 2021) method is used to reduce the dimension of the sub-images to 100 dimensions.

If this is just putting these features together obtained in different ways, it could lead to low recognition accuracy. For the obtained canonical correlation features, the following serial and parallel fusion strategies (Yang et al., 2003) can be used to calculate the typical correlation discriminant features Z, namely

$$Z_1 = \begin{pmatrix} \omega_x^T X \\ \omega_y^T Y \end{pmatrix} = \begin{pmatrix} \omega_x & 0 \\ 0 & \omega_y \end{pmatrix}^T \begin{pmatrix} X \\ Y \end{pmatrix}$$

$$Z_2 = \omega_x^T X + \omega_y^T Y = \begin{pmatrix} \omega_x \\ \omega_y \end{pmatrix}^T \begin{pmatrix} X \\ Y \end{pmatrix}$$

These two fusion strategies both have good fusion performance. Z_1 combines the two feature sets, and some information may be lost in the fusion process, while Z_2 splices the two feature sets together without losing any information. In this paper, the parallel fusion strategy Z_3 is used to fuse the feature set, and the K-nearest neighbour (KNN) (Zhang et al., 2017) classifier is used to classify the final pattern of the fusion feature.

In the experimental part, DOCCA is compared with LPCCA, DCCA, graph multi-view canonical correlation (GMCC) (Chen et al., 2019) and CCA respectively. The specific analysis method is as follows.

4.1 Experiments on the GT database

The GT image database collects a total of 50 people's images, each of which has 15 frontal photos with different angles and different expressions. In the GT image database, we select n (n = 4, 5, 6, 7) images as training samples, and the rest is utilised as test samples. A total of ten random trials are performed. Table 1 shows the average recognition rate of different training samples of diverse methods in this database.

Table 1 Experimental results on GT image database

	4Train	*5Train*	*6Train*	*7Train*
DOCCA	69.58 ± 1.47	72.22 ± 1.51	73.93 ± 1.08	75.05 ± 1.80
LPCCA	36.55 ± 3.00	45.26 ± 2.06	49.89 ± 3.68	56.18 ± 2.79
DCCA	53.27 ± 2.32	63.56 ± 2.77	67.80 ± 1.29	73.68 ± 1.71
GMCC	39.44 ± 2.18	47.82 ± 1.47	50.49 ± 1.65	55.58 ± 1.44
CCA	52.80 ± 2.79	59.08 ± 1.81	61.78 ± 1.35	66.22 ± 1.66

Notes: A ± B: A represents the average recognition rate and B represents the corresponding standard deviation of the recognition rate.

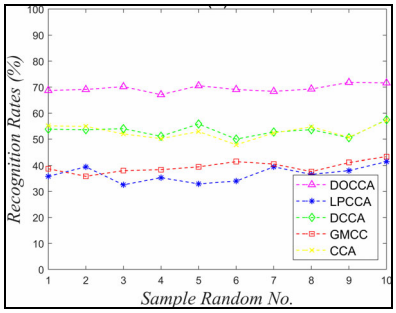
It can be seen from Table 1 that our proposed method DOCCA maintains the highest recognition rates. As the number of training samples decreases, the recognition rates of LPCCA, DCCA, GMCC and CCA all decrease in separate ranges, while the average recognition of DOCCA stabilises at about 70%. Especially when there are only four samples per class, the maximum recognition rate of DOCCA is at least 15% higher than other methods. The experiment shows that the recognition performance of LPCCA, DCCA, GMCC and CCA depends on the number of training samples to a great extent.

Figure 1 shows the variation on the recognition rate of DOCCA, LPCCA, DCCA, GMCC and CCA that samples random ten times under different number of training samples in the GT database. Under different training samples, DOCCA changes more smoothly compared with other method, which shows that our method has better robustness.

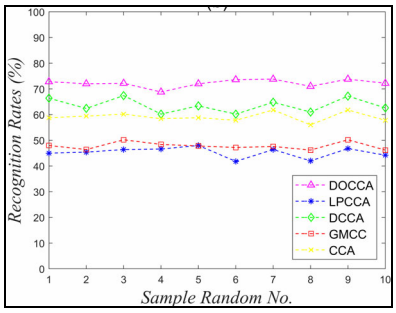
CCA method obtains all canonical correlations. It is meaningless that the extracted features contain too much redundant information. CCA only uses global structure information and does not commit full use of the class label information, so it cannot obtain the strong and effective supervision information. Both DOCCA and DCCA take full advantage of the class information in the view, thereby enhance the extracted supervision information. Compared to the other three methods, it can be clearly seen in Table 1 that DOCCA maintains a higher recognition rate which rises smoothly with the increase of the test images, regardless of the number of test images. Compared with DOCCA, it is difficult for DCCA to make full use of local structure information when there are few test images. DOCCA can ensure the orthogonality of the projection vector sets, eliminate the information redundancy between sample features and make the extracted feature information more

discriminative, so the recognition rate of DOCCA remains optimal.

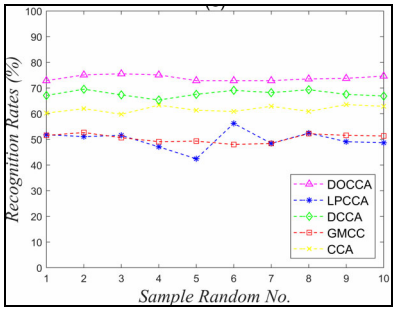
Figure 1 The recognition rate of DOCCA, LPCCA, DCCA, GMCC and CCA on the GT database for facial image features when the number of training samples n are 4, 5, 6, and 7, respectively (see online version for colours)



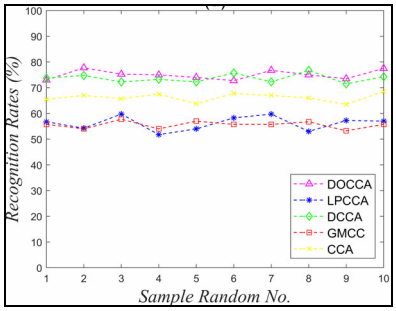
(a)



(b)



(c)

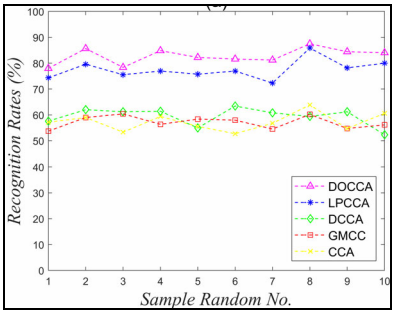


(d)

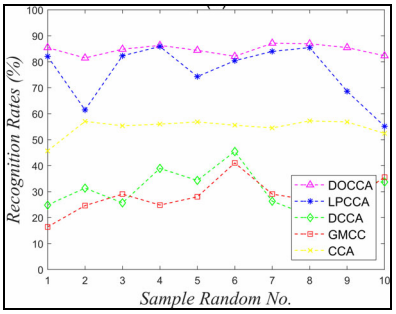
4.2 *Experiments on the Umist database*

The Umist database contains 575 pictures collected from 20 different people, covers a series of postures from the side to the front of each person. In Table 2, GMCC and CCA show poor recognition rate, while our proposed DOCCA method still maintains a high recognition rate. DCCA and LPCCA methods show uncommon amplitude of fluctuation. However, recognition rates of DOCCA change gently with the increase of training times, which indicate that our proposed method has good robustness and substantiated the advantage of orthogonal performance.

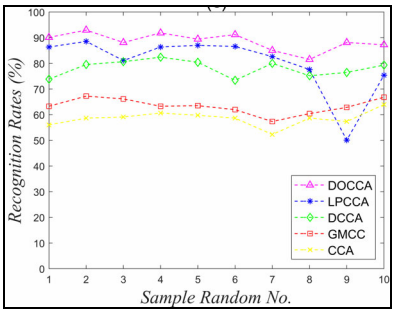
Figure 2 The recognition rate of DOCCA, LPCCA, DCCA, GMCC and CCA on the Umist database for facial image features when the number of training samples n are 4, 5, 6, and 7, respectively (see online version for colours)



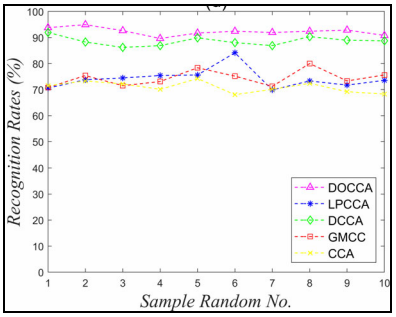
(a)



(b)



(c)



(d)

Table 2 Experimental results on Umist image database

	4Train	*5Train*	*6Train*	*7Train*
DOCCA	82.79 ± 3.08	84.65 ± 2.05	88.57 ± 3.40	92.32 ± 1.47
DCCA	77.56 ± 3.73	75.98 ± 10.81	80.18 ± 11.43	74.25 ± 3.97
LPCCA	59.43 ± 3.48	31.54 ± 7.36	78.13 ± 3.15	88.62 ± 1.76
GMCC	57.15 ± 2.39	28.11 ± 6.60	63.30 ± 3.01	74.46 ± 3.06
CCA	57.31 ± 3.45	54.76 ± 3.51	58.53 ± 3.02	70.94 ± 2.12

Notes: A ± B: A represents the average recognition rate and B represents the corresponding standard deviation of the recognition rate.

Figure 2 shows the random ten changes in the recognition rate of DOCCA, LPCCA, DCCA, GMCC and CCA under different training samples in the Umist database. As can be seen from Figure 2, DOCCA still has the best recognition performance. In ten random experiments with different training samples, LPCCA shows great instability, which is largely due to the redundant information. The overall recognition rate of DCCA and GMCC also drops with the increase of the number of training samples, which lacks robustness. It can be observed from Figure 2 that with the increase of training samples, the overall recognition performance of DOCCA is increasing, and the performance of ten random times is more stable. Our DOCCA method has less variation range and better robustness. These further shows that the embedding of class information can effectively improve the recognition rates of the low number of training samples.

4.3 Experiments on the Yale database

The Yale face database contains 15 persons who have 11 facial images respectively, so totally there are 165 images. These images are derived from different angle of view, having change of expression and illumination. In Table 3, CCA shows poor recognition rates, the overall recognition rates of LPCCA and GMCC are low and have not grown significantly with the increase of training samples, while our proposed DOCCA method still maintains the highest recognition rate. More specifically, when the number of training samples of CCA, DCCA, GMCC and LPCCA reaches 8, the overall recognition rates decrease. In addition to the feature information in the sample data, the inevitable redundant information will affect the recognition performance. We attenuate the influence of this redundant information by embedding orthogonal structures. The recognition rate of DOCCA changes smoothly with the increase of training samples, and the robustness advantage brought by orthogonal constraints is more obvious.

The above experiments show that our DOCCA method has better recognition performance and DOCCA method is insensitive to changes in the number of training samples. This further indicates that the addition of orthogonality has better discrimination and robustness in the case of less training samples.

Table 3 Experimental results on Yale image database

	4Train	*5Train*	*6Train*	*7Train*
DOCCA	57.11 ± 3.29	58.27 ± 4.58	60.10 ± 2.66	60.34 ± 3.00
DCCA	49.01 ± 5.61	53.00 ± 5.91	57.17 ± 5.13	56.14 ± 4.86
LPCCA	48.60 ± 7.58	49.82 ± 5.71	46.16 ± 7.66	40.23 ± 6.99
GMCC	51.40 ± 3.61	53.00 ± 3.59	50.00 ± 3.54	48.64 ± 2.97
CCA	38.76 ± 3.98	39.82 ± 5.05	38.48 ± 2.99	33.64 ± 6.39

Notes: A ± B: A represents the average recognition rate and B represents the corresponding standard deviation of the recognition rate.

5 Conclusions

As a classical cross-modal feature extraction method, classical CCA integrated the relevant features of multimodal data and effectively improves discrimination. But classical CCA did not consider the class label information of the samples and could not find the discrimination information embedded in the data samples. On the other hand, the features obtained by non-orthogonal projection contained a lot of redundant information. The remaining information was significantly affected by the number of samples and the number of dimensions.

For this reason, we proposed cross-modal correlation feature extraction with orthogonality redundancy reduction and discriminant structure constraint method, called DOCCA. This method added class label information to the objective function and constructs orthogonal constraints at the same time, embedding it into the relevant theory of classical CCA, and then obtained the optimisation model of DOCCA. Orthogonality could ensure that the extracted features are fixed as distinct as possible and more discriminative. The targeted experiments were designed on GT image database, Umist image database and Yale image database. Satisfactory experimental results showed that DOCCA is an effective feature extraction method. However, our modalities required pairing. When there are modalities losses, how to effectively use semi pairing will be our next research direction.

References

Al-Sharman, M.K., Zweiri, Y., Jaradat, M.A.K. et al. (2019) 'Deep-learning-based neural network training for state estimation enhancement: application to attitude estimation', *IEEE Transactions on Instrumentation and Measurement*, Vol. 69, No. 1, pp.24–34.

Andrew, G., Arora, R., Bilmes, J. et al. (2013) 'Deep canonical correlation analysis', *International Conference on Machine Learning*, PMLR, pp.1247–1255.

Bao, Z., Hu, J., Pan, G. et al. (2019) 'Canonical correlation coefficients of high-dimensional Gaussian vectors: finite rank case', *The Annals of Statistics*, Vol. 47, No. 1, pp.612–640.

Bhowmik, B., Tripura, T., Hazra, B. et al. (2020) 'Real time structural modal identification using recursive canonical correlation analysis and application towards online structural damage detection', *Journal of Sound and Vibration*, Vol. 468, p.115101.

Cai, D., He, X., Han, J. et al. (2006) 'Orthogonal Laplacian faces for face recognition', *IEEE Transactions on Image Processing*, Vol. 15, No. 11, pp.3608–3614.

Chellappan, R., Satheeskumaran, S., Venkatesan, C. et al. (2021) 'Discrete stationary wavelet transform and SVD-based digital image watermarking for improved security', *International Journal of Computational Science and Engineering*, Vol. 24, No. 4, pp.354–362.

Chen, J., Wang, G. and Giannakis, G.B. (2019) 'Graph multiview canonical correlation analysis', *IEEE Transactions on Signal Processing*, Vol. 67, No. 11, pp.2826–2838.

Chen, J., Wang, G., Shen, Y. et al. (2018) 'Canonical correlation analysis of datasets with a common source graph', *IEEE Transactions on Signal Processing*, Vol. 66, No. 16, pp.4398–4408.

Elmadany, N.E.D., He, Y. and Guan, L. (2016) 'Multiview learning via deep discriminative canonical correlation analysis', *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp.2409–2413.

Hu, L. and Zhang, W. (2020) 'Orthogonal neighborhood preserving discriminant analysis with patch embedding for face recognition', *Pattern Recognition*, Vol. 106, p.107450.

Jiang, J., Ma, J., Chen, C. et al. (2018) 'SuperPCA: a superpixelwise PCA approach for unsupervised feature extraction of hyperspectral imagery', *IEEE Transactions on Geoscience and Remote Sensing*, Vol. 56, No. 8, pp.4581–4593.

Jiang, Y., Liang, W., Tang, J. et al. (2021) 'A novel data representation framework based on nonnegative manifold regularisation', *Connection Science*, Vol. 33, No. 2, pp.136–152.

Kheyrinataj, F. and Nazemi, A. (2020) 'Fractional power series neural network for solving delay fractional optimal control problems', *Connection Science*, Vol. 32, No. 1, pp.53–80.

Kokiopoulou, E. and Saad, Y. (2007) 'Orthogonal neighborhood preserving projections: a projection-based dimensionality reduction technique', *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 29, No. 12, pp.2143–2156.

Lisanti, G., Masi, I. and Del Bimbo, A. (2014) 'Matching people across camera views using kernel canonical correlation analysis', *Proceedings of the International Conference on Distributed Smart Cameras*, pp.1–6.

Peng, Y., Zhang, D. and Zhang, J. (2010) 'A new canonical correlation analysis algorithm with local discrimination', *Neural Process. Lett.*, Vol. 31, No. 1, pp.1–15.

Qian, J., Zhao, R., Wei, J. et al. (2020) 'Feature extraction method based on point pair hierarchical clustering', *Connection Science*, Vol. 32, No. 3, pp.223–238.

Roweis, S.T. and Saul, L.K. (2000) 'Nonlinear dimensionality reduction by locally linear embedding', *Science*, Vol. 290, No. 5500, pp.2323–2326.

Samat, A., Persello, C., Gamba, P. et al. (2017) 'Supervised and semi-supervised multi-view canonical correlation analysis ensemble for heterogeneous domain adaptation in remote sensing image classification', *Remote Sensing*, Vol. 9, No. 4, p.337.

Shen, X.B., Sun, Q.S. and Yuan, Y.H. (2013) 'Orthogonal canonical correlation analysis and its application in feature fusion', *Proceedings of the 16th International Conference on Information Fusion*, pp.151–157.

Shu, H., Wang, X. and Zhu, H. (2020) 'D-CCA: a decomposition-based canonical correlation analysis for high-dimensional datasets', *Journal of the American Statistical Association*, Vol. 115, No. 529, pp.292–306.

Srivastava, V. and Biswas, B. (2020) 'CNN-based salient features in HSI image semantic target prediction', *Connection Science*, Vol. 32, No. 2, pp.113–131.

Sun, Q.S., Zeng, S.G., Liu, Y. et al. (2005) 'A new method of feature fusion and its application in image recognition', *Pattern Recognition*, Vol. 38, No. 12, pp.2437–2448.

Sun, T. and Chen, S. (2007) 'Locality preserving CCA with applications to data visualization and pose estimation', *Image and Vision Computing*, Vol. 25, No. 5, pp.531–543.

Sun, T., Chen, S., Yang, J. et al. (2008) 'A supervised combined feature extraction method for recognition', *Proceedings of the IEEE International Conference on Data Mining*, Pisa, Italy, pp.1043–1048.

Wang, F. and Zhang D. (2013) 'A new locality-preserving canonical correlation analysis algorithm for multi-view dimensionality reduction', *Neural Processing Letters*, Vol. 37, No. 2, pp.135–146.

Wang, H.T., Smallwood, J., Mourao-Miranda, J. et al. (2020) 'Finding the needle in a high-dimensional haystack: canonical correlation analysis for neuroscientists', *NeuroImage*, Vol. 216, 116745.

Waqas, U.A., Khan, M. and Batool, S.I. (2020) 'A new watermarking scheme based on Daubechies wavelet and chaotic map for quick response code images', *Multimedia Tools and Applications*, Vol. 79, No. 9, pp.6891–6914.

Wu, Z., Gao, Y., Li, L. et al. (2019) 'Semantic segmentation of high-resolution remote sensing images using fully convolutional network with adaptive threshold', *Connection Science*, Vol. 31, No. 2, pp.169–184.

Xu, H., Lin, T., Xie, Y. et al. (2018) 'Enriching the random subspace method with margin theory – a solution for the high-dimensional classification task', *Connection Science*, Vol. 30, No. 4, pp.409–424.

Yang, J., Yang, J., Zhang, D. et al. (2003) 'Feature fusion: parallel strategy vs. serial strategy', *Pattern Recognition*, Vol. 36, No. 6, pp.1369–1381.

Ying, L., Qiqi, L., Jiulun, F. et al. (2021a) 'Tyre pattern image retrieval – current status and challenges', *Connection Science*, Vol. 33, No. 2, pp.237–255.

Ying, L., Qian Nan, Z., Fu Ping, W. et al. (2021b) 'Adaptive weights learning in CNN feature fusion for crime scene investigation image classification', *Connection Science*, Vol. 33, No. 3, pp.1–16.

Yang, X., Weifeng, L., Liu, W. et al. (2021c) 'A survey on canonical correlation analysis', *IEEE Transactions on Knowledge and Data Engineering*, Vol. 33, No. 6, pp.2349–2368.

Yu, Q., Xu, H. and Liao, S. (2018) 'Coiflets solutions for Föppl-von Kármán equations governing large deflection of a thin flat plate by a novel wavelet-homotopy approach', *Numerical Algorithms*, Vol. 79, No. 4, pp.993–1020.

Yuan, Y.H. and Sun, Q.S. (2020) 'Graph regularized multiset canonical correlations with applications to joint feature extraction', *Pattern Recognition*, Vol. 47, No. 12, pp.3907–3919.

Zhang, G., Zou, W., Zhang, X. et al. (2018) 'Singular value decomposition based virtual representation for face recognition', *Multimedia Tools and Applications*, Vol. 77, No. 6, pp.7171–7186.

Zhang, L., Wang, L., Bai, Z. et al. (2020) 'A self-consistent-field iteration for orthogonal CCA', *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Zhang, S., Li, X., Zong, M. et al. (2017) 'Efficient kNN classification with different numbers of nearest neighbors', *IEEE Transactions on Neural Networks and Learning Systems*, Vol. 29, No. 5, pp.1774–1785.

Zheng, W. (2016) 'Multichannel EEG-based emotion recognition via group sparse canonical correlation analysis', *IEEE Transactions on Cognitive and Developmental Systems*, Vol. 9, No. 3, pp.281–290.