Exponential Multi-Modal Discriminant Feature Fusion for Small Sample Size

YANMIN ZHU\(^1\), TIANHAO PENG\(^1\), AND SHUZHI SU\(^2\)

\(^1\)School of Mechanical Engineering, Anhui University of Science and Technology, Huainan, Anhui 232001, China
\(^2\)School of Computer Science and Engineering, Anhui University of Science and Technology, Huainan, Anhui 232001, China

Corresponding author: Tianhao Peng (thp1909@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61806006 and Grant 51475001, and in part by the China Postdoctoral Science Foundation under Grant 2019M660149.

ABSTRACT Multi-modal Canonical Correlation Analysis (MCCA) is an important information fusion method, and some discriminant variations of MCCA have been proposed. However, the variations suffer from the Small Sample Size (SSS) problem and the absence of cross-modal discriminant scatters. Thus we propose a novel exponential multi-modal discriminant feature fusion method for a small amount of training samples, i.e. exponential multi-modal discriminant correlation analysis. In the method, we construct a discriminative integration scatter of all the modalities by constraining the aggregation towards cross-modal discriminative centroids. Besides, the method gives a decomposition-based matrix exponential strategy. The strategy can solve the SSS problem and improve the robustness of noises, and we further provide corresponding theoretical proofs and some intuitive analysis. The method can learn correlation fusion features with well discriminative power from a small amount of samples. Encouraging experimental results show the effectiveness and robustness of our method.

INDEX TERMS Canonical correlation analysis, feature fusion, discriminant subspace, matrix decomposition, structure integration.

I. INTRODUCTION

Data collected in real-world applications is divided into single-modal and multi-modal data depending on the number of data representations corresponding to the same target \([1]\). There is only one type of data representations for one single target, and this data is single-modal data. Unlike single-modal data, multi-modal data has several categories of data representations correlating to one same target. For example, we can gather face data, fingerprint data, gait data, and so on, which can simultaneously characterize the person from different viewpoints. These data can be viewed as the person’s multi-modal data. As compared to single-modal data, multi-modal data contains more information on a single target. Multi-modal fusion is divided into three different fusion levels \([2]\): pixel-level fusion, feature-level fusion, and decision-level fusion. In the three-level fusions, pixel-level fusion \([3]\) is located at the lowest level, which can retain the original information to the greatest extent. Pixel-level fusion is generally divided into four steps: preprocessing, transformation, synthesis, and inverse transformation. Decision-level fusion \([4]\) mainly focuses on the integration of multiple classifiers. Although feature-level fusion \([5]\) starts later than other fusion methods, its advantages are obvious. Feature level fusion not only extracts the discriminant information of each modality but also reduces the complementary information among different modalities as much as possible in learned fusion subspaces, which is crucial for classification tasks. Our research of this paper also focuses on feature-level fusion.

In all multi-modal feature fusion methods, Canonical Correlation Analysis (CCA) \([6]\) plays an important role. CCA aims at learning a correlation fusion subspace of two modalities. In the subspace, the between-modal correlation is maximal. CCA has been used in several real-world applications, including fault detection \([7]\), multi-modal biometric analysis \([8]\), and process monitoring \([9]\). CCA has some limitations as an unsupervised linear feature fusion method. To preserve the raw data’s geometry structure information, local CCA \([10]\) discovers the hidden local common information by a novel metric and develops a data-driven correlation model based on the metric. Gao et al. \([11]\) developed locality preserving CCA of two-dimension matrix data, which
can capture intrinsic structures with the help of local neighbor information. Moreover, class information is important discriminant information that can significantly improve the classification performance of many feature fusion methods. To embed class labels into correlation analysis theories, generalized CCA [12] optimizes the within-modal discriminative information when the across-modal correlation is maximized. Inspired by the regularization technique and multi-modal discriminant analysis [13], Han et al. [14] proposed a supervised correlation analysis by considering the discrimination information and correlation information of paired data. Local discriminant CCA [15] can extract nonlinear features with the discriminative structures of neighbor graphs. Similarly, supervised local-preserving CCA [16], locality preserving randomized CCA [17], local discrimination CCA [18], and hypergraph CCA [19] exploit local structure information and class label information under the correlation analysis framework, and the effectiveness of these methods has been demonstrated in image recognition tasks.

The above correlation-based methods only deal with two-modal data, but the number of modalities is usually more than two in many real-world applications. With the development of correlation analysis theories, Multi-modal CCA (MCCA) [20] appears, and the method can be also called multi-set CCA, multi-view CCA, and multi-way CCA in different researches. MCCA is a generalized modality extension of CCA, and CCA can be treated as an exceptional example of MCCA. Thus, MCCA possesses a broader range of implementations and improved theoretical benefits. Up to now, research subjects of many scholars have changed from CCA-related methods and MCCA-related methods. Discriminative Multiple CCA (DMCCA) [21] maximizes intra-class correlations and reduces inter-class correlations of various modalities to improve the discriminative power of correlation fusion features. Graph Regularized Multi-set Canonical Correlations (GrMCC) [22] builds an object optimization function with within-modal discriminative scatters using the graph regularization technique, allowing learned correlation fusion features to maintain discriminative geometry structures. Yuan et al. [23] suggested a Laplacian Multi-view Canonical Correlation Analysis (LapMCCA) method that can learn nonlinear correlation fusion features by using weight constraints of nearest neighbor graphs. Shen et al. [24] devised a generalized graph embedding CCA framework that can embed various graphs into multi-modal correlation theories, and three graph embedding correlation analysis methods were also given by means of three different graphs. To enhance the discriminative power, Hessian MCCA [25] employs the Hessian technique to capture intrinsic locality geometry structures concealed in raw multi-modal data.

Multi-modal correlation theories with supervised information are important and hot research subjects in multi-modal feature fusion. Although some supervised MCCA-related methods have been proposed, the existing methods ignore between-modal discriminative scatter information that can significantly improve the class separability of low-dimensional fusion features, especially those that are easily confused. Furthermore, the existing supervised MCCA-related methods still suffer from the well-known Small Sample Size (SSS) problem [26]. The SSS problem is that the sample dimensionality is more than the number of samples, which will lead to singular matrices. The singularity of matrices is usually fatal to the optimization of feature fusion methods. To solve these issues, we propose a novel Exponential Multi-modal Discriminant Correlation Analysis (EMDCA) method. In the method, we construct a discriminative integration scatter of all the modalities, and the scatter further integrates cross-modal discriminant scatter information on the basis of within-modal discriminant scatter information, which can better reveal the global discriminant structure of all the modalities. Besides, the method gives a decomposition-based matrix exponential strategy to elegantly address the SSS problem and improve the robustness of noises, and we further provide corresponding theoretical proofs and some intuitive analysis. To our best knowledge, the method is the first exponential-based method for solving the SSS problem of multi-modal correlation theories. By minimizing the exponential discriminative integration scatter and simultaneously maximizing exponential between-modal correlations, the method can obtain correlation fusion features with well discriminative power from a small number of raw data. We design extensive experiments on five visible image datasets, two infrared image datasets, and Reuters multilingual text dataset. Promising findings on the experiments indicate the effectiveness and robustness of our method.

II. BRIEF REVIEW OF MCCA

Suppose that \((Z^{(p)}) = \{z_{1}^{(p)}, z_{2}^{(p)}, \ldots, z_{N}^{(p)}\} \in \mathbb{R}^{d_{p} \times N}_{1=1}^{M}\) are the \(M\) modality datasets and \(z_{k}^{(p)} \in \mathbb{R}^{d_{p}}\) \((k = 1, 2, \ldots, N)\) represents the \(k\)th samples of the \(p\)th modality dataset \(Z^{(p)}\). MCCA belongs to a multi-modal feature fusion method that aims at finding a correlation projection direction \(\phi^{(p)} \in \mathbb{R}^{d_{p} \times 1}\) correlating to \(Z^{(p)}\) \((p = 1, 2, \ldots, M)\). In more detail, MCCA’s optimization function is as follows:

\[
\max_{\phi^{(p)}} \frac{\sum_{p=1}^{M} \sum_{q=1}^{M} \phi^{(p)T} C^{(pq)} \phi^{(q)}}{\sum_{p=1}^{M} \sum_{q=1}^{M} \phi^{(p)T} C^{(pq)} \phi^{(q)}}
\]

where \(C^{(pq)} = \frac{1}{N} \sum_{k=1}^{N} (z_{k}^{(p)} - \bar{z}^{(p)}) (z_{k}^{(q)} - \bar{z}^{(q)})^{T}\) denotes the covariance matrix of \(Z^{(p)}\) and \(Z^{(q)}\) \((p, q = 1, 2, \ldots, M)\), and \(\bar{z}^{(p)} = \frac{1}{N} \sum_{k=1}^{N} z_{k}^{(p)}\) indicates the sample mean of \(Z^{(p)}\). Besides, \(C^{(pp)} = \frac{1}{N} \sum_{k=1}^{N} (z_{k}^{(p)} - \bar{z}^{(p)}) (z_{k}^{(p)} - \bar{z}^{(p)})^{T}\) is the variance matrix of \(Z^{(p)}\) \((p = 1, 2, \ldots, M)\). For the convenience of description, we utilize the above representations for the same conception in the whole paper. As in [27], the optimization function of Eq. (1) can be transformed into the following constraint-based optimization problem:

\[
\max_{\phi^{(p)}} \sum_{p=1}^{M} \sum_{q=1}^{M} \phi^{(p)T} C^{(pq)} \phi^{(q)}
\]
employ the centroid \( \bar{z}_g \) to construct the discriminative integration scatter \( \rho_w \) of all the modalities in the subspace:

\[
\rho_w = \sum_{p=1}^{M} \left( \frac{1}{N} \sum_{g=1}^{G} \sum_{k=1}^{N_g} \left( \frac{\psi(p)^T \bar{z}_g}{z_{gk}} - \bar{z}_g \right) \right) \times \left( \frac{\psi(p)^T \bar{z}_g}{z_{gk}} - \bar{z}_g \right)^T
\] (4)

The physical meaning of the scatter is the Euclidean distance metric between the correlation features and their centroids. Although the scatter is a simple form of within-modal scatters, the centroids of the scatter are composed of the intra-class samples from all the modalities. Thus, the scatter also includes the intra-class samples from different modalities on the basis of the within-modal intra-class samples. However, since the centroid \( \bar{z}_g \) is based on the correlation projection directions of all the modalities, the optimization problem with Eq. (4) is difficult to solve the optimal projection directions under correlation analysis theories. Thus, the scatter in Eq. (4) is necessary to implement some equivalent derivations. For the convenience of derivations, we define the class mean \( z_{gk}^{(p)} \) in the \( g \)th class of \( Z(p) \), i.e. \( z_{g}^{(p)} = (1/N_g) \sum_{k=1}^{N_g} z_{gk}^{(p)} \). Then the equivalent derivation of Eq. (4) can be as follows:

\[
\rho_w = \sum_{p=1}^{M} \left( \frac{1}{N} \sum_{g=1}^{G} \sum_{k=1}^{N_g} \left( \frac{\psi(p)^T \bar{z}_g}{z_{gk}} - \bar{z}_g \right) \right) \times \left( \frac{\psi(p)^T \bar{z}_g}{z_{gk}} - \bar{z}_g \right)^T
\] (4)

\[
= \frac{1}{N} \sum_{p=1}^{M} \sum_{g=1}^{G} \sum_{k=1}^{N_g} \psi(p)^T \bar{z}_g \bar{z}_g^T \psi(p)
\]

III. EXPONENTIAL MULTI-MODAL DISCRIMINANT CORRELATION ANALYSIS

A. MOTIVATION

Existing supervised correlation-based methods focus on constraining within-modal discriminant scatters under the correlation analysis framework. These methods have two issues: ignoring cross-modal discriminant scatters and suffering from the SSS problem. Therefore, we try to construct a discriminative integration scatter of all the modalities by enhancing the aggregation towards cross-modal discriminative centroids of all the correlation features. However, it is challenging to compute the cross-modal discriminative centroids because the sample dimensionalities of different modalities can be different. For this issue, we follow the idea of subspace feature fusion to reverse thinking. By projecting high-dimensional data into a correlation fusion subspace, correlation features are obtained, and the correlation features from different modalities have a beneficial property, i.e. the same dimensionality. Thus, in the correlation fusion subspace, we can first construct cross-modal discriminative centroids of all modalities and then use specific derivations to obtain an implicit mathematical representation of the centroids in the high-dimensional sample space. Then the discriminative integration scatter will be constructed using the centroids. Besides, the scatter also usually suffers from the SSS problem, and thus we develop a decomposition-based matrix exponential strategy. The strategy can solve the SSS problem, which can be proved in theory.

B. FORMULATION OF EMDCA

Suppose that \( z_{gk}^{(p)} \) denotes the \( k \)th sample of the \( g \)th \((g = 1, 2, \ldots, G)\) class in the \( p \)th modality dataset \( Z(p) \), and \( N_g \) represents the sample number in the \( g \)th class of each modality dataset. Besides \( \psi(p)^T z_{gk}^{(l)} \) is the corresponding correlation feature of \( z_{gk}^{(l)} \) in a discriminative integration correlation fusion subspace. We firstly construct a cross-modal discriminative centroid \( \bar{z}_g \) of the \( g \)th class:

\[
\bar{z}_g = \frac{1}{NM_g} \sum_{p=1}^{M} \sum_{k=1}^{N_g} \psi(p)^T z_{gk}^{(p)}
\] (3)

The centroid is the mean of all the samples from the \( g \)th class in the subspace, and the discriminative information of all the modalities is also integrated in the subspace. Motivated by the global modality idea [13], [14], we try to enhance the aggregation between the correlation features and their corresponding cross-modal discriminative centroids. Thus we
between-modal samples. Similar to the physical meaning [29] of minimizing the within-modal intra-class scatter, the minimum of the discriminative integration scatter can enhance the aggregation towards the cross-modal neighbor-consistent centroids, and the aggregation further constrains the consistency of cross-modal discriminative structures, which will reveal better the intrinsic neighbor relationships. Then, by minimizing the cross-modal discriminative scatters and meanwhile maximizing the pair-wise correlations of all the modalities, we construct the objective optimization problem as follows:

$$\max \varphi^T R \varphi$$

$$\text{s.t. } \varphi^T H \varphi = 1$$

(6)

where $\varphi = [\varphi^{(1)}; \ldots; \varphi^{(M)}]$, $R = [C^{(1)} \ldots C^{(M)}; \ldots; C^{(M)} C^{(M)}]$, and $H = [H_w^{(11)} \ldots H_w^{(M)}; \ldots; H_w^{(M1)} \ldots H_w^{(MM)}]$. $H$ is called the discriminative integration matrix of all the modalities, and $R$ is named as the correlation matrix.

In Eq. (6), the discriminative integration matrix $H$ cannot avoid the SSS problem, especially for high-dimensional samples, which will usually be fatal to the optimization of many feature fusion methods. Thus, we exploit the decomposition-based matrix exponential strategy to solve the SSS problem. More concretely, we decompose the matrix $H$ by eigenvalue decomposition [30]:

$$H = \mathcal{U} \Delta \mathcal{U}^T$$

(7)

where $\mathcal{U} = [u_1, u_2, \ldots, u_d]$ is the eigenvector matrix of $H$, and $\Delta$ represents the diagonal matrix of eigenvalues, i.e. $\Delta = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_d)$. $\hat{d}$ is equal to $\sum_{i=1}^M d_i$. Inspired by the eigenvalue correction idea [31], [32], we construct an exponential discriminative integration matrix $\tilde{H}$ as follows:

$$\tilde{H} = \mathcal{U} \exp(\Delta) \mathcal{U}^T$$

(8)

where $\exp(\Delta)$ represents the exponential eigenvalue matrix. Besides, $\exp(\Delta)$ is defined as follows:

$$\exp(\Delta) = I + \Delta + \frac{\Delta^2}{2!} + \frac{\Delta^3}{3!} + \ldots + \frac{\Delta^s}{s!} + \ldots$$

(9)

where $I$ is the identity matrix, $2!$ denotes the factorial of 2, and $\Delta^2$ is the second power of $\Delta$, i.e. $\Delta^2 = \Delta \Delta$. Concretely, the exponential of matrices will be computed by the scaling and squaring method [33], [34]. Similar to Eq. (8), the exponential correlation matrix $\tilde{R}$ of $R$ is constructed as follows:

$$\tilde{R} = \mathcal{V} \exp(\wedge) \mathcal{V}^T$$

(10)

where $\mathcal{V}$ and $\wedge$ are respectively the eigenvector matrix and eigenvalue matrix of the correlation matrix $R$, and $\exp(\wedge)$ has the same definition with $\exp(\Delta)$ of Eq.(9). As a result, the EMDCA objective optimization problem is represented as follows:

$$\max_{\varphi} \varphi^T \mathcal{U} \exp(\wedge) \mathcal{U}^T \varphi$$

$$\text{s.t. } \varphi^T \mathcal{U} \exp(\Delta) \mathcal{U}^T \varphi = 1$$

(11)

When the sample dimensionality is larger than the sample number, the constraint matrices will singular in many optimization problems. This SSS problem is avoided in the objective optimization problem of Eq. (11). For theoretical proofs in avoiding this SSS problem, we define the matrix exponential of the discriminative integration matrix $H$:

$$\exp(H) = I + H + \frac{H^2}{2!} + \frac{H^3}{3!} + \ldots + \frac{H^s}{s!} + \ldots$$

(12)

The matrix exponential of the correlation matrix $R$ is as follows:

$$\exp(C) = I + C + \frac{C^2}{2!} + \frac{C^3}{3!} + \ldots + \frac{C^s}{s!} + \ldots$$

(13)

Theorem 1: $\exp(H)$ is equal to $\mathcal{U} \exp(\Delta) \mathcal{U}^T$, and $\exp(R) = \mathcal{V} \exp(\wedge) \mathcal{V}^T$ is right.

Proof: $u_k (k = 1, \ldots, d)$ is the eigenvector of $H$, and $\lambda_k$ is the eigenvalue of $H$. According to properties of eigenvalue decomposition, we obtain the following equations:

$$I u_k = \lambda_k u_k$$

$$H u_k = \lambda_k u_k$$

$$\frac{H^2}{2!} u_k = \frac{\lambda_k^2}{2!} u_k$$

$$\vdots$$

$$\frac{H^s}{s!} u_k = \frac{\lambda_k^s}{s!} u_k$$

By summing the above equations, we can obtain

$$\left( I + H + \frac{H^2}{2!} + \frac{H^3}{3!} + \ldots + \frac{H^s}{s!} + \ldots \right) u_k = (1 + \lambda_k + \frac{\lambda_k^2}{2!} + \frac{\lambda_k^3}{3!} + \ldots + \frac{\lambda_k^s}{s!} + \ldots) u_k$$

(14)

Based on the definition of the matrix exponential, Eq. (14) can be reformulated into

$$\exp(H) u_k = \exp(\lambda_k) u_k$$

(15)

From Eq. (14), we can find that $H$ and $\exp(H)$ have the same eigenvector matrix and $\exp(\lambda_k)$ is the eigenvalue of $\exp(H)$. Therefore, we can obtain

$$\exp(H) = \mathcal{U} \exp(\Delta) \mathcal{U}^T$$

(16)

Besides, the proof in $\exp(R) = \mathcal{V} \exp(\wedge) \mathcal{V}^T$ is the same with the above proof. Thus, the theorem is turned out to be correct.

Property 1: On any A and B matrices of the same size, $\exp(A + B) = \exp(A) + \exp(B)$.
Theorem 2: \(\exp(H)\) must have an inverse matrix (i.e., \(\exp(H)^{-1}\)), and \(\exp(H)^{-1} = \exp(-H)\).

Proof: Since the matrix exponential of the zero matrix \(O\) is the identity matrix \(I\) (i.e., \(\exp(O) = I\)), the following equation is correct:

\[
\exp(H - H) = I
\]

(17)

Due to Property 1, we can obtain

\[
\exp(H)\exp(-H) = I
\]

(18)

Thus \(\exp(-H)\) is the inverse matrix of \(\exp(H)\), and \(\exp(H)\) must be a non-singular matrix. The theorem is proved.

According to Theorem 1, the objective optimization problem of Eq. (11) can be reformulated into

\[
\max \varphi^T \exp(R)\varphi \quad \text{s.t.} \quad \varphi^T \exp(H)\varphi = 1
\]

(19)

Theorem 2 has proved the non-singularity of \(\exp(H)\), which can ensure the evasion of the SSS problem in theory. Additionally, the computation of inverse matrices is essential in the optimization of Eq. (19), and the time cost in computing inverse matrices is usually very high. However, Theorem 2 also proves that the inverse matrix of \(\exp(H)\) is equal to \(\exp(-H)\), which will significantly reduce the time cost in computing its inverse matrix.

C. OPTIMIZATION

To obtain the analytical solutions of Eq. (14), we first construct the Lagrange multiplier function:

\[
\mathcal{L} (\varphi) = \varphi^T \exp(R)\varphi - \eta \varphi^T \exp(H)\varphi - 1
\]

(20)

where \(\eta\) is a Lagrange multiplier. Then the partial derivative of \(\mathcal{L} (\varphi)\) to \(\varphi\) is set as zero, and we can obtain

\[
\exp(R)\varphi = \eta \exp(H)\varphi
\]

(21)

From Theorem 2, we know that the inverse matrix of \(\exp(H)\) must exist and \(\exp(H)^{-1} = \exp(-H)\). By multiplying \(\exp(H)^{-1}\) to the left of Eq. (21), we can derive

\[
\exp(-H)\exp(R)\varphi = \eta \varphi
\]

(22)

Based on Property 1, we can obtain

\[
\exp(R - H)\varphi = \eta \varphi
\]

(23)

Theorem 3: The Lagrange multiplier \(\eta\) is equal to the objective function of (19), i.e., \(\eta = \varphi^T \exp(R)\varphi\).

Proof: By multiplying \(\varphi^T\) to the left of Eq. (21), we can obtain

\[
\varphi^T \exp(R)\varphi = \eta \varphi^T \exp(H)\varphi
\]

(24)

According to Eq. (19), it can be observed that \(\varphi^T \exp(H)\varphi\) is equal to one. Then Eq. (24) can be equivalently translated to

\[
\varphi^T \exp(R)\varphi = \eta
\]

(25)

Thus the theorem is proved.

Eq. (23) is an eigenvalue decomposition problem, and \(\eta\) can also be named the eigenvalue. The optimization problem in Eq. (19) is reformulated into the computation problem of the eigenvector corresponding to the largest eigenvalue in Eq. (19) using Theorem 3 and the above optimization derivations. Thus the optimal solutions of correlation projection directions are the eigenvectors of \(\exp(R - H)\) corresponding to the \(d\) largest eigenvalues, i.e., \(\eta_k = [\eta_k^1, \eta_k^2, \ldots, \eta_k^M]^T\). Then we can create the correlation direction matrices \(\Phi(p) = [\eta_1^{(p)}, \eta_2^{(p)}, \ldots, \eta_d^{(p)}] \in \mathbb{R}^{d \times d}\) corresponding to \(Z(p) \in \mathbb{R}^{k \times N}^M\) \(\eta = 1\), and the correlation features of \(Z(p) \in \mathbb{R}^{d \times N}^M\) \(\eta = 1\) can be directly computed by the form of \(\Phi(p)^T Z(p) \in \mathbb{R}^{d \times d}\).

For the samples \([z_k^{(p)}]_{p=1}^M(k = 1, 2, \ldots, N)\), they correspond to the same target and their correlation features \([\eta_k]_{p=1}^M\) are necessary to be fused in the final recognition stage. Based on the simple fusion strategy [35], [36], the correlation fusion feature \(y_k\) can be obtain

\[
y_k = \sum_{p=1}^M \Phi(p)^T z_k^{(p)}
\]

(26)

where \(y_k\) belongs to the correlation fusion feature training set \(Y = [y_1, y_2, \ldots, y_N] \in \mathbb{R}^d \times N\). For the testing sample sets \([Z(p) = [z_1^{(p)}, z_2^{(p)}, \ldots, z_N^{(p)}] \in \mathbb{R}^{d \times N}^M\) \(\eta = 1\), the \(M\) testing samples \([z_k^{(p)}]_{p=1}^M\) also correspond to the same target. \(N\) denotes their sample number. By means of the correlation projection matrices \(\{\Phi(p)^T\}^M_{p=1}\), we can directly get the correlation features \(\Phi(p)^T Z(p) \in \mathbb{R}^{d \times d}\) \(\eta = 1\), and then the correlation fusion feature testing set \(\tilde{Y} = [\tilde{y}_1, \tilde{y}_2, \ldots, \tilde{y}_N]\) can be also computed as follows:

\[
\tilde{y}_k = \sum_{p=1}^M \Phi(p)^T \tilde{z}_k^{(p)}
\]

(27)

where \(\tilde{y}_k\) is the \(k\) th correlation fusion features of \(\tilde{Y}\). Finally, the nearest neighbor classifier will be utilized in the correlation fusion features, and the recognition results of the corresponding raw high-dimensional multi-modal data. Besides, we give detailed descriptions of the methods steps in Algorithm 1, and the flowchart of EMDCA is also exhibited in Fig. 1.

D. ANALYSIS OF EMDCA

Our method considers the within-modal and between-modal discriminative structures by constraining the discriminative integration scatter of all the modalities. To further understand the discriminative integration scatter of Eq. (5), we equivalently reformulate it into the below form:

\[
\rho_w = \sum_{g=1}^G \left( \frac{1}{2MN_g} \sum_{p=1}^M \sum_{k=1}^N \sum_{l=1}^N \Phi(p)^T \left( Z_{gk}^{(p)} - \tilde{Z}_{gk}^{(p)} \right) \left( Z_{gl}^{(p)} - \tilde{Z}_{gl}^{(p)} \right)^T \Phi(p) \right)
\]

\[
\times \left( \tilde{z}_{gk}^{(p)} - \tilde{z}_{gl}^{(p)} \right)
\]
Figure 1. The flowchart of EMDCA.

Algorithm 1: The EMDCA Method

Input: The training dataset \( \{Z^{(p)}\} = \{z_1^{(p)}, z_2^{(p)}, \ldots, z_N^{(p)}\} \in R^{d_p \times N} \) 

The testing dataset \( \{Z^{(p)}\} = \{\tilde{z}_1^{(p)}, \tilde{z}_2^{(p)}, \ldots, \tilde{z}_N^{(p)}\} \in R^{d_p \times N} \)

Output: The correlation fusion feature training set \( Y = [y_1, y_2, \ldots, y_N] \in R^{d \times N} \) 

The correlation fusion feature testing set \( \tilde{Y} = [\tilde{y}_1, \tilde{y}_2, \ldots, \tilde{y}_N] \)

1: Construct the discriminative integration matrix \( H \) by Eq. (5) and Eq. (6).
2: Construct the correlation matrix \( R \) by Eq. (6).
3: Compute the matrix exponential of \( R - H \), i.e. \( \exp(R - H) \).
4: Solve the correlation projection matrices \( \{\Phi^{(p)}\} \in R^{d_p \times d^M} \) by Eq. (23).
5: Obtain the correlation fusion feature set \( Y \) and \( \tilde{Y} \) by Eq. (26) and Eq. (27)

\[
\sum_{g=1}^{G} \left( \frac{1}{2MN} \sum_{p=1}^{M} \sum_{q=1}^{M} \sum_{k=1}^{N_g} \sum_{l=1}^{N_l} \phi^{(p)T} \phi^{(q)} \right) \\
x \left( z_{gk}^{(p)} - z_{gl}^{(q)} \right) \left( z_{gk}^{(p)} - z_{gl}^{(q)} \right)^T \phi^{(q)} 
\]

(28)

From Eq. (28), we can observe that the discriminative integration scatter \( \rho_o \) can capture the intra-class variations of within-modal and cross-modal correlation features. As \( \rho_o \) is reduced, the second term of Eq. (28) will compact intra-class correlation features from different modalities, and the first term of Eq. (28) can render the intra-class correlation features from the same modalities as similar as possible.

How to avoid the SSS problem is an important research subject in feature fusion of raw high-dimensional data. Our decomposition-based matrix exponential strategy effectively solves the SSS problem of EMDCA and the corresponding theoretical proofs have been given in Section 3.2. Next, we further discuss the robustness of EMDCA to noises of raw high-dimensional data. As pointed out in [37], noises of raw data focus on influencing small eigenvalues of the scatter and correlation matrices based on the raw data. However, by means of matrix exponential, the eigenvalues \( \lambda_1, \lambda_2, \ldots, \lambda_d \) are changed to \( \exp(\lambda_1), \exp(\lambda_2), \ldots, \exp(\lambda_d) \). According to properties of matrix exponential, the exponentials of the large eigenvalues are larger, but for the small eigenvalues, their exponentials are relatively smaller. To make it easy to understand, we give an intuitive illustration in Fig. 2. From Fig. 2, we can find that the smallest eigenvalue \( \lambda_1 \) is 8% in the total proportion of eigenvalues, and its exponential \( \exp(\lambda_1) \) will be changed to 2% of all the eigenvalue exponentials. However, the smaller the eigenvalues are, the smaller their exponentials are. For the large eigenvalue \( \lambda_6 \), the proportion of its exponential is changed from 26% to 56% in Fig. 2. This means that with the help of matrix exponential, the small eigenvalues with much noise information are weakened, and the large eigenvalues are enhanced. Thus the decomposition-based matrix exponential strategy ensures the non-singularity of matrices in theory and improves the robustness of noises.

IV. EXPERIMENTAL RESULTS AND ANALYSIS

In this section, we compare our method with six representative multi-modal feature fusion methods including Graph Multi-view Canonical Correlation Analysis (GMCCA) [38], Labeled Multiple Canonical Correlation analysis (LMCCA) [39], Discriminative Multiple Canonical Correlation Analysis (DMCCA) [21], Multi-view...
Discriminant Analysis (MvDA) [13], Laplacian Multi-set Canonical Correlation Analysis (LapMCCA) [23], and Graph Regularized Multi-set Canonical Correlations (GrMCC) [22]. This section evaluates the recognition performance on visible image datasets, infrared image datasets, and Reuters multilingual text dataset. For image datasets, we employ a common multi-modal strategy [22], [40] to obtain three modalities of each image. In all the methods, the nearest neighbor classifier with Euclidean distances [41] is used in the final recognition stage, and the best recognition rates are reported in the tables below for all possible dimensionalities. For all the experiments, codes of the methods is written by the MATLAB programming language, and the running environment of the codes is MATLAB2012a under the operating system of Windows 10.

### A. EXPERIMENTS ON THE VISUAL IMAGE DATASETS

The visible images are common in real-world applications, and this subsection employs five representative visible image datasets, i.e. Coil20 object visible image dataset\(^1\), Georgia Tech (GT) face visible image dataset\(^2\), ORL visible image dataset\(^3\), YaleB visible image dataset\(^4\), and Semeion handwritten digit image dataset\(^5\). For the datasets, we randomly choose four images per class for training, and the remaining images are treated as testing samples. The random sample experiments are repeated ten times in all, and the average recognition rates are tabulated in Table 1.

|        | EDMCA | LMCCA | DMCCA | MvDA | LapMCCA | GMCCA | GrMCC |
|--------|-------|-------|-------|------|---------|-------|-------|
| Coil20 | 83.89 | 81.91 | 79.18 | 81.71 | 73.10   | 64.35 | 56.17 |
| GT     | 66.84 | 61.73 | 53.36 | 61.64 | 64.20   | 39.80 | 49.65 |
| ORL    | 93.33 | 88.54 | 88.96 | 92.17 | 90.79   | 77.21 | 88.1  |
| YaleB  | 62.11 | 52.25 | 55.97 | 59.73 | 52.02   | 47.76 | 44.91 |
| Semeion| 67.55 | 66.05 | 63.02 | 63.63 | 55.11   | 50.35 | 47.04 |

On the basis of correlation analysis theories, LMCCA further considers the intra-class scatter of every modality. DMCCA maximizes the between-modal intra-class correlations and meanwhile reduces the between-modal inter-class correlations. MvDA is a multi-modal extension of linear discriminant analysis [29], and a discriminant common subspace is learned by optimizing a generalized Rayleigh quotient. The three methods only consider the embedding of class labels from different viewpoints, and their recognition rates are also similar in Table 1. In LapMCCA and GrMCC, the correlation features preserve the nonlinear structure information hidden in discriminative neighbor graphs as much as possible. For GMCCA, the prior knowledge of common sources is encoded by graphs, and graph-induced geometry structures are considered based on minimizing distances between correlation features and common low-dimensional representations. The three graph-based methods further preserve the nonlinear geometry structures based on raw samples. Due to the sample noises, the sample-based geometry structures usually derive from real geometry structures hidden in the raw samples. In most cases, the derivation reduces the class separability of correlation features, which is why the three compared graph-based approaches have lower recognition rates than the other methods. Using the discriminative integration scatter, our method minimizes the distances between the correlation features and cross-modal discriminative centroids and meanwhile maximizes the pair-wise correlations between various modalities. Besides, the decomposition-based matrix exponential strategy effectively solves the SSS problem of our method and enhances the robustness for noises of the raw samples. In Table 1, our method also exhibits the best recognition rates. The experimental results in the above table are the average recognition rates of ten sample randomness experiments. The smaller the variation of the recognition rates under different sample randomness experiments is, the better the robustness of the sample randomness is. The good robustness for the sample randomness will enhance the adaptability of the methods in real-world applications. For analyzing the sample robustness, Fig.3 displays the recognition rates under each sample randomness experiment of each method, and “Sample Random No.” of Fig.3 represents the sequence number of the ten sample randomness experiments. In most instances, as seen in Fig. 3, our method outperforms the compared methods in terms of sample randomness robustness.

### V. EXPERIMENTS ON THE INFRARED IMAGE DATASETS

Infrared cameras take infrared images, and we did some experiments on the CBRS NIR image dataset\(^6\) and the IRIS thermal image dataset\(^7\). The IRIS thermal image dataset takes 4228 images from 30 persons, and the images have five illuminations according to the combinations of on/off lights. Our experiments employ thermal images under two illuminations of all light off and all light on. Fig.5 shows some representative thermal images of one person. Four images from each class are used for training on the two infrared image datasets, while the remainder of the images is used as testing samples. The random sample experiments are run

---

1. Available at: https://www.cs.columbia.edu/CAVE/software/softlib/coil-20.php
2. Available at: http://www.anefian.com/research/face_reco.htm
3. Available at: http://www.cad.zju.edu.cn/home/dengcai/Data/FaceData.html
4. Available at: http://vision.ucsd.edu/~iskwak/ExtYaleDatabase/ExtYaleB.html
5. Available at: http://archive.ics.uci.edu/ml/datasets/Semeion+Handwritten+Digit
6. http://vcip-okstate.org/pbvs/bench/Data/07/download.html
7. http://vcip-okstate.org/pbvs/bench/Data/02/download.html
FIGURE 3. Recognition rates under the ten sample randomness experiments of each method on the (a) Coil20, (b) GT, (c) ORL, (d) YaleB, and (e) Semeion datasets.

TABLE 2. The recognition rates on the infrared image datasets.

|        | EDMCA | LMCCA | DMCCA | MvDA | LapMCCA | GMCCA | GrMCC |
|--------|-------|-------|-------|------|---------|-------|-------|
| CBSR   | 95.35 | 92.85 | 91.85 | 94.19| 91.63   | 76    | 82.13 |
| IRIS   | 62.44 | 57    | 52.07 | 58.19| 62.23   | 42.37 | 46.65 |

ten times, and the average recognition rates are reported in Table 2. LMCCA and MvDA still have similar recognition rates, and the two methods are superior to DMCCA in Table 2, which is consistent with the experimental results in the above visible image datasets. In Table 2, GMCCA and GrMCC display lower recognition rates than the other compared methods. On the infrared image datasets, EDMCA still has the best recognition performance, and the sample robustness of EDMCA is better in most cases, which can be found in Fig. 4.

A. EXPERIMENTS ON THE REUTERS MULTILINGUAL TEXT DATASET

On the Reuters multilingual text dataset, each document has five language versions, and different languages of one docu-
ment describe the same content. Thus each language description can be treated as one modality of the document. The five languages are English (abbreviated as EN), French (FR), German (GR), Italian (IT), and Spanish (SP). In the subsection, we utilize three modalities, and ten combinations of different modalities can be obtained. Ten samples in each class are chosen randomly as teaching samples, while the remaining samples are used for testing. Ten random sample experiments are performed ten times, and the average recognition rates and standard deviations are tabulated in Tables 3 and 4, respectively. Table 3 shows that the differences between different methods are relatively minor compared to the experimental results on visible and infrared image datasets. LMCCA and LapMCCA have similar recognition rates on the Reuters multilingual text dataset.

![FIGURE 4. Recognition rates under the ten sample randomness experiments of each method (a) on the CBSR dataset, and (b) on the IRIS dataset.](image)

![FIGURE 5. Representative images of one person on the IRIS dataset.](image)

**TABLE 3.** The recognition rates on the reuters multilingual text dataset.

|          | EDMCA | LMCCA | DMCCA | MvDA | LapMCCA | GMCCA | GrMCC |
|----------|-------|-------|-------|------|---------|-------|-------|
| EN-FR-GR | 57.67 | 57.08 | 53.38 | 53.59| 56.25   | 54.48 | 54.26 |
| EN-FR-IT | 58.96 | 55.08 | 52.20 | 55.51| 56.18   | 55.35 |       |
| EN-FR-SP | 57.96 | 57.94 | 54.26 | 56.23| 54.52   | 54.25 |       |
| EN-GR-IT | 58.80 | 57.96 | 54.64 | 57.39| 56.17   | 55.30 |       |
| EN-GR-SP | 58.29 | 57.83 | 54.52 | 57.23| 54.05   | 53.23 |       |
| EN-IT-SP | 59.10 | 58.96 | 55.97 | 56.58| 55.00   | 54.51 |       |
| FR-GR-IT | 58.06 | 57.94 | 53.46 | 56.96| 54.65   | 55.75 |       |
| FR-GR-SP | 57.13 | 57.47 | 53.42 | 56.42| 53.30   | 53.96 |       |
| FR-IT-SP | 58.54 | 58.34 | 54.72 | 56.90| 54.78   | 55.13 |       |
| GR-IT-SP | 58.14 | 58.27 | 54.67 | 57.63| 54.95   | 55.90 |       |

**TABLE 4.** The standard deviations on the reuters multilingual text dataset.

|          | EDMCA | LMCCA | DMCCA | MvDA | LapMCCA | GMCCA | GrMCC |
|----------|-------|-------|-------|------|---------|-------|-------|
| EN-FR-GR | 1.72  | 2.24  | 2.45  | 3.61 | 2.33    | 2.29  | 2.59  |
| EN-FR-IT | 1.65  | 2.61  | 3.32  | 3.70 | 1.78    | 2.68  | 2.17  |
| EN-FR-SP | 2.10  | 2.30  | 2.18  | 3.24 | 2.71    | 2.91  | 2.98  |
| EN-GR-IT | 2.15  | 2.75  | 3.58  | 4.74 | 2.52    | 2.79  | 3.42  |
| EN-GR-SP | 2.35  | 1.91  | 2.99  | 3.63 | 3.33    | 3.15  | 3.29  |
| EN-IT-SP | 2.23  | 2.85  | 1.96  | 4.66 | 3.47    | 2.84  | 3.33  |
| FR-GR-IT | 2.14  | 2.76  | 2.85  | 4.56 | 2.53    | 2.17  | 3.06  |
| FR-GR-SP | 2.15  | 2.73  | 2.47  | 2.71 | 2.73    | 1.97  | 3.18  |
| FR-IT-SP | 2.65  | 2.80  | 2.20  | 3.27 | 2.92    | 1.41  | 2.56  |
| GR-IT-SP | 2.59  | 2.57  | 4.31  | 4.13 | 2.79    | 2.48  | 4.01  |
Encouraging experimental findings show the effectiveness of the method, not only providing corresponding theoretical evidence. The standard deviation of Table 4 reveals the robustness of the sample randomness. The sample robustness improves as the standard deviation decreases. Compared with the compared methods, our method still possesses smaller standard deviations in most cases, which can reflect the excellent adaptability of EDMCA in real-world applications. The text dataset needs the transformation from the word-based documents to the vector samples, and all the methods are implemented on the vector samples. The transformation of the text dataset loses much effective information of original documents, which is a potential reason why the recognition rates of the text dataset is lower than those of the image-based datasets. The extensive experimental results on the above three categories of datasets reveal the effectiveness and robustness of our method in the recognition and classification tasks.

VI. CONCLUSION

Many discriminant variations of MCCA have also been proposed for utilizing class information to improve the discriminative power of correlation fusion features. Although the variations have been widely applied to many real-world applications, the Small Sample Size (SSS) problem and the absence of cross-modal discriminant scatter are inevitable in the variations. Aiming at these issues, we construct the discriminative integration scatter of all the modalities by enhancing the aggregation towards cross-modal discriminative centroids, and we give a decomposition-based matrix exponential strategy that can solve the SSS problem and improve the robustness of noises. Then we further develop the novel EDMCA method by minimizing the exponential discriminative integration scatter and simultaneously maximizing exponential between-modal correlations, and the method can learn correlation fusion features with good discriminative power from a few raw high-dimensional multi-modal data. For the method, we not only provide corresponding theoretical proofs and some intuitive analysis, but also design extensive experiments on the visible image datasets, the infrared image datasets, and the Reuters multilingual text dataset. Encouraging experimental findings show the effectiveness and robustness of the method.

REFERENCES

[1] X. Li, M. Jia, M. T. Islam, L. Yu, and L. Xing, “Self-supervised feature learning via exploiting multi-modal data for retinal disease diagnosis,” IEEE Trans. Med. Imag., vol. 39, no. 12, pp. 4023–4033, Dec. 2020.

[2] L. Liu, G. Deng, D. Beutemp, and X.-P. Zhang, “Re-synchronization using the hand preceding model for multi-modal fusion in automatic continuous cued speech recognition,” IEEE Trans. Multimedia, vol. 23, pp. 292–305, 2021.

[3] S. C. Kulkarni and P. P. Rege, “Pixel level fusion techniques for SAR and optical images: A review,” Inf. Fusion, vol. 59, pp. 13–29, Jul. 2020.

[4] Y. Wei, D. Wu, and J. Terpenny, “Decision-level data fusion in quality control and predictive maintenance,” IEEE Trans. Autom. Sci. Eng., vol. 18, no. 1, pp. 184–194, Jan. 2021.

[5] P. S. Lamba, D. Virmani, and O. Castillo, “Multimodal human eye blink recognition method using feature level fusion for exigency detection,” Soft Comput., vol. 24, no. 22, pp. 16829–16845, May 2020.

[6] X. Yang, W. Liu, W. Liu, and D. Tao, “A survey on canonical correlation analysis,” IEEE Trans. Knowl. Data Eng., vol. 33, no. 6, pp. 2349–2368, Jun. 2021.

[7] Z. Chen, Y. Cao, K. Zhang, T. Koenings, T. Peng, C. Yang, W. Gui, and S. X. Ding, “A distributed canonical correlation analysis-based fault detection method for plant-wide process monitoring,” IEEE Trans. Ind. Informat., vol. 15, no. 5, pp. 2710–2720, May 2019.

[8] Q. Zhu, N. Xu, Z. Zhang, D. Guan, R. Wang, and D. Zhang, “Cross-spectral palmprint recognition with rank-four canonical correlation analysis,” Multimedia Tools Appl., vol. 79, nos. 45–46, pp. 33771–33792, Dec. 2019.

[9] Y. Liu, B. Liu, X. Zhao, and M. Xie, “A mixture of variational canonical correlation analysis for nonlinear and quality-relevant process monitoring,” IEEE Trans. Ind. Electron., vol. 65, no. 8, pp. 6478–6486, Aug. 2018.

[10] O. Yair and R. Talmon, “Local canonical correlation analysis for nonlinear common variables discovery,” IEEE Trans. Signal Process., vol. 65, no. 5, pp. 1101–1115, Mar. 2017.

[11] X. Gao, Q. Sun, H. Xu, and Y. Li, “2D-LPCCA and 2D-SPCCA: Two new canonical correlation methods for feature extraction, fusion and recognition,” Neurocomputing, vol. 284, pp. 148–159, Apr. 2018.

[12] Q.-S. Sun, Z.-D. Liu, P.-A. Heng, and D.-S. Xia, “A theorem on the generalized canonical projective vectors,” Pattern Recognit., vol. 38, no. 3, pp. 449–452, Mar. 2005.

[13] M. Kan, S. Shan, H. Zhang, S. Lao, and X. Chen, “Multi-view discriminant analysis,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 38, no. 1, pp. 188–194, Jan. 2015.

[14] L. Han, X.-Y. Jing, and F. Wu, “Multi-view local discrimination and canonical correlation analysis for image classification,” Neurocomputing, vol. 275, pp. 1087–1098, Jan. 2018.

[15] X. Huang, B. Zhang, H. Qiao, and X. Nie, “Local discriminant canonical correlation analysis for supervised PolSAR image classification,” IEEE Geosci. Remote Sens. Lett., vol. 14, no. 11, pp. 2102–2106, Nov. 2017.

[16] J. Yang and X. Zhang, “Feature-level fusion of fingerprint and fingerprint vein for personal identification,” Pattern Recognit. Lett., vol. 33, no. 5, pp. 423–428, Apr. 2012.

[17] P. Wu, X. Zhang, J. He, S. Lou, and J. Gao, “Locality preserving randomized canonical correlation analysis for real-time nonlinear process monitoring,” Process Saf. Environ. Protection, vol. 147, pp. 1088–1100, Mar. 2021.

[18] Y. Peng, D. Zhang, and J. Zhang, “A new canonical correlation analysis algorithm with local discrimination,” Neural Process. Lett., vol. 31, no. 1, pp. 1–15, Nov. 2009.

[19] Y. Wang, P. Li, and C. Yao, “Hypergraph canonical correlation analysis for multi-label classification,” Signal Process., vol. 105, pp. 258–267, Dec. 2014.

[20] N. M. Correa, T. Eichele, T. Adahi, Y.-O. Li, and V. D. Calhoun, “Multiset canonical correlation analysis for the fusion of concurrent single trial ERP and functional MRI,” NeuroImage, vol. 50, no. 4, pp. 1438–1445, Mar. 2010.

[21] L. Gao, L. Qi, E. Chen, and L. Guan, “Discriminative multiple canonical correlation analysis for image classification,” IEEE Trans. Image Process., vol. 27, no. 4, pp. 1951–1965, Apr. 2018.

[22] Y.-H. Yuan and Q.-S. Sun, “Graph regularized multiset canonical correlations with applications to joint feature extraction,” Pattern Recognit., vol. 47, no. 12, pp. 3907–3919, Dec. 2014.

[23] Y.-H. Yuan, Y. Li, X.-B. Shen, Q.-S. Sun, and J.-L. Yang, “Laplacian multiset canonical correlations for multiview feature extraction and image recognition,” Multimedia Tools Appl., vol. 76, no. 1, pp. 731–755, Nov. 2015.

[24] X. Shen, Q. Sun, and Y. Yuan, “A unified multiset canonical correlation analysis framework based on graph embedding for multiple feature extraction,” Neurocomputing, vol. 148, pp. 397–408, Jan. 2015.

[25] W. Liu, X. Yang, D. Tao, J. Cheng, and Y. Tang, “Multiview dimension reduction via Hessian multiset canonical correlations,” Inf. Fusion, vol. 41, pp. 119–128, May 2018.

[26] R. Van de Schoot and M. Miocevic, Small Sample Size Solutions: A Guide to Applied Researchers and Practitioners, Exeter, U.K.: Taylor & Francis, 2020.

[27] X. Shen and Q. Sun, “Orthogonal multiset canonical correlation analysis based on fractional-order and its application in multiple feature extraction and recognition,” Neural Process. Lett., vol. 42, no. 2, pp. 301–316, May 2014.
[28] S. Wang, J. Lu, X. Gu, C. Shen, R. Xia, and J. Yang, “Canonical principal angles correlation analysis for two-view data,” J. Vis. Commun. Image Represent., vol. 35, pp. 209–219, Feb. 2016.

[29] N. Han, J. Wu, X. Fang, J. Wen, S. Zhan, S. Xie, and X. Li, “Transferable linear discriminant analysis,” IEEE Trans. Neural Netw. Learn. Syst., vol. 31, no. 12, pp. 5630–5638, Dec. 2020.

[30] S. Su, G. Zhu, and Y. Zhu, “An orthogonal locality and globality dimensionality reduction method based on twin eigen decomposition,” IEEE Access, vol. 9, pp. 55714–55725, 2021.

[31] Y.-H. Yuan, Q.-S. Sun, and H.-W. Ge, “Fractional-order embedding canonical correlation analysis and its applications to multi-view dimensionality reduction and recognition,” Pattern Recognit., vol. 47, no. 3, pp. 1411–1424, Mar. 2014.

[32] Y. Zhu, T. Peng, S. Su, and C. Li, “Multi-modal subspace fusion via Cauchy multi-set canonical correlations,” IEEE Access, vol. 8, pp. 115228–115239, 2020.

[33] N. J. Higham, “The scaling and squaring method for the matrix exponential revisited,” SIAM J. Matrix Anal. Appl., vol. 26, no. 4, pp. 1179–1193, Jan. 2005.

[34] S.-J. Wang, H.-L. Chen, X.-J. Peng, and C.-G. Zhou, “Exponential locality preserving projections for small sample size problem,” Neurocomputing, vol. 74, no. 17, pp. 3654–3662, Oct. 2011.

[35] S. Su, X. Fang, G. Yang, B. Ge, and P. Zheng, “Clustering adaptive canonical correlations for high-dimensional multi-modal data,” J. Vis. Commun. Image Represent., vol. 71, pp. 1–12, Aug. 2020.

[36] S. Su, H. Ge, and Y. Tong, “Multigraph embedding discriminative correlation feature learning for image recognition,” Signal Process., Image Commun., vol. 60, pp. 173–182, Feb. 2018.

[37] C. Peng, Y. Chen, Z. Kang, C. Chen, and Q. Cheng, “Robust principal component analysis: A factorization-based approach with linear complexity,” Inf. Sci., vol. 513, pp. 581–599, Mar. 2020.

[38] J. Chen, G. Wang, and G. B. Giannakis, “Graph multiview canonical correlation analysis,” IEEE Trans. Signal Process., vol. 67, no. 11, pp. 2826–2838, Jun. 2019.

[39] L. Gao, R. Zhang, L. Qi, E. Chen, and L. Guan, “The labeled multiple canonical correlation analysis for information fusion,” IEEE Trans. Multimedia, vol. 21, no. 2, pp. 375–387, Jul. 2018.

[40] S. Su, X. Fang, G. Yang, B. Ge, and Y. Zhu, “Self-balanced multi-view orthogonality correlation analysis for image feature learning,” Inf. Phys. Technol., vol. 100, pp. 44–51, Aug. 2019.

[41] X. Zhuang, Z. Yang, and D. Cordes, “A technical review of canonical correlation analysis for neuroscience applications,” Hum. Brain Mapping, vol. 41, no. 13, pp. 3807–3833, Jun. 2020.

YANMIN ZHU received the M.S. degree from the School of Mechanical and Electronic Engineering, Shandong Agricultural University, Taian, China. She is currently pursuing the Ph.D. degree with the School of Mechanical Engineering, Anhui University of Science and Technology, Huainan, China. Her research interests include multi-modal information fusion, correlation projection theory, image processing, and bearing fault diagnosis.

TIANHAO PENG received the Ph.D. degree from Zhejiang University, Hangzhou, China. He is currently a Professor with the School of Mechanical Engineering, Anhui University of Science and Technology, Huainan, China. His research interests include data fusion, intelligent control, and higher-dimensional data processing.

SHUZHI SU received the Ph.D. degree from the School of Internet of Things Engineering, Jiangnan University, Wuxi, China. He is currently an Associate Professor with the School of Computer Science and Engineering, Anhui University of Science and Technology, Huainan, China. His research interests include multimodal pattern recognition, information fusion, feature learning, and image processing.

* * *