IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 27, 2020, accepted June 7, 2020, date of publication June 16, 2020, date of current version July 2, 2020.

*Digital Object Identifier 10.1109/ACCESS.2020.3002810*

# Multi-Modal Subspace Fusion via Cauchy Multi-Set Canonical Correlations

**YANMIN ZHU[1], TIANHAO PENG[1], SHUZHI SU[2], AND CHANGPENG LI[1]**
[1]School of Mechanical Engineering, Anhui University of Science and Technology, Huainan 232001, China
[2]School of Computer Science and Engineering, Anhui University of Science and Technology, Huainan 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, in part by the Natural Science Research Project of Colleges and Universities in Anhui Province under Grant KJ2018A0083, and in part by the China Postdoctoral Science Foundation under Grant 2019M660149.

**ABSTRACT** Multi-set canonical correlation analysis (MCCA) is a famous multi-modal coherent subspace learning method. However, sample-based between-modal and within-modal covariance matrices of MCCA usually deviate from real covariance matrices due to noise information and limited sample size. The deviation will weaken the performance of MCCA, especially in image recognition. Aiming at this challenging issue, we correct singular values of sample covariance matrices with the employment of Cauchy estimate theory and further obtain Cauchy covariance matrices that are closer to real covariance matrices. On the basis of Cauchy covariance matrices, we develop a novel multi-modal subspace fusion method, i.e. Cauchy multi-set canonical correlations. By maximizing Cauchy correlations between different modalities and constraining Cauchy scatters of within-modal data, the method can learn a Cauchy coherent fusion subspace with well discriminative power from a few images. Experiment results have shown the effectiveness of the proposed method, promising to the aims of this research.

**INDEX TERMS** Multi-modal subspace fusion, coherent fusion subspace learning, correlation projection theory, image recognition.

## I. INTRODUCTION

One object usually possesses multiple data representations in real-world applications. For instance, we can collect face data, fingerprint data, and iris data of one person, so these data can be treated as multi-modal data [1] of a person. Raw multi-modal data are usually high-dimensional data with a lot of noise and redundant information, and how to effectively fuse these multi-modal data is still a challenging task. Canonical correlation analysis (CCA) [2] is a classical multi-modal subspace fusion method that can achieve the multi-modal fusion in a learned coherent fusion subspace. CCA aims at simultaneously learning correlation projection directions of two-modal data on the basis of maximal between-modal correlations, and then raw multi-modal data can be projected into the coherent fusion subspace. Up to now, many variants of CCA have been proposed for different applications, such as visual evoked potential classification [3], process monitoring [4], blind source separation [5], stress recognition [6], defect prediction [7], brain data processing [8] and so on.

As an unsupervised subspace fusion method, CCA is difficult to enhance the discriminative power of the coherent subspace through class labels. Thus, generalized CCA (GCCA) [9] utilizes class labels to constrain the coherent subspace learning by further minimizing the within-modal scatter of each modality, and the fused low-dimensional data obtained by the method show well class separability in image recognition. Besides, Sun *et al.* constructed discriminative between-modal correlations with the help of supervised covariance matrices, and proposed a novel discriminative CCA (DCCA) [10] method. From different viewpoints, unsupervised discriminant CCA based on spectral clustering (UDCCASC) [11] utilizes the class information obtained by spectral clustering, which can achieve self-supervised subspace learning of multi-modal high-dimensional data on the unsupervised cases. By using class information, GCCA, DCCA, and UDCCASC can improve the fusion performance between two-modal datasets. However, the three supervised methods belong to the linear subspace fusion method, which

The associate editor coordinating the review of this manuscript and approving it for publication was Michele Nappi.

will have difficulty capturing nonlinear structure information hidden in raw multi-modal data. Aiming at this issue, locality preserving CCA [12] exploits local geometry structures from neighbor graphs, which will make the coherent fusion subspace preserve the geometry structure information as soon as possible. Based on the graph regularization technique, Chen *et al.* [13] proposed graph CCA that encodes the underlying manifolds of a common source graph on the basis of the maximal between-modal correlation. In image-based pose estimation, there is a pixel modality and a pose modality, and the pixel modality can be divided into many local patches. Inspired by patch alignment idea, a novel local CCA alignment [14] method firstly implements CCA to learn the coherent fusion subspace from each type of patches, and then low-dimensional local data of the coherent fusion are aligned to obtain global fusion data. Besides supervised variants and graph-based variants, some other CCA-related variants have been proposed, such as dual multi-kernel discriminating correlation analysis [15], complete CCA [16], fractional-order embedding CCA [17], and uncertain CCA [18] and so on.

CCA can only deal with two-modal data, but the number of modalities corresponding to one same object is usually more than two in real-world applications. Aiming at three-modal image data, color image CCA (CICCA) [19] constructs a correlation coherent subspace learning model of three color components, and the optimization problem of the model is theoretically derived to solve three equations. The model with analytical solutions is a classical correlation coherent subspace learning model, and its effectiveness has been shown in real-world image recognition tasks. For better fusing multi-modal data, multi-set CCA (MCCA) [20] can simultaneously learn a coherent fusion subspace from more than two modalities using correlation analysis theory. In essence, CCA and CICCA are special examples of MCCA, and MCCA is a multi-modal extension of CCA and CICCA. Similar to CCA and CICCA, MCCA is also an unsupervised linear subspace fusion method.

To effectively constrain the class separability of the coherent fusion subspace using class labels, discriminative multiple CCA (DMCCA) [21] constructs a label-induced coherent fusion subspace learning model, and multi-modal data in the learned subspace possess the maximal between-modal discriminative correlation. Also, DMCCA is a multi-modal extension of DCCA to some extent. Different from DMCCA, labeled MCCA [22] considers the intra-class scatter matrix of each modality under the multi-modal correlation analysis framework, and its effectiveness has been verified in image recognition tasks.

Graph technique has been widely applied to MCCA-related variants. Graph regularized multi-set canonical correlations [23] constrains the local discriminative scatters of the same modalities under the maximal between-modal correlations. Similarly, graph regularized MCCA [24] minimizes the distances between wanted low-dimensional correlation data and projected low-dimensional correlation data, and low-dimensional fusion data in the coherent fusion subspace will embed the geometry structure of the common sources by regularization of a common graph. According to the advantages of Hessian, Liu *et al.* [25] developed a novel Hessian multi-set canonical correlations method, which exploits graph-based geometry manifolds of raw multi-modal data and effectively integrates the manifold structure into the coherent fusion subspace. To effectively reduce the redundant information of the learned coherent subspace, fractional-order orthogonal MCCA [26] embeds orthogonality constraints into the correlation analysis framework with fractional-order correction, and the effectiveness of the method has been verified in image recognition tasks. Based on different purposes and different applications, multi-set globality locality preserving CCA [27], two-dimensional MCCA [28], view-consistent collaborative multi-set correlation projection [29], and multi-modal hybrid centroid CCA [30] have been proposed.

Above mentioned correlation analysis methods are based on between-modal and within-modal covariance matrices constructed by training samples. However, due to noise information and insufficient sample size, these sample covariance matrices usually deviate from real covariance matrices, which will affect the performance of the correlation analysis methods in many applications, especially recognition tasks of high-dimensional data. To solve the challenging problem, we correct the singular values of these matrices with the help of Cauchy estimate theory [31] and develop novel Cauchy between-modal and within-modal covariance matrices that are closer to real covariance matrices. As far as we know, it is novel for correcting singular values of sample covariance matrices using Cauchy estimate theory. Besides, Cauchy covariance matrices can effectively reduce the derivation of sample covariance matrices, which is beneficial to improve the recognition performance of correlation analysis methods. In this paper, we embed Cauchy covariance matrices into the correlation analysis framework and then propose a novel Cauchy subspace fusion method for multi-modal high-dimensional data, i.e. Cauchy multi-set canonical correlations (CauMCCs). By maximizing Cauchy correlations between different modalities and simultaneously constraining Cauchy scatters of within-modal data, the proposed method can learn a Cauchy coherent fusion subspace with well discriminative power from a few training high-dimensional samples. In the Cauchy coherent fusion subspace, multi-modal data can be effectively fused, and the fused low-dimensional data possess well class separability, which is beneficial to final recognition tasks. Cauchy covariance matrices of our method may be also utilized to improve the recognition performance of other methods with covariance matrices, such as locality preserving projections, principal component analysis, and linear discriminant analysis. On one synthetic dataset and three real-world image datasets, the proposed method exhibits the superior recognition performance, and the experimental results and the parameter influence are analyzed in detail.

Next, MCCA is briefly reviewed in Section II, and then our proposed method is described in detail. In Section IV, we

discuss multiple aspects of experiments. Finally, conclusions are given in Section V.

## II. PRINCIPLE OF MCCA

Suppose that $\left\{Z^{(p)} = [z_1^{(p)}, z_2^{(p)}, \ldots, z_N^{(p)}] \in R^{d_p \times N}\right\}_{p=1}^{M}$ are the $M$ modality datasets corresponding to $N$ objects and $d_p$ represents the dimension of samples. $z_k^{(p)}$ is the $k$th ($k = 1, 2, \ldots, N$) sample of the $p$th modality dataset ( i.e. $Z^{(p)}$), and $\left\{z_k^{(p)}\right\}_{p=1}^{M}$ correspond to the same object. $\eta^{(p)} \in R^{d_p \times 1}$ is a correlation projection direction of $Z^{(p)}(p = 1, 2, \ldots, M)$. MCCA aims at learning the correlation projection directions of $\left\{Z^{(p)}\right\}_{p=1}^{M}$ so that the projected low-dimensional data $\left\{\eta^{(p)T} Z^{(p)}\right\}_{p=1}^{M}$ possess maximal between-modal correlations. Concretely, the objective function of MCCA is as follows:

$$\max_{\{\eta^{(p)}\}_{p=1}^{M}} \sum_{p=1}^{M} \sum_{p=1}^{M} \frac{\eta^{(p)T} S^{(pq)} \eta^{(q)}}{\sqrt{\eta^{(p)T} S^{(pp)} \eta^{(p)}} \sqrt{\eta^{(q)T} S^{(qq)} \eta^{(q)}}} \quad (1)$$

where $S^{(pq)} = \frac{1}{N} \sum_{k=1}^{N} (z_k^{(p)} - \bar{z}^{(p)})(z_k^{(q)} - \bar{z}^{(q)})^T$ is the between-modal ($p \neq q$) or within-modal ($p = q$) covariance matrix. Additionally, $\bar{z}^{(p)} = \frac{1}{N} \sum_{k=1}^{N} z_k^{(p)}$ is the sample mean of $Z^{(p)}$, and $\bar{z}^{(p)}$ and $\bar{z}^{(q)}$ have the same definition. As pointed in [32], the between-modal covariance matrix $S^{(pq)}$ is capable of revealing the correlation between $Z^{(p)}$ and $Z^{(q)}$, and the within-modal covariance matrix $S^{(pp)}$ can capture the within-modal scatter information of $Z^{(p)}$. Therefore, the objection function of Eq. (1) can be treated as the maximum of the between-modal correlations and minimum of the within-modal scatters.

## III. CAUCHY MULTI-SET CANONICAL CORRELATIONS

### A. MOTIVATION

The correlation analysis methods are common multi-modal subspace learning methods that can achieve the multi-modal fusion in a learned coherent fusion subspace. Covariance matrices are essential and crucial to the correlation analysis methods. In the correlation analysis methods, between-modal covariance matrices can reveal the correlations between different modalities, and within-modal covariance matrices can capture the scatter information of each modality. Besides, samples are usually employed to construct sample covariance matrices that will take the place of real covariance matrices, since real covariance matrices cannot be obtained in many real-world applications. However, samples from real-world applications have a lot of redundant information and noises, which will cause that sample covariance matrices deviate from real covariance matrices. The deviation can weaken the correlation and scatter structures of the coherent fusion subspace. Cauchy estimate theory has been used to correct some distortion and deviation problems of data with noises. From different viewpoints, we try to correct singular values of sample covariance matrices by Cauchy estimate theory so that the corrected covariance matrices are closer to real covariance matrices. Besides, we further construct a novel multi-modal coherent subspace learning model on the basis of the corrected covariance matrices.

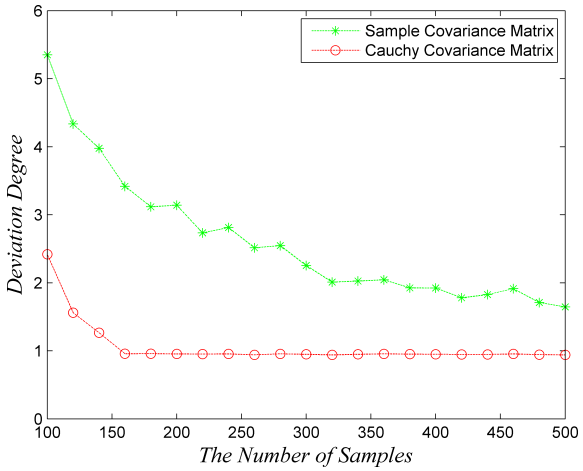


**FIGURE 1.** The deviation degree versus the number of samples.

### B. DISCUSSION OF COVARIANCE MATRICES

Covariance matrices are vital to many pattern recognition methods. The deviation of Covariance matrices weakens the performance of many pattern recognition methods, shown in [33], [34]. For evaluating covariance matrices based on samples, we give a deviation degree $\varepsilon$ between sample covariance matrices and real ones:

$$\varepsilon = \left\| H - \tilde{H} \right\|_F^2 / \left\| \tilde{H} \right\|_F^2 \quad (2)$$

where $H$ represents a sample covariance matrix and $\tilde{H}$ is the corresponding real covariance matrix. The smaller the deviation degree is, the closer the sample covariance matrix is to the real one. In Fig.1 and Fig. 2, we exhibit the deviation degree of sample covariance matrices based on the same synthetic samples, and it can be seen that sample covariance matrices deviate from real ones, especially on the cases of high dimension and small sample size. In real-world image recognition tasks, the dimension of image samples is very large, but the number of training image samples is often small.

As a result, covariance matrices based on image samples will have a great deviation degree, which is an important reason why MCCA and many feature learning methods based on covariance matrices show the bad performance of image recognition.

### C. CONSTRUCTION OF CAUCHY COVARIANCE MATRICES

The deviation degree of sample covariance matrices is crucial to the discriminative power of the low-dimensional data learned by MCCA. As pointed out in [17], the smaller the deviation degree is, the better MCCA holds the between-modal correlations and the within-modal scatter information, which will be beneficial to enhance the recognition performance of MCCA and MCCA-related methods. To decrease the deviation degree of sample covariance

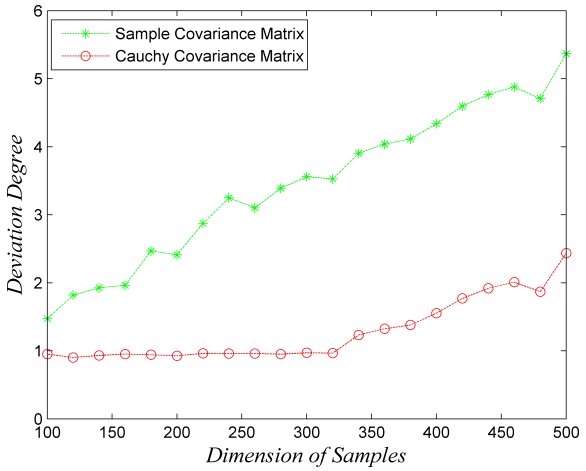


**FIGURE 2.** The deviation degree versus dimension of samples.

matrices, we construct Cauchy covariance matrices with the help of Cauchy estimate theory. Taking the within-modal covariance matrices as an example, we describe how to construct Cauchy within-modal covariance matrices. Firstly, we decompose within-modal sample covariance matrices by singular value decomposition [3]:

$$S^{(pp)} = P^{(pp)}\Lambda^{(pp)}Q^{(pp)T} \tag{3}$$

In Eq. (3), $P^{(pp)}$ (or $Q^{(pp)}$) is the left (or right) singular matrix. $\Lambda^{(pp)} = diag(\lambda_1^{(pp)}, \lambda_2^{(pp)}, \ldots, \lambda_{d_p}^{(pp)})$ denotes the diagonal matrix of singular values. Then, the singular values are corrected using the Cauchy estimate theory, and we can further construct the Cauchy within-modal covariance matrix $R_{Cau}^{(pp)}$:

$$R_{Cau}^{(pp)} = P^{(pp)}\Lambda_{Cau}^{(pp)}Q^{(pp)T} \tag{4}$$

where $\Lambda_{Cau}^{(pp)} = diag(log(1 + (\frac{\lambda_1^{(pp)}}{c})^2), \log(1 + (\frac{\lambda_2^{(pp)}}{c})^2), \ldots, log(1 + (\frac{\lambda_{d_p}^{(pp)}}{c})^2))$, and $c$ is a scale parameter.

*Property 1:* The rank of Cauchy covariance matrices is equivalent to that of sample covariance matrices, i.e. $\text{rank}(R_{Cau}^{(pp)}) = \text{rank}(S^{(pp)})$.

If $\lambda_k^{(pp)} = 0\,(k = 1, 2, \ldots, d_p)$, $log(1 + (\lambda_k^{(pp)}/c)^2)$ is also equal to zero. Additionally, $log(1 + (\lambda_k^{(pp)}/c)^2) \neq 0$ when $\lambda_k^{(pp)}$ is not equal to zero. Thus $R_{Cau}^{(pp)}$ and $S^{(pp)}$ have the same rank, i.e. rank $(R_{Cau}^{(pp)})$ = rank $(S^{(pp)})$, which reveals that Cauchy covariance matrices still preserve the basis property of sample covariance matrices.

Similar to the definition of Cauchy within-modal covariance matrix, Cauchy between-modal covariance matrix $R_{Cau}^{(pq)}$ is

$$R_{Cau}^{(pq)} = P^{(pq)}\Lambda_{Cau}^{(pq)}Q^{(pq)T} \tag{5}$$

In Eq. (5), $S^{(pq)} = \frac{1}{N}\sum_{k=1}^{N}(z_k^{(p)} - \bar{z}^{(p)})(z_k^{(q)} - \bar{z}^{(q)})^T$ is the between-modal covariance matrix of the two modality datasets $Z^{(p)}$ and $Z^{(q)}(p \neq q)$, and $P^{(pq)}$ and $Q^{(pq)}$ are the left and right singular matrices of $S^{(pq)}$. $\Lambda_{Cau}^{(pq)} = diag(\log(1 + (\frac{\lambda_1^{(pq)}}{c})^2), log(1 + (\frac{\lambda_2^{(pq)}}{c})^2), \ldots, log(1 + (\frac{\lambda_{\tau_{pq}}^{(pq)}}{c})^2))$, where $\Lambda^{(pq)} = diag(\lambda_1^{(pq)}, \lambda_2^{(pq)}, \ldots, \lambda_{\tau_{pq}}^{(pq)})$ is the diagonal matrix of singular values corresponding to $S^{(pq)}$. $\tau_{pq}$ is the rank of $S^{(pq)}$, i.e. $\tau_{pq} = \text{rank}(S^{(pq)})$. As far as we know, it is novel for correcting singular values of sample covariance matrices using Cauchy estimate theory. Besides, Cauchy between-modal and within-modal covariance matrices are closer to real covariance matrices, which can effectively enhance the recognition performance of many multi-modal pattern recognition methods.

### D. ANALYSIS OF CAUCHY COVARIANCE MATRICES

Besides the deviation degree of sample covariance matrices, Fig.1 and Fig.2 also intuitively exhibit the deviation degree of Cauchy covariance matrices. When the dimension of samples is increasing, the deviation degree of Cauchy covariance matrices has an increasing trend. Also, with the decrease of samples, the increasing trend of the deviation degree appears in Cauchy covariance matrices. However, Cauchy covariance matrices are always closer to real covariance matrices than sample covariance matrices.

To further analyze the reasons why Cauchy covariance matrices are closer to real ones, we implement the singular value decomposition on real covariance matrices, sample covariance matrices, and Cauchy covariance matrices. For the three categories of covariance matrices, singular values are arranged from small to large, and Fig.3 exhibits these ordered singular values. In Fig.3, we observe that Cauchy covariance matrices correct singular values and the corrected singular values are closer to singular values of real covariance matrices, which is an important reason that Cauchy covariance matrices are superior to sample ones on the deviation degree.

### E. FORMULATION AND OPTIMIZATION OF CauMCCs

Based on Cauchy between-modal and within-modal covariance matrices and the correlation analysis theory, we further propose a regularized Cauchy correlation function of CauMCCs as (6), as shown at the bottom of this page where $\tau \geq 0$ denotes a regularization parameter. From Property 1, we can observe that the correction of the singular values has not been changed the rank of covariance matrices. That is,

$$\max_{\{\eta^{(p)}\}_{p=1}^{M}} \sum_{p=1}^{M}\sum_{q=1}^{M} \frac{\eta^{(p)T}R_{\text{Cau}}^{(pq)}\eta^{(q)}}{\sqrt{\eta^{(p)T}R_{\text{Cau}}^{(pp)}\eta^{(p)} + \tau\eta^{(p)T}\eta^{(p)}}\sqrt{\eta^{(q)T}R_{\text{Cau}}^{(qq)}\eta^{(q)} + \tau\eta^{(q)T}\eta^{(q)}}} \tag{6}$$

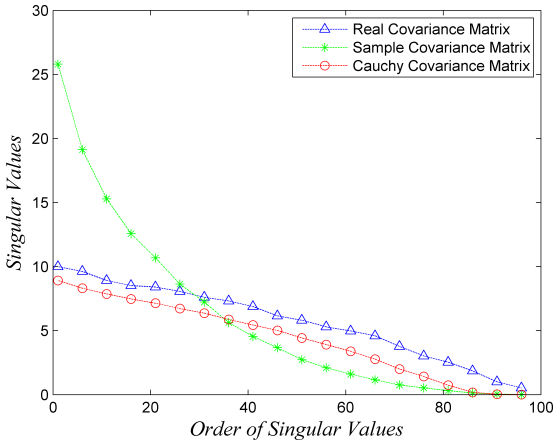


**FIGURE 3.** Singular values of the real covariance matrix, sample covariance matrix, and Cauchy covariance matrix.

Cauchy within-modal covariance matrices may be singular matrices, especially for a small amount of high-dimensional samples. Thus the projection regularization factor of Eq. (6) is embedded for enhancing the robustness of the small sample size problem [30]. The regularization parameter is set as 0.001 in this paper. Since correlation projection directions possess the scale invariance [35], solutions of Eq. (6) are unchanged when $\eta^{(p)T}R^{(pp)}\eta^{(p)}$ is constrained to one. Thus, Eq. (6) is equivalently reformulated into the below Cauchy multi-set correlation optimization problem:

$$\begin{aligned} &\max_{\{\eta^{(p)}\}_{p=1}^{M}} \sum_{p=1}^{M}\sum_{q=1}^{M} \eta^{(p)T}R_{\text{Cau}}^{(pq)}\eta^{(q)} \\ &s.t.\ \eta^{(p)T}R_{\text{Cau}}^{(pp)}\eta^{(p)} + \tau\eta^{(p)T}\eta^{(p)} = 1, \quad p = 1, 2, \ldots, M \end{aligned} \tag{7}$$

Using the Lagrange multiplier technique [36], Eq. (7) can be solved by optimizing a multivariate eigenvalue problem [23]. However, the multivariate eigenvalue problem has no exact solutions. Aiming at the issue, the constraint relaxation strategy [23] is utilized for constructing a relaxed version of Eq. (7):

$$\begin{aligned} &\max_{\{\eta^{(p)}\}_{p=1}^{M}} \sum_{p=1}^{M}\sum_{q=1}^{M} \eta^{(p)T}R_{\text{Cau}}^{(pq)}\eta^{(q)} \\ &s.t.\ \sum_{p=1}^{M}\left(\eta^{(p)T}R_{\text{Cau}}^{(pp)}\eta^{(p)} + \tau\eta^{(p)T}\eta^{(p)}\right) = 1 \end{aligned} \tag{8}$$

For the optimization of Eq. (8), its Lagrange function $\mathcal{L}\left(\eta^{(p)}\right)$ is first constructed as follows:

$$\begin{aligned} \mathcal{L}\left(\eta^{(p)}\right) = &\sum_{p=1}^{M}\sum_{q=1}^{M} \eta^{(p)T}R_{\text{Cau}}^{(pq)}\eta^{(q)} \\ &-\frac{\lambda}{2}\left(\sum_{p=1}^{M}\left(\eta^{(p)T}R_{\text{Cau}}^{(pp)}\eta^{(p)} + \tau\eta^{(p)T}\eta^{(p)}\right) - 1\right) \end{aligned} \tag{9}$$

where $\frac{\lambda}{2}$ is a Lagrange multiplier and $\lambda$ is the corresponding Lagrange multiplier factor. By setting $\frac{\partial L\left(\eta^{(p)}\right)}{\partial \eta^{(p)}} = 0$, we can obtain

$$\begin{aligned} \frac{\partial L\left(\eta^{(p)}\right)}{\partial \eta^{(p)}} &= \sum_{q=1}^{M} R_{\text{Cau}}^{(pq)}\eta^{(q)} - \lambda\left(R_{\text{Cau}}^{(pp)} + \tau I^{(p)}\right)\eta^{(p)} \\ &= 0 \quad (p = 1, 2, \ldots, M) \end{aligned} \tag{10}$$

where $I^{(p)} \in R^{d_p \times d_p}$ denotes the identity matrix. Eq. (10) can be equally translated into

$$\sum_{q=1}^{M} R_{\text{Cau}}^{(pq)}\eta^{(q)} = \lambda\left(R_{\text{Cau}}^{(pp)} + \tau I^{(p)}\right)\eta^{(p)} \quad (p = 1, \ldots, M) \tag{11}$$

We integrate the above $M$M equations of Eq. (11) into the following integration form:

$$\begin{aligned} &\begin{bmatrix} R_{\text{Cau}}^{(11)} & R_{\text{Cau}}^{(12)} & \cdots & R_{\text{Cau}}^{(1M)} \\ R_{\text{Cau}}^{(21)} & R_{\text{Cau}}^{(22)} & \cdots & R_{\text{Cau}}^{(2M)} \\ \vdots & \vdots & \ddots & \vdots \\ R_{\text{Cau}}^{(M1)} & R_{\text{Cau}}^{(M2)} & \cdots & R_{\text{Cau}}^{(MM)} \end{bmatrix} \begin{bmatrix} \eta^{(1)} \\ \eta^{(2)} \\ \vdots \\ \eta^{(M)} \end{bmatrix} \\ &= \lambda \begin{bmatrix} R_{\text{Cau}}^{(11)} + \tau I^{(1)} & & & \\ & R_{\text{Cau}}^{(22)} + \tau I^{(2)} & & \\ & & \ddots & \\ & & & R_{\text{Cau}}^{(MM)} + \tau I^{(M)} \end{bmatrix} \\ &\quad \times \begin{bmatrix} \eta^{(1)} \\ \eta^{(2)} \\ \vdots \\ \eta^{(M)} \end{bmatrix} \end{aligned} \tag{12}$$

The above equation is a generalized eigenvalue decomposition problem [35], and the Lagrange multiplier factor $\lambda$ can be also referred to as the eigenvalue in Eq. (12).

*Property 2:* The Lagrange multiplier factor $\lambda$ is equivalent to the optimized object, i.e. $\lambda = \sum_{p=1}^{M}\sum_{q=1}^{M} \eta^{(p)T}R_{\text{Cau}}^{(pq)}\eta^{(q)}$.

*Proof:* Both sides of Eq. (11) are multiplied by $\eta^{(p)T}$ at the same time. We can obtain

$$\begin{aligned} \sum_{q=1}^{M} \eta^{(p)T}R_{\text{Cau}}^{(pq)}\eta^{(q)} = \lambda\eta^{(p)T}\left(R_{\text{Cau}}^{(pp)} + \tau I^{(p)}\right)\eta^{(p)} \\ (p = 1, 2, \ldots, M) \end{aligned} \tag{13}$$

The $M$ equations of Eq. (13) can be integrated into

$$\begin{aligned} &\sum_{p=1}^{M}\sum_{q=1}^{M} \eta^{(p)T}R_{\text{Cau}}^{(pq)}\eta^{(q)} \\ &\quad = \lambda\sum_{p=1}^{M}\left(\eta^{(p)T}R_{\text{Cau}}^{(pp)}\eta^{(p)} + \tau\eta^{(p)T}\eta^{(p)}\right) \end{aligned} \tag{14}$$

From Eq. (8), we can find that $\sum_{p=1}^{M}\left(\eta^{(p)T}R_{\text{Cau}}^{(pp)}\eta^{(p)} + \tau\eta^{(p)T}\eta^{(p)}\right)$ is one. Eq. (14) can be equivalently converted into

$$\sum_{p=1}^{M}\sum_{q=1}^{M} \eta^{(p)T}R_{\text{Cau}}^{(pq)}\eta^{(q)} = \lambda \tag{15}$$

Thus this property is turned out to be right.

According to Property 2, the optimization object of $\left\{\eta^{(p)}\right\}_{p=1}^{M}$ can equivalently translate from maximizing $\sum_{p=1}^{M}\sum_{q=1}^{M} \eta^{(p)T}R_{Cau}^{(pq)}\eta^{(q)}$ to maximizing $\lambda$. In other words, the optimal solutions to $\left\{\eta^{(p)}\right\}_{p=1}^{M}$ are eigenvectors corresponding to the largest $d$ eigenvalues in Eq. (12). Thus, the first $d$ pairs of correlation projection directions will be got by solving Eq. (12), i.e. $\eta_k^T = \left[\eta_k^{(1)T}, \eta_k^{(2)T}, \ldots, \eta_k^{(M)T}\right] \in$

**TABLE 1. The detailed steps of cauchy multi-set canonical Correlations (CauMCCs).**

Input: The training datasets $\{Z^{(p)} = [z_1^{(p)}, z_2^{(p)}, \ldots, z_N^{(p)}] \in R^{d_p \times N}\}_{p=1}^{M}$
The testing datasets $\{\tilde{Z}^{(p)} = [\tilde{z}_1^{(p)}, \tilde{z}_2^{(p)}, \ldots, \tilde{z}_n^{(p)}] \in R^{d_p \times n}\}_{p=1}^{M}$
Output: The fused training low-dimensional datasets $\{Y^{(p)} = [y_1^{(p)}, y_2^{(p)}, \ldots, y_N^{(p)}]\}_{p=1}^{M}$
The fused testing low-dimensional datasets $\{\tilde{Y}^{(p)} = [\tilde{y}_1^{(p)}, \tilde{y}_2^{(p)}, \ldots, \tilde{y}_n^{(p)}]\}_{p=1}^{M}$
1: for $p = 1, \ldots, M$ do
2: for $q = 1, \ldots, M$ do
3: Construct the Cauchy covariance matrices $R_{\text{Cau}}^{(pp)}$ and $R_{\text{Cau}}^{(pq)}$ by Eq. (4) and Eq. (5);
4: end for
5: end for
6: Compute the correlation projection matrices $\{\Theta^{(p)}\}_{p=1}^{M}$ by solving Eq. (12);
7: Obtain the fused low-dimensional datasets $\{Y^{(p)}\}_{p=1}^{M}$ and $\{\tilde{Y}^{(p)}\}_{p=1}^{M}$ by Eq. (16) and Eq. (17)

$R^{\left(\sum_{p=1}^{M} d_p\right) \times 1}$, $k = 1, 2, \ldots, d$. According to $\left\{\eta_k^{(p)T}\right\}_{p=1}^{M}$, we can construct the correlation projection matrix $\Theta^{(p)} = \left[\eta_1^{(p)}, \eta_2^{(p)}, \ldots, \eta_d^{(p)}\right] \in R^{d_p \times d}$ of $Z^{(p)} (p = 1, 2, \ldots, M)$, and then projected low-dimensional data of $Z^{(p)}$ can be obtained by $\Theta^{(p)T} Z^{(p)} \in R^{d \times N}$.

These projected low-dimensional data in the Cauchy coherent fusion subspace have the maximal between-modal Cauchy correlations. That is, different low-dimensional modality data corresponding to the same object have the good consistency of the fusion structure and the fusion information. Thus, in the Cauchy coherent fusion subspace, the projected low-dimensional data of the training datasets can be fused with

the help of the parallel fusion strategy [25]:

$$y_k^{(p)} = \sum_{p=1}^{M} \Theta^{(p)T} z_k^{(p)} \tag{16}$$

where $z_k^{(p)}$ is the $k$th ($k = 1, 2, \ldots, N$) projected low-dimensional data of $Z^{(p)}$, and $y_k^{(p)}$ is the $k$th fused low-dimensional data of the fused training low-dimensional dataset $Y^{(p)} = [y_1^{(p)}, y_2^{(p)}, \ldots, y_N^{(p)}]$ and $\left\{z_k^{(p)}\right\}_{p=1}^{M}$ are a group of multi-modal training samples corresponding to the $k$th ($k = 1, 2, \ldots, N$) object. For the testing datasets $\left\{\tilde{Z}^{(p)} = [\tilde{z}_1^{(p)}, \tilde{z}_2^{(p)}, \ldots, \tilde{z}_n^{(p)}]\right\}_{p=1}^{M}$, the fused testing low-dimensional dataset $\left\{\tilde{Y}^{(p)} = [\tilde{y}_1^{(p)}, \tilde{y}_2^{(p)}, \ldots, \tilde{y}_n^{(p)}]\right\}_{p=1}^{M}$ can be obtained:

$$\tilde{y}_k^{(p)} = \sum_{p=1}^{M} \Theta^{(p)T} \tilde{z}_k^{(p)} \tag{17}$$

Finally, classification and recognition of multi-modal data will be realized by implementing the classifier on these fused low-dimensional data. To intuitively exhibit our proposed method, we give the detailed steps of CauMCCs in Table 1.

## IV. EXPERIMENTS

On one synthetic dataset and three real-world image datasets, we design some experiments for evaluating the image recognition performance of our method. CBSR NIR image dataset, Georgia Tech Face dataset, and Semeion handwritten image dataset are three frequently-used image datasets in image recognition, and our experiments are implemented in these datasets. In essence, these image datasets belong to single-modal datasets. Therefore, the modality strategy [25] is utilized to obtain three modalities corresponding to each image. More concretely, three low-frequency sub-images of each image can be obtained with the help of Symlets, Daubechies, and Coiflets orthonormal wavelet transforms, and then K–L transform is employed to reduce the dimension of sub-images to 150 for each category of sub-images on the CBSR and Georgia Tech image datasets. The dimension of sub-images are reduced to 100 on the Semeion dataset because the dimension of some sub-images on the Semeion dataset is less than 150. The three-category low-dimensional data (i.e. three modality data) can be obtained from each image. Our method contains a scale parameter $c$, and the parameter is chosen from the range of 0.1 to 2 with an interval of 0.1. We compare our method with two representative methods, i.e. MCCA and CICCA. In the final recognition stage, recognition rates of each method are determined by the nearest neighbor classifier [37], and the tables of this section tabulate the best recognition rates under all the possible dimensions.

### A. EXPERIMENTS ON SYNTHETIC DATA VISUALIZATION

To visually exhibit and analyze the differences between our method and these compared methods, we construct a three-modal synthetic dataset, and data distribution of projected data are visualized in a three-dimensional correlation coherent subspace. The three-modal synthetic dataset $\left\{F^{(p)} = [F_1^{(p)}, F_2^{(p)}]\right\}_{p=1}^{3}$ includes 400 samples with two classes, where $F_1^{(p)}$ is the sample set of the first class with 200 samples and $F_2^{(p)}$ represents the 200 samples of the second class. More concretely, $F_1^{(1)}$ and $F_2^{(1)}$ respectively follow Gaussian distributions $N(\mu_1, \sigma_1)$ and $N(\mu_2, \sigma_2)$, where $\mu_1 = [20, -20]^T$, $\sigma_1 = \begin{bmatrix} 15 & 3.75 \\ 3.75 & 2 \end{bmatrix}$, $\mu_2 = [10, -5]^T$, and

**TABLE 2.** The experimental results on the CBSR NIR image dataset.

| | 6 training samples per class | 8 training samples per class | 10 training samples per class | 12 training samples per class |
|---|---|---|---|---|
| CauMCCs | 94.76±0.90 | 97.56±0.84 | 98.70±0.69 | 98.83±0.73 |
| CICCA | 87.26±4.52 | 93.14±1.42 | 94.73±2.08 | 94.96±0.89 |
| MCCA | 84.90±1.550 | 96.53±0.82 | 97.73±0.89 | 98.29±0.95 |

A±B : A denotes the average recognition rate (%) and B is the corresponding standard deviation

$\sigma_2 = \begin{bmatrix} 15 & 3.75 \\ 3.75 & 2 \end{bmatrix}$. The second modal data $F_1^{(2)}$ and $F_2^{(2)}$ are obtained by the following transformations: $F_1^{(2)} = \omega_1^T F_1^{(1)} + \xi$ and $F_2^{(2)} = \omega_2^T F_2^{(1)} + \xi$, where $\omega_1 = \begin{bmatrix} 0.6 & -\sqrt{0.5} \\ 0.8 & -\sqrt{0.5} \end{bmatrix}$, $\omega_2 = \begin{bmatrix} -0.5 & \sqrt{1.5} \\ -\sqrt{1.5} & -0.5 \end{bmatrix}$, and $\xi$ is a Gaussian noise matrix. The third modal data $F_1^{(3)}$ and $F_2^{(3)}$ will be generated by $F_1^{(3)} = \omega_1^T F_1^{(2)} + \xi$ and $F_2^{(3)} = \omega_2^T F_2^{(2)} + \xi$.

Our method and these compared methods are implemented on the synthetic dataset, and Fig.4 visualizes the first data pair $(\Theta^{(1)T} F^{(1)}, \Theta^{(2)T} F^{(2)}, \Theta^{(3)T} F^{(3)})$ extracted by different methods in the three-dimensional space. In Fig.4, the projected data of the three-dimensional space show linear relationships, which can reveal linear correlations of the three-modal datasets. However, different methods exhibit different discriminative performance in Fig.4. The projected data of CICCA and MCCA have some overlaps in two classes. For CauMCCs, two classes are completely separated, and there exists a clear decision boundary. In a word, how to learn a discriminative space from samples with noises is a challenging task under unsupervised cases, and our method can effectively improve the discriminative power to achieve better class separability, which can be intuitively observed.

### B. EXPERIMENTS ON THE CBSR NIR IMAGE DATASET

In the CBSR NIR image dataset, a NIR camera with active NIR lighting took pictures of 197 persons, and 3940 face images with 480 by 640 pixels were obtained. In our experiments, we utilize a subset of the dataset, i.e. all the NIR images of the first thirty individuals. Each person randomly chooses $b(b = 6, 8, 10, 12)$ images for training, and the remaining images are referred to as testing images. The random experiments are independently implemented ten times, and Table 2 tabulates the average recognition rates and the corresponding standard deviations of ten random experiments.

As covariance-based multi-modal subspace fusion methods, MCCA and CICCA have the same objective function and similar constraints. For CICCA, the within-modal global scatter of each modality is constrained to one. The correlation optimization problem with these constraints is usually

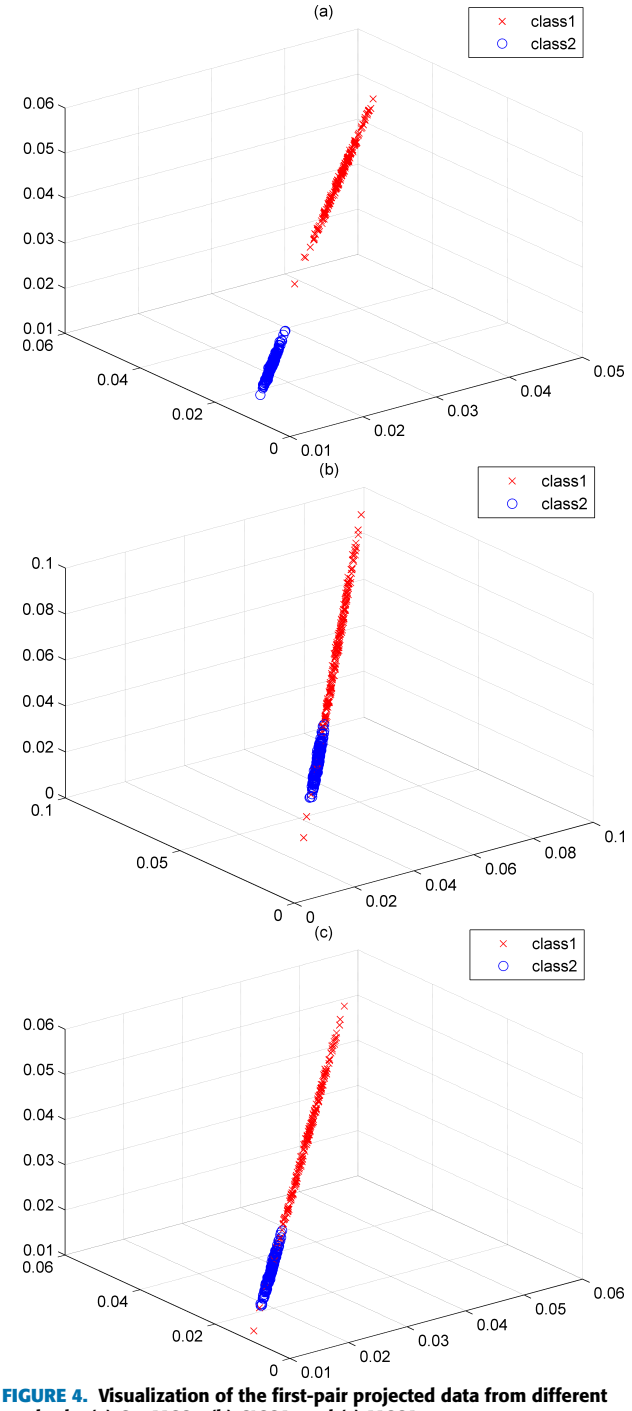


**FIGURE 4.** Visualization of the first-pair projected data from different methods: (a) CauMCCs, (b) CICCA, and (c) MCCA.

transformed into a multivariate eigenvalue problem that has no exact solutions. However, when the number of modalities is limited to three, the three-modal correlation optimization problem is derived into three general eigenvalue problem, which is capable of obtaining exact solutions. Different from CICCA, MCCA employs a relaxed version of constraints, i.e. the sum of the within-modal global scatters of all the modalities is equivalent to one. The correlation optimization problem of MCCA can be obtained exact solutions of correlation projection directions. Although CICCA solves the

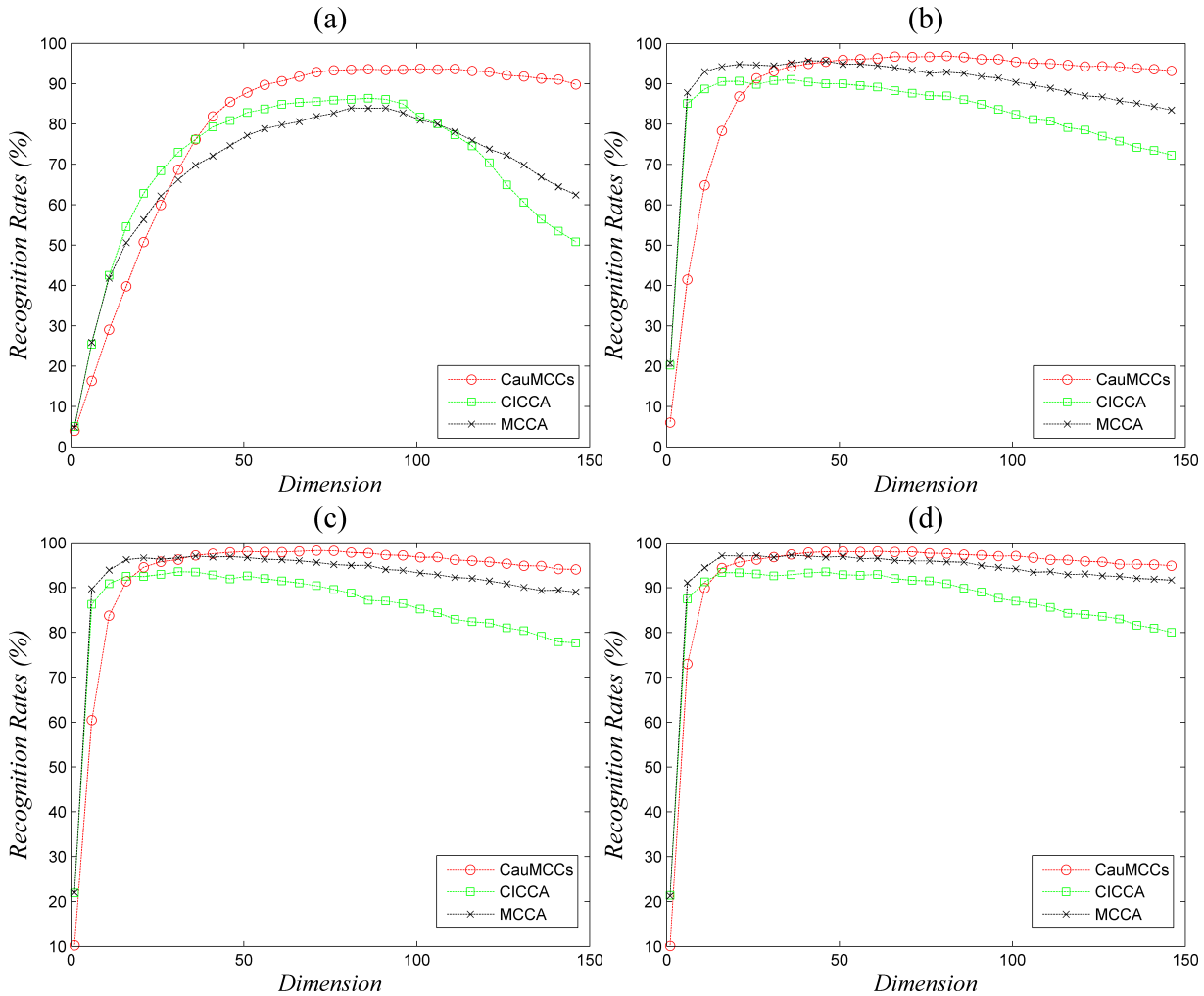


**FIGURE 5.** Curves of recognition rates under different dimensions when the number of training samples in each class is respectively (a) six, (b) eight, (c) ten, and (d) twelve on the CBSR NIR image dataset.

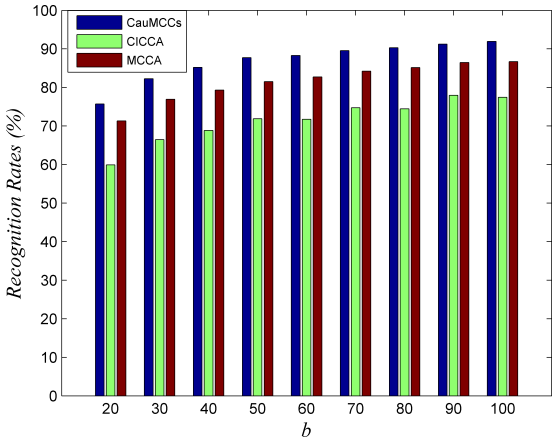


**FIGURE 6.** average recognition rates under the values of c (i.e. the number of the training samples per class) on the Semeion handwritten image dataset.

issue of inexact solutions through theoretical derivation, different optimization models of MCCA and CICCA lead to different recognition performance. For the CBSR NIR image dataset, MCCA possesses higher recognition rates than CICCA in most cases of Table 2. In the two methods, within-modal covariance matrices capture the global scatter information of each modality, and between-modal covariance matrices reveal correlations between different modalities. Between-modal and within-modal covariance matrices are constructed by training samples, and noise information and insufficient sample size will cause the deviation of these covariance matrices, which will weaken the recognition performance of MCCA and CICCA. Our method exploits the relaxed version of constraints, and the correlation projection directions of our method also have exact solutions by the theoretical derivation. Different from the two methods, our method corrects between-modal and within-modal covariance matrices using the Cauchy estimate theory, and the Cauchy covariance matrices from our method are closer to real covariance matrices. Also, our method constructs a novel Cauchy correlation optimization modal with Cauchy covariance matrices, and the Cauchy coherent fusion subspace with well discriminative power is learned by our method. These are important reasons why our method has higher recognition rates than these compared methods.

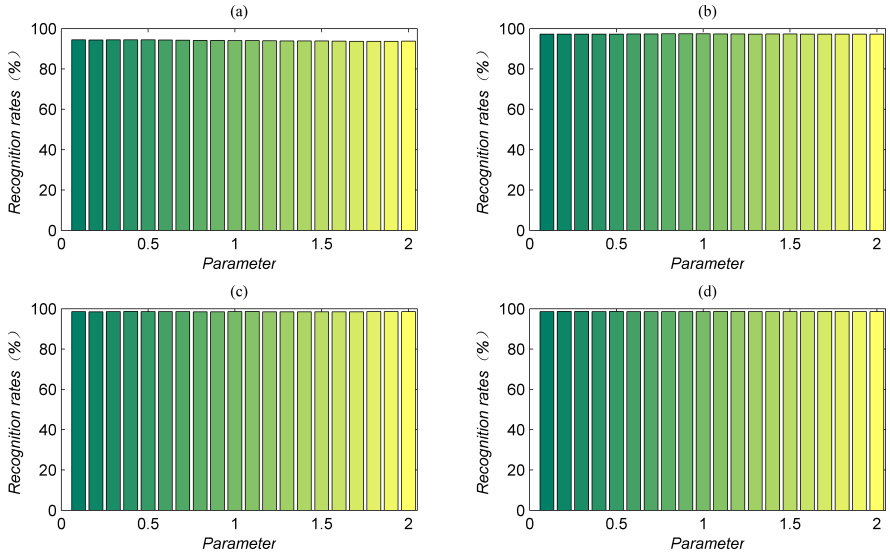


**FIGURE 7.** recognition rates versus the scale parameter c on the CBSR NIR image dataset when the number of the training samples per class is (a) six, (b) eight, (c) ten, and (d) twelve.

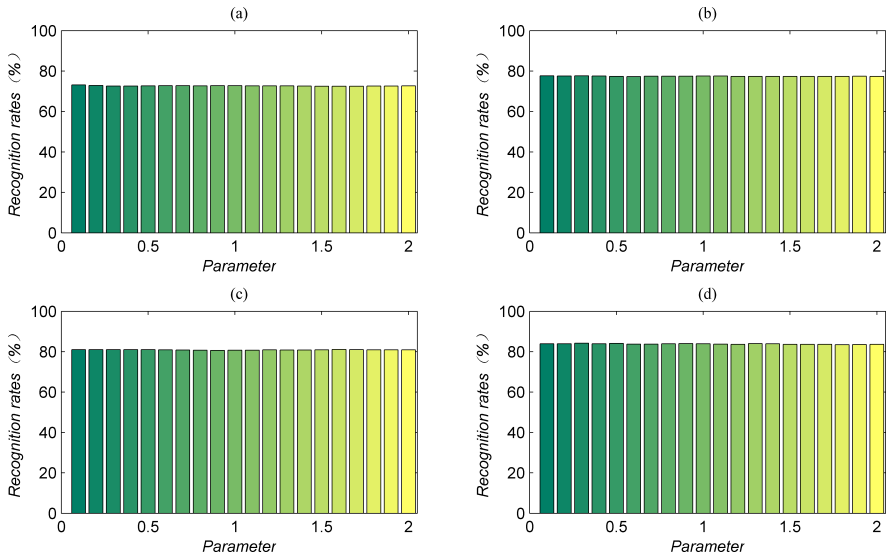


**FIGURE 8.** recognition rates versus the scale parameter c on the Georgia Tech Face dataset when the number of the training samples per class is (a) six, (b) eight, (c) ten, and (d) twelve.

To analyze the influence of dimension on recognition rates, we further show the recognition performance of each method under different dimensions in Fig.5. Besides, the dimension of covariance matrices in Fig.2 is determined by the raw training samples. The higher the dimension of the raw training samples is, the more the noise and redundancy information hidden in the covariance matrices will be. Different from the dimension of Fig.2, the dimension in Fig.5 is the number of the correlation projection directions, i.e. the dimension of the learned coherent subspace. The learned discriminative structures will be more complete with the increasing of the correlation projection directions. The two categories of dimensions in Fig.2 and Fig.5 are different. The small dimension of the raw training samples and the small number of the correlation projection directions don't have a causal relationship of the recognition rates. Next, we focus on the analysis of Fig.5. For all the methods of Fig.5, recognition rates sharply improve when the dimension begins to increase. With the continuous increase of the dimension, recognition rates of our method trend to be more stable than those of the compared methods. Although the dimension cannot be exactly determined, the stabilization of the recognition rates corresponding to the relatively high dimensions gives the possibility that the dimension can be directly set as a large constant in the real-world image recognition tasks. Thus our method possesses better practicality and operability for real-world image recognition.

### C. EXPERIMENTS ON THE GEORGIA TECH FACE DATASET

Images of the Georgia Tech Face dataset have different lighting conditions and different facial expressions. Resolutions of these images are normalized to 70 by 50 pixels. The

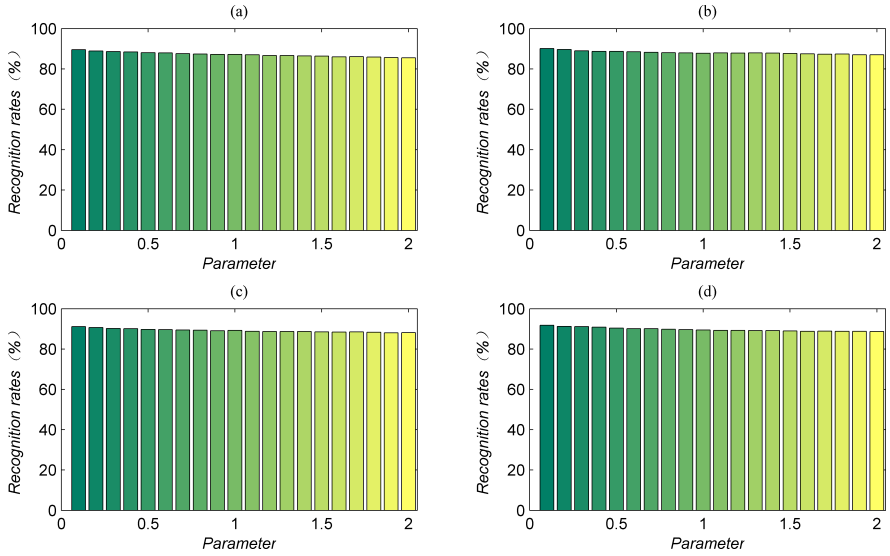


**FIGURE 9.** recognition rates versus the scale parameter c on the Semeion handwritten image dataset when the number of the training samples per class is (a) seventy, (b) eighty, (c) ninety, and (d) One hundred.

images are taken from fifty individuals. Table 3 reports the average recognition rates (%) of ten random experiments when training samples are random $b(b = 6, 8, 10, 12)$ images per class. In Table 3, CICCA still shows the worse recognition performance than MCCA, and our method has the highest recognition rates, which are consistent with the experimental results of the CBSR dataset. The standard deviation in the random experiments reflects the difference degree of the recognition rates under different training samples. The smaller the standard deviation is, the more stable the recognition performance is for different training samples. The stability for random training samples is beneficial to many real-world applications, and thus we further tabulate the standard deviation corresponding to the average recognition rates in Table 3. In this table, the recognition performance of our method is more stable than that of the compared methods under different random experiments, which reveals that our method possesses the better robustness than the other methods in sample random.

### D. EXPERIMENTS ON THE SEMEION HANDWRITTEN IMAGE DATASET

The Semeion handwritten image dataset is a handwritten digit image dataset widely used in image recognition, and the handwritten digit images are written by 80 individuals. Fig.6 intuitively exhibits the average recognition rates (%) of ten sample random experiments when the training samples are random $b(b = 20, 30, 40, 50, 60, 70, 80, 90, 100)$ samples of each class. The experimental results in Fig.6 are consistent with those on the above datasets. As the number of the training samples is increasing, each method shows an increasing tendency in recognition rates, and the recognition rates of our method are always the highest. In summary, extensive experimental results on all the datasets can give a reasonable observation that our method can improve the recognition performance of images.

**TABLE 3.** The experimental results on the Georgia Tech Face dataset.

| | 6 training samples per class | 8 training samples per class | 10 training samples per class | 12 training samples per class |
|---|---|---|---|---|
| CauMCCs | 73.36±1.52 | 78.06±1.24 | 81.64±1.38 | 84.73±1.76 |
| CICCA | 47.98±4.39 | 56.26±4.06 | 60.24±3.22 | 63.27±3.07 |
| MCCA | 60.42±1.71 | 66.91±1.56 | 71.00±1.65 | 73.80±1.91 |

A±B： A denotes the average recognition rate (%) and B is the corresponding standard deviation

### E. ANALYSIS OF THE SCALE PARAMETER

In pattern recognition, the exact determination of parameters is usually an open problem. Our method also includes a parameter, i.e. the scale parameter $c$. The parameter is crucial to the construction of Cauchy covariance matrices. In Fig.7, Fig.8, and Fig.9, we give the recognition rates corresponding to the different values of the parameter. The variation of the recognition rates under the different parameter values can reflect the impact of the parameter on the recognition rates. From the three figures, we can find that the recognition rates have a small variation with the change of the parameter. Although the parameter values cannot be exactly determined on different datasets, the parameter has a small impact on the recognition rates. The relative stability of the recognition rates on the parameter not only provides the possibility that the parameter is directly set as a fixed value but also remedies

the disadvantages that it is difficult to determine the exact values of the parameter in real-world applications.

## V. CONCLUSION

Image samples belong to high-dimensional data with a lot of redundant information and noises, and the number of image samples is limited in the training stage of many real-world applications. Thus, between-modal and within-modal covariance matrices based on image samples will seriously deviate from real ones, which will weaken the image recognition performance of correlation analysis methods. To re-estimate covariance matrices, we correct the singular values of sample covariance matrices by the Cauchy estimate theory. As far as we know, it is novel for this singular value correction, and the corrected Cauchy covariance matrices are closer to real covariance matrices, which is beneficial to improve the recognition performance of the correlation analysis methods. Then, by maximizing Cauchy correlations between different modalities and constraining Cauchy scatters of within-modal data, we further propose the novel multi-modal subspace fusion method, i.e. CauMCCs. In our method, the Cauchy coherent fusion subspace with well class separability can be learned from a small number of images. We design some experiments on one synthetic dataset and three biometric image datasets, and the good experimental results reveal the superiority of our method. Besides, Cauchy covariance matrices of our method may be also utilized in other methods with covariance matrices, such as locality preserving projections, principal component analysis, and linear discriminant analysis. In the future, we will explore how to improve the performance of these methods using Cauchy covariance matrices.

## REFERENCES

[1] Z. Fu, Y. Zhao, Y. Xu, L. Xu, and J. Xu, "Gradient structural similarity based gradient filtering for multi-modal image fusion," *Inf. Fusion*, vol. 53, pp. 251–268, Jan. 2020.

[2] X. Yang, L. Weifeng, W. Liu, and D. Tao, "A survey on canonical correlation analysis," *IEEE Trans. Knowl. Data Eng.*, early access, Dec. 9, 2019, doi: 10.1109/TKDE.2019.2958342.

[3] Q. Liu, Y. Jiao, Y. Miao, C. Zuo, X. Wang, A. Cichocki, and J. Jin, "Efficient representations of EEG signals for SSVEP frequency recognition based on deep multiset CCA," *Neurocomputing*, vol. 378, pp. 36–44, Feb. 2020.

[4] Q. Jiang and X. Yan, "Multimode process monitoring using variational Bayesian inference and canonical correlation analysis," *IEEE Trans. Autom. Sci. Eng.*, vol. 16, no. 4, pp. 1814–1824, Oct. 2019.

[5] A. Bertrand and M. Moonen, "Distributed canonical correlation analysis in wireless sensor networks with application to distributed blind source separation," *IEEE Trans. Signal Process.*, vol. 63, no. 18, pp. 4800–4813, Sep. 2015.

[6] Z. Li, X.-Y. Jing, F. Wu, X. Zhu, B. Xu, and S. Ying, "Cost-sensitive transfer kernel canonical correlation analysis for heterogeneous defect prediction," *Automated Softw. Eng.*, vol. 25, no. 2, pp. 201–245, Jun. 2018.

[7] K. Hong, G. Liu, W. Chen, and S. Hong, "Classification of the emotional stress and physical stress using signal magnification and canonical correlation analysis," *Pattern Recognit.*, vol. 77, pp. 140–149, May 2018.

[8] A. de Cheveigné, G. M. Di Liberto, D. Arzounian, D. D. E. Wong, J. Hjortkjær, S. Fuglsang, and L. C. Parra, "Multiway canonical correlation analysis of brain data," *NeuroImage*, vol. 186, pp. 728–740, Feb. 2019.

[9] 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.

[10] T. Sun, S. Chen, J. Yang, and P. Shi, "A supervised combined feature extraction method for recognition," in *Proc. IEEE Int. Conf. Data Mining*, Pisa, Italy, Dec. 2008, pp. 1043–1048.

[11] S. Wang, J. Lu, X. Gu, B. A. Weyori, and J.-Y. Yang, "Unsupervised discriminant canonical correlation analysis based on spectral clustering," *Neurocomputing*, vol. 171, pp. 425–433, Jan. 2016.

[12] T. Sun and S. Chen, "Locality preserving CCA with applications to data visualization and pose estimation," *Image Vis. Comput.*, vol. 25, no. 5, pp. 531–543, May 2007.

[13] J. Chen, G. Wang, Y. Shen, and G. B. Giannakis, "Canonical correlation analysis of datasets with a common source graph," *IEEE Trans. Signal Process.*, vol. 66, no. 16, pp. 4398–4408, Aug. 2018.

[14] L. Wei and F. Xu, "Local CCA alignment and its applications," *Neurocomputing*, vol. 89, pp. 78–88, Jul. 2012.

[15] Q. Liu and C. Wang, "Within-component and between-component multikernel discriminating correlation analysis for colour face recognition," *IET Comput. Vis.*, vol. 11, no. 8, pp. 663–674, Dec. 2017.

[16] X. Xing, K. Wang, T. Yan, and Z. Lv, "Complete canonical correlation analysis with application to multi-view gait recognition," *Pattern Recognit.*, vol. 50, pp. 107–117, Feb. 2016.

[17] 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.

[18] W.-P. Li, J. Yang, and J.-P. Zhang, "Uncertain canonical correlation analysis for multi-view feature extraction from uncertain data streams," *Neurocomputing*, vol. 149, pp. 1337–1347, Feb. 2015.

[19] X. Jing, S. Li, C. Lan, D. Zhang, J. Yang, and Q. Liu, "Color image canonical correlation analysis for face feature extraction and recognition," *Signal Process.*, vol. 91, no. 8, pp. 2132–2140, Aug. 2011.

[20] N. M. Correa, T. Eichele, T. Adalı, 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, May 2010.

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

[22] 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, Feb. 2019.

[23] 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.

[24] 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.

[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] 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, Oct. 2015.

[27] N. E. D. Elmadany, Y. He, and L. Guan, "Information fusion for human action recognition via biset/multiset globality locality preserving canonical correlation analysis," *IEEE Trans. Image Process.*, vol. 27, no. 11, pp. 5275–5287, Nov. 2018.

[28] N. Desai, A.-K. Seghouane, and M. Palaniswami, "Algorithms for two dimensional multi set canonical correlation analysis," *Pattern Recognit. Lett.*, vol. 111, pp. 101–108, Aug. 2018.

[29] H.-K. Ji, Q.-S. Sun, Y.-H. Yuan, and Z.-X. Ji, "C 2 DMCP: View-consistent collaborative discriminative multiset correlation projection for data representation," *J. Vis. Commun. Image Represent.*, vol. 40, pp. 393–405, Oct. 2016.

[30] N. E. D. Elmadany, Y. He, and L. Guan, "Multimodal learning for human action recognition via Bimodal/Multimodal hybrid centroid canonical correlation analysis," *IEEE Trans. Multimedia*, vol. 21, no. 5, pp. 1317–1331, May 2019.

[31] C. Xu, D. Tao, and C. Xu, "Multi-view intact space learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 12, pp. 2531–2544, Dec. 2015.

[32] 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.

[33] 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.
[34] P. Xu, G. N. Brock, and R. S. Parrish, "Modified linear discriminant analysis approaches for classification of high-dimensional microarray data," *Comput. Statist. Data Anal.*, vol. 53, no. 5, pp. 1674–1687, Mar. 2009.
[35] S. Su, X. Fang, G. Yang, B. Ge, and Y. Zhu, "Self-balanced multi-view orthogonality correlation analysis for image feature learning," *Infr. Phys. Technol.*, vol. 100, pp. 44–51, Aug. 2019.
[36] Y. Qin, L. Bruzzone, B. Li, and Y. Ye, "Cross-domain collaborative learning via cluster canonical correlation analysis and random walker for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 57, no. 6, pp. 3952–3966, Jun. 2019.
[37] H. Tan, Y. Gao, J. Du, and S. Yang, "Eigenspectrum regularization on Grassmann discriminant analysis with image set classification," *IEEE Access*, vol. 7, pp. 150792–150804, 2019.

**YANMIN ZHU** received the M.S. degree from the School of Mechanical and Electronic Engineering, Shandong Agricultural University, Tai'an, 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, and image processing.

**TIANHAO PENG** received the Ph.D. degree from Zhejiang University, Hangzhou, China. He is 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 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.

**CHANGPENG LI** received the M.S. degree from the Anhui University of Science and Technology, Huainan, China, in 2018, where he is currently pursuing the Ph.D. degree in mine electromechanical engineering. His research interests include coal rock recognition, signal processing, and intelligent control.