IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received May 5, 2020, accepted June 16, 2020, date of publication June 19, 2020, date of current version July 1, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.3003619

Supervised Fractional-Order Embedding Geometrical Multi-View CCA (SFGMCCA) for Multiple Feature Integration

KEISUKE MAEDA[1], (Member, IEEE), YOSHIKI ITO[2], (Member, IEEE), TAKAHIRO OGAWA[1], (Senior Member, IEEE), AND MIKI HASEYAMA[1], (Senior Member, IEEE)

[1]Faculty of Information Science and Technology, Hokkaido University, Sapporo 060-0814, Japan
[2]Graduate School of Information Science and Technology, Hokkaido University, Sapporo 060-0814, Japan

Corresponding author: Keisuke Maeda (maeda@lmd.ist.hokudai.ac.jp)

This work was supported in part by the MIC/SCOPE under Grant 181601001.

ABSTRACT Techniques for integrating different types of multiple features effectively have been actively studied in recent years. Multiset canonical correlation analysis (MCCA), which maximizes the sum of pairwise correlations of inter-view (i.e., between different features), is one of the powerful methods for integrating different types of multiple features, and various MCCA-based methods have been proposed. This work focuses on a supervised MCCA variant in order to construct a novel effective feature integration framework. In this paper, we newly propose supervised fractional-order embedding geometrical multi-view CCA (SFGMCCA). This method constructs not only the correlation structure but also two types of geometrical structures of intra-view (i.e., within each feature) and inter-view simultaneously, thereby realizing more precise feature integration. This method also supports the integration of small sample and high-dimensional data by using the fractional-order technique. We conducted experiments using four types of image datasets, i.e., MNIST, COIL-20, ETH-80 and CIFAR-10. Furthermore, we also performed an fMRI dataset containing brain signals to verify the robustness. As a result, it was confirmed that accuracy improvements using SFGMCCA were statistically significant at the significance level of 0.05 compared to those using conventional representative MCCA-based methods.

INDEX TERMS Feature integration, multi-view, canonical correlation analysis, fractional-order technique, geometrical structure.

I. INTRODUCTION

In research fields on computer vision, pattern recognition and data visualization, many researchers have studied techniques to improve their accuracies by using different types of features. This is because multi-modal feature representation can handle a larger variety of characteristics than uni-modal feature representation can. The method to handle multi-modal data is called multi-view learning, and it is divided into two representative methods [1], latent subspace learning and co-training [2]. Among the methods, canonical correlation analysis (CCA) [3], one of the latent subspace learning methods, has been widely used for the integration of two kinds of features. By using CCA, two kinds of features can be integrated in a lower-dimensional canonical space where a pair correlation between them is maximized even if the numbers of their dimensions are different. CCA realizes effective feature integration by projecting both of them into the space.

Several extended approaches based on CCA have been reported. In order to handle high-dimensional data compared with the number of samples, regularized CCA (RCCA) [4] with a regularization term in the objective function was proposed. RCCA can suppress over-fitting by using the term. In order to integrate features more accurately, two types of methods based on unique approaches were proposed, i.e., locality preserving CCA (LPCCA) [5], [6] and discriminative CCA (DCCA) [7]. For instance, LPCCA is a method extending CCA by locality preserving projection (LPP) [8], which takes into account the neighbor structure (e.g., sample number-based k-neighbor and distance-based ϵ-neighbor) for

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

each sample. By introducing LPP, it is possible to obtain projections considered non-linear structures in the original spaces because samples placed close together in the spaces are placed close together in the canonical one as well. LPCCA constructs a graph-based geometrical structure rather than a conventional correlation structure. Moreover, supervised LPCCA (SLPCCA) [9] preserves the neighbor structure in the same class only. SLPCCA can consider inter-class separability by setting a new constraint. On the other hand, DCCA is a method extending CCA by linear discriminant analysis (LDA) [10]–[12]. This method calculates projections by simultaneously taking into account minimization of intra-class scatter and maximization of inter-class scatter. In addition, by using LPCCA and DCCA collaboratively, they have been extended to discriminative locality preserving CCA (DLPCCA) [13]. Furthermore, for considering the non-linear structure directly, kernel CCA [14] was proposed, and Sun *et al.* derived kernelized version of DCCA (KDCCA) [7]. Moreover, kernelized version of DLPCCA (KDLPCCA) [15] which can consider the non-linear structure, locality preserving and discriminant analysis was proposed.

Methods that can integrate not only two features but also three or more features have been required. For instance, in research on multimedia analysis, we extract various data, such as video, audio, text and sensor signals, and integrate them comprehensively to let the computer recognize and understand the semantic meaning of media information. In order to develop such technologies and applications, several methods that can effectively integrate three or more features have been proposed. Multiset CCA (MCCA) [16], [17], which maximizes the sum of three or more kinds of pairwise correlations, has recently been proposed.[1] MCCA has been applied to a higher level of research fields since the method can integrate not only two kinds of features but also three or more kinds of features. In order to suppress over-fitting, regularized MCCA (RMCCA) [18] with a Ridge type of regularization term was proposed. On the other hand, MCCA has been extended to kernel MCCA (KMCCA) [19] for considering the non-linear structure. However, since KMCCA only takes advantage of one non-linear function in each feature, Yuan *et al.* presented a multiple nonlinear function–based MCCA [20], which combines multiple kernels in each feature. Recently, Chen *et al.* proposed a kernel MCCA method with a single graph regularizer (GKMCCA) [21] for multiple feature integration, which seeks for a common low-dimensional representation shared by all features. Some kernel-based approaches considering the multi-view features and discriminant analysis have been proposed. Multi-view LDA (MLDA) [22], which combines CCA and LDA, can maximize the discrimination and correlation between these features. By an extension of MLDA, multi-view uncorrelated LDA (MULDA) [23] based on ULDA [24] can extract effective features with minimum redundancy, and its extended version, kernel multi-view discriminant analysis [25] providing more discrimination and generalization ability, has been proposed. Furthermore, MCCA has been extended to Laplacian multiset canonical correlations (LapMCCs) [26] using LPP as well as CCA extended to LPCCA. Pairwise correlation is basically used as a way for calculation of the methods described above. Tensor CCA (TCCA) [27], which calculates a high order correlation by constructing a covariance tensor using all features simultaneously, has also been proposed.

Moreover, a creative and powerful method, supervised multi-view CCA (sMVCCA) [28], to improve the accuracy by using class information and the other features collaboratively was reported. In this method, the class information is used as a matrix obtained from one-hot vectors, and it is handled as equivalent to the other features. This method can successfully calculate more accurate canonical features by maximizing the sum of all pairwise correlations as with MCCA. Actually, it has been reported that some feature integration methods [29] based on sMVCCA was effective. We also focused on the integration capability and previously proposed a method with better performance based on sMVCCA, or supervised fractional-order embedding multi-view CCA (SFEMCCA) [30]. SFEMCCA is a supervised method that can deal with data with noise, a small number of samples and a large number of dimensions. Furthermore, LPbSCCA combining the semi-supervised approach and multi-view CCA was proposed [31]. LPbSCCA introduces a label propagation algorithm based on sparse representation to infer label information for unlabeled data. Furthermore, discriminative multiple canonical correlation analysis (DMCCA) [32], which can be constructed by adding the LDA-based approach to the MCCA-based approach was proposed. From the above, several unique and interesting supervised CCA-based methods for multi-view data have been proposed. Many CCA methods such as MCCA and sMVCCA assume calculation under the ideal conditions of the number of samples being sufficiently larger than the number of dimensions and little noise being included in the data. However, in eigenvalue decomposition (EVD) and singular value decomposition (SVD) of a covariance matrix calculated from non-ideal data, it has been established that the eigenvalues and the singular values are greater than those calculated from ideal data [33]–[35]. By applying the ''fractional-order technique'' [36]–[38] to sMVCCA, we solved the problem and showed the effectiveness of SFEMCCA compared to that of sMVCCA, SLPCCA (LapMCCs) and DLPCCA from experimental results.[2]

However, a subspace learning method collaboratively utilizing correlation structure such as CCA, MCCA, and sMVCCA and a geometrical structure such as LPCCA and LapMCCs had not been proposed until graph regularized MCCs (GrMCCs) [39] was proposed. The method constructs not only the correlation structure but also the geometrical

[1] ''Multiset'' has the same meaning as ''multiview'' and ''multi-view''.

[2] In the experiments, SLPCCA can be regarded to be the same as LapMCCs since two kinds of features and a label feature were used.

structure, which enables locality structure preservation and discriminant analysis, and embeds them into the objective function collaboratively. This idea became innovative among conventional methods and resulted in more successful integration capability. However, the geometrical structure proposed in [39] is constructed by only intra-view. This means that locality structure preservation and discriminant analysis are performed by only intra-view, and the geometrical relationship of inter-view is ignored. In contrast, LapMCCs constructs only the geometrical structure of inter-view and does not construct the correlation structure and the geometrical structure of intra-view. Furthermore, a unified multiset canonical correlation analysis framework based on graph embedding for dimensionality reduction (GbMCC-DR) [40] was proposed. In this method, it was verified that a combination approach of both GbMCC and the simplest supervised approach such as LDA was effective. Meanwhile, a method combining a supervised approach such as sMVCCA and SFEMCCA with a geometrical approach such as GrMCCs and LapMCCs has not been proposed.

Because of this background, we newly propose supervised fractional-order embedding geometrical multi-view CCA (SFGMCCA) in this paper. The biggest contribution of SFGMCCA is realization of the architecture considering both the correlation structure and two kinds of the geometrical structures for intra-view and inter-view as a novel subspace learning method. Specifically, since previous methods focus on the following three points, i) multi-view integration, ii) a supervised approach and iii) a fractional-order technique, we introduce a novel viewpoint iv) construction of the correlation and geometrical structures into their approaches. A combination of points (i)–(iii) and point (iv) is the strong point in this paper. Especially, since points (ii)–(iv) are also new points for methods based on a geometrical approach such as LapMCCs and GrMCCs, SFGMCCA is expected to improve integration capability compared to that of the previously reported methods. Furthermore, SFGMCCA can be considered as an extended version of our previous method, SFEMCCA [30], which considers only points (i)–(iii). By introducing the new point (iv) for SFEMCCA, locality structure preservation and discriminant analysis can be performed in both intra-view and inter-view.

The rest of this paper is organized as follows: First, we briefly introduce related works in Section **II**. In Section **III**, we describe the model of SFGMCCA and how to integrate all features in a canonical space. In Section **IV**, we show experimental results to confirm the effectiveness of our method using five types of public datasets. Finally, conclusions are given in Section **V**.

II. RELATED WORKS

A. CANONICAL CORRELATION ANALYSIS AND ITS REGULARIZATION

First, we explain canonical correlation analysis (CCA) [3]. Given two types of matrices $\boldsymbol{X} = \{\boldsymbol{x}^1, \cdots, \boldsymbol{x}^N\} \in \mathbb{R}^{D_x \times N}$ and $\boldsymbol{Y} = \{\boldsymbol{y}^1, \cdots, \boldsymbol{y}^N\} \in \mathbb{R}^{D_y \times N}$, covariance matrices $\boldsymbol{C}_{mn}$ $(m, n \in \{x, y\})$ are calculated as follows:

$$\boldsymbol{C}_{mn} = \frac{1}{N} \sum_{i=1}^{N} \boldsymbol{m}^i \boldsymbol{n}^i, \tag{1}$$

where N is the number of samples used for integration, and D_x and D_y are the dimensions of $\boldsymbol{X}$ and $\boldsymbol{Y}$, respectively. Note that we assume that all of the features defined in this paper are normalized. In CCA, two optimal projection vectors $\hat{\boldsymbol{w}}_x \in \mathbb{R}^{D_x}$ and $\hat{\boldsymbol{w}}_y \in \mathbb{R}^{D_y}$ are calculated by the following function:

$$\begin{aligned} \left\{\hat{\boldsymbol{w}}_x, \hat{\boldsymbol{w}}_y\right\} = \arg \max_{\boldsymbol{w}_x, \boldsymbol{w}_y} \boldsymbol{w}_x^{\mathrm{T}} \boldsymbol{C}_{xy} \boldsymbol{w}_y \\ \text{s.t. } \boldsymbol{w}_x^{\mathrm{T}} \boldsymbol{C}_{xx} \boldsymbol{w}_x = \boldsymbol{w}_y^{\mathrm{T}} \boldsymbol{C}_{yy} \boldsymbol{w}_y = 1. \end{aligned} \tag{2}$$

Then we can calculate the projection vectors in such a way that the correlation of canonical variates $\hat{\boldsymbol{w}}_x^{\mathrm{T}} \boldsymbol{X}$ and $\hat{\boldsymbol{w}}_y^{\mathrm{T}} \boldsymbol{Y}$ is maximized. For the constraints of Eq. (2), a method adding regularization terms such as $\boldsymbol{C}_{mm} \rightarrow \boldsymbol{C}_{mm} + \epsilon \boldsymbol{I}_{D_m}$ is called regularized canonical correlation analysis (RCCA) [4], where ϵ is a small parameter and $\boldsymbol{I}_{D_m} \in \mathbb{R}^{D_m \times D_m}$ is the identity matrix. The regularization terms can suppress over-fitting.

B. MULTISET CANONICAL CORRELATION ANALYSIS

Next, we explain multiset CCA (MCCA) [17]. Given matrices $\boldsymbol{X}_m = \{\boldsymbol{x}_m^1, \boldsymbol{x}_m^2, \cdots, \boldsymbol{x}_m^N\} \in \mathbb{R}^{D_m \times N}$ from M kinds of features used for integration, where $m \in \{m_1, m_2, \cdots, m_M\}$, covariance matrices are calculated in the same manner as Eq. (1). Note that $\boldsymbol{x}_m^i$ is a vector in the ith sample of $\boldsymbol{X}_m$ and D_m is the dimension of $\boldsymbol{X}_m$. Next, M kinds of optimal projection vectors $\hat{\boldsymbol{w}}_m \in \mathbb{R}^{D_m}$ are calculated by the following function:

$$\begin{aligned} \left\{\hat{\boldsymbol{w}}_{m_1}, \hat{\boldsymbol{w}}_{m_2}, \cdots, \hat{\boldsymbol{w}}_{m_M}\right\} = \arg \max_{\forall \boldsymbol{w}_m} \sum_m \sum_{n \neq m, n} \boldsymbol{w}_m^{\mathrm{T}} \boldsymbol{C}_{mn} \boldsymbol{w}_n \\ \text{s.t. } \sum_m \boldsymbol{w}_m^{\mathrm{T}} \boldsymbol{C}_{mm} \boldsymbol{w}_m = 1. \end{aligned} \tag{3}$$

By maximizing the sum of all pairwise correlations, CCA is extended to a method that can integrate three or more kinds of multiple features.

III. SUPERVISED FRACTIONAL-ORDER EMBEDDING GEOMETRICAL MULTI-VIEW CCA

In this section, we present our novel method: supervised fractional-order embedding geometrical multi-view CCA (SFGMCCA). First, we explain how to construct the correlation structure using the fractional-order technique in **III-A**, and then we explain how to construct the geometrical structures in **III-B**. There are two types of geometrical structures: locality intra-view structure and locality inter-view structure. In **III-C**, we solve the optimization problem and multiple feature integration can finally be realized.

A. CORRELATION STRUCTURE

First, given matrices for integration $\boldsymbol{X}_m = \{\boldsymbol{x}_m^1, \boldsymbol{x}_m^2, \cdots, \boldsymbol{x}_m^N\} \in \mathbb{R}^{D_m \times N}$, where $m \in \{m_1, m_2, \cdots, m_M, L\}$, autocovariance

matrices $\boldsymbol{C}_{mm}$ and covariance matrices $\boldsymbol{C}_{mn}$ $(m \neq n)$ are calculated in the same manner as Eq. (1).

SFGMCCA is a supervised CCA method that handles class information as one of the features equivalently; hence, we use label features "$\boldsymbol{X}_L$" besides M kinds of features. $\boldsymbol{X}_L$ is a matrix obtained from D_L-dimensional one-hot vectors corresponding to each sample, where D_L is the number of classes. Next, SVD is performed to the autocovariance matrices and the covariance matrices as follows:

$$\begin{aligned} \boldsymbol{C}_{mn} &= \boldsymbol{P}_{mn}\boldsymbol{\Lambda}_{mn}\boldsymbol{Q}_{mn}^{\mathrm{T}} \\ \boldsymbol{\Lambda}_{mn} &= \mathrm{diag}(\lambda_{mn,1}, \lambda_{mn,2}, \cdots, \lambda_{mn,D_{mn}}). \end{aligned} \tag{4}$$

In the above equations, $D_{mn} = \mathrm{rank}(\boldsymbol{C}_{mn})$, $\lambda_{mn,1} \geq \lambda_{mn,2} \geq \cdots \geq \lambda_{mn,D_{mn}}$, and column vectors of $\boldsymbol{P}_{mn}$ and $\boldsymbol{Q}_{mn}$ are the left singular vectors and the right singular vectors of $\boldsymbol{C}_{mn}$, respectively. In order to suppress the increase of eigenvalues calculated by data with noise, a small number of samples and a large number of dimensions, two types of covariance matrices are reconstructed by using two kinds of fractional-order parameters ξ and η $(0 < \xi < 1, 0 < \eta < 1)$ as follows:

$$\tilde{\boldsymbol{\Lambda}}_{mn} = \begin{cases} \mathrm{diag}\left(\lambda_{mn,1}^{\xi}, \lambda_{mn,2}^{\xi}, \cdots, \lambda_{mn,D_{mn}}^{\xi}\right) & \text{if } m = n \\ \mathrm{diag}\left(\lambda_{mn,1}^{\eta}, \lambda_{mn,2}^{\eta}, \cdots, \lambda_{mn,D_{mn}}^{\eta}\right) & \text{if } m \neq n \end{cases}, \tag{5}$$

$$\tilde{\boldsymbol{C}}_{mn} = \boldsymbol{P}_{mn}\tilde{\boldsymbol{\Lambda}}_{mn}\boldsymbol{Q}_{mn}^{\mathrm{T}}. \tag{6}$$

This approach is called the fractional-order technique. The technique suppresses the increase of singular values by raising fractional-order parameters (between 0 to 1) for singular values obtained by SVD of the covariance matrix. The fractional-order technique has been used in [30], [36]–[38] as described earlier, and it is applied to the correlation structure of SFGMCCA in the same way.

B. TWO TYPES OF GEOMETRICAL STRUCTURES

SFEMCCA [30] only uses the above reconstructed covariance matrices. SFGMCCA newly introduces two types of geometrical structures: locality intra-view structure and locality inter-view structure. The application of this approach to SFEMCCA is the biggest contribution of this paper. By introducing these geometrical structures, CCA can use locality structure preservation and discriminant analysis. GrMCCs [39] and LapMCCs [26] have introduced only a locality intra-view structure and a locality inter-view structure, respectively, as geometrical ones, but SFGMCCA takes both of them into account. They will be described in **III-B1** and **III-B2**.

First, we calculate k_1 neighbor-based similarity matrix $\boldsymbol{S}_m \in \mathbb{R}^{N\times N}$ and k_2 neighbor-based similarity matrix $\overline{\boldsymbol{S}}_m \in \mathbb{R}^{N\times N}$ using $\boldsymbol{X}_m$ as follows:

$$s_m^{ij} = \begin{cases} \exp\left(\dfrac{-||\boldsymbol{x}_m^i - \boldsymbol{x}_m^j||^2}{t_m}\right) & \text{if } \boldsymbol{x}_L^i = \boldsymbol{x}_L^j \wedge \boldsymbol{x}_m^j \in \Omega_{\boldsymbol{x}_m^i} \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

and

$$\bar{s}_m^{ij} = \begin{cases} \exp\left(\dfrac{-||\boldsymbol{x}_m^i - \boldsymbol{x}_m^j||^2}{t_m}\right) & \text{if } \boldsymbol{x}_L^i \neq \boldsymbol{x}_L^j \wedge \boldsymbol{x}_m^j \in \Omega_{\boldsymbol{x}_m^i} \\ 0 & \text{otherwise} \end{cases}, \tag{8}$$

where s_m^{ij} $(\bar{s}_m^{ij})$ is an (i, j) element of $\boldsymbol{S}_m$ $(\overline{\boldsymbol{S}}_m)$ and $\Omega_{\boldsymbol{x}_m^i}$ is the set of k_1 neighbors of $\boldsymbol{x}_m^i$ in Eq. (7) and the set of k_2 neighbors of $\boldsymbol{x}_m^i$ in Eq. (8). t_m is an average norm of a full-connected graph constructed by $\boldsymbol{X}_m$ as follows:

$$t_m = \frac{2}{N(N-1)}\sum_{p=1}^{N}\sum_{q=p+1}^{N}||\boldsymbol{x}_m^p - \boldsymbol{x}_m^q||^2. \tag{9}$$

Moreover, we calculate the diagonal matrix $\boldsymbol{D}_m \in \mathbb{R}^{N\times N}$ with $d_m^{ii} = \sum_j s_m^{ij}$ in a diagonal element using $\boldsymbol{S}_m$. The diagonal matrix $\overline{\boldsymbol{D}}_m$ is also calculated from $\overline{\boldsymbol{S}}_m$ in the same way.

1) LOCALITY INTRA-VIEW STRUCTURE

By using the above similarity matrix and diagonal matrix, two types of locality intra-view structures are constructed in this section. Specifically, SFGMCCA constructs locality intra-class intra-view structure $\boldsymbol{G}^{(w)}$ by using $\boldsymbol{S}_m$ and $\boldsymbol{D}_m$ and locality inter-class intra-view structure $\overline{\boldsymbol{G}}^{(w)}$ by using $\overline{\boldsymbol{S}}_m$ and $\overline{\boldsymbol{D}}_m$. For example, $\boldsymbol{G}^{(w)}$ is calculated using $\boldsymbol{G}_{mm}$ as

$$\begin{aligned} \boldsymbol{G}_{mm} &= \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N} s_m^{ij}(\boldsymbol{w}_m^{\mathrm{T}}\boldsymbol{x}_m^i - \boldsymbol{w}_m^{\mathrm{T}}\boldsymbol{x}_m^j) \\ &\quad \times (\boldsymbol{w}_m^{\mathrm{T}}\boldsymbol{x}_m^i - \boldsymbol{w}_m^{\mathrm{T}}\boldsymbol{x}_m^j)^{\mathrm{T}} \\ &= \boldsymbol{w}_m^{\mathrm{T}}\left\{\sum_{i=1}^{N} d_m^{ii}\boldsymbol{x}_m^i(\boldsymbol{x}_m^i)^{\mathrm{T}} \right. \\ &\qquad \left. - \sum_{i=1}^{N}\sum_{j=1}^{N} s_m^{ij}\boldsymbol{x}_m^i(\boldsymbol{x}_m^j)^{\mathrm{T}}\right\}\boldsymbol{w}_m \\ &= \boldsymbol{w}_m^{\mathrm{T}}(\boldsymbol{X}_m\boldsymbol{D}_m\boldsymbol{X}_m^{\mathrm{T}} - \boldsymbol{X}_m\boldsymbol{S}_m\boldsymbol{X}_m^{\mathrm{T}})\boldsymbol{w}_m \\ &= \boldsymbol{w}_m^{\mathrm{T}}\boldsymbol{X}_m\boldsymbol{\mathcal{L}}_m\boldsymbol{X}_m^{\mathrm{T}}\boldsymbol{w}_m, \end{aligned} \tag{10}$$

where $\boldsymbol{w}_m \in \mathbb{R}^{D_m}$ is a projection vector to project $\boldsymbol{X}_m$ into a canonical space. Next, all of $\boldsymbol{G}_{mm}$ are integrated to $\boldsymbol{G}^{(\mathrm{w})}$ as follows:

$$\begin{aligned} \boldsymbol{G}^{(w)} &= \sum_m \boldsymbol{G}_{mm} \\ &= \sum_m \boldsymbol{w}_m^{\mathrm{T}}(\boldsymbol{X}_m\boldsymbol{\mathcal{L}}_m\boldsymbol{X}_m^{\mathrm{T}})\boldsymbol{w}_m \\ &= \boldsymbol{w}^{\mathrm{T}}\begin{bmatrix} \boldsymbol{X}_{m_1}\boldsymbol{\mathcal{L}}_{m_1}\boldsymbol{X}_{m_1}^{\mathrm{T}} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \vdots & \ddots & \vdots \\ \boldsymbol{0} & \cdots & \boldsymbol{X}_L\boldsymbol{\mathcal{L}}_L\boldsymbol{X}_L^{\mathrm{T}} \end{bmatrix}\boldsymbol{w} \\ &= \boldsymbol{w}^{\mathrm{T}}\boldsymbol{\Delta}^{(\mathrm{w})}\boldsymbol{w}, \end{aligned} \tag{11}$$

where $\boldsymbol{w} \in \mathbb{R}^{D_{\mathrm{sum}}}$ $(D_{\mathrm{sum}} = \sum_m D_m)$ is a vector obtained by the concatenation of $\boldsymbol{w}_m$, and $\boldsymbol{\Delta}^{(w)}$ is a matrix with $\boldsymbol{X}_m\boldsymbol{\mathcal{L}}_m\boldsymbol{X}_m^{\mathrm{T}}$ in a diagonal block and a zero matrix in a non-diagonal block.

$\overline{\boldsymbol{G}}_{mm}$ and $\overline{\boldsymbol{G}}^{(\mathrm{w})}$ are also calculated in the same way. Finally, in order to minimize intra-class scatter and maximize inter-class scatter simultaneously, we integrate these structures by the trade-off parameter $\delta \in [0, 1]$ as follows:

$$\boldsymbol{\Gamma}^{(w)} = (1-\delta)\boldsymbol{G}^{(w)} - \delta\overline{\boldsymbol{G}}^{(w)}. \tag{12}$$

In this way, the locality intra-view structure is calculated. Finally, $\boldsymbol{\Gamma}^{(w)}$ will be minimized in the objective function described in **III-C**.

2) LOCALITY INTER-VIEW STRUCTURE

In the previous section, we calculated the locality intra-view structure. In this section, we calculate two types of locality inter-view structures. Specifically, SFGMCCA constructs locality intra-class inter-view structure $\boldsymbol{G}^{(b)}$ by using $\boldsymbol{S}_m$, $\boldsymbol{S}_n$, $\boldsymbol{D}_m$, and $\boldsymbol{D}_n$ and constructs locality inter-class inter-view structure $\overline{\boldsymbol{G}}^{(b)}$ by using $\overline{\boldsymbol{S}}_m$, $\overline{\boldsymbol{S}}_n$, $\overline{\boldsymbol{D}}_m$, and $\overline{\boldsymbol{D}}_n$. As in the previous section, we calculate $\boldsymbol{G}^{(b)}$ using $\boldsymbol{G}_{mn} = \boldsymbol{w}_m^{\mathrm{T}}\boldsymbol{X}_m\boldsymbol{\mathcal{L}}_{mn}\boldsymbol{X}_n^{\mathrm{T}}\boldsymbol{w}_n$. Note that $\boldsymbol{\mathcal{L}}_{mn} = \boldsymbol{\mathcal{L}}_m \circ \boldsymbol{\mathcal{L}}_n$, where the symbol "$\circ$" means the Hadamard product. All of $\boldsymbol{G}_{mn}$ are integrated to $\boldsymbol{G}^{(\mathrm{b})}$ as follows:

$$\begin{aligned}\boldsymbol{G}^{(b)} &= \sum_m \sum_{n\neq m,n} \boldsymbol{G}_{mn} \\ &= \boldsymbol{w}^{\mathrm{T}} \begin{bmatrix} \boldsymbol{0} & \cdots & \boldsymbol{X}_{m_1}\boldsymbol{\mathcal{L}}_{m_1 L}\boldsymbol{X}_L^{\mathrm{T}} \\ \boldsymbol{X}_{m_2}\boldsymbol{\mathcal{L}}_{m_2 m_1}\boldsymbol{X}_{m_1}^{\mathrm{T}} & \cdots & \boldsymbol{X}_{m_2}\boldsymbol{\mathcal{L}}_{m_2 L}\boldsymbol{X}_L^{\mathrm{T}} \\ \vdots & \ddots & \vdots \\ \boldsymbol{X}_L\boldsymbol{\mathcal{L}}_{Lm_1}\boldsymbol{X}_{m_1}^{\mathrm{T}} & \cdots & \boldsymbol{0} \end{bmatrix} \boldsymbol{w} \\ &= \boldsymbol{w}^{\mathrm{T}}\boldsymbol{\Delta}^{(\mathrm{b})}\boldsymbol{w}, \end{aligned} \tag{13}$$

where $\boldsymbol{\Delta}^{(b)}$ is a matrix with a zero matrix in a diagonal block and $\boldsymbol{X}_m\boldsymbol{\mathcal{L}}_{mn}\boldsymbol{X}_n^{\mathrm{T}}$ in a non-diagonal block. $\overline{\boldsymbol{G}}_{mn}$ and $\overline{\boldsymbol{G}}^{(\mathrm{b})}$ are also calculated in the same way. Finally, we integrate these structures by δ as follows:

$$\boldsymbol{\Gamma}^{(b)} = (1-\delta)\boldsymbol{G}^{(b)} - \delta\overline{\boldsymbol{G}}^{(b)}. \tag{14}$$

In this way, the locality inter-view structure is also calculated. Finally, $\boldsymbol{\Gamma}^{(b)}$ will be maximized in the objective function described in **III-C**.

C. SOLUTION OF OPTIMIZATION PROBLEM

SFGMCCA calculates the optimal projection vectors $\hat{\boldsymbol{w}}_m \in \mathbb{R}^{D_m}$ to project all of the features into a canonical space using $\tilde{\boldsymbol{C}}_{mn}$ in Eq. (6), $\boldsymbol{\Gamma}^{(w)}$ in Eq. (12) and $\boldsymbol{\Gamma}^{(b)}$ in Eq. (14). In the correlation structure, the sum of all pairwise correlations is maximized by using $\tilde{\boldsymbol{C}}_{mn}$. On the other hand, in the geometrical structures, locality intra-class structure $\boldsymbol{\Gamma}^{(w)}$ is minimized and locality inter-class structure $\boldsymbol{\Gamma}^{(b)}$ is maximized. Therefore, the optimization problem is written as follows:

$$\begin{aligned}&\left\{\hat{\boldsymbol{w}}_{m_1}, \hat{\boldsymbol{w}}_{m_2}, \cdots, \hat{\boldsymbol{w}}_{m_M}, \hat{\boldsymbol{w}}_L\right\} \\ &\quad = \arg\max_{\forall \boldsymbol{w}_m} \sum_m \sum_{n\neq m,n} \boldsymbol{w}_m^{\mathrm{T}}\tilde{\boldsymbol{C}}_{mn}\boldsymbol{w}_n + \alpha\left(\boldsymbol{\Gamma}^{(b)} - \boldsymbol{\Gamma}^{(w)}\right) \\ &\qquad \text{s.t.} \sum_m \boldsymbol{w}_m^{\mathrm{T}}(\tilde{\boldsymbol{C}}_{mm} + \epsilon\boldsymbol{I}_{D_m})\boldsymbol{w}_m = 1, \end{aligned} \tag{15}$$

where ϵ is a small parameter and $\boldsymbol{I}_{D_m} \in \mathbb{R}^{D_m \times D_m}$ is the identity matrix. This regularization term can suppress overfitting. Next, the Lagrange multiplier method is applied to Eq. (15), and the following Lagrange function is obtained:

$$\begin{aligned}\boldsymbol{L}(\boldsymbol{w}, \zeta) = &\sum_m \sum_{n\neq m,n} \boldsymbol{w}_m^{\mathrm{T}}\tilde{\boldsymbol{C}}_{mn}\boldsymbol{w}_n + \alpha\left(\boldsymbol{\Gamma}^{(b)} - \boldsymbol{\Gamma}^{(w)}\right) \\ &- \zeta\left\{\sum_m \boldsymbol{w}_m^{\mathrm{T}}(\tilde{\boldsymbol{C}}_{mm} + \epsilon\boldsymbol{I}_{D_m})\boldsymbol{w}_m - 1\right\}, \end{aligned} \tag{16}$$

where ζ is a Lanrange multiplier. The Lagrange function is transformed by partial differentiations; that is, we get the gradient of Eq. (16) by calculating the extrema as

$$\begin{aligned}\frac{\partial \boldsymbol{L}}{\partial \boldsymbol{w}_m} = &\left\{\sum_{n\neq m,n} \tilde{\boldsymbol{C}}_{mn} + \alpha\left(\boldsymbol{\Delta}^{(b)} - \boldsymbol{\Delta}^{(w)}\right)\right\}\boldsymbol{w}_n \\ &- 2\zeta\left\{(\tilde{\boldsymbol{C}}_{mm} + \epsilon\boldsymbol{I}_{D_m})\boldsymbol{w}_m\right\}. \end{aligned} \tag{17}$$

Accordingly, the optimization problem in Eq. (15) is transformed into the following generalized eigenvalue problem:

$$\begin{aligned}&\begin{bmatrix} \boldsymbol{\Psi}_{m_1 m_1} & \boldsymbol{\Psi}_{m_1 m_2} & \cdots & \boldsymbol{\Psi}_{m_1 L} \\ \boldsymbol{\Psi}_{m_2 m_1} & \boldsymbol{\Psi}_{m_2 m_2} & \cdots & \boldsymbol{\Psi}_{m_2 L} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Psi}_{L m_1} & \boldsymbol{\Psi}_{L m_2} & \cdots & \boldsymbol{\Psi}_{LL} \end{bmatrix} \begin{bmatrix} \boldsymbol{w}_{m_1} \\ \boldsymbol{w}_{m_2} \\ \vdots \\ \boldsymbol{w}_L \end{bmatrix} \\ &= \zeta_d \left(\begin{bmatrix} \tilde{\boldsymbol{C}}_{m_1 m_1} & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \tilde{\boldsymbol{C}}_{m_2 m_2} & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{0} & \boldsymbol{0} & \cdots & \tilde{\boldsymbol{C}}_{LL} \end{bmatrix} + \epsilon\boldsymbol{I}_{D_{sum}} \right) \begin{bmatrix} \boldsymbol{w}_{m_1} \\ \boldsymbol{w}_{m_2} \\ \vdots \\ \boldsymbol{w}_L \end{bmatrix}, \end{aligned} \tag{18}$$

where $D_{\mathrm{sum}} = \sum_m D_m$, $\zeta_1 \geq \zeta_2 \geq \cdots \geq \zeta_d \geq \cdots > \zeta_D$, and $D \leq min\{D_m\}$. Note that the ζ_d obtained by solving the above generalized eigenvalue problem are eigenvalues. $\boldsymbol{\Psi}_{mn} \in \mathbb{R}^{D_m \times D_n}$ of the left side in Eq. (18) are block matrices described as

$$\boldsymbol{\Psi}_{mn} = \begin{cases} -\alpha\boldsymbol{X}_m\left\{(1-\delta)\boldsymbol{\mathcal{L}}_m - \delta\overline{\boldsymbol{\mathcal{L}}}_m\right\}\boldsymbol{X}_m^{\mathrm{T}} & \text{if } m = n \\ \tilde{\boldsymbol{C}}_{mn} + \alpha\boldsymbol{X}_m\left\{(1-\delta)\boldsymbol{\mathcal{L}}_{mn} - \delta\overline{\boldsymbol{\mathcal{L}}}_{mn}\right\}\boldsymbol{X}_n^{\mathrm{T}} & \text{if } m \neq n. \end{cases} \tag{19}$$

Next, we can obtain the optimal projection matrix for each feature as follows:

$$\hat{\boldsymbol{W}}_m = \left[\hat{\boldsymbol{w}}_m^1, \hat{\boldsymbol{w}}_m^2, \cdots, \hat{\boldsymbol{w}}_m^D\right] \in \mathbb{R}^{D_m \times D}, \tag{20}$$

where the dth largest eigenvalue ζ_d corresponds to a projection vector $\hat{\boldsymbol{w}}_m^d$. By using the projection matrix, each feature is projected to a lower-dimensional canonical space to obtain the following canonical features:

$$\hat{\boldsymbol{X}}_m = \hat{\boldsymbol{W}}_m^{\mathrm{T}}\boldsymbol{X}_m \in \mathbb{R}^{D \times N}. \tag{21}$$

In summary, $\boldsymbol{X}_m$ is projected as $\hat{\boldsymbol{X}}_m$ to the D-dimensional canonical space. Finally, all features are integrated as $\sum_m \hat{\boldsymbol{X}}_m$ in the space, and the integrated canonical feature $\hat{\boldsymbol{X}}$ is obtained.

Algorithm 1 Multiple Feature Integration Using SFGMCCA

Require: $\boldsymbol{X}_m(m \in \{m_1, \cdots, m_M, L\}), \xi, \eta, k_1, k_2, \epsilon, \alpha, \delta, D$
Ensure: $\hat{\boldsymbol{X}}$

```
Normalize all of X_m
for m = m_1, ..., m_M, L
    Compute L_m and L̄_m by solving Eqs. (7)–(10).
end for
for m = m_1, ..., m_M, L
    for n = m_1, ..., m_M, L
        Compute C̃_mn based on the fractional-order
        technique by solving Eqs. (1), (4)–(6).
        if m ≠ n
            Compute L_mn and L̄_mn.
        end if
        Compute Ψ_mn by Eq. (19).
    end for
end for
Compute Eq. (18), and obtain Ŵ_m by Eq. (20).
Obtain X̂ by integrating all of X̂_m in the canonical
space.
```

The summarized procedure of SFGMCCA is shown in Algorithm 1. SFGMCCA, which takes into account the four points mentioned at the same time, is more effective than the conventional methods. If the sum of $\boldsymbol{w}^{\mathrm{T}}\boldsymbol{\Delta}^{(\mathrm{b})}\boldsymbol{w}$ in Eq. (13) is maximized, the objective function is the same as LapMCCs. On the other hand, if the sum of all pairwise correlations using Eq. (1) is maximized and $\boldsymbol{\Gamma}^{(\mathrm{w})}$ in Eq. (12) is minimized, the objective function is the same as GrMCCs.

IV. EXPERIMENTS

Experiments were carried out to verify the performance of our method by applying our method to public toy datasets for image classification. First, an overview of datasets and the feature representation are shown in **IV-A**. Next, we explain comparative methods and experimental conditions in **IV-B**. Finally, we show experimental results and discuss them in **IV-C**.

A. DATASETS AND FEATURE REPRESENTATION

In this section, four public toy datasets for image classification used in the experiments, i.e., MNIST [41], COIL-20 [42], ETH-80 [43] and CIFAR-10 [44], and features extracted from them are described. The number of classes and samples of the datasets are shown in Table 1. For MNIST and CIFAR-10, note that 1,000 images are selected from data opened for test ones at random under the constraint that the number of samples per class (i.e., "N/C" in Table 1) is constant so as not to make imbalanced data. Furthermore, we adopted not only image datasets but also the public fMRI dataset provided by [45] in order to verify the robustness of SFGMCCA. The details of the datasets will be explained below.

MNIST is a dataset with ten kinds of handwritten digits from "0" to "9". The public dataset consists of 60,000 images for training data and 10,000 images for test data. In the experiments, 100 images per class were randomly selected from the test data, and it is used as a dataset composed of 1,000 images. All images are grayscale and the image size is 28 × 28. Ten sample images of handwritten digits are shown in Fig. 1(a).

TABLE 1. Number of classes (C) and samples (N) of datasets.

Datasets	C	N/C	N
MNIST [41]	10	100	1,000
COIL-20 [42]	20	72	1,440
ETH-80 [43]	8	410	3,280
CIFAR-10 [44]	10	100	1,000
fMRI [45]	150	8	1,200

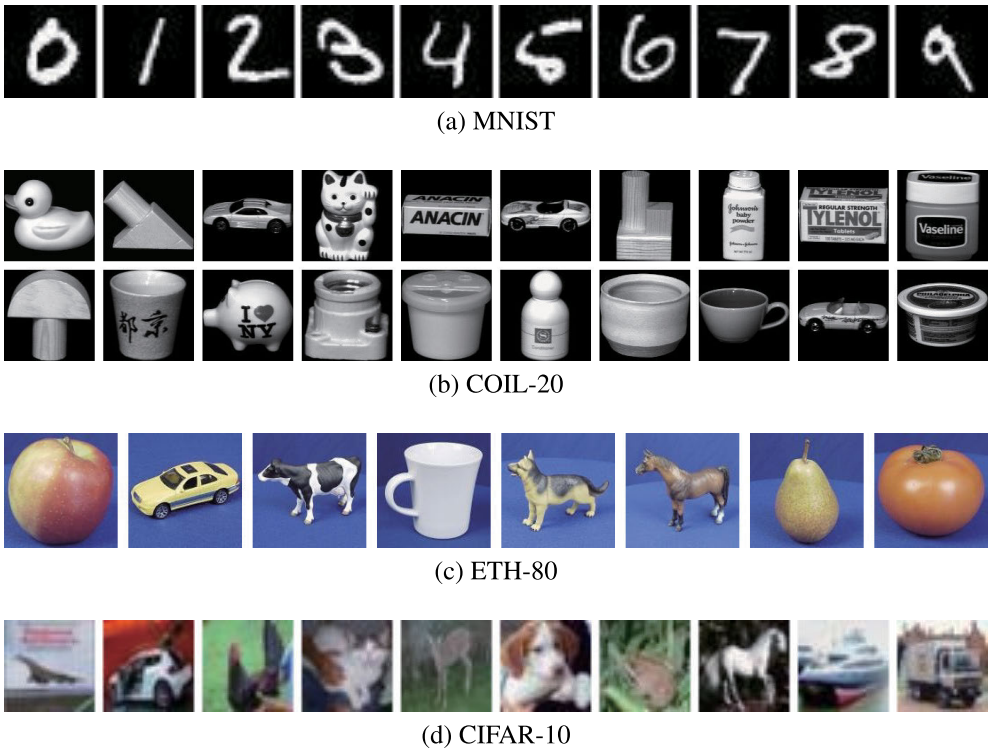
COIL-20 is a dataset with 1,440 images consisting of 20 kinds of objects. Each of the objects has been imaged by rotating it at five-degree intervals from the same depression angle. Therefore, each class consists of 72 images ($360/5 = 72$). All images are grayscale and the image size is 128 × 128. Twenty sample images of objects are shown in Fig. 1(b).

ETH-80 is a dataset with 3,280 images consisting of eight kinds of objects (i.e., apple, car, cow, cup, dog, horse, pear, and tomato). The objects have been imaged from various depression angles. The categories are further divided into ten kinds, but they were regarded as belonging to the same category in the experiments. Therefore, they are used as eight classes as stated above, and each class consists of 410 images. All of the images are color and the image size is 128 × 128. Eight sample images of objects are shown in Fig. 1(c).

CIFAR-10 is divided into 10 categories (i.e., airplane, automobile, bird, cat, deer, dog, frog, horse, ship, and truck). The public dataset consists of 60,000 images for training data and 10,000 images for test data. As with MNIST, 100 images per class were randomly selected from the test data, and it is used as a dataset composed of 1,000 images. All images are color and the image size is 32 × 32. Ten sample images of objects are shown in Fig. 1(d).

The fMRI dataset contains brain signal features based on fMRI activity obtained from some participants while viewing images collected from the ImageNet dataset [46], and other feature integration approaches [47] often used it in the experiment. The fMRI dataset consists of 1,200 images from 150 categories (each category has eight images) and corresponding brain signal features from participants. In this experiment, we adopted visual features obtained from ImageNet and corresponding brain signal features calculated from three participants, and totally four-view features are applied to all methods.

For these images other than the fMRI dataset, histograms of oriented gradients (HOG) [48] and GIST [49] are extracted, and Gabor filtering [50] is applied. Note that all features are generated as 100 dimensions by applying principal component analysis (PCA) [51] to the



(a) MNIST

(b) COIL-20

(c) ETH-80

(d) CIFAR-10

FIGURE 1. (a) Ten sample images of handwritten digits in MNIST are shown in ascending order from left "0" to right "9". (b) Twenty sample images of objects in COIL-20 are shown. (c) Eight sample images of objects in ETH-80 are shown in order of apple, car, cow, cup, dog, horse, pear, and tomato from left. (d) Ten sample images of objects in CIFAR-10 are shown in order of airplane, automobile, bird, cat, deer, dog, frog, horse, ship, and truck from left.

extracted features. HOG is a method of local feature representation that calculates the gradient intensity and direction of the luminance for each pixel of the image, converts them into histograms in each cell region, and normalizes for each block of several cells. On the other hand, GIST is a method of global feature representation that calculates the edge strength by transforming into a frequency domain after dividing an image into several blocks and performing filtering for each block. In addition, Gabor filtering can extract the direction of the edges included in an image by applying the signal defined as the product of the sine wave and the Gaussian function to the image. Since these features can represent different characteristics for the images as described above and can be applied to both grayscale and color images, we adopted them as a set of heterogeneous features in the experiments. On the other hand, we extracted visual features from the ImageNet dataset based on a VGG19 model [52], which is one of convolutional neural network models. Furthermore, we used brain signal features provided in the dataset [45]. Note that all features are generated as 300 dimensions by applying PCA to the extracted features.

B. EXPERIMENTAL CONDITIONS

Here we explain the comparative methods used in the experiments and the conditions including parameter settings. First, the methods and their details are shown in Table 2. We used the following seven kinds of CCA as comparative methods: MCCA, RMCCA, TCCA, LapMCCs, GrMCCs, sMVCCA, DMCCA and SFEMCCA. As shown in the table, the checkmark "✓" is put in accordance with the kind of correlation (i.e., pairwise or high order), use of the regularization term, class information, locality structure preservation and discriminant analysis in intra-view and inter-view, and the fractional-order technique.

Next, we explain experimental conditions. The division ratios for learning and testing data were set as 1:4 for MNIST, 1:11 for COIL-20, 1:9 for ETH-80, and 1:4 for CIFAR-10, 1:7 for the fMRI dataset. We performed cross-validation in the experiment. Specifically, we divided the dataset, and the rates of learning data and test data were 1:4, 1:11, 1:9 and 1:4 for MNIST, COIL-20, ETH-80 and CIFAR-10. Then we performed 5, 12, 10 and 5 fold cross-validation for each dataset. The above verification approach is often adopted in the previous feature integration methods such as [40]. Note that 8 fold cross-validation was performed when the fMRI dataset was used. When the methods were applied to the learning data, the dimensionality of the space after integration, or the canonical space, can be roughly divided into the following two types according to the method, and the dimension giving the best accuracy was adopted. For the five methods, MCCA, RMCCA, TCCA, LapMCCs and

TABLE 2. Methods used in the experiments and their details. "RT" means that the regularization term is used in the method to suppress over-fitting. "LP" and "DA" mean that the method considers locality structure preservation and discriminant analysis, respectively. "FT" means that the fractional-order technique is used in the method.

Methods	Correlation		RT	Class info.	Intra-view		Inter-view		FT
	Pairwise	Tensor			LP	DA	LP	DA	
MCCA [17]	✓								
RMCCA [18]	✓		✓						
TCCA [27]		✓	✓						
LapMCCs [26]			✓	✓			✓		
GrMCCs [39]	✓		✓	✓	✓	✓			
sMVCCA [28]	✓		✓	✓					
DMCCA [32]	✓		✓	✓		✓		✓	
SFEMCCA [30]	✓		✓	✓					✓
SFGMCCA	✓		✓	✓	✓	✓	✓	✓	✓

TABLE 3. Best classification accuracy in each method. $a \pm b$ (c) means that a is classification accuracy (%), b is standard deviation, and c is dimensionality.

Methods	MNIST	COIL-20	ETH-80	CIFAR-10	fMRI
MCCA	57.5±0.3 (90)	65.0±2.7 (90)	77.4±1.3 (60)	27.6±3.5 (50)	29.3±4.2 (149)
RMCCA	76.4±1.8 (70)	79.2±0.9 (100)	77.6±1.3 (60)	37.6±3.5 (40)	41.4±2.1 (150)
TCCA	73.3±2.1 (100)	85.3±1.7 (100)	76.7±1.3 (100)	32.1±3.4 (100)	32.4±2.0 (150)
LapMCCs	77.1±2.0 (60)	80.5±1.3 (20)	78.0±1.4 (60)	38.0±2.6 (30)	34.5±2.0 (150)
GrMCCs	83.4±1.2 (10)	80.9±1.4 (20)	82.5±1.2 (100)	41.6±2.6 (10)	44.7±3.9 (150)
sMVCCA	82.4±1.5 (9)	81.8±1.1 (19)	81.1±1.2 (7)	41.7±3.3 (9)	41.1±4.7 (150)
DMCCA	82.2±1.3 (9)	80.2±1.3 (19)	80.9±1.0 (7)	41.4±3.5 (9)	40.9±4.6 (150)
SFEMCCA	83.5±2.0 (9)	85.0±1.1 (19)	82.3±1.0 (7)	42.1±2.9 (9)	49.8±3.4 (150)
SFGMCCA	**86.3±1.2** (9)	**85.8±1.0** (19)	**87.0±1.0** (7)	**48.2±3.2** (9)	**51.6±3.2** (150)

GrMCCs, searches were performed in increments of 10 dimensions, $10, 20, \cdots, 100$ when MNIST, COIL-20 ETH-80 and CIFAR-10 were used. When the fMRI dataset was used, these methods searched in increments of 20 dimensions, $50, 70, \cdots, 149$. On the other hand, for three methods, sMVCCA, DMCCA, SFEMCC and our proposed SFGMCCA, the maximum dimensionality of the space in the experimental settings is a minimum value among the dimensionality of HOG, GIST, Gabor filtering and "the number of classes $(C) - 1$".[3] In these experiments, we searched in increments of two dimensions, $1, 3, 5, \cdots$ when MNIST, COIL-20 ETH-80 and CIFAR-10 were used. On the other hand, we searched in increments of 20 dimensions, $50, 70, 90, \cdots$ when the fMRI dataset was used. The linear-support vector machine (SVM) [53] was used as a classifier, and classification accuracy was used as an evaluation index. Tucker decomposition [54] for low rank approximation required by TCCA was performed by the Higher-order SVD (HOSVD) algorithm [55].

Parameter settings are described below. The regularization parameter was set to 0.01 in all methods. The number of neighborhoods (k) for locality structure preservation was set in LapMCCs and GrMCCs, where $k \in \{10, 20, 50, 100, 200\}$ for MNIST, $k \in \{10, 20, 50, 120\}$ for COIL-20, $k \in \{10, 20, 50, 100, 200, 328\}$ for ETH-80, $k \in \{10, 20, 50, 100, 300, 500, 800\}$ for CIFAR-10 and $k \in \{10, 20, 50, 100\}$ for the fMRI dataset, and k giving the best accuracy was selected. The graph constructed by the highest value in the settings corresponds to a weighted complete graph using all samples of the learning data. On the other hand, for two methods (GrMCCs and SFGMCCA) that can set different number of neighborhoods to compute both the intra-class and the inter-class graphs, the same value was used in order to simplify the experiments. Moreover, there are two important parameters in addition to them for GrMCCs and SFGMCCA. First, the relative weight (α) of the geometrical structures for the correlation structure was selected among $\alpha \in \{0.5, 1.0, 2.0\}$. Second, the trade-off ratio (δ) between intra-class and inter-class for discriminant analysis was selected among $\delta \in \{0.00, 0.25, 0.50, 0.75, 1.00\}$. The smaller the value of δ is, the greater is the weight of intra-class. Two fractional-order parameters [30], [36]–[38] used by SFEMCCA and SFGMCCA, ξ and η, were both set to 0.9.

[3]When sMVCCA, DMCCA, SFEMCC and SFGMCCA are applied to MNIST for instance, the maximum dimensionality of the canonical space is $\min\{100, 100, 100, 10 - 1\} = 9$.

C. EXPERIMENTAL RESULTS AND DISCUSSION

1) COMPARISON OF THE METHODS

The best classification accuracy by each method is shown in Table 3 and more detailed results for accuracy differences

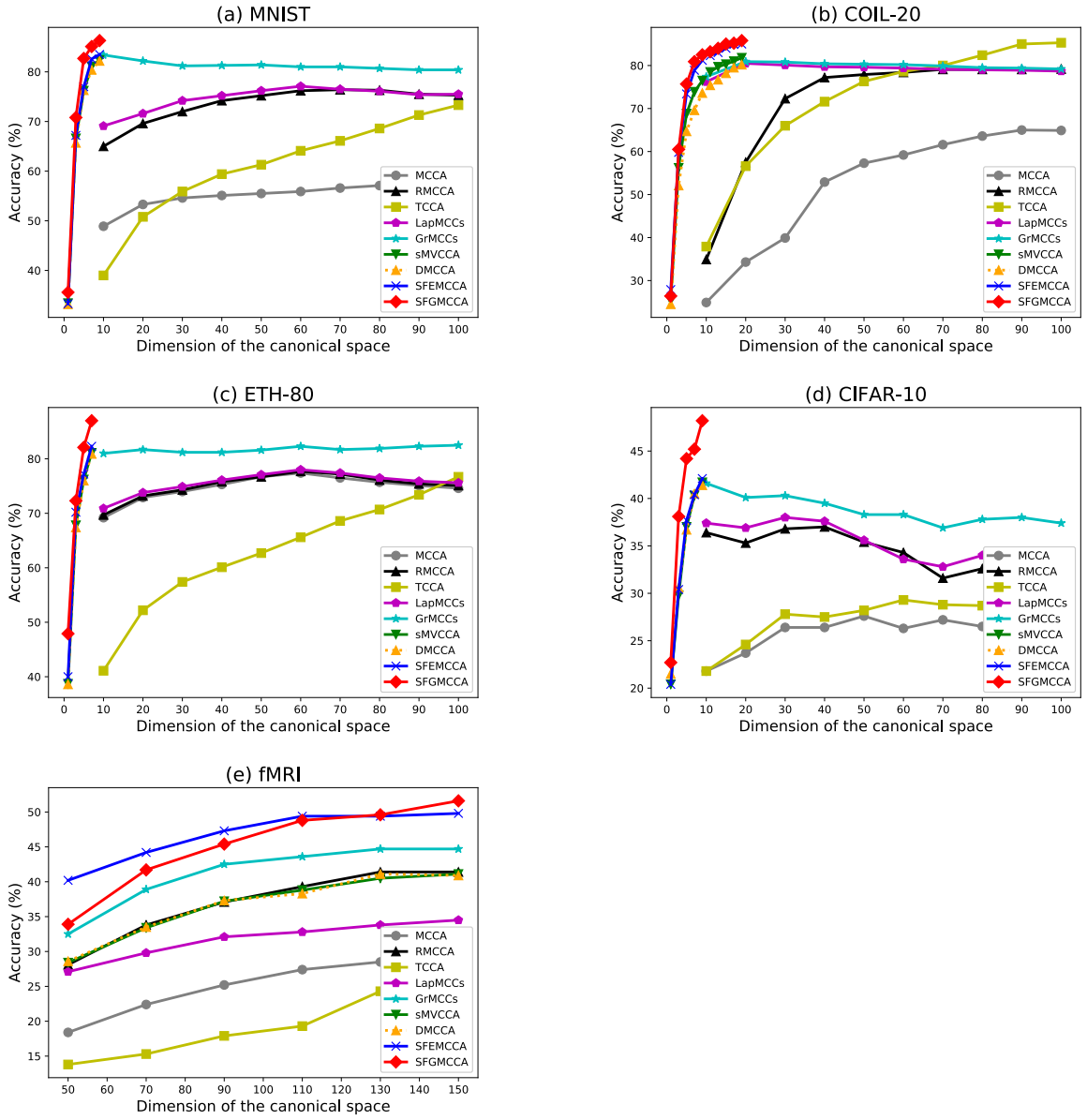


FIGURE 2. Accuracies corresponding to the change of dimension in the canonical space are shown for all datasets, (a) MNIST, (b) COIL-20, (c) ETH-80, (d) CIFAR-10 and (e) fMRI. The numbers of classes in (a)–(e) are 10, 20, 8, 10 and 150, respectively. The maximum number of dimensions of the canonical space that can be constructed by sMVCCA, DMCCA, SFEMCCA, and SFGMCCA can be expressed as the minimum value of "the number of dimensions of each feature" and "the number of classes (C) -1". Therefore, they become 9, 19, 7, 9 and 149 in the order in experiment settings.

with dimensionality of the canonical space are shown in Fig. 2. In these figures, the classification accuracies have been plotted for each method by the step size of dimensions as described earlier.

MCCA and RMCCA are the most fundamental benchmarks among the methods. According to the results of TCCA, which can maximize a high-order correlation, the results for MNIST, ETH-80 and CIFAR-10 became worse than those for RMCCA, but the results for COIL-20 reached a high accuracy rate (85.3%) in the 100-dimensional space as shown in Fig. 2(b). Therefore, almost no achievement of tensor-type covariance was obtained. In addition, the accuracy of classification by TCCA generally improved in all datasets as the dimensionality of the canonical space became greater. If the representation capability of data improves as the dimensionality increases, the accuracy in a high-dimensional canonical space may improve as well. However, the tensor covariance needs to be unfolded by Tucker decomposition. First, the higher-mode tensor is decomposed into a row of a mode (a feature such as HOG) and a column consisting of the total

product of the dimensionality of the other modes.[4] Next, SVD is performed for the unfolded matrix. For the procedure, the influence on calculation cost due to the increase in the number of features or their dimensionality is remarkable compared with other methods maximizing the sum of all pairwise correlations. Therefore, TCCA is an impractical method if the integration of such data is required. It can be concluded that RMCCA and LapMCCs are superior to TCCA since they recorded the same level of integration capability at lower dimensions.

Although LapMCCs has improved accuracies in all datasets compared with RMCCA, they have almost no effect. Therefore, it is difficult to conclude that effective feature integration can be achieved only by using a locality intra-class inter-view structure. On the other hand, GrMCCs, which introduced both locality structure preservation and discriminant analysis in the geometrical structure for intra-view besides the correlation structure, recorded great improvements of accuracies compared with these methods. More effective integration was achieved by the introduction of discriminant analysis than by the introduction of locality structure preservation. Moreover, GrMCCs kept the accuracies at a high level from low to high dimensions in all datasets. Therefore, using not only the correlation structure but also the geometrical structure for intra-view is effective for the following two points: the overall high classification accuracy and the high representation capability in a low-dimensional space. The average of all datasets with the highest classification accuracy was 72.1%, and the accuracy increased by +4.4% compared with RMCCA and by +3.7% compared with LapMCCs.

As mentioned earlier, when four methods including the proposed method (sMVCCA, DMCCA, SFEMCCA and SFGMCCA) were used in the experimental settings, the maximum dimensionality of the canonical space that can be constructed was greatly restricted unlike the above five methods. Under the constraints, the accuracies using the three methods increased in proportion to the dimensionality in all datasets. The accuracies using the three methods is equivalent to the highest accuracy of RMCCA and LapMCCs in just a three-dimensional space and to that of GrMCCs in just a five-dimensional space. SMVCCA, which uses class information as an independent feature, showed that it has integration capability equivalent to that of GrMCCs, although it does not consider locality structure preservation and discriminant analysis. Moreover, the results obtained by using SFEMCCA, which introduced the fractional-order technique to sMVCCA, exceeded those obtained by using sMVCCA in all datasets, and its effectiveness was confirmed. Furthermore, the results obtained by using SFGMCCA exceeded those obtained by using sMVCCA and SFEMCCA in almost all dimensions. In particular, the results obtained by using CIFAR-10 showed a great improvement. By exceeding the highest classification accuracy of all comparative methods in just a five-dimensional space, it showed high expressive capability even in a low-dimensional space, and it further extended its accuracy to a nine-dimensional space. However, the results obtained by using SFGMCCA in COIL-20 showed an increase of only 0.8% compared with SFEMCCA and 0.5% compared to TCCA. The reason is that COIL-20 is a dataset imaged by rotating the object from the same depression angle, unlike the other datasets. As a result of feature extraction, a large number of dependent vectors occurred, and effective utilization of the geometrical structures was restricted due to the substantial decrease in the number of dependent samples. In MNIST, ETH-80 and CIFAR-10, the results obtained by using SFEMCCA, for which data with a small number of samples and high-dimensional data can be handled, showed only slight improvements (+0.9%) compared with sMVCCA, whereas it showed a great improvement in COIL-20 (+3.2%). DMCCA introduces the within-class and between-class correlations based on labels instead of the use of label features. We confirmed that the results of DMCCA were close to those of sMVCCA, which is also one of supervised approaches. On the other hand, since the results of DMCCA for all datasets were lower than those of SFGMCCA, the effectiveness of SFGMCCA was verified. We used Welch's t-test to verify whether the differences between the classification accuracies using SFGMCCA and those using the other comparative methods were significant or not. As a result, it was confirmed that accuracy improvements using SFGMCCA were statistically significant at the significance level of 0.05 compared to all of the other methods.

Moreover, as shown in Table 3, the results of SFGMCCA exceeded those of other comparative methods. Since DMCCA and SFEMCCA achieved higher performance than baseline approaches such as MCCA and RMCCA, the validity of the introduction of the local structure preserving, the discriminant analysis and the fractional-order embedding was verified. From the above, since SFGMCCA was also effective for the fMRI dataset, which consists of brain signal features, it was verified that SFGMCCA could be applied to the dataset other than image datasets. Furthermore, Fig. 2 (e) indicated that although the result of SFGMCCA was lower than SFEMCCA when the dimension in the canonical space was low, the performance was increased as the dimension was high. Since the maximum dimensionality of SFGMCCA, SFEMCCA and sMVCCA are restricted due to the use of label features as one modality, the maximum dimensionality is set to 150. Thus, in the range of the possible values of dimensions, SFGMCCA achieved the highest performance among all comparative methods.

2) EFFECTS OF PARAMETERS OF SFGMCCA ON ACCURACY

Figure 3 shows three-dimensional bar graphs corresponding to the classification accuracies with two parameters of SFGMCCA, α and δ. For all datasets, setting both α and δ to large values had a tendency to decrease accuracies. In other

[4]When three kinds of features composed of 100 dimensions are used like this setting, SVD for an unfolded 100×10000 matrix is performed three times, or the number of the tensor rank.

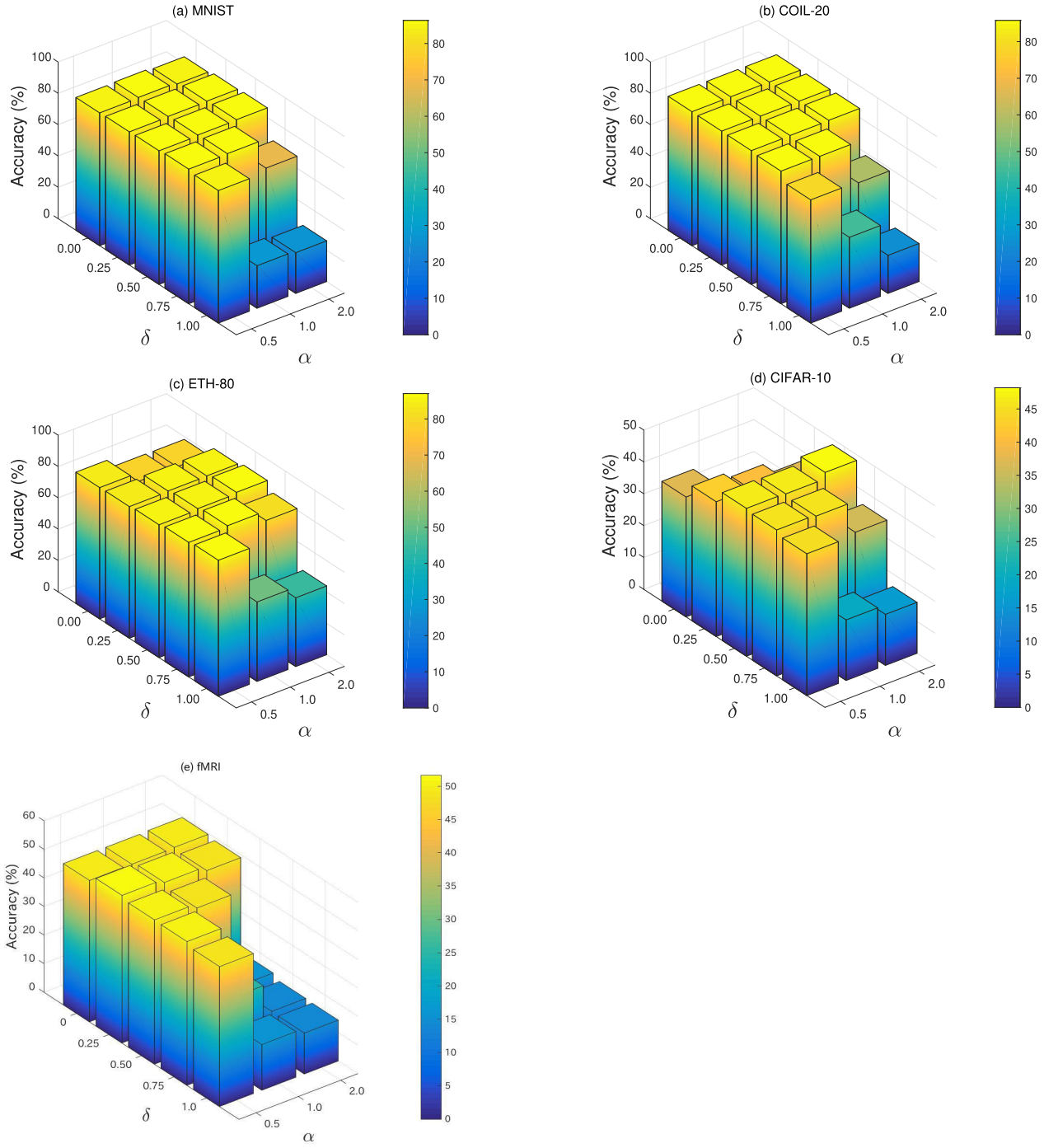


FIGURE 3. Classification accuracies corresponding to the changes of two parameters using SFGMCCA, α and δ, are shown as three-dimensional bar graphs for all datasets, (a) MNIST, (b) COIL-20, (c) ETH-80, (d) CIFAR-10 and (e) fMRI. When both parameters were simultaneously set to large values, the accuracy greatly decreased. Whereas, if two types of balances that can be set by both parameters are equalized; that is, if they are set to around $(\alpha, \delta) = (1.0, 0.50)$, there is a tendency to stabilize with high accuracy.

words, a decrease occurred when the geometrical structures were considered to be more important than the correlation structure and inter-class scattering was considered to be more important than intra-class scattering. On the other hand, when δ is smaller, only (d) showed a significant decrease in accuracy regardless of the size of α. Stable high accuracies were recorded by a medium setting, around $(\alpha, \delta) = (1.0, 0.50)$. From the above, when SFGMCCA is utilized, it is desirable to equalize two types of balances that can be set by both parameters and then search for a set of values that conforms to a target dataset. Moreover, the larger the number of neighborhoods k is, the higher the accuracies generally are in any dimension. The parameters with the best accuracy (k, α, δ) were (200, 1.0, 0.75) for MNIST, (120, 2.0, 0.50) for

COIL-20, (200, 1.0, 0.75) for ETH-80, and (300, 2.0, 0.50) for CIFAR-10. Moreover, the tendency of the result used in the fMRI dataset was similar to those in other datasets since the performance was decreasing when both α and δ were high. The parameters with the best accuracy (k, α, δ) were (100, 0.5, 0.25) for the fMRI dataset.

3) COMPUTATIONAL COMPLEXITY

When $D = \max(D_1, D_2, \ldots, D_M, L)$ is given, time complexity of calculation of the covariance matrix, the eigenvalue problem and the scatter matrix are $O((MD)^2N)$, $O((MD)^3)$ and $O(MN^2D)$, respectively. Note that M and N are the number of views and the number of samples, respectively. The time complexity of MCCA and RMCCA is $O((MD)^2N + (MD)^3)$. This is a baseline of the complexity. That of TCCA is $O(\mu_{\text{tensor}}\kappa D_1\, D_2 \ldots D_M L)$. Specifically, $\kappa < \min(D_1, D_2, \ldots, D_M, L)$ is a parameter of the alternating least squares (ALS) algorithm used for the calculation of a TCCA-based projection matrix, and μ_{tensor} is the number of iterations for calculation of the ALS algorithm. Thus, since TCCA has more time complexity as the number of views is increasing, TCCA is not computationally preferable in the consideration of the multi-view analysis. Furthermore, the complexity of LapMCCs and GrMCCs incur the time complexity $O((MD)^2N + \mu_{\text{lapm}}(MN^2D + (MD)^3))$ and $O((MD)^2N + \mu_{\text{grm}}(MN^2D + (MD)^3))$. Note that μ_{lapm} and μ_{grm} are the number of iterations for searching parameters of LapMCCs and GrMCCs, respectively. Since sMVCCA is constructed by using label features as one modality, sMVCCA has the same complexity as MCCA and RMCCA. Also, since DMCCA can be constructed by adding the scatter matrix to the MCCA-based approach, the complexity of DMCCA is $O((MD)^2N + MN^2D + (MD)^3)$. Moreover, the time complexity of SFEMCCA is $O((MD)^2N + \mu_{\text{sfem}}(MD)^3 + (MD)^3)$, where μ_{sfem} is the number of iterations for searching fractional order parameters. Finally, since SFGMCCA can be derived by combining SFEMCCA and the geometrical approach, SFGMCCA incurs the time complexity $O((MD)^2N + \mu_{\text{sfgm}}(MN^2D + (MD)^3))$, where μ_{sfgm} is the number of iterations for searching parameters including a weight parameter α and a trade-off parameter δ. Although SFGMCCA has higher complexity than MCCA and RMCCA, the complexity is almost the same as GrMCCs. Furthermore, SFGMCCA can achieve the high accurate classification performance, and it was therefore verified that SFGMCCA was effective feature integration approach.

V. CONCLUSIONS

In this paper, we presented supervised fractional-order embedding multi-view geometrical canonical correlation analysis (SFGMCCA). We previously proposed a supervised approach, i.e., SFEMCCA, that can avoid noise disturbance and handle data with a small number of samples and a large number of dimensions. In order to improve the integration capability of heterogeneous features, we newly introduce two kinds of geometrical structures for intra-view and inter-view besides the conventional correlation structure. The intra-view structure constructs graphs for performing locality structure preservation and discriminant analysis within each feature, and the inter-view structure similarly constructs them between heterogeneous features. From experiments using typical toy datasets for image classification and the public fMRI dataset, we showed high integration capability of SFGMCCA compared to that of several benchmarking methods. We are now studying the theory to construct geometrical structures with higher precision with the aim of developing a method with a level of feature integration.

Recently, cross-modal retrieval has attracted attention. Specifically, hashing methods have gained increasing interests in facilitating large-scale cross-view retrieval tasks [56], [57]. For example, semi-paired discrete hashing [56] constructs the cross-view similarity graph with paired data to preserve the similarities of semi-paired data in the constructed common latent subspace. Furthermore, multi-view discrete hashing [57] for integrating several visual features has been proposed. Since these methods focus on the construction of the canonical space, it can be expected that SFGMCCA can be extended to the cross-modal retrieval. Therefore, we try to extend to the cross-modal retrieval approach in the future work.

REFERENCES

[1] S. Sun, "A survey of multi-view machine learning," *Neural Comput. Appl.*, vol. 23, nos. 7–8, pp. 2031–2038, Dec. 2013.
[2] A. Blum and T. Mitchell, "Combining labeled and unlabeled data with co-training," in *Proc. 11th Annu. Conf. Comput. Learn. Theory (COLT)*, 1998, pp. 92–100.
[3] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, nos. 3–4, pp. 321–377, Dec. 1936.
[4] D. R. Hardoon, S. Sandor, and S.-T. John, "Canonical correlation analysis: An overview with application to learning methods," *Neural Comput.*, vol. 16, no. 12, pp. 2639–2664, 2004.
[5] 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.
[6] F. Wang and D. Zhang, "A new locality-preserving canonical correlation analysis algorithm for multi-view dimensionality reduction," *Neural Process. Lett.*, vol. 37, no. 2, pp. 135–146, Apr. 2013.
[7] T.-K. Sun, S.-C. Chen, Z. Jin, and J.-Y. Yang, "Kernelized discriminative canonical correlation analysis," in *Proc. Int. Conf. Wavelet Anal. Pattern Recognit.*, Nov. 2007, pp. 1283–1287.
[8] X. He and P. Niyogi, "Locality preserving projections," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2004, pp. 153–160.
[9] J. Yang and X. Zhang, "Feature-level fusion of fingerprint and finger-vein for personal identification," *Pattern Recognit. Lett.*, vol. 33, no. 5, pp. 623–628, Apr. 2012.
[10] K. Fukunaga, *Introduction to Statistical Pattern Recognition*. New York, NY, USA: Academic, 1990.
[11] R. A. Fisher, "The use of multiple measurements in taxonomic problems," *Ann. Eugenics*, vol. 7, no. 2, pp. 179–188, Sep. 1936.
[12] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, *Eigenfaces Vs. Fisherfaces: Recognition Using Class Specific Linear Projection*. New Haven, CT, USA: T.R. Yale Univ., 1997.
[13] X. Zhang, N. Guan, Z. Luo, and L. Lan, "Discriminative locality preserving canonical correlation analysis," in *Proc. Chin. Conf. Pattern Recognit. (CCPR)*, 2012, pp. 341–349.
[14] T. Melzer, M. Reiter, and H. Bischof, "Appearance models based on kernel canonical correlation analysis," *Pattern Recognit.*, vol. 36, no. 9, pp. 1961–1971, Sep. 2003.
[15] R. Sawata, T. Ogawa, and M. Haseyama, "Novel audio feature projection using KDLPCCA-based correlation with EEG features for favorite music classification," *IEEE Trans. Affect. Comput.*, vol. 10, no. 3, pp. 430–444, Jul. 2019.

[16] J. R. Kettenring, "Canonical analysis of several sets of variables," *Biometrika*, vol. 58, no. 3, pp. 433–451, 1971.
[17] A. A. Nielsen, "Multiset canonical correlations analysis and multispectral, truly multitemporal remote sensing data," *IEEE Trans. Image Process.*, vol. 11, no. 3, pp. 293–305, Mar. 2002.
[18] Y. Takane, H. Hwang, and H. Abdi, "Regularized multiple-set canonical correlation analysis," *Psychometrika*, vol. 73, no. 4, pp. 753–775, 2008.
[19] R. Arora and K. Livescu, "Multi-view CCA-based acoustic features for phonetic recognition across speakers and domains," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, May 2013, pp. 7135–7139.
[20] Y.-H. Yuan, Y. Li, J. Liu, C.-F. Li, X.-B. Shen, G. Zhang, and Q.-S. Sun, "Learning multi-kernel multi-view canonical correlations for image recognition," *Comput. Vis. Media*, vol. 2, no. 2, pp. 153–162, Jun. 2016.
[21] 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.
[22] A. Sharma, A. Kumar, H. Daume, and D. W. Jacobs, "Generalized multiview analysis: A discriminative latent space," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2160–2167.
[23] M. Yang and S. Sun, "Multi-view uncorrelated linear discriminant analysis with applications to handwritten digit recognition," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2014, pp. 4175–4181.
[24] J. Ye, R. Janardan, Q. Li, and H. Park, "Feature reduction via generalized uncorrelated linear discriminant analysis," *IEEE Trans. Knowl. Data Eng.*, vol. 18, no. 10, pp. 1312–1322, Oct. 2006.
[25] S. Sun, X. Xie, and M. Yang, "Multiview uncorrelated discriminant analysis," *IEEE Trans. Cybern.*, vol. 46, no. 12, pp. 3272–3284, Dec. 2016.
[26] 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, Jan. 2017.
[27] Y. Luo, D. Tao, K. Ramamohanarao, C. Xu, and Y. Wen, "Tensor canonical correlation analysis for multi-view dimension reduction," *IEEE Trans. Knowl. Data Eng.*, vol. 27, no. 11, pp. 3111–3124, Nov. 2015.
[28] G. Lee, A. Singanamalli, H. Wang, M. D. Feldman, S. R. Master, N. N. C. Shih, E. Spangler, T. Rebbeck, J. E. Tomaszewski, and A. Madabhushi, "Supervised multi-view canonical correlation analysis (sMVCCA): Integrating histologic and proteomic features for predicting recurrent prostate cancer," *IEEE Trans. Med. Imag.*, vol. 34, no. 1, pp. 284–297, Jan. 2015.
[29] K. Maeda, S. Takahashi, T. Ogawa, and M. Haseyama, "Multi-feature fusion based on supervised multi-view multi-label canonical correlation projection," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 3936–3940.
[30] Y. Ito, T. Ogawa, and M. Haseyama, "Sfemcca: Supervised fractional-order embedding multiview canonical correlation analysis for video preference estimation," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2018, pp. 3086–3090.
[31] X. Shen and Q. Sun, "A novel semi-supervised canonical correlation analysis and extensions for multi-view dimensionality reduction," *J. Vis. Commun. Image Represent.*, vol. 25, no. 8, pp. 1894–1904, 2014.
[32] 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.
[33] O. Ledoit and M. Wolf, "A well-conditioned estimator for large-dimensional covariance matrices," *J. Multivariate Anal.*, vol. 88, no. 2, pp. 365–411, Feb. 2004.
[34] A. Hendrikse, L. Spreeuwers, and R. Veldhuis, "A bootstrap approach to eigenvalue correction," in *Proc. 9th IEEE Int. Conf. Data Mining*, Dec. 2009, pp. 818–823.
[35] K. Yata and M. Aoshima, "Effective PCA for high-dimension, low-sample-size data with noise reduction via geometric representations," *J. Multivariate Anal.*, vol. 105, no. 1, pp. 193–215, Feb. 2012.
[36] J. Liu, S. Chen, and X. Tan, "Fractional order singular value decomposition representation for face recognition," *Pattern Recognit.*, vol. 41, no. 1, pp. 378–395, Jan. 2008.
[37] 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.
[38] Y.-H. Yuan and Q.-S. Sun, "Fractional-order embedding multiset canonical correlations with applications to multi-feature fusion and recognition," *Neurocomputing*, vol. 122, pp. 229–238, Dec. 2013.
[39] 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.
[40] 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.
[41] Y. LeCun. *The MNIST Database of Handwritten Digits*. Accessed: Jan. 17, 2018. [Online]. Available: http://yann.lecun.com/exdb/mnist/
[42] S. A. Nene, S. K. Nayar, and H. Murase, "Columbia object image library (COIL-20)," Dept. Comput. Sci., Columbia Univ., New York, NY, USA, Tech. Rep. CUCS-005-96, 1996.
[43] B. Leibe and B. Schiele, "Analyzing appearance and contour based methods for object categorization," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2003, pp. II–409.
[44] A. Krizhevsky and G. Hinton, *Learning Multiple Layers of Features From Tiny Images*. Toronto, ON, USA: T.R. Univ., 2009.
[45] T. Horikawa and Y. Kamitani, "Generic decoding of seen and imagined objects using hierarchical visual features," *Nature Commun.*, vol. 8, no. 1, pp. 1–15, Aug. 2017.
[46] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.
[47] Y. Akamatsu, R. Harakawa, T. Ogawa, and M. Haseyama, "Estimating viewed image categories from human brain activity via semi-supervised fuzzy discriminative canonical correlation analysis," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 1105–1109.
[48] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2005, pp. 886–893.
[49] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *Int. J. Comput. Vis.*, vol. 42, no. 3, pp. 145–175, 2001.
[50] J. G. Daugman, "Uncertainty relation for resolution in space, spatial frequency, and orientation optimized by two-dimensional visual cortical filters," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 2, no. 7, pp. 1160–1169, 1985.
[51] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics Intell. Lab. Syst.*, vol. 2, nos. 1–3, pp. 37–52, 1987.
[52] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556
[53] C. Cortes and V. Vapnik, "Support-vector networks," *Mach. Learn.*, vol. 20, no. 3, pp. 273–297, 1995.
[54] L. R. Tucker, "Some mathematical notes on three-mode factor analysis," *Psychometrika*, vol. 31, no. 3, pp. 279–311, Sep. 1966.
[55] L. De Lathauwer, B. De Moor, and J. Vandewalle, "A multilinear singular value decomposition," *SIAM J. Matrix Anal. Appl.*, vol. 21, no. 4, pp. 1253–1278, 2000.
[56] X. Shen, F. Shen, Q.-S. Sun, Y. Yang, Y.-H. Yuan, and H. T. Shen, "Semi-paired discrete hashing: Learning latent hash codes for Semi-paired cross-view retrieval," *IEEE Trans. Cybern.*, vol. 47, no. 12, pp. 4275–4288, Dec. 2017.
[57] X. Shen, F. Shen, L. Liu, Y.-H. Yuan, W. Liu, and Q.-S. Sun, "Multiview discrete hashing for scalable multimedia search," *ACM Trans. Intell. Syst. Technol.*, vol. 9, no. 5, pp. 1–21, Jul. 2018.

KEISUKE MAEDA (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in electronics and information engineering from Hokkaido University, Japan, in 2015, 2017, and 2019, respectively. He is currently a Specially Appointed Assistant Professor with the Office of Institutional Research, Hokkaido University. His research interests include multimodal signal processing and machine learning and its applications. He is a member of IEICE.

YOSHIKI ITO (Member, IEEE) received the B.S. and M.S. degrees in information science and technology from Hokkaido University, Japan, in 2017 and 2019, respectively. His research interest includes computer vision and its applications. He is a member of IEICE.

TAKAHIRO OGAWA (Senior Member, IEEE) received the B.S., M.S., and Ph.D. degrees in electronics and information engineering from Hokkaido University, Japan, in 2003, 2005, and 2007, respectively. He joined the Graduate School of Information Science and Technology, Hokkaido University, in 2008. He is currently an Associate Professor with the Faculty of Information Science and Technology, Hokkaido University. His research interests include AI, the IoT, and big data analysis for multimedia signal processing and its applications. He is a member of ACM, IEICE, and ITE. He was a Special Session Chair of the IEEE ISCE2009, a Doctoral Symposium Chair of ACM ICMR2018, an Organized Session Chair of the IEEE GCCE2017-2019, a TPC Vice Chair of the IEEE GCCE2018, a Conference Chair of the IEEE GCCE2019, and so on. He has been also an Associate Editor of *ITE Transactions on Media Technology and Applications*.

MIKI HASEYAMA (Senior Member, IEEE) received the B.S., M.S., and Ph.D. degrees in electronics from Hokkaido University, Japan, in 1986, 1988, and 1993, respectively. She joined the Graduate School of Information Science and Technology, Hokkaido University, as an Associate Professor, in 1994. From 1995 to 1996, she was a Visiting Associate Professor with Washington University, USA. She is currently a Professor with the Faculty of Information Science and Technology, Hokkaido University. Her research interests include image and video processing and its development into semantic analysis. She is a fellow of ITE and a member of IEICE and ASJ. She has been a Vice-President of the Institute of Image Information and Television Engineers (ITE), Japan. She has been the Editor-in-Chief of *ITE Transactions on Media Technology and Applications*. She has been the Director of the International Coordination and Publicity, Institute of Electronics, Information and Communication Engineers (IEICE).