

Regularized graph-embedded covariance discriminative learning for image set classification

Hengliang Tan,* Ying Gao, Jiao Du, and Shuo Yang

Guangzhou University, School of Computer Science and Cyber Engineering, Guangzhou, China

Abstract. Riemannian manifold has attracted an increasing amount of attention for visual classification tasks, especially for video or image set classification. Covariance matrices are the natural second-order statistics of image sets. However, nonsingular covariance matrices, known as symmetric positive defined (SPD) matrices, lie on the non-Euclidean Riemannian manifold (SPD manifold). Covariance discriminative learning (CDL) is an effective discriminative learning method that employs the Riemannian manifold in the SPD kernel space. However, in practice, the discriminative learning of CDL often suffers from the problems of poor generalization and overfitting caused by a finite number of training samples and noise corruption. Hence, we propose to address these problems by importing eigenspectrum regularization and graph-embedded frameworks. Discriminative learning with SPD manifold is generalized by the graph-embedded framework, which combines with eigenspectrum regularization in the SPD kernel space. Three local Laplacian graphs of graph-embedded framework and two eigenspectrum regularized models are incorporated to the proposed method. Comprehensive mathematical deduction of the proposed method is depicted with the “kernel tricks.” Experimental results on set-based face recognition and object categorization tasks reveal the effectiveness of the proposed method. © The Authors. Published by SPIE under a Creative Commons Attribution 4.0 Unported License. Distribution or reproduction of this work in whole or in part requires full attribution of the original publication, including its DOI. [DOI: [10.1117/1.JEI.29.4.043018](https://doi.org/10.1117/1.JEI.29.4.043018)]

Keywords: symmetric positive defined manifold; covariance discriminative learning; eigenspectrum regularization; graph-embedded framework; image set classification.

Paper 200283 received Apr. 19, 2020; accepted for publication Jul. 22, 2020; published online Aug. 4, 2020.

1 Introduction

The development of intelligent video surveillance, social networks, and electronic commerce enables a probe image set to be matched against all gallery image sets that becomes an image set classification task.¹ Image sets can be extract from videos or albums. Each probe image set and gallery image set contains multiple images that belong to the same class, which allows the extraction of considerably more discriminative information than is possible in the traditional single image classification task.² Image set classification has achieved widespread success in face recognition^{3–6} and object categorization.^{7–11}

Recently, many studies have indicated that numerous particular visual features lie on a Riemannian manifold.¹² The subspaces of image sets form the Grassmann manifold, the symmetric positive definite (SPD) matrices form the SPD manifold,¹³ and the two-dimensional shapes lie on Kendall shape spaces.¹⁴ The use of Riemannian manifold to model image set and build the corresponding classifier for classification is popular in recent years.¹⁵ Subspace and covariance matrix are two typical representations for modeling image sets on the Riemannian manifold. The subspaces of image sets form the Grassmann manifold,⁸ and the nonsingular covariance matrices form the SPD manifold.¹⁰ Linear subspace is a popular choice for modeling image sets due to its excellent accommodation of image variations. Hence, the Grassmann manifold formed by subspaces is widely used for image set classification. However, linear subspace-based modeling has the limitation that it incorporates only relatively weak information (such as the subspace angles) about the location and boundary of the samples in the input space.¹⁰ The second-order statistic feature known as the (nonsingular) covariance matrices of image sets that

*Address all correspondence to Hengliang Tan, E-mail: tanheng2013@gmail.com

form the SPD manifold characterizes the set structure more faithfully.¹⁰ Many studies^{10,12,16} have shown the effectiveness of SPD manifold for image set classification and that the covariance descriptor is robust to noise and illumination variations.

Covariance discriminative learning (CDL)¹⁰ is one of the most representative methods that uses covariance descriptor for image set classification. Covariance descriptor provides a natural representation for an image set, which makes no assumption about the set data distribution. Hence, covariance descriptor characterizes the set structure more faithfully, and the representation possesses stronger resistance to outliers.¹⁰ The SPD manifold formed by covariance matrices is mapped to a high-dimensional reproducing kernel Hilbert spaces (RKHS), where Euclidean geometry applies. Subsequently, linear discriminant analysis (LDA)¹⁷ is applied to perform discriminative learning with the “kernel trick,” which is known as kernel discriminant analysis.¹⁸ CDL has achieved considerable results on set-based face recognition and objection categorization tasks.

In this work, we focus on the discriminative learning problem of SPD manifold on the mapped RKHS. Due to the conventional problems of linear discriminative learning, such as the singularity of within-class scatter matrix and the instability of its inverse caused by the finite number of training samples,¹⁷ CDL may suffer from overfitting and poor generalization since conventional problems may also occur during discriminative learning in the kernel space.¹⁹ To address the conventional problems of LDA, numerous approaches, such as the Fisherface LDA,¹⁷ direct LDA,²⁰ and null space LDA²¹ on linear Euclidean space, have been proposed. For the conventional problems in the kernel space, kernel methods of kernel Fisherface LDA,²² null space kernel LDA,¹⁹ and kernel direct-LDA²³ in the nonlinear kernel space exist. However, these approaches usually discard a subspace (either the principal space or null space) to circumvent the singularity before discriminant learning, which causes a loss of discriminative information.²⁴ Although dual-subspace LDA²⁵ considers the contributions of both subspaces, the associated average scaling factor may not be a suitable choice for information in the principal subspace. To address these problems, the eigenfeature regularization and extraction (ERE)²⁴ and complete discriminant evaluation and feature extraction (CDEFE)²⁶ approaches were proposed to address these problems in a linear flat space and nonlinear kernel Euclidean space, respectively. ERE considers that the entire eigenspace of the within-class scatter matrix \mathbf{S}_W should be retained for discriminant analysis and regularized by the eigenspectrum regularization weighting function. The entire eigenspace is partitioned into three parts according to the median operation, and three different strategies according to the eigenspectrum of \mathbf{S}_W are devised for regularization.²⁴ CDEFE tackles these problems in the kernel space by nonlinear mapping; it decomposes the kernel within-class variation matrix into principal and noise dominated subspaces. A weighting function that is based on the ratios of the successive eigenvalues of the eigenspectrum was proposed to circumvent the undue scaling of projection vectors.²⁶ Discriminative vectors by applying predicted eigenvalues²⁷ combined the eigenspectrum regularization models of ERE and CDEFE. Recently, regularized locality preserving discriminant embedding²⁸ and locality regularization embedding (LRE)²⁹ were proposed; these methods generalized the eigenfeature extraction of ERE by the graph-embedded framework to better preserve data locality. An adaptive locality preserving regulation model was devised for eigenspectrum regularization. The experimental results have demonstrated the effectiveness of these eigenspectrum regularization techniques.²⁹

Inspired by eigenspectrum regularization, in this work, we aim to address the conventional problems of CDL in discriminative learning by exploiting the eigenspectrum regularization with the graph-embedded framework in the RKHS, which is mapped from the SPD manifold. We refer to the proposed method as regularized graph-embedded covariance discriminative learning (RGCDL). Figure 1 shows the conceptual illustration of the proposed method. The main contributions of this paper are presented as follows.

1. We circumvent the instability, overfitting, and poor generalization on CDL¹⁰ with kernel eigenspectrum regularization architecture. The input elements in high-dimensional SPD manifold are reduced to lower-dimensional SPD manifold by principal component analysis (PCA).
2. We incorporate the graph-embedded framework with three local Laplacian graphs into the Riemannian kernel eigenspectrum regularization architecture to better preserve data

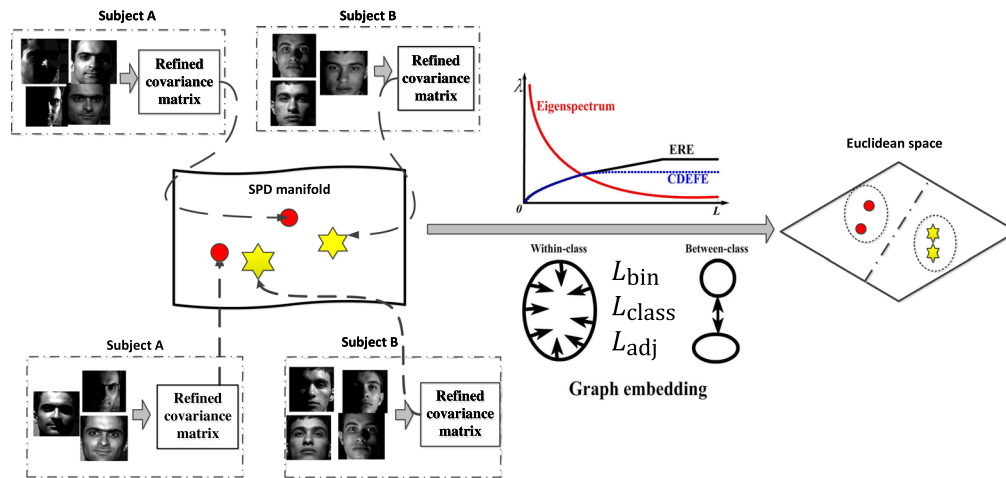


Fig. 1 Conceptual illustration of the proposed RGCDL. Image sets of subjects A and B can be described by refined covariance matrices, and the lower-dimensional SPD manifold is formed. Then, the eigenspectrum regularization models of ERE and CDEFE are incorporated with the graph-embedded framework for discriminative learning on the mapped RKHS, where Euclidean geometry applies. The eigenspectrum regularization circumvents the instability, overfitting, or poor generalization of discriminative learning, and the graph-embedded retains the local properties and increases the discriminatory power between classes.

locality. We give the systematic derivation of the graph-embedded framework that incorporates with the eigenspectrum in Riemannian kernel space.

3. We evaluate the advantages of the proposed method with two eigenspectrum regularization models on face recognition and object categorization tasks. The experimental results show the stability of the extracted features, robustness to noise-corrupted data, and high classification rate to numerous set-based classification methods.

The rest of this paper is organized as follows. We present the works related to image set classification according to the image set representations in Sec. 2. The original CDL method and the architecture of eigenspectrum regularization are introduced in Sec. 3. Then, the RGCDL approach is presented in Sec. 4. Experimental evaluation and discussions are presented in Sec. 5. Finally, Sec. 6 concludes this paper.

2 Related Work

In this paper, we aim to use the proposed method to solve the image set classification task. The major issues of image set classification focus on how to represent image set and measure the distance or similarity between two sets.⁵ Various techniques have been proposed to represent image set, such as the statistical distribution,^{30,31} affine/convex hull model,⁴ spare representation,⁵ subspace,^{7,32} and covariance matrix.^{10,12}

The methods^{30,31} that model each image set by statistical distribution are one of the earliest approaches employed for image set classification. They measure the similarities between pairs of distributions of two sets and achieve considerable results. However, if the set data have no strong statistical correlations for parameter estimation, these methods often fail to work.⁵ The most representative affine/convex hull-based methods are the affine/convex hull-based image set distance (AHISD/CHISD);⁴ AHISD/CHISD represents images as points in a linear or affine feature space and computes the distance of convex geometric region spanned by its feature points. Hu et al.⁵ incorporated the sparse representation to regularize the affine hull model. Zhu et al.⁶ employed the collaborative representation technique to utilize the discrimination information between gallery sets. The affine/convex hull approaches actually aim to find the synthetic nearest points between image sets.¹¹ However, these hull models usually cannot handle the complex appearance variations caused by multiple views and extreme illumination.

Subspace is a popular and effective approach for modeling image sets. Mutual subspace method (MSM)³² is one of the earliest classic subspace-based method for image set classification. MSM models all image sets by linear subspaces, and the similarity between pairs of subspaces is measured by canonical correlation analysis (CCA).³³ Fukui and Yamaguchi³⁴ and Fukui and Maki³⁵ projected the linear subspaces to a “difference subspace,” which can extract the disparity between two subspaces. Kim et al.⁷ incorporated discriminative learning into subspace-based set classification according to canonical correlations (DCC). DCC attempts to obtain a linear transformation that maximizes the canonical correlations of within-class subspaces and minimizes the canonical correlations of between-class subspaces. Arandjelovic³⁶ extended CCA to an extended version (ECCA) by extracting the most similar models of variability within two sets and exploited the discriminative learning architecture to train a classifier (DECCA).

Subspaces can also be treated as points that lie on a special type of Riemannian manifold, which is known as the Grassmann manifold. The method in Ref. 3 represents an image set as multiple local linear subspaces and treats them as points on the Grassmann manifold; then, the manifold-to-manifold distance (MMD) is defined between two manifolds of two image sets. Manifold discriminant analysis³⁷ was proposed to learn an embedding space by maximizing the manifold margin of the MMD. Grassmann manifold can also be mapped to an RKHS, where Euclidean geometry applies. Grassmann discriminant analysis (GDA)⁸ implements LDA on the mapped RKHS by the Grassmannian kernel. GDA is generated to kernel GDA (KGDA) using Gaussian kernel principal subspaces.³⁸ Graph-embedding Grassmann discriminant analysis (GGDA)⁹ is another counterpart to the GDA method; it exploits the graph-embedded framework to implement discriminant analysis on the mapped RKHS. Grassmann nearest points (GNP)¹¹ finds the nearest Grassmann points on the mapped vector space using the affine hull. More recently, regularized Grassmann discriminant analysis (RGDA)² was proposed to circumvent the conventional problems of LDA, when the training sets are insufficient. However, as previously mentioned, the linear subspace-based methods have the limitation of using weak information to measure the similarity.¹⁰ Modeling visual features as covariance matrices for visual classification has become popular in recent years^{10,12,39} since the nonsingular covariance matrix (as known as SPD matrix) can form a special Riemannian manifold, which is referred to as SPD manifold.¹² Previous studies employed covariance matrices to characterize local regions within an image, which is named the region covariance.³⁹ Different from the region covariance descriptor, CDL is the crucial method that models the whole image set by the covariance descriptor for addressing the image set classification with SPD manifold. Huang et al.⁴⁰ proposed log-Euclidean metric learning to learn a tangent mapping from the original tangent space of the SPD manifold to a new discriminative space. Tan and Gao¹⁶ proposed a patch-based principal covariance discriminative learning (PPCDL) method, in which the image set is partitioned into several local maximum linear patches by a hierarchical divisive clustering method, the local patches are modeled by covariance matrices, and the final discriminative learning is similar to CDL. Discriminant analysis on Riemannian manifold of Gaussian distributions (DARG)⁴¹ models the image set with a Gaussian mixture model (GMM) and derives a series of kernels for Gaussians discriminative learning on SPD manifold. Symmetric positive definite manifold learning¹² learns an orthonormal projection from the high-dimensional SPD manifold to a low-dimensional, more discriminative manifold.

3 Preliminaries

In this section, we first review the theory of CDL¹⁰ and then present the architecture of eigenspectrum regularization according to LRE.²⁹

3.1 Covariance Discriminative Learning

CDL uses a natural methodology to characterize image sets by the covariance descriptor. Let $\mathbf{X} = [x_1, x_2, \dots, x_n]$ denote the data matrix of an image set with n image vectors, where $x_i \in \mathbb{R}^D$ in the D -dimensional vector space. The covariance descriptor can be expressed as

$$\mathbf{B} = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})(x_i - \bar{x})^T, \tag{1}$$

where \bar{x} denotes the mean of image vectors in \mathbf{X} . The covariance matrix of \mathbf{B} represents one image set, which is rather simple to derive and compute. It is worth noting that, due to the high dimensionality of visual features and insufficient samples within set, the covariance matrix of an image set is usually singular (when the number of image samples is less than the dimensions of the vector space). A simple way to circumvent this problem is to introduce a small perturbation to the covariance matrix.¹⁰ This perturbation can be denoted as $\mathbf{B}^* = \mathbf{B} + \eta \mathbf{I}$, where \mathbf{I} is the identity matrix and η is a scaling parameter. Hence, the nonsingular covariance matrix becomes a $D \times D$ SPD matrix sym_D^+ , which is an element on Riemannian manifold. In the following paper, we still use \mathbf{B} to denote the nonsingular covariance matrix for simplicity. After modeling the image sets as multiple SPD matrices, CDL explores a Riemannian kernel that is induced by the Riemannian metric, such as the log-Euclidean distance (LED)⁴² to map the sym_D^+ to an Euclidean space. The Riemannian metric of LED defines a true geodesic on the Riemannian manifold, as it is induced by a positive definite kernel,⁴² and the manifold structure can be preserved as much as possible. The metric of LED is defined as

$$d_{\text{LED}}(\mathbf{B}_1, \mathbf{B}_2) = \|\log(\mathbf{B}_1) - \log(\mathbf{B}_2)\|_F, \tag{2}$$

where $\|\cdot\|_F$ is the matrix Frobenius norm and $\log(\cdot)$ denotes the principal matrix logarithm operation. The eigendecomposition of an SPD matrix \mathbf{B} is given by $\mathbf{B} = \mathbf{U}\mathbf{\Sigma}\mathbf{U}^T$, and it can compute the principal matrix logarithm of \mathbf{B} as

$$\log(\mathbf{B}) = \mathbf{U} \log(\mathbf{\Sigma}) \mathbf{U}^T, \tag{3}$$

where $\log(\mathbf{\Sigma})$ is easily calculated using the logarithms of the eigenvalues in the diagonal matrix $\mathbf{\Sigma}$. CDL implements image set classification in an extrinsic manner by first mapping the Riemannian manifold to an Euclidean space. The mapping induced by the LED metric can be defined as $\phi: \mathcal{M} \rightarrow \mathcal{H}$, where \mathcal{M} denotes the manifold spanned by the SPD matrices and the vector space \mathcal{H} is the inner product space on RKHS, which can be viewed as an Euclidean space \mathbb{R} . Subsequently, the kernel function induced by the LED metric, $k_{\log}: (\mathcal{M} \times \mathcal{M}) \rightarrow \mathbb{R}$ is used to define the inner product on RKHS. For two sym_D^+ matrices of \mathbf{B}_1 and \mathbf{B}_2 , the LED Riemannian kernel function can be formulated as

$$k_{\log}(\mathbf{B}_1, \mathbf{B}_2) = \text{tr}[\log(\mathbf{B}_1) \log(\mathbf{B}_2)]. \tag{4}$$

The kernel function k_{\log} is shown to be an SPD kernel^{10,13} that obeys Mercer's theorem.⁴³ Therefore, the manifold structure can be preserved by the LED Riemannian kernel.

The explicit kernel feature mapping allows application of any standard vector space learning algorithms. The discriminative learning of CDL is conducted by the kernel LDA¹⁸ with the kernel trick. The mapping of Riemannian manifold to an Euclidean space is defined by the function $\phi(\cdot)$. Therefore, if L points of the specified Riemannian manifold are spanned by the sym_D^+ matrices $\{\mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_L\}$, the mapped feature points on Euclidean space can be denoted as $\{f(\mathbf{B}_1), f(\mathbf{B}_2), \dots, f(\mathbf{B}_L)\}$. With the inner product $\langle \phi(\mathbf{B}_i), \phi(\mathbf{B}_j) \rangle = k_{\log}(\mathbf{B}_i, \mathbf{B}_j)$, CDL seeks to solve the following optimization:¹⁰

$$\boldsymbol{\alpha}_{\text{opt}} = \arg \max \frac{\boldsymbol{\alpha}^T \mathbf{K} \mathbf{W} \mathbf{K} \boldsymbol{\alpha}}{\boldsymbol{\alpha}^T \mathbf{K} \mathbf{K} \boldsymbol{\alpha}}, \tag{5}$$

where $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \dots, \alpha_L]^T$, \mathbf{K} is the kernel Gram matrix with elements $K_{ij} = k_{\log}(\mathbf{B}_i, \mathbf{B}_j)$, and \mathbf{W} is the connection matrix with element

$$W_{ij} = \begin{cases} \frac{1}{N_c}, & \text{if } \mathbf{B}_i \in C_c \text{ and } \mathbf{B}_j \in C_c, \\ 0, & \text{otherwise} \end{cases}, \tag{6}$$

where N_c is the number of sets in the c 'th class, we denote the c 'th class as C_c in this paper. Here, $\mathbf{B}_i \in C_c$ indicates that the label of \mathbf{B}_i belongs to class C_c . The optimal projection matrix is given by the largest $C - 1$ (C is the number of training classes) eigenvectors of solving the eigenproblem $\mathbf{K}\mathbf{W}\mathbf{K}\alpha = \lambda\mathbf{K}\mathbf{K}\alpha$, which is denoted as $\mathbf{A} = [\alpha_1, \alpha_2, \dots, \alpha_{C-1}]$. Finally, for a given testing sym_D^+ matrix $\mathbf{B}^{te} \in \mathcal{M}$ in the input manifold space. The projected feature z^{te} in the new discriminant Euclidean subspace can be obtained by

$$z^{te} = \mathbf{A}^T \mathbf{K}^{te}, \quad \mathbf{K}^{te} = [k_{\log}(\mathbf{B}_1, \mathbf{B}^{te}), \dots, k_{\log}(\mathbf{B}_L, \mathbf{B}^{te})]^T. \quad (7)$$

3.2 Eigenspectrum Regularization Technique

Eigenspectrum regularization^{24,26,29} was originally proposed to address the conventional problems (problems caused singularity of \mathbf{S}_W and the numerical instability of its inverse) of LDA on the linear Euclidean space. In this section, we introduce LRE²⁹ as an instance since it is the prototype of the proposed method in this paper.

Consider n samples of training data $\mathbf{X} = [x_1, x_2, \dots, x_n]$ with $\{x_i \in \mathbb{R}^D | i = 1, 2, \dots, n\}$. In LRE, intrinsic data structure is modeled to regularize the directions of data locality $\mathbf{X}\mathbf{L}_{\text{loc}}\mathbf{X}^T$.²⁹ The eigenspectrum and directions can be obtained by decomposing

$$\lambda = \mathbf{V}^T \mathbf{X}\mathbf{L}_{\text{loc}}\mathbf{X}^T \mathbf{V}, \quad (8)$$

where \mathbf{L}_{loc} is the local Laplacian matrix, which manifests the manifold through local geometry preservation,²⁹ and λ is a diagonal matrix whose diagonal elements are the eigenvalues $[\lambda_1, \lambda_2, \dots, \lambda_D]$ in descending order. The plot of the eigenvalues λ_k against the index k is referred to as the eigenspectrum. $\mathbf{V} = [v_1, \dots, v_D]$ contains the eigenvectors (directions) of the locality preserving matrix $\mathbf{X}\mathbf{L}_{\text{loc}}\mathbf{X}^T$ corresponding to λ . The locality preserving matrix $\mathbf{X}\mathbf{L}_{\text{loc}}\mathbf{X}^T$ has been shown to be exactly equal to the within-class scatter matrix with equal weights on the edges of adjacent data pairs of \mathbf{L}_{loc} .²⁸

LRE decomposes the entire eigenspace \mathbf{V} into two subspaces: (1) the disparity subspace $\mathbf{V}_{\text{disparity}} = [v_1, v_2, \dots, v_q]$, which corresponds to lower locality preservation, and (2) the principal subspace $\mathbf{V}_{\text{principal}} = [v_{q+1}, v_{q+2}, \dots, v_D]$ for higher locality preservation. LRE indicates that the first few eigenvectors of the eigenspace correspond to large eigenvalues that provide lower locality preserving capability, whereas the eigenvectors that correspond to smaller eigenvalues provide higher locality preserving capability. Hence, larger weights are imposed on the subspace with higher locality preservation, whereas smaller weights are assigned to the subspace with lower locality preservation. A method is devised by determining "fences" to separate the disparity subspace and principal subspace, and then regularize these two subspace according to an adaptive eigenspectrum regularization model. The fences is defined by a split point on eigenspectrum $\lambda_{\text{disparity}} = \gamma(Q3 + 1.5 \times IQR)$, where $Q3$ is the third quartile (cutting off the highest 75% or lowest 25% of the sum of λ), and γ is a parameter for adaptively scaling the separating value. The definition of IQR is $IQR = Q3 - Q1$, where $Q1$ is the first quartile.

This adaptive eigenspectrum regularization model finds the q 'th split eigenvalue that satisfies $\lambda_q = \max\{\forall \lambda_i | \lambda_i \leq \lambda_{\text{disparity}}\}$. The piecewise regularization function of LRE is defined as

$$w_k^{\text{LRE}} = \begin{cases} \lambda_k^{-1/2}, & 1 \leq k \leq q \\ \lambda_q^{-1/2}, & q < k \leq D \end{cases}. \quad (9)$$

The regularization function is imposed on the corresponding eigenvectors to form a full-dimensional transformation matrix

$$\tilde{\mathbf{V}} = [w_k^{\text{LRE}} v_k]_{k=1}^D. \quad (10)$$

Then, LRE can obtain a more localized feature by transforming the original training data

$$\tilde{\mathbf{X}} = \tilde{\mathbf{V}}^T \mathbf{X}. \tag{11}$$

We indicate that there is no dimensional reduction has occurred in this transforming. The information of the original training data is preserved as much as possible.

In the subsequent step of LRE, feature extraction and dimensional reduction from the regularized and more compact data are performed. To further preserve the within-locality and between-locality power, a similarity weight matrix \mathbf{G}^{LRE} is utilized. The element of \mathbf{G}^{LRE} is defined as

$$G_{ij}^{\text{LRE}} = \begin{cases} \frac{1}{n_c} - \frac{1}{n}, & \text{if } x_i \in C_c \text{ and } x_j \in C_c, \\ -\frac{1}{n}, & \text{otherwise} \end{cases}, \tag{12}$$

where n_c is the number of samples in the c 'th class. The within-locality graph edges are weighted with positive-valued coefficients that quantify the intraclass similarity, whereas the between-locality graph edges are weighted with negative-valued coefficients that characterize discriminative features among different class samples.²⁹ The final objective function of LRE is defined as

$$\mathbf{U}^* = \arg \max_{\mathbf{U}^T \mathbf{U} = \mathbf{I}} \mathbf{U}^T \tilde{\mathbf{X}} \mathbf{G}^{\text{LRE}} \tilde{\mathbf{X}}^T \mathbf{U}. \tag{13}$$

This problem can be easily solved by converting it to a generalized eigenvalue problem $\tilde{\mathbf{X}} \mathbf{G}^{\text{LRE}} \tilde{\mathbf{X}}^T \mathbf{u}_i = \varphi_i \mathbf{u}_i$. By retaining d eigenvectors $\mathbf{U} = [u_1, u_2, \dots, u_d]$ ($d \leq D$), which correspond to the largest d eigenvalues, the projection matrix $\mathbf{Z}^{\text{LRE}} = \tilde{\mathbf{V}} \mathbf{U}$ is used for the final lower-dimensional eigenfeature extraction.

4 Proposed Method

In this section, the proposed RGCDL is presented. To incorporate the eigenspectrum regularization and graph-embedded framework with SPD manifold in the kernel space, the algorithm of RGCDL is quite different from the original CDL algorithm. Generally, the algorithm of our RGCDL mainly comprises of two steps. The first step is eigenspectrum regularization, and the second step is feature extraction and dimensional reduction.

4.1 Representation of SPD Manifold

At first, according to Harandi et al.¹² and Tan and Gao,¹⁶ the computational cost of the Riemannian kernel with a high-dimensional SPD matrix is quite high. Several strategies are available to lower the dimensionality of the SPD matrix and reduce the computational cost of constructing the Riemannian kernel matrix.¹² Here, we combine all the training data in different training sets to collaboratively produce the dimensional reduction projection matrix by PCA.

Consider L training image sets $\chi = \{\mathbf{X}_i\}_{i=1}^L$, each set contains L_i images $\mathbf{X}_i = [x_1, x_2, \dots, x_{L_i}]$. We combine all images of all sets to build a sample data collection

$$\chi = \{x_j\}_{j=1}^N, \quad N = \sum_{i=1}^L L_i. \tag{14}$$

The dimensional reduction projection matrix can be obtained by decomposing the following sample covariance matrix:

$$\mathbf{\Pi} = \frac{1}{N} \sum_{j=1}^N (x_j - \bar{x})(x_j - \bar{x})^T, \tag{15}$$

where \bar{x} is the sample mean. We select d_1 ($d_1 \leq D$) orthonormal eigenvectors that correspond to the d_1 largest eigenvalues of $\mathbf{\Pi}$ to form the dimensional reduction projection matrix $\mathbf{\Gamma}$.

All images in each set are transformed to a low-dimensional feature space, and the j 'th sample in the low-dimensional feature space is calculated as

$$y_j = \Gamma x_j. \quad (16)$$

This simple PCA that is applied to all training sets not only alleviates the problem of the high computational complexity of constructing the SPD kernel matrix but also better preserves the main variations in the set data to build the covariance matrices, which form the SPD manifold. This operation of refining the high-dimensional SPD matrices can also be viewed as a transformation from the high-dimensional manifold to a low-dimensional manifold.¹⁶

4.2 Eigenspectrum Regularization with SPD Manifold

The i 'th dimensional reduced set can be represented as $\mathbf{Y}_i = [y_1^i, y_2^i, \dots, y_{L_i}^i]$. \mathbf{Y}_i can be modeled by the covariance descriptor [Eq. (1)] and represented as $\mathbf{B}_i \in \mathbb{R}^{d_i \times d_i}$. To ensure that \mathbf{B}_i is non-singular to form the SPD manifold, a small perturbation is added to the covariance matrix $\mathbf{B}_i + \eta \mathbf{I}$. Hence, the perturbed \mathbf{B}_i is an SPD matrix $\text{sym}_{d_i}^+$.

For L image sets of C classes, they can be denoted as a collection of $\text{sym}_{d_i}^+$ matrices $\mathbf{B} = \{\mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_L\}$ that form an SPD manifold. By defining the Riemannian mapping $\phi: \mathcal{M} \rightarrow \mathcal{H}$, we can obtain the samples $\Phi(\mathbf{B}) = [\phi(\mathbf{B}_1), \phi(\mathbf{B}_2), \dots, \phi(\mathbf{B}_L)]$ on the RKHS \mathcal{H} , which is homeomorphic to Euclidean space.

To further preserve the local structure, we incorporate the graph-embedded framework into our proposed method. The local Laplacian matrix \mathbf{L}_{loc} is utilized to preserve the locality information, whereas the global Laplacian matrix is adjacent regardless of the class membership of all vertices.²⁹ In this step, we aim to obtain the eigenspectrum and directions of the local structure in the SPD Riemannian kernel space. They can be implemented by decomposing the locality preserving matrix $\Phi(\mathbf{B})\mathbf{L}_{\text{loc}}\Phi(\mathbf{B})^T$ on the mapped space; we denote $\Phi(\mathbf{B})$ as Φ for simplicity. Then, we have

$$\lambda = \mathbf{V}^T \Phi \mathbf{L}_{\text{loc}} \Phi^T \mathbf{V}, \quad (17)$$

where \mathbf{V} constructs the kernel eigenspace of $\Phi \mathbf{L}_{\text{loc}} \Phi^T$, and the eigenvalues in λ define the kernel eigenspectrum. The local Laplacian matrix \mathbf{L}_{loc} can be specified into different local Laplacian graphs. In this work, we employ the binary local Laplacian \mathbf{L}_{bin} , intraclass local Laplacian $\mathbf{L}_{\text{class}}$, and adjustable local Laplacian $\mathbf{L}_{\text{adjloc}}$ for instances. \mathbf{L}_{bin} is a simple-minded Laplacian matrix in which intraclass vertices are adjacent with equal weight of each edge. $\mathbf{L}_{\text{class}}$ is the Laplacian graph that satisfies

$$\mathbf{L}_{\text{class}} = \begin{cases} 1 - W_{ij}, & i = j, \\ -W_{ij}, & i \neq j, \text{ if } \mathbf{B}_i \in C_c \text{ and } \mathbf{B}_j \in C_c, \\ 0, & \text{otherwise} \end{cases} \quad (18)$$

where W_{ij} is the connection weight of the i 'th and j 'th sets, and it has the same definition as Eq. (6) in CDL.¹⁰ The locality preserving matrix $\Phi \mathbf{L}_{\text{class}} \Phi^T$ can be proved to be equal to the kernel within-class scatter matrix \mathbf{S}_W^Φ with equal weights on the edges of adjacent data pairs of $\mathbf{L}_{\text{class}}$.²⁹

Unlike the edge weights in \mathbf{L}_{bin} and $\mathbf{L}_{\text{class}}$, which are fixed in values, the edge weights in \mathbf{L}_{adj} are variables that are based on different similarity definitions, such as the heat kernel in locality preserving projections⁴⁴ and neighborhood reconstruction coefficients in neighborhood preserving embedding.⁴⁵ \mathbf{L}_{adj} is a Laplacian matrix that is computed by

$$\mathbf{L}_{\text{adj}} = \mathbf{D} - \mathbf{W}, \quad (19)$$

where \mathbf{D} is a diagonal matrix calculated by $D_{ii} = \sum_j W_{ij}$. In this paper, we compute the edge weights of \mathbf{W} in Eq. (19) based on the heat kernel, which is calculated by Gaussian distribution. The edge weight W_{ij} of \mathbf{L}_{adj} can be computed as

$$W_{ij} = \begin{cases} \exp\left(-\frac{\|\phi(\mathbf{B}_i) - \phi(\mathbf{B}_j)\|^2}{\sigma}\right), & \text{if } \mathbf{B}_i \in C_c \text{ and } \mathbf{B}_j \in C_c, \\ 0, & \text{otherwise} \end{cases}, \quad (20)$$

where σ is the kernel width parameter. The Euclidean distance of the mapped feature $\phi(\mathbf{B}_i)$ and $\phi(\mathbf{B}_j)$ can be easily transformed to $K_{ii} - 2K_{ij} + K_{jj}$, where K_{ij} can be calculated by the Riemannian kernel, such as Eq. (4).

As known by linear algebra, the projection direction of \mathbf{V} in Eq. (17) can be represented as a linear combination of the eigenspace on the mapped space

$$\mathbf{V} = \Phi\alpha. \quad (21)$$

By substituting Eq. (21) into Eq. (17), we obtain $\lambda = \alpha^T \Phi^T \Phi \mathbf{L}_{\text{loc}} \Phi^T \Phi \alpha$. We use the Riemannian kernel function [e.g., Eq. (4)] to build the kernel Gram matrix $\mathbf{K} = \Phi^T \Phi$, Eq. (17) can be rewritten as

$$\lambda = \alpha^T \mathbf{K} \mathbf{L}_{\text{loc}} \mathbf{K} \alpha. \quad (22)$$

Equation (22) can be solved by the eigendecomposition of $\mathbf{K} \mathbf{L}_{\text{loc}} \mathbf{K}$ subject to $\alpha^T \alpha = 1$.

In the theory of eigenspectrum regularization, we need to regularize the whole feature space \mathbf{V} [see Eq. (21)] on the mapped space. Assume that, the regularization can be generalized to the weighting function, which is defined as

$$\mathbf{w} = [w_k]_{k=1}^L, \quad (23)$$

where L is the number of all training sets. The full-dimensional feature space \mathbf{V} contains L vectors $[v_1, v_2, \dots, v_L]$ since the dimensions of matrix $\mathbf{K} \mathbf{L}_{\text{loc}} \mathbf{K}$ is $L \times L$. Hence, the regularized eigenspace can be computed as

$$\tilde{\mathbf{V}} = [w_k v_k]_{k=1}^L. \quad (24)$$

According to Eq. (21), we have $v_k = \Phi\alpha_k$. By defining

$$\rho = [w_k \alpha_k]_{k=1}^L \quad (25)$$

the regularized eigenspace of Eq. (24) can be rewritten as

$$\tilde{\mathbf{V}} = \Phi\rho. \quad (26)$$

The regularized eigenspace is known as a transformation matrix²⁴ that can transform the original feature data to an intermediate feature vector space. It is worth noting that the transformation matrix $\tilde{\mathbf{V}}$ is a full-dimensional matrix with size $L \times L$. Hence, the mapped data $\Phi(\mathbf{B})$ from SPD manifold can be transformed to the new feature vector space $\tilde{\Phi}(\mathbf{B})$ with no dimensional reduction, which can preserve information as much as possible. We denote $\tilde{\Phi}(\mathbf{B})$ as $\tilde{\Phi}$ for simplicity. The transformation is depicted as

$$\tilde{\Phi} = \tilde{\mathbf{V}}^T \Phi. \quad (27)$$

Although Φ is implicitly defined, the transformed feature $\tilde{\Phi}$ can be explicitly expressed by the kernel trick. According to Eqs. (26) and (27), we can denote the transformed $\tilde{\Phi}$ by $\tilde{\Phi} = \rho^T \Phi^T \Phi$. As $\mathbf{K} = \Phi^T \Phi$, Eq. (27) can be rewritten as

$$\tilde{\Phi} = \rho^T \mathbf{K}. \quad (28)$$

In this aspect, according to the previous mathematical deduction, by defining the important regularized eigenspace ρ [see Eq. (25)], the regularization of eigenspace \mathbf{V} is turned into the

regularization of the eigenspace α . In other words, the effectiveness of eigenspectrum regularization model on the eigenspace α is equivalent to the eigenspace \mathbf{V} .

The selection of a suitable eigenspectrum regularization model is a critical aspect of the proposed method. The proper eigenspectrum regularization model ensures that the regularized data can be very close to the real population variances.⁴⁶ The eigenspectrum regularization of LRE is an adaptive model that estimates the optimal parameter γ using training data. However, this process is usually time-consuming, and the performance decays quickly when the training data are insufficient. In this paper, we employ the data-independent eigenspectrum regularization models of ERE and CDEFE to regularize the eigenspace α , which are more general and robust. The first model is the eigenspectrum regularization model of ERE.²⁴ The heuristic theory of ERE for designing the eigenspectrum regularization model is the median operation. The weighting function applied to the eigenspace α is defined as

$$w_k^{\text{ERE}} = \begin{cases} \lambda_k^{-1/2} & k < m_1 \\ \left(\frac{a}{k+b}\right)^{-1/2} & m_1 \leq k \leq r, \\ \left(\frac{a}{r+1+b}\right)^{-1/2} & r < k \leq L \end{cases}, \quad (29)$$

where m_1 is the m_1 'th eigenvalue of λ in descending order, which satisfies

$$\lambda_{m_1} = \max\{\forall \lambda_k | \lambda_k < [\lambda_{\text{med}} + \mu(\lambda_{\text{med}} - \lambda_r)]\}, \quad (30)$$

where λ_{med} is the median value computed by $\text{median}\{\forall \lambda_k | k \leq r\}$. μ is a constant with a recommendation value of 1.²⁴ r is the rank of $\mathbf{KL}_{\text{loc}}\mathbf{K}$. The parameters of a and b are calculated as

$$\begin{aligned} a &= \frac{\lambda_1 \lambda_{m_1} (m_1 - 1)}{\lambda_1 - \lambda_{m_1}}, \\ b &= \frac{m_1 \lambda_{m_1} - \lambda_1}{\lambda_1 - \lambda_{m_1}}. \end{aligned} \quad (31)$$

The second regularization model is taken from CDEFE.²⁶ The eigenspectrum regularization model of CDEFE regularizes the eigenspace in a Gaussian kernel space, which may have a special effect on the proposed RGCDL in the Riemannian kernel space.

The second regularization model aims to find the minimum eigenratio from the eigenspectrum of $\mathbf{KL}_{\text{loc}}\mathbf{K}$, which is formed by the eigenvalues [see Eq. (17)] in descending order. Let δ_k denote the ratio of two adjacent eigenvalues λ_k and λ_{k+1} in the eigenspectrum, we have

$$\delta_k = \lambda_k / \lambda_{k+1}. \quad (32)$$

The minimum eigenratio can be formulated as

$$\delta_s = \min\{\forall \delta_k, 1 \leq k < r\}, \quad (33)$$

where s is the index of the minimum eigenratio, and r is the rank of the locality matrix $\mathbf{KL}_{\text{loc}}\mathbf{K}$. The eigenspectrum λ is split by the point of the m_2 'th eigenvalue, and λ_s is defined as $\lambda_s = \max\{\forall \lambda_k, k \geq m_2\}$. Thus, the final regularized weighting function can be defined as

$$w_k^{\text{CDEFE}} = \begin{cases} \lambda_k^{-1/2}, & 1 \leq k \leq m_2 \\ \lambda_{m_2}^{-1/2}, & m_2 < k \leq L. \end{cases} \quad (34)$$

4.3 Feature Extraction and Dimensional Reduction

The new feature vectors $\tilde{\Phi}$ are compact and full dimensional, the eigenfeature of $\tilde{\Phi}$ should be decorrelated and dimension reduced for classification. According to Jiang et al.,²⁴ PCA is exploited to extract the final discriminative eigenfeatures since it is less sensitive to different

training databases. However, the class affinity is not considered in Ref. 24, which may cause the missing discriminative information. In this work, we employ a graph-embedded framework to extract the final discriminative features, which incorporates a similarity weight matrix to form the final scatter matrix. Although LRE had extended the graph-embedded framework to eigenfeature extraction, our method is designed to address the problems in a Riemannian kernel space.

According to Eq. (13), the eigenfeature extraction and dimensional reduction of RGCDL can be achieved by solving the following eigendecomposition problem on the mapped space:

$$\mathbf{U}^* = \arg \max_{\mathbf{U}^T \mathbf{U} = \mathbf{I}} \mathbf{U}^T \tilde{\Phi} \mathbf{G} \tilde{\Phi}^T \mathbf{U}, \quad (35)$$

where \mathbf{G} is the similarity weight matrix, and the affinity between set \mathbf{X}_i and set \mathbf{X}_j is defined as

$$\mathbf{G}_{ij} = \begin{cases} \frac{1}{N_c} - \frac{1}{L}, & \text{if } \mathbf{X}_i \in C_c \text{ and } \mathbf{X}_j \in C_c, \\ -\frac{1}{L}, & \text{otherwise} \end{cases}, \quad (36)$$

where N_c is the number of sets in the c 'th class. This similarity weight matrix allows the intra-class samples to be more compact and allows the interclass samples to be more separated. Clearly, the problem of Eq. (35) can be solved by decomposing matrix $\tilde{\Phi} \mathbf{G} \tilde{\Phi}^T$. The projection matrix \mathbf{U} consists of the eigenvectors that correspond to eigenvalues in descending order. We retain the first d_2 eigenvectors $\mathbf{U} = [u_1, u_2, \dots, u_{d_2}]$, where $d_2 \leq L$, for the final dimension of the extracted feature. Hence, the final regularized projection matrix of RGCDL can be defined as

$$\mathbf{Z} = \tilde{\mathbf{V}} \mathbf{U}. \quad (37)$$

Obviously, \mathbf{Z} does not have an explicit expression since $\tilde{\mathbf{V}} = \Phi \rho$, and Φ is the vector space mapped by the Riemannian mapping, which is implicitly defined.

However, an explicit expression can be provided by the kernel trick when calculated with the test samples. For a given test nonsingular covariance matrix \mathbf{B}^{te} , which is an element of the SPD manifold, we use ϕ^{te} to denote the test feature vector that is mapped by the Riemannian mapping. Subsequently, we can extract the discriminative feature \mathbf{F} by the transformation

$$\mathbf{F} = \mathbf{Z}^T \phi^{te} = \mathbf{U}^T \tilde{\mathbf{V}}^T \phi^{te} \quad (38)$$

Substitute $\tilde{\mathbf{V}}$ by Eq. (26), and calculate a kernel Gram matrix \mathbf{K}^{te} by the Riemannian kernel function [e.g., Eq. (4)], the final extracted eigenfeature can be rewritten as

$$\mathbf{F} = (\rho \mathbf{U})^T \mathbf{K}^{te}, \quad \mathbf{K}^{te} = [k_{\log}(\mathbf{B}_1, \mathbf{B}^{te}), \dots, k_{\log}(\mathbf{B}_L, \mathbf{B}^{te})]^T. \quad (39)$$

Here, \mathbf{F} is constructed by feature vectors on the mapped space. Hence, various distance metrics and classification methods that designed in Euclidean space, such as the nearest neighbor (NN) classifier, can be applied for classification.

4.4 Complete RGCDL Algorithm

The steps of RGCDL algorithm are given in Algorithm 1.

5 Experimental Results

Experiments were conducted on set-based face recognition and object categorization tasks. First, we compare the proposed RGCDL to the original CDL method when using different numbers of extracted features. Second, we show the advantages of RGCDL over the recent RGDA method.

Algorithm 1 RGCDL algorithm.

At the training stage:

1. Collaboratively reduce L training sets $\{\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_L\}$ by Eq. (16) to $\{\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_L\}$, and model them as $\text{sym}_{d_1}^+$ by the covariance descriptor [Eq. (1)], denoted as $\{\mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_L\}$.
2. Use Riemannian mapping $\phi: \mathcal{M} \rightarrow \mathcal{H}$ to map the $\text{sym}_{d_1}^+$ matrices $\{\mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_L\}$ to the RKHS, and designate them $\{\phi_1, \phi_2, \dots, \phi_L\}$.
3. Obtain the optimal coefficient α and the eigenspectrum λ by solving the eigendecomposition problem of Eq. (22); switch the local Laplacian matrix L_{loc} by L_{bin} , L_{class} , and L_{adj} , respectively.
4. Regularize the eigenspace \mathbf{V} using the weighting functions of Eqs. (29) and (34), respectively, and construct the full-dimensional transformation matrix \mathbf{V} using Eq. (24).
5. Transform the mapped feature Φ to an intermediate full-dimensional feature space using Eqs. (25) and (28).
6. Solve the eigendecomposition problem of Eq. (35), and compute the final projection matrix \mathbf{Z} by Eq. (37).

At the testing stage:

1. Given a test image set \mathbf{X}^{te} , project it to a lower-dimensional feature space and extract the covariance feature \mathbf{B}^{te} .
 2. Map \mathbf{B}^{te} to the RKHS by the Riemannian mapping, and designate as ϕ^{te} .
 3. Project ϕ^{te} to the final projection matrix \mathbf{Z} to extract the discriminative features by Eq. (39).
 4. Measure the extracted features between the training and test sets, and classify the label by the classifier (e.g., NN).
-

Last, we evaluate the recognition performance of our RGCDL, and compare it to numerous image set-based classification methods.

5.1 Dataset and Parameter Settings

We employed the Extended Yale face database B (ExtYaleB)⁴⁷ for face recognition task and the RGB-D object database⁴⁸ for object categorization task. The ExtYaleB database is the extension of the Yale face database B; it contains 16,128 images of 28 human subjects with 64 illumination conditions and 9 poses for each subject. According to the 9 poses of each subject, we built 9 image sets (~60 images per set), which correspond to 9 poses for each subject. We utilized a cascaded face detector⁴⁹ to collect faces from each image frame. The captured faces were then converted to grayscale and resized to 20×20 pixels. Some example images are shown in Fig. 2. We selected 2 to 5 image sets of 9 poses for discriminative training (103 sets) and employed the remaining sets for testing (149 sets). The experiments were repeated 10 times by randomly choosing the reference sets for training and the test sets for probe.

The RGB-D object database is a large-scale dataset of 300 common household objects that are organized into 51 categories (classes). Each category has 3 to 14 objects that belong to the same category. For each object, 3 video sequences are recorded with a camera that is mounted at different heights so that the object is viewed from different angles with the horizon. The video sequences were captured by placing each object on a turntable for a whole rotation using a Kinect style three-dimensional camera. More than 100 images were extracted for each object's video sequences; they involve RGB color channels and a depth channel. We removed the depth images in this study to ensure fair comparisons. We built image sets according to each object, forming a total of 300 image sets (102 sets for training and 198 sets for testing). Grayscale and resized images of 20×20 pixels were adopted for RGB-D dataset. Some example objects are shown in Fig. 3. To obtain more general results, we also conducted 10 cross-validation experiments by randomly choosing different combinations of training sets and test sets. The NN classifier was applied for all evaluations to ensure fair comparisons.

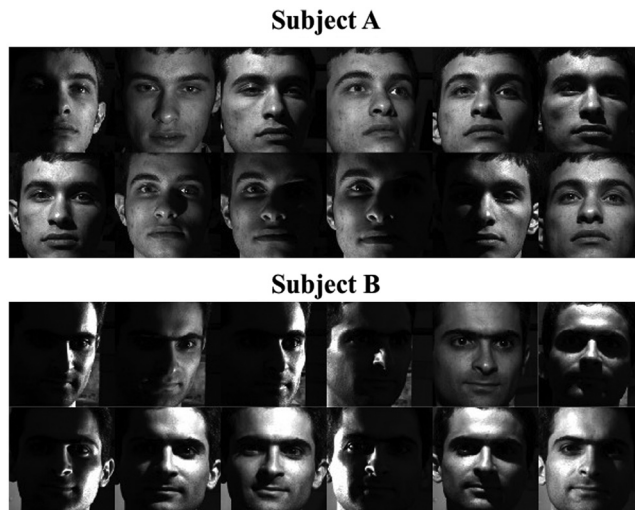


Fig. 2 Captured face examples from the ExtYaleB dataset.

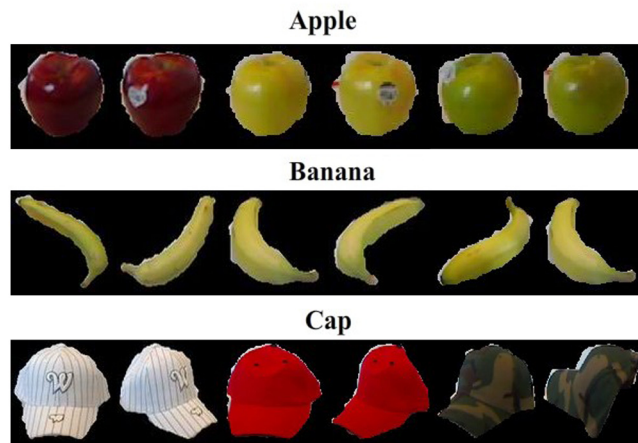


Fig. 3 Some example objects from the RGB-D dataset.

5.2 Stability of Extracted Features

In this section, we evaluated the stability of the extracted features from RGCDL. We show that by applying eigenspectrum regularization, the features extracted by RGCDL are more stable than those extracted by the original CDL method.

As described in Sec. 4, the proposed RGCDL aims to extract features from the whole regularized eigenfeature space. As the number of final extracted features increases [controlled by varying d_2 dimensions of \mathbf{U} in Eq. (37)], a higher performance can be achieved by our RGCDL, whereas the original CDL algorithm cannot retain this characteristic. To confirm this assumption and provide evidence, we employed real data of face and object datasets to conduct these experiments.

The collaborative dimensional reduction of each image set is set to 100 dimensions, that is, the dimension-reduced covariance matrix of each image set is 100×100 . Attributed to the covariance descriptor, the computational cost of constructing the Riemannian kernel matrices is not associated with the number of images within a set. The Riemannian kernel induced by the LED in Eq. (4) was employed for the RGCDL and CDL methods. We vary the final feature dimensions of RGCDL and CDL to perform a comprehensive comparison. The comparison results are shown in Figs. 4 and 5. Each figure consists of the recognition rates against the number of final extracted features. The recognition rates are the average results of

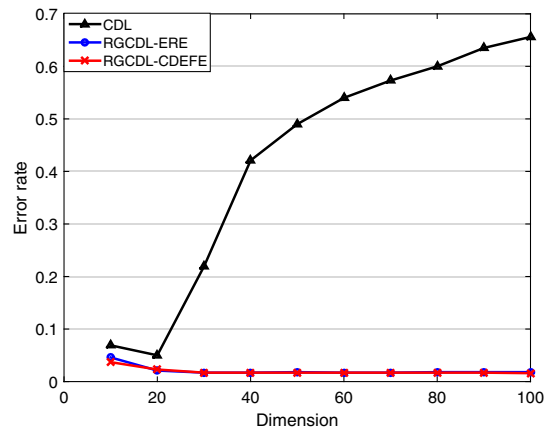


Fig. 4 Error rates using different numbers of features on the ExtYaleB dataset.

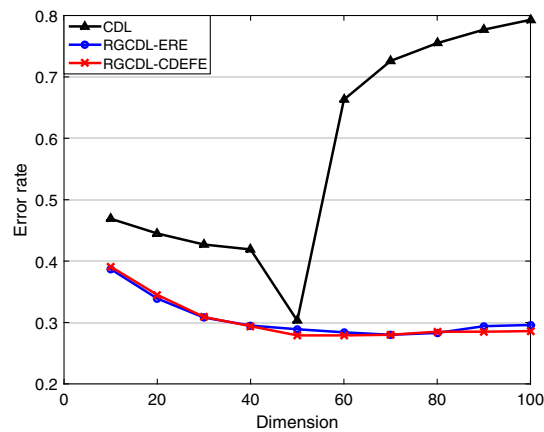


Fig. 5 Error rates using different numbers of features on the RGB-D dataset.

10 cross-validation experiments. Two eigenspectrum regularization models of ERE and CDEFE were evaluated for the proposed RGCDL.

As shown in Figs. 4 and 5, with the increasing dimensions of the final extracted features, the proposed RGCDL methods with two regularization models generally produce low error rates, whereas the original CDL degrades rapidly from the dimension $C - 1$. The dimension of $C - 1$ represents the optimal performance number of features in LDA.¹⁷ The degradation of CDL is caused by the incorrectly scaled null kernel space of the within-class scatter matrix,²⁴ which causes overfitting and poor generalization. These results reveal that, with the eigenspectrum regularization models, the conventional problems (e.g., the singularity of within-class scatter matrix) caused by limited training samples in CDL can be alleviated. Since the new feature space is properly scaled, the estimated eigenvalues obey the true variances of the population;²⁴ hence, better generalization can be achieved. The final extracted features are learned from the regularized full-dimensional transformation matrix $\tilde{\mathbf{V}}$ [Eq. (26)], which can preserve discriminative information as much as possible. As a result, using an increasing number of features, the recognition rates of RGCDL preserve the stable performance.

5.3 Performance Evaluation against RGDA

RGDA is the preliminary work of this paper, however, RGDA was proposed to solve the overfitting and poor generalization problems of Grassmann discriminative learning, whereas our RGCDL solves these problems against CDL on SPD manifold. Moreover, we further employed different local Laplacian graphs to analyze the locality preserving ability and improve the performance; the locality preserving is evaluated in Sec. 5.4. We discovered that the performance of

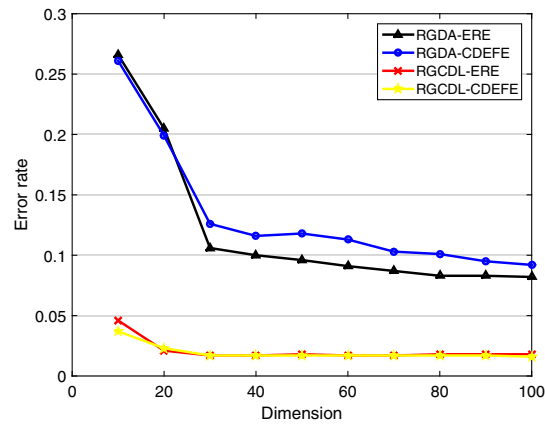


Fig. 6 Performance comparisons of RGCDL and RGDA on the ExtYaleB dataset.

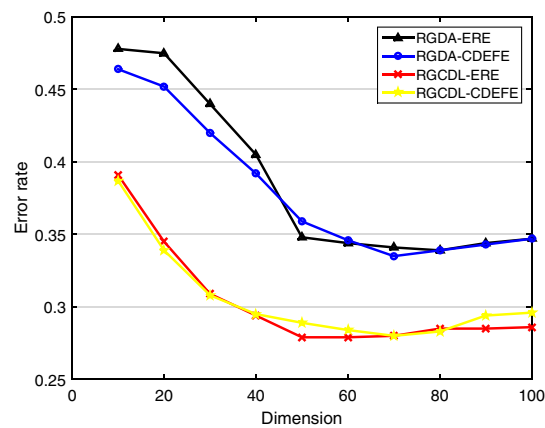


Fig. 7 Performance comparisons of RGCDL and RGDA on the RGB-D dataset.

the SPD manifold with the eigenspectrum regularization techniques is better than that of the Grassmann manifold. We conducted two experiments to evaluate the advantages of the proposed method. First, we compared the classification ability of RGDA and RGCDL when different dimensions of the extracted features were applied. As shown in Figs. 6 and 7, for both eigenspectrum regularization models in two datasets, the error rates of RGCDL with different number of features always lower than those of the RGDA. Moreover, the error rate curves of RGCDL are smoother and steadier for a different number of features, and RGCDL can achieve a lower error rate even with a low number of features, particularly for the ExtYaleB dataset. The RGDA method usually cannot achieve high performance using lower dimensions of features. This finding demonstrates that the SPD manifold formed by the covariance matrices has better discriminative information preservation ability than the Grassmann manifold formed by subspaces.

Subsequently, we conducted noisy set data to evaluate the robustness of the proposed method. Image sets may contain noisy data in real-world applications, for example, outliers of other categories or subjects within sets exist, which may degrade the performance of classifiers. Here, we show that the SPD manifold-based RGCDL method is more robust than the Grassmann manifold-based RGDA method. We conducted experiments by systematically corrupting the training (gallery) sets or test (probe) sets. The corruption is implemented by adding images from other classes. The data with no noise are denoted as “clean,” the data with noise in the gallery sets are denoted as “N_G,” and the data with noise in the probe sets are denoted as “N_P.” Experiments were evaluated using both face recognition and object categorization.

The average classification rates of several cross-validations with different noise-corrupted datasets are shown in Figs. 8 and 9. The classification rates of our RGCDL always outperform

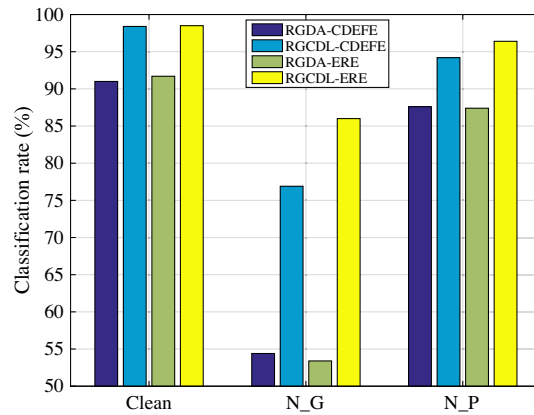


Fig. 8 Evaluated performances on the noisy ExtYaleB dataset.

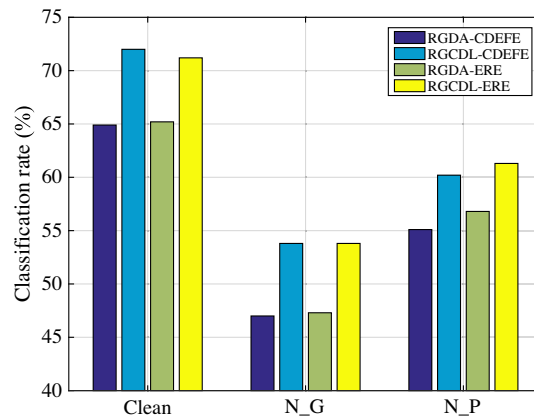


Fig. 9 Evaluated performances on the noisy RGB-D dataset.

RGDA in clean and different corrupted data. Especially in the gallery corrupted data N_G, RGCDL-ERE and RGCDL-CDEFE exhibit great advantages than RGDA-ERE and RGDA-CDEFE. Once again, these results demonstrate that the SPD manifold formed by the second-order statistic covariance matrices can be able to account for the noisy set data better than the Grassmann manifold formed by subspaces; it reveals the robustness of RGCDL when dealing with noisy set data.

5.4 Performance Comparison to Other Set-Based Classification Methods

We further evaluated the proposed RGCDL compared with other set-based classification methods. Multiple image set-based classification methods were evaluated for comprehensive comparison. The compared methods include the subspace-based methods DCC⁷ and ECCA,³⁶ the Grassmann manifold methods of GDA,⁸ KGDA,³⁸ GGDA,⁹ MMD,³ GNP,¹¹ and RGDA;² the SPD Riemannian manifold methods of CDL,¹⁰ PPCDL,¹⁶ and DARG.⁴¹

The parameter settings of different methods are depicted as follows. The final feature dimension of GDA, KGDA, GGDA, and RGDA was established as the recommendation of Ref. 2, and only the projection kernel⁸ was employed for Grassmannian mapping. The parameters (such as the nonlinearity score and number of NNs of data points) of MMD were tuned to be optimal with the code provided by the authors in our datasets. For CDL and PPCDL, the final feature dimension was set as the recommended value $C - 1$. The dimension of the input covariance matrices of DARG and PPCDL was set to 100×100 , which is the same as the proposed RGCDL for fairness. The Riemannian kernel induced by LED was applied for CDL, PPCDL, and our RGCDL.

Table 1 Average classification rates and standard deviations (%) on ExtYaleB and RGB-D datasets.

Method	ExtYaleB	RGB-D
MMD	58.9 ± 4.1	54.4 ± 2.8
KGDA	86.7 ± 2.7	66.5 ± 2.6
GGDA	73.4 ± 4.1	61.9 ± 2.6
GDA	89.7 ± 2.1	66.3 ± 3.6
ECCA	68.2 ± 2.5	56.3 ± 2.7
DCC	79.9 ± 2.6	67.3 ± 2.6
CDL	97.4 ± 1.2	69.7 ± 3.4
GNP	80.1 ± 1.6	62.1 ± 1.1
PPCDL	97.2 ± 1.3	51.8 ± 2.7
DARG	91.8 ± 2.4	55.3 ± 4.1
RGDA-ERE	91.7 ± 2.4	65.2 ± 3.0
RGDA-CDEFE	91.0 ± 1.1	64.9 ± 3.2
RGCDL-ERE-L_{bin}	98.5 ± 1.3	71.9 ± 3.3
RGCDL-ERE-L_{class}	98.5 ± 1.3	72.1 ± 3.3
RGCDL-ERE-L_{adj}	98.5 ± 1.3	72.0 ± 2.2
RGCDL-CDEFE-L_{bin}	98.1 ± 1.1	71.4 ± 3.3
RGCDL-CDEFE-L_{class}	98.3 ± 1.1	71.1 ± 2.2
RGCDL-CDEFE-L_{adj}	98.5 ± 1.1	71.5 ± 2.2

Note: The bold values denote the proposed methods and the highest classification rates.

We chose kernel-based DARG and the good performance of MD + LED⁴¹ distance matrix for evaluation. We fixed the dimension to 150 for DCC by preapplying PCA to the data.⁷ The number of canonical correlations of DCC and ECCA was set to 20, which is the same as the Grassmannian dimension of the GDA, KGDA, GGDA, and RGDA methods. For RGDA and our RGCDL, the eigenspectrum regularization models of ERE and CDEFE were applied for comparison. Three local Laplacian matrices of L_{bin} , L_{class} , and L_{adj} were evaluated in the proposed RGCDL.

Experiments were evaluated on the ExtYaleB and RGB-D datasets. The experimental results are formed by the average classification rates and standard deviations over 10-fold cross-validations. As shown in Table 1, the proposed RGCDL with regularization models of ERE and CDEFE achieves the best classification results among all methods. The SPD manifold with covariance matrices of CDL, PPCDL, DARG, and our RGCDL approaches usually achieves better performance than other methods in ExtYaleB, which has shown the better accommodative ability of the second-order statistic of covariance matrix on handling the illumination-varying face recognition. The inferior performances of PPCDL and DARG on RGB-D dataset may cause by the conventional problems of discriminative learning and the improperly estimated GMM. Benefitting from the eigenspectrum regularization with the graph-embedded framework, the proposed RGCDL with different models outperforms all other methods. For the evaluation of different local Laplacian graphs, the L_{class} achieves the best results with the ERE regularization model. However, the performance of adjustable Laplacian matrix L_{adj} is also outstanding in ERE model, and it achieves the best results with the CDEFE model. The adjustable Laplacian matrix L_{adj} performs stable in different regularization models, it has revealed the good locality preserving

ability. Obviously, L_{adj} is not the best local Laplacian matrix for locality preserving, better affinity matrix can be designed according to suitable theories.

6 Conclusion

In this paper, we proposed a regularized graph-embedded CDL method, which is referred to as RGCDL. The eigenspectrum regularization and graph-embedded framework are collaboratively employed to attenuate the overfitting and poor generalization problems of the original CDL method. Comprehensive mathematical deduction in SPD manifold kernel space is given to exhibit the combination of these techniques. The experimental results of evaluating a different number of extracted features show that the proposed method can maintain stable and lower error rates throughout all dimensions of the extracted features. This result manifests the stability of the eigenspectrum regularization to linear discriminative learning in the SPD manifold kernel space. The graph-embedded framework benefits by preserving compact within-class affinity relations and achieves higher performance. Compared with the more recent RGDA method, our RGCDL achieves higher and steadier performance when different number of features are employed. Moreover, our RGCDL exhibits more robust ability than RGDA when the gallery or probe sets are corrupted by noise. According to the plentiful comparisons with other set-based classification methods, our RGCDL has shown considerable results. The local Laplacian matrix reflects the local structure of intraclass, how to devise the similarity of intraclass vertex pairs to better preserve locality information is one of our future works.

Acknowledgments

This work was supported in part by the National Natural Science Foundation of China under Grant Nos. 61701126, 61802148, and 61802079, the Research projects in Guangzhou University, China, under Grant No. RP2020123, and the Scientific Research Program of Guangzhou under Grant No. 201904010493.

References

1. H. Hu, "Face recognition with image sets using locally Grassmannian discriminant analysis," *IEEE Trans. Circuits Syst. Video Technol.* **24**(9), 1461–1474 (2014).
2. H. Tan et al., "Eigenspectrum regularization on Grassmann discriminant analysis with image set classification," *IEEE Access* **7**, 150792 (2019).
3. R. Wang et al., "Manifold–manifold distance and its application to face recognition with image sets," *IEEE Trans. Image Process.* **21**(10), 4466–4479 (2012).
4. H. Cevikalp and B. Triggs, "Face recognition based on image sets," in *IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 2567–2573 (2010).
5. Y. Q. Hu, A. S. Mian, and R. Owens, "Face recognition using sparse approximated nearest points between image sets," *IEEE Trans. Pattern Anal. Mach. Intell.* **34**(10), 1992–2004 (2012).
6. P. F. Zhu et al., "Image set-based collaborative representation for face recognition," *IEEE Trans. Inf. Forensics Secur.* **9**(7), 1120–1132 (2014).
7. T. K. Kim, J. Kittler, and R. Cipolla, "Discriminative learning and recognition of image set classes using canonical correlations," *IEEE Trans. Pattern Anal. Mach. Intell.* **29**(6), 1005–1018 (2007).
8. J. Hamm and D. D. Lee, "Grassmann discriminant analysis: a unifying view on subspace-based learning," in *Int. Conf. Mach. Learn.*, pp. 376–383 (2008).
9. M. T. Harandi et al., "Graph embedding discriminant analysis on Grassmannian manifolds for improved image set matching," in *IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 2705–2712 (2011).
10. R. Wang et al., "Covariance discriminative learning: a natural and efficient approach to image set classification," in *IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 2496–2503 (2012).

11. H. L. Tan et al., "Grassmann manifold for nearest points image set classification," *Pattern Recognit. Lett.* **68**, 190–196 (2015).
12. M. Harandi, M. Salzmann, and R. Hartley, "Dimensionality reduction on SPD manifolds: the emergence of geometry-aware methods," *IEEE Trans. Pattern Anal. Mach. Intell.* **40**(1), 48–62 (2018).
13. S. Jayasumana et al., "Kernel methods on the Riemannian manifold of symmetric positive definite matrices," in *IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 73–80 (2013).
14. D. G. Kendall, "Shape manifolds, procrustean metrics, and complex projective spaces," *Bull. London Math. Soc.* **16**(2), 81–121 (1984).
15. R. Vemulapalli, J. K. Pillai, and R. Chellappa, "Kernel learning for extrinsic classification of manifold features," in *IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 1782–1789 (2013).
16. H. Tan and Y. Gao, "Patch-based principal covariance discriminative learning for image set classification," *IEEE Access* **5**, 15001–15012 (2017).
17. P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. Fisherfaces: recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.* **19**(7), 711–720 (1997).
18. G. Baudat and F. Anouar, "Generalized discriminant analysis using a kernel approach," *Neural Comput.* **12**(10), 2385–2404 (2000).
19. W. Liu et al., "Null space-based kernel Fisher discriminant analysis for face recognition," in *IEEE Int. Conf. Autom. Face and Gesture Recognit.*, pp. 369–374 (2004).
20. H. Yu and H. Yang, "A direct LDA algorithm for high-dimensional data—with application to face recognition," *Pattern Recognit.* **34**(10), 2067–2070 (2001).
21. L. F. Chen et al., "A new LDA-based face recognition system which can solve the small sample size problem," *Pattern Recognit.* **33**(10), 1713–1726 (2000).
22. M. H. Yang, "Kernel eigenfaces vs. kernel Fisherfaces: face recognition using kernel methods," in *IEEE Int. Conf. Autom. Face and Gesture Recognit.* (2002).
23. J. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Face recognition using kernel direct discriminant analysis algorithms," *IEEE Trans. Neural Networks* **14**, 117–126 (2003).
24. X. D. Jiang, B. Mandal, and A. Kot, "Eigenfeature regularization and extraction in face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.* **30**(3), 383–394 (2008).
25. X. Wang and X. Tang, "Dual-space linear discriminant analysis for face recognition," in *IEEE Conf. Comput. Vision and Pattern Recognit.* (2004).
26. X. Jiang, B. Mandal, and A. Kot, "Complete discriminant evaluation and feature extraction in kernel space for face recognition," *Mach. Vision Appl.* **20**(1), 35–46 (2009).
27. B. Mandal et al., "Prediction of eigenvalues and regularization of eigenfeatures for human face verification," *Pattern Recognit. Lett.* **31**(8), 717–724 (2010).
28. P. Y. Han, A. B. J. Teoh, and F. S. Abas, "Regularized locality preserving discriminant embedding for face recognition," *Neurocomputing* **77**(1), 156–166 (2012).
29. Y. H. Pang, A. B. J. Teoh, and F. S. Hiew, "Locality regularization embedding for face verification," *Pattern Recognit.* **48**(1), 86–102 (2015).
30. X. Liu and T. Cheng, "Video-based face recognition using adaptive hidden markov models," in *IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 340–345 (2003).
31. M. Kim et al., "Face tracking and recognition with visual constraints in real-world videos," in *IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 1787–1794 (2008).
32. O. Yamaguchi, K. Fukui, and K. I. Maeda, "Face recognition using temporal image sequence," in *IEEE Int. Conf. Autom. Face and Gesture Recognit.*, pp. 318–323 (1998).
33. H. Hotelling, "Relation between two sets of variable," *Biometrika* **28**, 321–377 (1936).
34. K. Fukui and O. Yamaguchi, "Face recognition using multi-viewpoint patterns for robot vision," in *Int. Symp. Rob. Res.*, pp. 192–201 (2005).
35. K. Fukui and A. Maki, "Difference subspace and its generalization for subspace-based methods," *IEEE Trans. Pattern Anal. Mach. Intell.* **37**(11), 2164–2177 (2015).
36. O. Arandjelovic, "Discriminative extended canonical correlation analysis for pattern set matching," *Mach. Learn.* **94**(3), 353–370 (2014).
37. R. Wang and X. Chen, "Manifold discriminant analysis," in *IEEE Conf. Comput. Vision and Pattern Recognit.*, pp. 429–436 (2009).

38. T. S. Wang and P. F. Shi, "Kernel Grassmannian distances and discriminant analysis for face recognition from image sets," *Pattern Recognit. Lett.* **30**(13), 1161–1165 (2009).
39. O. Tuzel, F. Porikli, and P. Meer, "Pedestrian detection via classification on Riemannian manifolds," *IEEE Trans. Pattern Anal. Mach. Intell.* **30**(10), 1713–1727 (2008).
40. Z. H. Huang et al., "Log-Euclidean metric learning on symmetric positive definite manifold with application to image set classification," in *Int. Conf. Mach. Learn.*, pp. 720–729 (2015).
41. W. Wang et al., "Discriminant analysis on Riemannian manifold of Gaussian distributions for face recognition with image sets," *IEEE Trans. Image Process.* **27**, 151–163 (2018).
42. V. Arsigny et al., "Geometric means in a novel vector space structure on symmetric positive definite matrices," *SIAM J. Matrix Anal. Appl.* **29**(1), 328–347 (2007).
43. B. Scholkopf and A. J. Smola, *Learning with Kernels*, MIT Press, Cambridge, Massachusetts (2002).
44. X. He and P. Niyogi, "Locality preserving projections," in *Adv. Neural Inf. Process. Syst.*, pp. 234–241 (2003).
45. X. He et al., "Neighborhood preserving embedding," in *IEEE Int. Conf. Comput. Vision* (2005).
46. X. D. Jiang, "Linear subspace learning-based dimensionality reduction," *IEEE Signal Process Mag.* **28**(2), 16–26 (2011).
47. A. S. Georgiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Pattern Anal. Mach. Intell.* **23**(6), 643–660 (2001).
48. K. Lai et al., "A large-scale hierarchical multi-view RGB-D object dataset," in *IEEE Int. Conf. Rob. and Autom.*, pp. 1817–1824 (2011).
49. P. Viola and M. J. Jones, "Robust real-time face detection," *Int. J. Comput. Vision* **57**(2), 137–154 (2004).

Hengliang Tan received his BE degree from Foshan University, Foshan, China, in 2006, and his ME and PhD degrees from Sun Yat-sen University, Guangzhou, China, in 2011 and 2016, respectively. He joined the School of Computer Science and Cyber Engineering, Guangzhou University, Guangzhou, in 2016. His current research interests include machine learning, pattern recognition, and manifold learning.

Ying Gao received his PhD from the South China University of Technology, Guangzhou, China, in 2002. He is currently a professor at the School of Computer Science and Cyber Engineering, Guangzhou University, Guangzhou. His main research interests include intelligent optimization algorithms, pattern recognition, and signal processing.

Jiao Du received her MS and PhD degrees from the Chongqing University of Posts and Telecommunications in 2013 and 2017, respectively. She is currently a lecturer at the School of Computer Science and Cyber Engineering, Guangzhou University, Guangzhou, China. Her research interests include pattern recognition and image processing.

Shuo Yang received his master's degree in software engineering from Dalian Jiaotong University, China, in 2013. He was awarded a doctorate degree in software engineering, University of Macau, in 2017. He is currently a lecturer at the School of Computer Science and Cyber Engineering, Guangzhou University, Guangzhou, China. His research interests include semantic interoperability and semantic inference with artificial intelligence technology, mainly applied to the fields of e-commerce, e-marketplace, and clinical area.