



## Generalized Multi-manifold Graph Ensemble Embedding for Multi-View Dimensionality Reduction

<sup>1</sup>Sumet Mehta, <sup>1</sup>Muhammad Sohail, <sup>2</sup>Jasjit Singh Samagh, <sup>2</sup>Silki Baghla.

<sup>1</sup>School of Computer Science and Communication Engineering, Jiangsu University, Zhenjiang, 212013, China.

<sup>2</sup>Department of Computer Science and Engineering, Chaudhry Devi Lal University, Sirsa, 125055, India

<sup>2</sup>Department of Electronics and Communication, JCDM Collage of Engineering, Sirsa, Haryana, 125055, India

msumet@outlook.com

### Abstract

In this paper, we propose a new dimension reduction (DR) algorithm called ensemble graph-based locality preserving projections (EGLPP); to overcome the neighbourhood size  $k$  sensitivity in locally preserving projections (LPP). EGLPP constructs a homogeneous ensemble of adjacency graphs by varying neighborhood size  $k$  and finally uses the integrated embedded graph to optimize the low-dimensional projections. Furthermore, to appropriately handle the intrinsic geometrical structure of the multi-view data and overcome the dimensionality curse, we propose a generalized multi-manifold graph ensemble embedding framework (MLGEE). MLGEE aims to utilize multi-manifold graphs for the adjacency estimation with automatically weight each manifold to derive the integrated heterogeneous graph. Experimental results on various computer vision databases verify the effectiveness of proposed EGLPP and MLGEE over existing comparative DR methods.

**Keywords:** manifold learning, multi-view dimension reduction, graph embedding, pattern recognition

### 1. INTRODUCTION

Researchers in various fields such as deep learning and machine learning, meets with a variety of data, such as texts, images, audios, and videos in very high dimensions. However, the recognition performance of many learning tasks for example face recognition, text categorization and signal processing, is seriously degrades when directly using this high dimensional data as input. To diminish the “curse of dimensionality”[1]

and improve the computational efficiency, dimensionality reduction (DR) techniques have been projected to preprocess these high dimensional data.

According to a general framework [2] these DR methods can be placed into two categories: linear and non-linear, where most of them are seen as an alternatives of graph embedding (GE). This GE scheme

has been generalized by evolving an integrated revolving uniform DR framework [3], and the extensive growth of GE has been presented in [4].

Among the linear algorithms, principal component analysis (PCA) [5, 6] and linear discriminant analysis (LDA) [7, 8] have been the two most recognized because of their comparative efficiency and accessibility. PCA, an unsupervised method preserves the global geometry whereas LDA, a supervised method detects the global discriminative information of the high-dimensional data. Further, by applying kernel tricks [9], the nonlinear ones of PCA and LDA are introduced to realize the intrinsic non-linear geometry of high-dimensional data. Recently many extensions of PCA [10-12] and LDA [13-15] have been developed in machine learning applications, yet PCA often seem like incapable to well detect the discrimination whereas LDA can often unable to resolve the small sample size problem and both fails to process nonlinear data.

Moreover, enormous manifold-based methods have been proposed to explicitly discover the nonlinear structure hidden in the data [16]. It is based on the idea that data usually denotes from a low-dimensional manifold that is embedded in a high-dimensional space. These methods include isomap [17], local linear embedding (LLE) [18], multidimensional scaling (MDS) [19], and laplacian eigenmaps (LE) [20]. Most of them attempt to implant the original high-dimension data into sub-local manifold based on the assumption that; a nearest-neighbor graph can modeled the manifold for preserving actual neighborhood properties of the data. Nevertheless, these non-linear manifold methods have some issues to be resolved with reference to machine learning applications. Such as their classification is not directly linked while preserving locality of data [21]. Also these techniques have out-of-sample problem, as the embedding realized from training data cannot be directly applied to a fresh data point [22].

In order to solve above problems, various improved linear manifold GE methods have been presented to reveal the local manifold structure of the data. The most representative linear ones are locality preserving projections (LPP) [23], neighborhood preserving embedding (NPE) [24] and locality sensitive discriminant analysis (LSDA)[25]. LPP and NPE are the linear versions of the non-linear LE and LLE. In many recent researches, LPP has been investigated for DR with new modifications [26-33]. Such as in [34], a recent supervised form of LPP called double adjacency graphs-based discriminant neighborhood embedding (DDNE) is presented. Xu et al. projected the LPP solution schemes for face recognition [28] and Cai et al.

proposed orthogonal locality preserving projection (OLPP) algorithm [35]. Although these approaches have shown improved performances than LPP, still these are very sensitive to neighborhood parameter  $k$ , because the KNN rule is employed for constructing the adjacent graphs.

In this article, being inspired from our weighted neighborhood preserving ensemble embedding (WNPEE) [36] framework, a new DR algorithm called ensemble graph-based locality preserving projections (EGLPP) is proposed. In EGLPP, we consider an ensemble of adjacent graphs by varying neighborhood size parameter unlike LPP which considers only one adjacent graph. EGLPP aims to improve recognition performance of LPP by preserving more local neighborhood information that best determines the essential manifold structure of high dimensional data while making LPP less sensitive to neighborhood size  $k$ .

Furthermore, these days several real-world databases are composed from numerous bases or characterized by multi-views and each single view refers to diverse perception of data. For example, images from various view-points of the identical object can be characterized as multi-views, one text could also be taken into various languages, and therefore the interpretation in every language will be measured as a view. We generally cannot train all multi-views into one specific view for pattern recognition, as every single view has its individual statistical assets. Recently, several multi-view DR approaches have been developed. For instance, Bilinear Models (BLM) [37], Partial Least Squares (PLS) [38] and the canonical correlation analysis (CCA) [39] are capable to realize mutual features from multi-view databases. RCCA [40] and LPCCA [41] are two extensions of CCA. Also numerous multi-view DR methods based on graph-embedding are known as multi-graph based algorithms. Similar to graph-based learning, the structure of these approaches also comprises of the smoothness constraint and the fitting constraint. Additionally, some of these DR rules are projected to assign suitable weights to distinct graphs. Although the current DR approaches have certain promising contributions in recognizing patterns, the following issues further need to be explored:

- The restriction of most of current DR approaches is that they are ineffective to evaluate the essential spatial structure of the original data.
- Most of the persisting DR approaches use only one adjacent graph for manifold embedding, which is unable to preserve complete discrimination and makes them inefficient for multi-view databases.

Therefore in this study, inspired from our ensemble graph framework, we propose a generalized multi-view DR framework called multi-manifold graph ensemble embedding (MLGEE) as in Fig. 1. MLGEE uses an ensemble regularization term to consider the multi-manifolds statistics for multi-views. Specifically, we use different manifolds to construct adjacent heterogeneous graphs to present the essential manifold structure for different views. This makes MLGEE to fuse the essential manifold learning through these heterogeneous graphs. Finally the objective function is derived by unification of the ensemble regularization term with linear embedding analysis.

The contributions of our proposed ensemble framework are summarized as:

- In first contribution, the proposed EGLPP unlike LPP constructs an ensemble of homogeneous adjacent graphs with an efficient weighting scheme. It makes EGLPP much less sensitive to neighborhood size  $k$  while obtaining projections with an optimum embedded graph pursuing in a joint optimization way. LPP can be seen as a special case of the EGLPP for a fixed neighborhood size, as rationale behind proposed EGLPP and LPP is same.
- In second contribution, a generalized multi-manifold graph ensemble embedding for multi-view DR (MLGEE). Multiple local manifold embeddings are effectively utilized in MLGEE to obtain local structures from multi-view databases. MLGEE regularization framework is employed to scale the local heterogeneous graphs of multi-views to estimate the essential manifold structure.

The experiments on four face databases demonstrate the supremacy of our EGLPP when compared to LPP and other related DR approaches. Also, experiments on four multi-view databases for handwritten numerals recognition, object recognition and face recognition show the superiority of MLGEE over that of proposed EGLPP, WNPEE, canonical correlation analysis (CCA), discriminant neighborhood embedding (DDNE), sparsity preserving projections (SPP), LPP, and NPE approaches.

We assemble the rest of the article, following to introduction in Section I as: in Section II, we present a brief review of related GE and multi-view DR approaches. In Section III, we frame the proposed ensemble-based DR approaches and finally the experimental results are expressed in Section IV. In Section V, we draw our conclusions and the future directions.

## 2. RELATED WORK

In this section, we briefly review the related representative DR approach; LPP and then discuss the recent growth on multi-view DR approaches.

### A. Locality Preserving Projections

In recent times, a number of DR approaches have been developed using different manifold ideas and methodology. However; the goal of these numerous approaches is same, that is, to obtain a reduced dimensional data on which further machine learning tasks can be easily performed with improved computational complexity. LPP is an unsupervised graph based embedding method [23]. LPP is designed with different objective principle from classical linear DR techniques. Due to locality preserving quality of LPP, it has been used in many information retrieval applications [35, 42, 43]. LPP can be simply functional everywhere unlike non-linear DR techniques. LPP could be carry out in the inventive extent or in the replicating kernel Hilbert extent into which data points are plotted.

For notational convenience, let us denote  $X = \{x_i \in \mathbb{R}^D\}_{i=1}^N$  a high  $D$ -dimensional data with  $N$  number of samples. Linear DR mapping is to assign the original high dimensional data to a low  $d$ -dimensional data i.e.  $Y = \{y_i \in \mathbb{R}^d\}_{i=1}^N$  ( $d \ll D$ ). If  $A$  is a transformation matrix, then linear DR is defined as

$$y_i = A^T x_i \text{ for } i = 1, \dots, N. \quad (1)$$

The functioning of the LPP consists of three steps. Firstly it constructs adjacent graph, then giving the weights to edges and finally obtains the projections by solving eigenvalue problem. The adjacent graph is usually constructed by  $\varepsilon$ -neighborhood or  $k$ -neighborhood. In  $\varepsilon$ -neighborhood: two nodes  $i$  and  $j$  are linked by an edge if  $\|x_i - x_j\|^2 < \varepsilon$ . In  $k$ -neighborhood: nodes  $i$  and  $j$  are linked by a verge if either of it presents in their  $k$  nearest neighbor. Weights  $W_{ij}$  are given to edges between nodes  $i$  and  $j$  by a symmetric weight matrix  $W$ , computed by uniform weight or Gaussian weight of Euclidean distance.  $W$  is often defined as

$$W_{ij} = \begin{cases} \frac{\exp\|x_i - x_j\|^2}{t}, & x_j \in \text{KNN of } x_i \text{ or } x_i \in \text{KNN of } x_j \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

where parameter  $t$  is a positive constant. Finally computing the eigenvectors and eigenvalues for the below eigenvector problem:

$$XLX^T \mathbf{a} = \lambda XD X^T \mathbf{a}. \quad (3)$$

where  $D$  is a diagonal matrix whose entities are column (or row) sum of  $W$ ,  $D_{ii} = \sum_j W_{ij}$ ,  $L = D - W$  is the Laplacian matrix and  $\mathbf{a}$  is a transformation vector.

### B. Multi-View DR

Multi-view DR aims to learn one function to represent each view and collectively optimizes all the functions to recover the generalization performance. One of the most common approach for multi-view DR is to consider sequence of all multiple views into one single view and applies single view learning algorithms directly. Such as graph based multi-view learning in which each view is represented by a similarity graph, and then different graphs are fused into an integrated one to optimally preserve the similarities of data. Finally the results are obtained by assembling (generally using  $k$ -means) on the integrated graph. Recently various graph-based multi-view DR approaches have been developed. Wang *et al.* proposed a general graph-based multi-view clustering (GMC) method by sufficiently considering weights of different views[44]. Hou *et al.* proposed a multi-view unsupervised feature selection with adaptive similarity and view weight (ASVW) by acknowledging the underlying common structures across different views[45]. Also a multi-view maximum entropy discriminant (MVMED<sup>+</sup>) model for learning with different numbers of views is proposed by Sun *et al.*[46]. Furthermore Wang *et al.* proposed a general graph-based system (GBS) for multi-view clustering by discussing and evaluating the impact of different graph metrics on the multi-view clustering performance[44].

Moreover, nonlinear relationships usually exist in real-world datasets, which have not been considered by most existing methods. In order to address these challenges, a novel model which simultaneously performs multi-view clustering task and learns similarity relationships in kernel spaces is proposed by Huang *et al.* [47]. Li *et al.* propose a generative bayesian model to not only jointly take the features and views into account, but also learn a discriminant representation across distinctive categories [48]. Geng *et al.* an ensemble manifold regularization (EMR) framework to approximate the intrinsic manifold by combining several initial guesses[49].

On the other hand, CCA is the most commonly used multi-view DR approach [50]. CCA tends to extract the couple of estimates with maximal connection within two sets of multi-view data [51]. A couple of estimates

$\mathbf{a}_x$  and  $\mathbf{a}_y$  is expressed by CCA which can be initiate in following objective function in order to exploit the connections between  $\mathbf{a}_x^T x$  and  $\mathbf{a}_y^T y$  [52, 53].

$$\begin{aligned} & \max_{\mathbf{a}_x, \mathbf{a}_y} \mathbf{a}_x^T X Y^T \mathbf{a}_y \\ \text{s. t. } & \mathbf{a}_x^T X X^T \mathbf{a}_x = 1, \quad \mathbf{a}_y^T Y Y^T \mathbf{a}_y = 1 \end{aligned} \quad (4)$$

### 3. PROPOSED METHODS

It has been proven that ensemble graph building in WNPEE [36], can best determines the essential local manifold assemble of high-dimensional data in NPE while keeping the projections less sensitive to neighborhood size  $k$ .

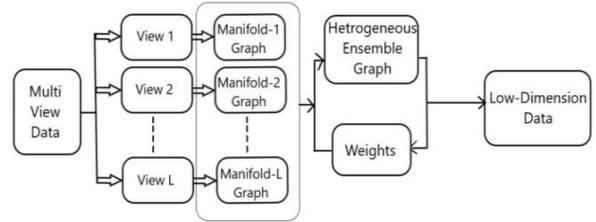


Figure 1: The proposed MLGEE DR method.

In this article, we proposed an extension of this ensemble graph building to LPP, to achieve better recognition rates with less sensitivity to neighborhood size  $k$ . A new DR technique termed as ensemble graph-based locality preserving projections (EGLPP) is introduced. EGLPP constructs an ensemble of adjacent *homogeneous* graphs with different weights while varying neighborhood size  $k$ . The cost function of the proposed EGLPP is given as:

$$\begin{aligned} & \arg \min_A \sum_{t=1}^T \alpha_t \sum_{ij} (y_i - y_j)^2 W_{t,ij} \\ \text{s. t. } & \sum_{t=1}^T \alpha_t = 1 \end{aligned} \quad (5)$$

where  $X = \{x_i \in \mathbb{R}^Q\}_{i=1}^N$  given data and  $A$  is a conversion matrix that plots these  $N$  points to  $Y = \{y_i \in \mathbb{R}^q\}_{i=1}^N$ , ( $q \ll Q$ ) such that  $y_i$  represents  $x_i$ , where  $y_i = A^T x_i$  in which locality of  $x_i$  is persevered by linear dimensionality reduction.  $W_{g,ij}$  represent the weights on the edges,  $\alpha_t$  is the weight coefficient for the adjacent homogeneous graphs and  $D$  is a diagonal matrix whose entities are column (or row) sum of  $W$ ,  $D_{ii} = \sum_j W_{t,ij}$ . Considering Laplacian matrix  $L = D - W_t$ , where matrix  $D$  provides a normal measure on the data points. The importance of  $y_i$  directly depends on how greater the value of  $D_{ii}$  (corresponding to  $y_i$ ). Consequently, we enforce a constraint as  $y^T D y = 1 \Rightarrow$

$\mathbf{a}^T X D X^T \mathbf{a} = 1$ . Finally the minimization problem reduces to

$$\arg \min_{\mathbf{a}^T X D X^T \mathbf{a} = 1} \mathbf{a}^T X \sum_{t=1}^T \alpha_t L X^T \mathbf{a} \quad (6)$$

Therefore, EGLPP using different  $k$  values to make  $T$  multiple adjacent graphs and integrated them by giving  $\alpha_t$  weights to obtain an optimal embedded graph. Finally the low dimensional projections are pursued in a joint optimization way by preserving more local neighborhood information.

Moreover, with additional approach of merging different local manifold embeddings; we propose a new generalized multi-manifold graph ensemble embedding framework (MLGEE). MLGEE preserves the similarity information of data sample in different views using different manifolds to construct the heterogeneous adjacent graphs. Then these heterogeneous adjacent graphs are weighted to build a unified representation for multi-view manifold learning. The objective function of MLGEE is give as follows:

$$\begin{aligned} \arg \min_w w^T X \sum_{g=1}^L \beta_g M_g X^T w \quad (7) \\ \text{s.t. } w^T X X^T w = 1 \text{ and } \sum_{g=1}^L \beta_g = 1 \end{aligned}$$

where  $w$  is the transformation vector,  $\beta = [\beta_1, \beta_2, \dots, \beta_L]$  symbolizes weight coefficient for the multi-manifolds normalization. Also  $\beta$  sentences the data points in the  $g^{\text{th}}$  local manifold assembly. Also,  $M_g$  is the  $g^{\text{th}}$  local manifold which is assembled from the local manifold embeddings such as LPP, NPE, SPP, LE or LSDA. Therefore, in MLGEE superiorities of different manifolds are combined to preserve more local representation of the multi-view high dimension data.

The transformation vector  $w$  that reduces the objective function is given by the minimum eigenvalue solution to the following generalized eigenvector problem while considering  $w^T X X^T w = 1$ :

$$X \sum_{g=1}^L \beta_g M_g X^T w = \lambda X X^T w \quad (8)$$

### A. Obtaining parameter $\alpha_t$ and $\beta_g$ in proposed methods

In this segment, we explore how to find the values of the considerations  $\alpha$  and  $\beta$ , that are weight coefficient vector for the several homogeneous graphs in EGLPP

and weight coefficient vector for the multiple heterogeneous local manifolds in MLGEE, respectively.

Considering having  $T$  adjacent homogeneous graphs:  $G_1, G_2, \dots, G_T$ , in EGLPP similar to [36]. We wish to obtain the value of  $\alpha_t$  which regulates the influence of  $t^{\text{th}}$  adjacent graph in EGLPP DR approach. Therefore, by adding Lagrange multiplier in regard to  $\alpha$  in (6), we obtain the following regularization framework:

$$\mathbf{a}^T X \sum_{t=1}^T \alpha_t L X^T \mathbf{a} + \lambda \left( 1 - \sum_{t=1}^T \alpha_t \right) \quad (9)$$

Although, from (9) it can be seen that cost function is linear in regard to  $\alpha$  and it achieves the optimal result at extreme ends, that is either  $\alpha_t = 0$  or  $\alpha_t = 1$ . Or we can say that it selects only one adjacent graph which is different from our idea of discovering the balanced multi-manifolds. Therefore, we consider  $\alpha_t$  as  $\alpha_t^r$ , where  $r$  is used to represent the weights variable for the multiple adjacent graphs. Further, if  $r > 1$ , the result will be built on a balance among multiple adjacent graphs [54]. Consequently the value of parameter  $\alpha_t$  can be calculated by pursuing partial derivative of (9) by using Leibniz's notation and simultaneously when projections are fixed and by considering the constrain  $\sum_{t=1}^T \alpha_t = 1$ , we can further derive that:

$$\alpha_t = \frac{(1/\text{Tr}(\mathbf{a}^T X L X^T \mathbf{a}))^{\frac{1}{r-1}}}{\sum_{t=1}^T (1/\text{Tr}(\mathbf{a}^T X L X^T \mathbf{a}))^{\frac{1}{r-1}}} \quad (10)$$

Similarly for MLGEE, considering  $L$  local manifolds:  $M_1, M_2, \dots, M_L$ , we can obtain the value of  $\beta$  that is to regulate the impact of  $g^{\text{th}}$  local manifold. Thus, by adding Lagrange multiplier in (3) in regard to  $\beta$  and derive the below cost function:

$$w^T X \sum_{g=1}^L \beta_g M_g X^T w + \lambda \left( 1 - \sum_{g=1}^L \beta_g \right) \quad (11)$$

Furthermore, by taking into account  $\beta_g$  as  $\beta_g^r$  and considering the partial derivatives while using the constraint  $\sum_{g=1}^L \beta_g = 1$ , we can obtain

$$\beta_g = \frac{(1/\text{Tr}(w^T X M_g X^T w))^{\frac{1}{r-1}}}{\sum_{g=1}^L (1/\text{Tr}(w^T X M_g X^T w))^{\frac{1}{r-1}}} \quad (12)$$

Therefore, by solving iteratively the value of weight parameters  $\alpha$  and  $\beta$ , the projection vectors  $\mathbf{a}$  and  $\mathbf{w}$  can be derived from Algorithms 1 and 2 respectively. The whole EGLPP procedure is given in Algorithm 1 whereas the MLGEE is detailed in Algorithm 2.

---

**Algorithm 1. The proposed EGLPP DR method.**

---

**Input:** High dimensional data  $X$ **Output:** projection vector  $\mathbf{a}$ **Parameter:**  $\alpha_t$ **Initialize:**- Construct the  $T$  adjacent homogeneous graphs using  $KNN$  with each  $\alpha = 1/T$  weights.**while** loss not converged **do**1. Computing the weights on edges with fix  $\alpha_t$  acc. to (5)2. Obtain projection vector  $\mathbf{a}$  acc. to (6)3. Fix  $\mathbf{a}$  and update  $\alpha_k$  acc.to (10)

4. Compute current loss

**End while**

---

**Algorithm 2. The proposed MLGEE DR method.**

---

**Input:** Multi-view data-sample  $X$ ,  $L$  no. of multi-manifolds.**Output:** Low-dimensional projection vector  $\mathbf{w}$ **Parameter:**  $\beta_g$ **Initialize:**- set =  $(\frac{1}{L}, \frac{1}{L}, \dots, \frac{1}{L})^T$ ,-Construct the  $L$  local heterogeneous manifolds  $(M_g)_{g=1}^L$ .**while** loss not converged **do**1. Find projection vector  $\mathbf{w}$  acc. to (8)2. Fix  $\mathbf{w}$  and update  $\beta_g$  acc.to (12)

3. Compute current loss

**End while**

---

## 4. EXPERIMENTS

In this section, four computer vision databases are used in our experiments. Firstly, two image databases Olivetti Research Laboratory (ORL) [55] and Georgia Tech (GT) [56] are used to compare recognition results of EGLPP variants to LPP. Then experiments are carried out on object recognition using Coil-100 [57], handwritten numeral recognition using Multi-Feature digit (MFD) [58] and face recognition using ORL databases are exploited to verify the recognition performance of MLGEE in comparison to WNPEE, EGLPP, LPP, NPE, SPP [16], DNPE and CCA methods.

The Georgia Tech (GT) face dataset has 750 frontal and/or tilted facial image samples from 50 persons, and each person has 15 samples. The Olivetti Research Laboratory (ORL) dataset comprises of 40 different persons of each taking ten distinct images giving a total of 400 instances. The Multi-Feature digit (MFD) dataset is built of features of handwritten numerals from 0 to 9. These features are taken from a Dutch utility maps collection and we regrouped the distinct features into a

set of 4-views (features 1, 2, 3, 4) shown in Table 1 similar to [59]. The COIL-100 database consists of hundred

---

*Table 1. Six groups of features in MFD database.*

---

1. fou : 76 Fourier coefficients of the character shapes feats.
  2. fac : 216 profile correlations feature
  3. kar : 64 Karhunen-Love coefficients feature
  4. pix : 240 pixel averages in  $2 \times 3$  windows feature
  5. zer : 47 Zernike moments
  6. mor : 6 morphological features
- 

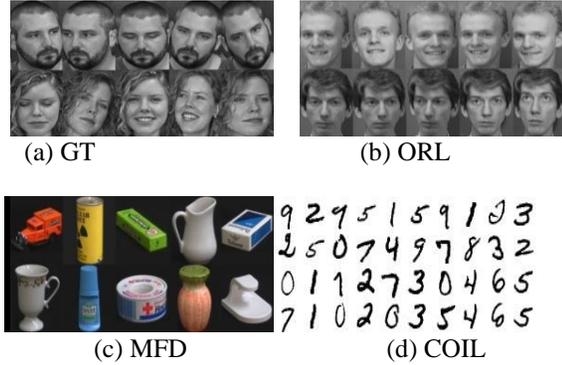


Figure 2: Sample images of databases. color images of distinct objects contributed by the Columbia University Image Library. Fig. 2(a-d) shows some sample images of the aforementioned databases.

The parameters of the state-of-the-art methods used in experiments are set in the following manner. CCA is set in conformity to the literature provided by Sharma *et al.* [60]. For the WNPEE and EGLPP, we have taken 7 adjacent graphs with  $k$  ranges from 1 to 13 with a step 1. In LPP and NPE, we set the neighborhood size to 5. Also in LPP, we measure the distance by heat kernel with the parameter  $t$  set to 0.5. The control parameter for the weights of the multiple adjacent graphs in EGLPP and multiple manifolds in MLGEE,  $r$  is chosen from (2, 4, 6, 8, 10 and 12). Three manifolds: LPP, LSDA and NPE are used in experiments such that  $L=3$ . To select the parameter values and optimal results, we use three-fold cross-validation technique for all training samples. The values of  $k$  are fixed as  $k=9$  in GT, ORL,  $k=11$  in MFD and  $k=7$  in Coil-100 databases, respectively. Also, number of training samples  $l$  was randomly preset to 7 on GT, 9 on ORL, 15 on MFD, and 7 on Coil-100, respectively.

### A. Comparison between EGLPP and LPP

Firstly, we compare our proposed EGLPP with the traditional LPP. In our experiments, we use EGLPP-1, EGLPP-2, and EGLPP-3 by constructing 7, 9 and 11 adjacent homogeneous graphs respectively. In each

variant, the value of neighborhood size is  $k = 1, 3, 5, \dots$  upto number of graphs.

We conceded out two cases of experiments, in first we took 60% of the data as training samples and rest 40% referred to testing samples in each dataset. In second case, we took 80% as the training data of complete database with balance 20% referred as testing data. The outcomes for EGLPP-1, EGLPP-2, EGLPP-3, and LPP are recorded in Table 2.

The experimental results in Table 2 illustrates that our proposed EGLPP variants attain substantial margins of progresses in classification accuracies as well as lowest DRs as compared to the traditional LPP in all the databases.

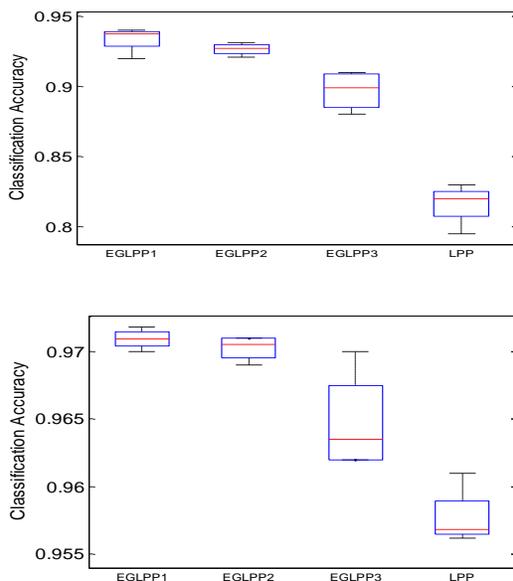
Table 2. Mean classification accuracies  $\pm$  standard deviations (%) and reduced dimensions.

Meth od	GT		ORL	
	60%	80%	60%	80%
<b>LPP</b>	81.20 $\pm$ 1.	84.73 $\pm$ 1.	95.56 $\pm$ 0.	96.12 $\pm$ 1.

	94 (32)	40 (32)	80 (43)	24 (45)
<b>EGL</b>	<b>92.63<math>\pm</math>0.</b>	93.07 $\pm$ 0.	<b>97.02<math>\pm</math>0.</b>	97.18 $\pm$ 0.
<b>PP-1</b>	<b>67 (23)</b>	60 (22)	<b>43 (35)</b>	09 (32)
<b>EGL</b>	92.02 $\pm$ 0.	<b>93.22<math>\pm</math>0.</b>	96.59 $\pm$ 1.	<b>97.29<math>\pm</math>0.</b>
<b>PP-2</b>	37 (24)	<b>94 (43)</b>	12 (30)	<b>57 (28)</b>
<b>EGL</b>	88.75 $\pm$ 0.	91.31 $\pm$ 0.	96.31 $\pm$ 0.	97.01 $\pm$ 0.
<b>PP-3</b>	94 (44)	42 (29)	13 (55)	38 (48)

Moreover, to validate their constancies in classification performance, we present the deviations through 18 random runs on the two image databases in Fig. 3. It can be observed from the figure that our EGLPP DR method sustain the least standard deviations for its all variants in most of the databases, while LPP shows comparatively higher. Particularly in the AR database, our EGLPP approaches attain a huge expansion in standard deviations as compared to LPP.

Further, in Fig. 4 we use different values of parameter  $k$  from 1 to 15 with a step size of one and plot the corresponding



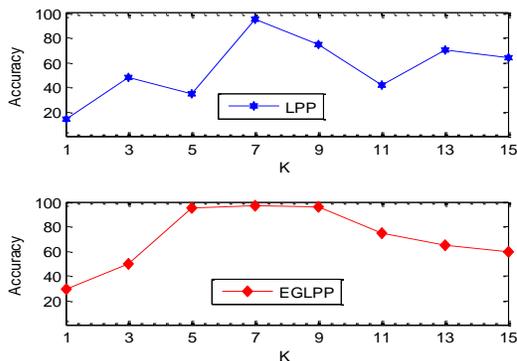
(a) GT And (b) ORL

Figure 3: Varying classification accuracies across 18 runs of EGLPP-1, EGLPP -2, EGLPP -3, and LPP on image databases.

average error rates to show the  $k$  value sensitivity. We can realize two important facts: first is that LPP accuracy continuously varies with  $k$  changes. On the other hand, EGLPP accuracy also changes w.r.t.  $k$ , but it becomes relatively stable after  $k$  reaches 7 and 9 in most of the databases. Another fact is that the proposed EGLPP has attained greater accuracy than LPP in both experiments.

## B. Experimental results for MLGEE

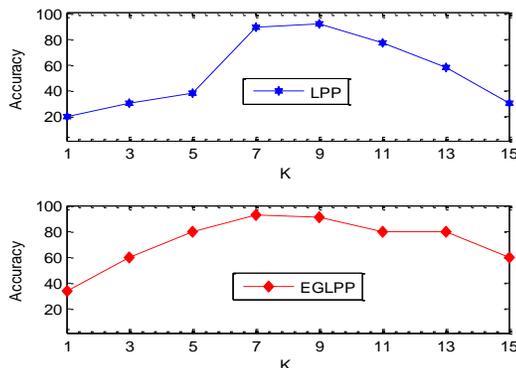
Recently *object recognition* is considered as one of the



(a) GT

various DR methods with respect to reduced dimensions in Fig. 5(a) for COIL-100 database. Fig. 5(a) clearly shows that most of the DR approaches other than DDNE and CCA are firstly goes to a sharp upswing before attaining their highest accuracies at reduced dimensions 10 and 20 in case of DDNE and CCA. The proposed MLGEE can be seen with the highest accuracy almost every dimension that demonstrates its superiority over other competitive DR approaches.

Handwritten numerals recognition is also a very challenging task in many real world machine learning applications. Therefore, we observe the handwritten



(b) ORL

Figure 4: Accuracy of LPP and proposed EGLPP corresponding to neighborhood size parameter  $k$  on (a) GT with  $l = 7$  and (b) ORL with  $l = 9$ , respectively.

most important task in machine learning especially when the data is in multi-views. Therefore, we determine the strength of our MLGEE DR algorithm in this trend by means of publically available database COIL-100.

The best average recognition accuracies with corresponding standard deviations (%) and reduced dimensions over 18 runs of MLGEE, WNPEE, NPE, DNPE, EGLPP, LPP, SPP and CCA are presented in Table 3. It can be viewed from the table that our MLGEE DR method performs best among the comparative methods with an average accuracy of 89.14% that is basically more than EGLPP by 0.94%, WNPEE by 4.08%, NPE by 5.36%, LPP by 6.4%, DNPE by 14.93%, SPP by 8.47% and CCA by 16.45%. In DR, MLGEE, EGLPP and WNPEE methods attain the superior accuracies over the other competitive approaches; on the other hand CCA shows the worst accuracies.

Additionally, to better elaborate the recognition results, we presents the classification accuracies of

numerals recognition performance of our MLGEE DR approach via

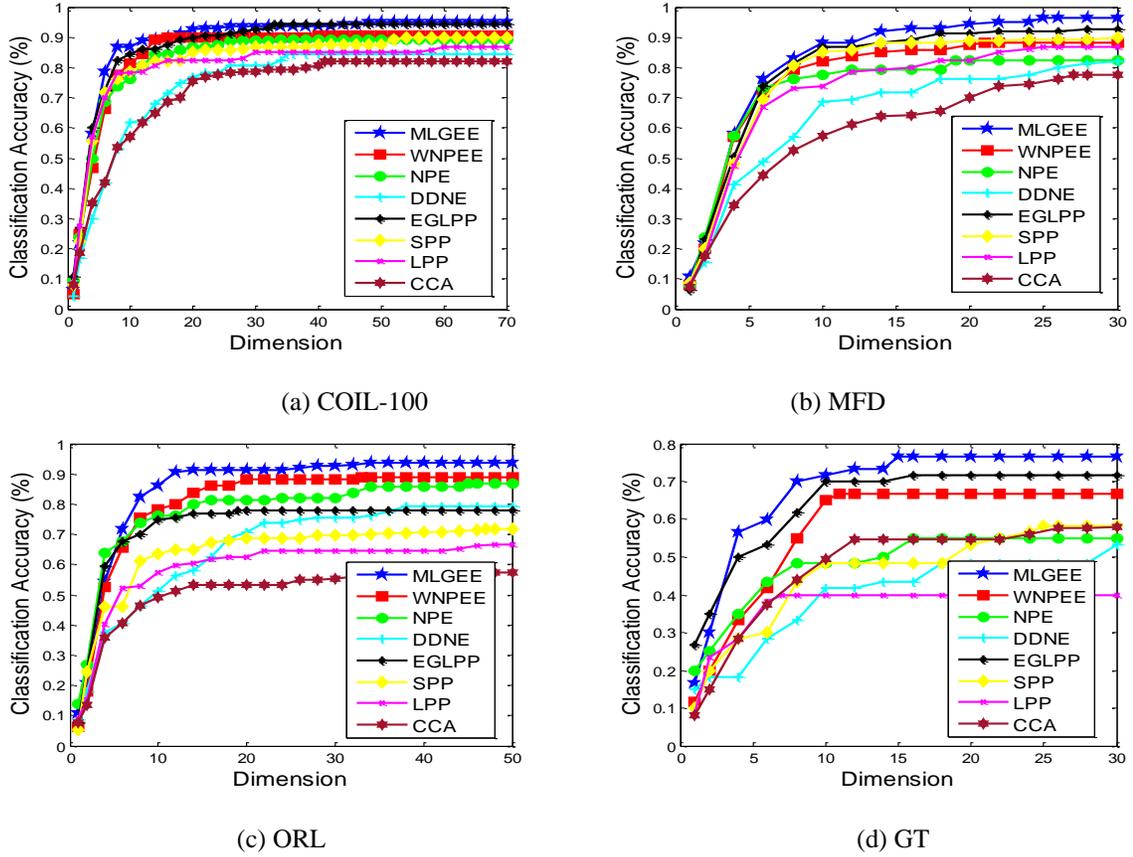


Figure 5: The classification accuracy rates of the comparative methods with varying dimensionality on (a) COIL-100 (b) MFD and (c) ORL and (d) GT databases.

Multi-Feature digit (MFD) database. Table 3 comprises the average recognition accuracies (%) with standard deviations and reduced dimensions of all comparative

DR approaches in MFD database. The proposed MLGEE attains the utmost mean

Table 3. The average classification accuracy rates (%) of all competitive methods on each dataset with their corresponding standard deviations and dimensionalities in parentheses.

	CCA	SPP	DNPE	LPP	NPE	WNPEE	EGLPP	MLGEE
<b>COIL-100</b>	72.69±0.1 6 (41)	80.67±0.1 3 (60)	74.21±0.1 8 (35)	82.74±0.1 4 (54)	83.78±0.1 5 (34)	85.06±0.1 6 (40)	88.21±0.1 5 (33)	<b>89.14±0.1 5 (48)</b>
<b>MFD</b>	51.13±0.1 0 (39)	59.36±0.1 2 (48)	66.04±0.1 7 (38)	64.21±0.1 3 (47)	78.14±0.1 4 (45)	80.70±0.1 8 (33)	73.33±0.1 3 (19)	<b>85.96±0.1 8 (33)</b>
<b>ORL</b>	59.75±0.1 8 (27)	72.92±0.2 0 (25)	65.04±0.1 9 (19)	78.08±0.2 1 (21)	73.15±0.1 8 (21)	77.00±0.2 1 (25)	79.9±0.22 (28)	<b>83.25±0.2 2 (30)</b>
<b>GT</b>	36.85±0.0 9 (26)	42.83±0.0 8 (18)	33.83±0.0 9 (28)	52.50±0.1 4 (25)	46.44±0.0 6 (14)	57.17±0.1 7 (12)	59.61±0.1 0 (15)	<b>60.28±0.1 6 (20)</b>

recognition accuracy of 85.96% that is more than EGLPP by 12.63%, WNPEE by 5.26%, NPE by 7.82%, LPP by 21.75%, SPP by 19.92%, DNPE by 26.6% and CCA-OM by 34.83%. Also, our MLGEE method obtains 33 reduced dimensions in DR, retaining following to EGLPP which gets the finest reduced

dimensions of 19. As well, we show the accuracies with respect to dimensions in all DR methods for MFD dataset in Fig. 5(b). From the figure it can be realized that our MLGEE method obtains the highest accuracies with WNPEE, EGLPP and SPP succeeding from the lowermost to the uppermost dimensions as

related to the competitive DR methods. Alternatively, CCA shows the lowest performance among related DR methods.

Experiments were also performed on the GT and ORL databases for the *face recognition* task. The effectiveness of the proposed MLGEE is estimated by performing sets of experiments with changing the dimensions. The dimension varies from 1 to 30 with a gap 1 in both databases.

Table 3 shows the maximum classification accuracies with the corresponding standard deviations (%) and reduced dimensions for all competitive DR approaches. The best result is displayed in bold-face. We can perceive from the outcomes that our introduced MLGEE, EGLPP and WNPEE methods mostly attain enhanced recognition performance as compared to DDNE, NPE, LPP, CCA and SPP. Especially, MLGEE attains the finest performance among the comparable methods. The proposed MLGEE has an average difference of 3.31% higher accuracy of both face databases. Additionally, in GT database, EGLPP considerably overtakes WNPEE, DDNE, LPP, NPE, SPP and CCA. Simultaneously, our presented method mostly attains superior recognition performance with much reduced dimensions as compared to other methods.

Furthermore the recognition accuracies of all the comparable DR techniques are performed with respect to dimensions are demonstrated in Fig. 5(c-d). The classification accuracies of each DR model can be detected that it first increases rapidly upto the dimensionality is less than 10 and then it slowly turn out to be constant. Also it is explicit from the figure that MLGEE, ELGPP and WNPEE more or less constantly achieve enhanced outcomes than NPE, LPP, DDNE, SPP and CCA. MLGEE frequently exceeds proposed EGLPP and WNPEE; also EGLPP and WNPEE achieve the comparable fine execution then other state-of-the-art approaches. In the meantime, the alterations of recognition accuracy are computably considerable between our proposed MLGEE and other DR approaches on all experimental databases. Therefore, from this face recognition experiment we can conclude that our method is an encouraging DR algorithm that can yield suitable recognition performances.

### C. Control parameter selection

To consider the effect of the control parameters  $r$  on proposed MLGEE method, we investigate with different values from 2 to 12 with a step of 2, for the ORL database. The mean recognition accuracies over 18 repetitive runs compared to different values of  $r$  are

shown in Fig. 6. It is clearly seen from the figure that the recognition accuracies of MLGEE are considerably inferior at  $r = 2, 10, 12$ ; while committing nearly comparable high accuracies at  $r = 4, 6, \text{ and } 8$ . Besides, when  $r = 2, 4$  and  $12$  the recognition accuracies have massive variances on the other hand minor variances are noted at  $r = 6, 8$  and  $10$ . Thus, for attaining the highest accuracy with maximum rate of constancy according to the recognition outcomes, we can choose the value of control parameter  $r = 6$  in our experiments.

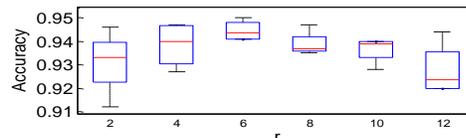


Figure 6: Recognition accuracies of proposed MLGEE DR method over different values of  $r$  on ORL dataset.

## 5. CONCLUSION

In this study, we proposed two graph ensemble embedding frameworks EGLPP and MLGEE. EGLPP is to improve recognition performance of LPP and make its projections less sensitive to neighborhood size  $k$ . On the other hand, we propose a generalized multi-manifold graph ensemble embedding for multi-view DR (MLGEE). MLGEE integrates the multi-manifolds heterogeneous graphs to build an intrinsic unified representation for multi-view manifold learning. Therefore, local manifold embeddings are efficiently used in MLGEE to attain local structures from multi-view databases.

Further we employ the widely used MFD, COIL-100, ORL and GT databases as test data to evaluate the performance of all compared methods; experimental results show that EGLPP and MLGEE outperforms several representative state-of-the-art DR methods. In future, we can extend our proposed MLGEE framework using sparse representation as well as using kernel tricks to well imitate the intrinsic structure of high dimensional data.

## REFERENCES

- [1] A. K. Jain, R. P. W. Duin, and J. Mao, "Statistical pattern recognition: A review," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 22, pp. 4-37, 2000.
- [2] S. Yan, D. Xu, B. Zhang, H. J. Zhang, Q. Yang, and S. Lin, "Graph embedding and extensions: a general framework for dimensionality reduction," *IEEE*

- Transactions on Pattern Analysis & Machine Intelligence*, vol. 29, p. 40, 2007.
- [3] Z. Lai, Y. Xu, J. Yang, L. Shen, and D. Zhang, "Rotational invariant dimensionality reduction algorithms," *IEEE transactions on cybernetics*, vol. 47, pp. 3733-3746, 2016.
- [4] H. Cai, V. W. Zheng, and K. C.-C. Chang, "A comprehensive survey of graph embedding: Problems, techniques, and applications," *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, pp. 1616-1637, 2018.
- [5] M. E. Tipping and C. M. Bishop, "Probabilistic principal component analysis," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 61, pp. 611-622, 1999.
- [6] I. T. Jolliffe, "Principal component analysis: a beginner's guide—II. Pitfalls, myths and extensions," *Weather*, vol. 48, pp. 246-253, 1993.
- [7] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, pp. 711-720, 1997.
- [8] H. C. Kan, T. Kwon, and R. J. Phaneuf, "Linear Discriminant Analysis in Document Classification," *IEEE Icdm Workshop on Text Mining*, vol. 77, pp. 998-1002, 2001.
- [9] J. Shawetaylor and N. Cristianini, *Kernel methods for pattern analysis* =, 2005.
- [10] Y. Qiang, W. Rong, X. Yang, N. L. Bing, and M. Yao, "Diagonal principal component analysis with non-greedy  $\ell_1$ -norm maximization for face recognition," *Neurocomputing*, vol. 171, pp. 57-62, 2016.
- [11] T. A. Abeo, X.-J. Shen, J.-P. Gou, Q.-R. Mao, B.-K. Bao, and S. Li, "Dictionary-induced least squares framework for multi-view dimensionality reduction with multi-manifold embeddings," *IET Computer Vision*, 2018.
- [12] W. Deng, J. Lu, J. Lu, and J. Guo, "Transform-Invariant PCA: A Unified Approach to Fully Automatic FaceAlignment, Representation, and Recognition," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 36, pp. 1275-1284, 2014.
- [13] H. Yan, J. Lu, and X. Zhou, "Prototype-Based Discriminative Feature Learning for Kinship Verification," *IEEE Transactions on Cybernetics*, vol. 45, pp. 2535-2545, 2017.
- [14] R. Saeidi, R. Astudillo, and D. Kolossa, "Uncertain LDA: Including observation uncertainties in discriminative transforms," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 38, pp. 1479-1488, 2015.
- [15] M. Kan, S. Shan, H. Zhang, S. Lao, and X. Chen, "Multi-view Discriminant Analysis," vol. 38, pp. 188-194, 2015.
- [16] L. Qiao, S. Chen, and X. Tan, "Sparsity preserving projections with applications to face recognition," *Pattern Recognition*, vol. 43, pp. 331-341, 2010.
- [17] J. B. Tenenbaum, V. De Silva, and J. C. Langford, "A global geometric framework for nonlinear dimensionality reduction," *science*, vol. 290, pp. 2319-2323, 2000.
- [18] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *science*, vol. 290, pp. 2323-2326, 2000.
- [19] T. F. Cox and M. A. Cox, *Multidimensional scaling*: Chapman and hall/CRC, 2000.
- [20] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural computation*, vol. 15, pp. 1373-1396, 2003.
- [21] H. Wang, S. Chen, Z. Hu, and W. Zheng, "Locality-preserved maximum information projection," *IEEE Transactions on Neural Networks*, vol. 19, pp. 571-585, 2008.
- [22] Y. Bengio, J.-f. Paiement, P. Vincent, O. Delalleau, N. L. Roux, and M. Ouimet, "Out-of-sample extensions for lle, isomap, mds, eigenmaps, and spectral clustering," in *Advances in neural information processing systems*, 2004, pp. 177-184.
- [23] X. He, *Locality preserving projections*, 2005.
- [24] X. He, D. Cai, S. Yan, and H.-J. Zhang, "Neighborhood preserving embedding," in *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, 2005, pp. 1208-1213.

- [25]D. Cai, X. He, K. Zhou, J. Han, and H. Bao, "Locality sensitive discriminant analysis," in *IJCAI*, 2007, pp. 1713-1726.
- [26]W. K. Wong and H. T. Zhao, *Supervised optimal locality preserving projection*, 2012.
- [27]B. Raducanu and F. Dornaika, "A supervised non-linear dimensionality reduction approach for manifold learning," *Pattern Recognition*, vol. 45, pp. 2432-2444, 2012.
- [28]Y. Xu, A. Zhong, J. Yang, and D. Zhang, "LPP solution schemes for use with face recognition," *Pattern Recognition*, vol. 43, pp. 4165-4176, 2010.
- [29]Q. X. Gao, H. Xu, Y. Y. Li, and D. Y. Xie, "Two-dimensional supervised local similarity and diversity projection," *Pattern Recognition*, vol. 43, pp. 3359-3363, 2010.
- [30]Q. Gao, J. Liu, H. Zhang, X. Gao, and K. Li, "Joint Global and Local Structure Discriminant Analysis," *IEEE Transactions on Information Forensics & Security*, vol. 8, pp. 626-635, 2013.
- [31]F. Song, F. Song, F. Ge, and Y. Zhao, "Short Communication: A novel local preserving projection scheme for use with face recognition," *Expert Systems with Applications An International Journal*, vol. 37, pp. 6718-6721, 2010.
- [32]G. Shikkenawis and S. K. Mitra, "On some variants of locality preserving projection," *Neurocomputing*, vol. 173, pp. 196-211, 2016.
- [33]X. Guo, T. Yun, L. Qi, and L. Guan, "A Novel Semi-Supervised Dimensionality Reduction Framework for Multi-manifold Learning," *IEEE Multimedia*, vol. 23, pp. 28-41, 2016.
- [34]C. Ding and Z. Li, "Double adjacency graphs-based discriminant neighborhood embedding," *Pattern Recognition*, vol. 48, pp. 1734-1742, 2015.
- [35]D. Cai, X. He, J. Han, and H.-J. Zhang, "Orthogonal laplacianfaces for face recognition," *IEEE transactions on image processing*, vol. 15, pp. 3608-3614, 2006.
- [36]S. Mehta, B.-S. Zhan, and X.-J. Shen, "Weighted Neighborhood Preserving Ensemble Embedding," *Electronics*, vol. 8, p. 219, 2019.
- [37]D. Shin, H. S. Lee, and D. Kim, "Illumination-robust face recognition using ridge regressive bilinear models," *Pattern Recognition Letters*, vol. 29, pp. 49-58, 2007.
- [38]A. Sharma and D. W. Jacobs, "Bypassing synthesis: PLS for face recognition with pose, low-resolution and sketch," in *Computer Vision & Pattern Recognition*, 2013.
- [39]D. R. Hardoon, S. Szedmak, and J. Shawe-Taylor, "Canonical Correlation Analysis: An Overview with Application to Learning Methods," *Neural Computation*, vol. 16, pp. 2639-2664, 2004.
- [40]L. Sun, S. Ji, and J. Ye, "Canonical Correlation Analysis for Multilabel Classification: A Least-Squares Formulation, Extensions, and Analysis," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 33, pp. 194-200, 2011.
- [41]T. Sun and S. Chen, "Locality preserving CCA with applications to data visualization and pose estimation," *Image & Vision Computing*, vol. 25, pp. 531-543, 2007.
- [42]Y. Chen, X.-h. Xu, and J.-h. Lai, "Optimal locality preserving projection for face recognition," *Neurocomputing*, vol. 74, pp. 3941-3945, 2011.
- [43]G. Yu, H. Peng, J. Wei, and Q. Ma, "Enhanced locality preserving projections using robust path based similarity," *Neurocomputing*, vol. 74, pp. 598-605, 2011.
- [44]H. Wang, Y. Yang, and B. Liu, "GMC: Graph-based Multi-view Clustering," *IEEE Transactions on Knowledge and Data Engineering*, 2019.
- [45]C. Hou, F. Nie, H. Tao, and D. Yi, "Multi-view unsupervised feature selection with adaptive similarity and view weight," *IEEE Transactions on Knowledge and Data Engineering*, vol. 29, pp. 1998-2011, 2017.
- [46]S. Sun, Y. Liu, and L. Mao, "Multi-view learning for visual violence recognition with maximum entropy discrimination and deep features," *Information Fusion*, vol. 50, pp. 43-53, 2019.
- [47]S. Huang, Z. Kang, I. W. Tsang, and Z. Xu, "Auto-weighted multi-view clustering via kernelized graph learning," *Pattern Recognition*, vol. 88, pp. 174-184, 2019.

- [48]J. Li, B. Zhang, G. Lu, and D. Zhang, "Generative multi-view and multi-feature learning for classification," *Information Fusion*, vol. 45, pp. 215-226, 2019.
- [49]B. Geng, D. Tao, C. Xu, L. Yang, and X.-S. Hua, "Ensemble manifold regularization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, pp. 1227-1233, 2012.
- [50]H. Hotelling, "Relations between 2 sets of variants," *Biometrika*, vol. 28, pp. 312-377, 1935.
- [51]X. Chen, S. Chen, H. Xue, and X. Zhou, "A unified dimensionality reduction framework for semi-paired and semi-supervised multi-view data," *Pattern Recognition*, vol. 45, pp. 2005-2018, 2012.
- [52]Y. Guo, X. Ding, and J.-H. Xue, "MiLDA: A graph embedding approach to multi-view face recognition," *Neurocomputing*, vol. 151, pp. 1255-1261, 2015.
- [53]F. Nie, J. Li, and X. Li, "Parameter-Free Auto-Weighted Multiple Graph Learning: A Framework for Multiview Clustering and Semi-Supervised Classification," in *IJCAI*, 2016, pp. 1881-1887.
- [54]M. Wang, X.-S. Hua, R. Hong, J. Tang, G.-J. Qi, and Y. Song, "Unified video annotation via multigraph learning," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 19, pp. 733-746, 2009.
- [55]F. Wang and D. Zhang, "A new locality-preserving canonical correlation analysis algorithm for multi-view dimensionality reduction," *Neural processing letters*, vol. 37, pp. 135-146, 2013.
- [56]L. Chen, H. Man, NEFIAN, and V. Ara, "Face recognition based on multi-class mapping of Fisher scores," *Pattern Recognition*, vol. 38, pp. 799-811, 2005.
- [57]S. Nayar, S. A. Nene, and H. Murase, "Columbia object image library (coil 100). Department of Comp," *Science, Columbia University, Tech. Rep. CUCS-006-96*, 1996.
- [58]L. Zong, X. Zhang, L. Zhao, H. Yu, and Q. Zhao, "Multi-view clustering via multi-manifold regularized non-negative matrix factorization," *Neural Networks*, vol. 88, pp. 74-89, 2017.
- [59]T. A. Abeo, X.-J. Shen, B.-K. Bao, Z.-J. Zha, and J. Fan, "A generalized multi-dictionary least squares framework regularized with multi-graph embeddings," *Pattern Recognition*, vol. 90, pp. 1-11, 2019.
- [60]A. Sharma, A. Kumar, H. Daume, and D. W. Jacobs, "Generalized multiview analysis: A discriminative latent space," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 2160-2167.