

Multiview Nonlinear Discriminant Structure Learning

Michael Shell, *Member, IEEE*, John Doe, *Fellow, OSA*, and Jane Doe, *Life Fellow, IEEE*

Abstract—Multiview subspace learning (MSL) that aims to project feature representations of each view into a latent subspace shared by all views, has been widely used in many practical applications. Despite the progress made, we think there are still two challenges to overcome. On the one hand, most of existing MSL methods indiscriminately utilize the helpful and defective information that are intermingled in each view. On the other hand, most recent methods are linear methods that could not work well when the dataset have weak linear separability. In this paper, we fully exploit the useful information of each view by local preserving and discriminant reconstruction (LPDR), then get the latent subspace with multiview nonlinear discriminant structure learning (MNDSL). The proposed method construct inter-view and intra-view weighted connections to explore discriminant structure and protect locality, under the complementary and uncorrelated principle. To meet the requirements of large-scale applications and get projections of new samples, an out-of-sample extension is introduced after that. Experiment results demonstrate the effectiveness and robustness of proposed method over other state-of-art methods.

Index Terms—Multiview subspace learning, nonlinear, locality and discriminant structure, complementary, uncorrelated, out-of-sample.

I. INTRODUCTION

MULTIVIEW learning is a long-standing and important area in the fields of machine learning and pattern recognition, which has been widely used in many aspects of machine learning [1]–[6]. In many real-world applications, objects can be characterized by various different viewpoints. For example, Internet news consists of text, images and video; images could be expressed by LBP feature [7], HOG feature [8] or SIFT feature [9]; Microsoft Kinect describes the environment with RGB view and Depth view. On the one hand, multiple views could provide complementary information for data description. On the other hand, multiview learning explores the consistency and complementary properties of different views to get an improved model. Instead, single view is relatively insufficient to characterize the full shape of objects [10], learning from a single view may not be robust and adequate. The main task of multiview learning is to make full use of rich information provided by multiple views and overcome their differences, following the principles of consensus and complementary [11].

Generally speaking, multiview learning methods largely fall into three categories [11]: 1) co-training, 2) multiple kernel learning, and 3) subspace learning. As one of the most

typical categories, multiview subspace learning (MSL) aims to project data of each view into a latent subspace shared by all views. Existing MSL methods can be grouped into two-view learning methods and multiview learning methods, according to the maximum number of views they can work on. These methods further run in supervised or unsupervised mode, based on whether class label information is exploited or not. Considerations of current methods include correlation maximization [12]–[15], category discriminant information balance [16]–[18], local neighborhood information protection [19]–[21], co-regularization based on support vector machine [22]–[24], entropy discriminant information maximization [25]–[27] and so on. Recently, multiview subspace learning is usually associated with sparse learning [28], low-rank decomposition [29], manifold learning [30] and many other directions.

However, most of existing MSL methods have the following disadvantages. First, helpful and defective information are intermingled in the multiple views, while most methods utilize them indiscriminately. For example, in some cases, outliers of one view may have normal distributions in another view, fixed samples in one view may be easy to separate in another view. Then the ideal strategy is to give up unique outliers of each view and refer to the view that can separate the fixed samples, following the complementary principle. But most of current MSL methods do not distinguish the given information is beneficial or not, they learn from the raw data almost blindly.

Second, most of existing MSL methods are linear and explicit methods, which means they could not work well when data have weak linear separability. Most methods seek for a linear transformer for each view to project the samples into the subspace. Kernel extension of some MSL methods may be suitable to address the nonlinear issue, but kernel functions would also increase the complexity of calculation, and sometimes it is difficult to choose an appropriate kernel function. Moreover, explicit MSL methods could get projections of new samples with the linear transformers, nonlinear MSL methods need a multiview out-of-sample extension to get projections of new samples, but few multiview out-of-sample methods were proposed. This makes nonlinear MSL methods difficult to adapt to large-scale applications.

To address the challenges above, in this paper, we present a multiview nonlinear discriminant structure learning framework that is trained on preprocessed data, and introduce its out-of-sample extension. The main contributions of this paper are summarized as below:

- 1) We reconstruct the feature representations of given views with local preserving and discriminant reconstruction

M. Shell was with the Department of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA, 30332 USA e-mail: (see <http://www.michaelshell.org/contact.html>).

J. Doe and J. Doe are with Anonymous University.

Manuscript received April 19, 2005; revised August 26, 2015.

(LPDR) to fully explore the helpful information provided by each view. LPDR minimizes distances between every sample and its within-class neighbors, as well as maximize the distances with its between-class neighbors. LPDR could characterize the connection between samples, reduce the influence of outliers peculiar to each view, and extract discriminant information of each view.

- 2) We get the embeddings of multiple views in the subspace with multiview nonlinear discriminant structure learning (MNDSL) which nonlinearly learns a discriminant structure and maintain the local neighborhood information in the subspace. MNDSL assigns a connection weight with each pair of inter-view or intra-view samples, following the principle of complementary. As inter-view samples have different dimensionality of feature representations, we introduce the global weight vector to measure their distance. Additionally, MNDSL impose the uncorrelated constrain that has been proved effective [21], [31] to further improve the performance.
- 3) An out-of-sample method that follows the principle of locality protection is introduced to get projections of new samples. The proposed method keeps the global nonlinearity and locality protection principle of MNDSL.

The remainder of this paper is organized as follows. Section II defines the common notations and reviews some related works. Section III introduces the formulations, optimizations of proposed methods in detail, and analyses the computation complexity. Section IV presents experimental results and quantitative evaluations. Finally, this paper is rounded up with a conclusion in Section V.

II. RELATED WORKS

In this section, we briefly review some works related to this paper. Numbers of multiview subspace learning methods have been proposed in recent years, we categorize them into four classes according to the primary objectives of them:

Maximize the cross correlations. This line of methods mainly origin from Canonical correlation analysis [12], a typical two-view learning method that finds two linear transforms for each view to maximize their cross correlation in the latent subspace. Basing on CCA, kernel CCA [14] employs kernel functions to address nonlinear projection, 2D CCA [13] directly sought relations between different sets of images without getting their feature vectors, sparse CCA [28] minimize the number of features in both projections while maximize their correlation. Partial least squares (PLS) [32] that shares similar objective with CCA regress samples from one view to the other, and kernel PLS [33] introduces kernel functions to PLS. In scenario of multiple views, many researches based on CCA were proposed as well. CCA is further generalized as multiview CCA (MCCA) [15] that maximize the correlation between each two views. And tensor CCA (TCCA) [34] analyses the high-order covariance tensor to maximize the correlations between each two views.

Balance the scatter discriminant information. Scatter discriminant information includes between-class, within-class, intra-view and inter-view discriminants here. Discriminative

CCA (DCCA) [17] and correlation discriminant analysis (CDA) [16] construct within-class similarity matrices and between-class similarity matrices based on CCA to separate sample pairs of different classes, while they have different formulations and constraints. Multiview uncorrelated discriminant analysis (MULDA) [35] characterizes DCCA by imposing an uncorrelated constraint. Several methods have been proposed for scenario of multiple views. Multiview discriminant analysis (MvDA) [18] jointly address the multiple linear transforms by maximizing between-class variations and minimizing within-class variations in a generalized Rayleigh quotient. Multiview local discrimination and canonical correlation analysis (MLDC²A) [36] makes use of correlation information, intra-view and inter-view discrimination information of paired views.

Protect the local neighborhood information. Locality preserving CCA (LPCCA) [37] incorporate locality idea into CCA. Laplacian MCCA (LapMCCA) [19] incorporate the locality into MCCA and discover the global nonlinear correlations for the multiple views. To preserve the locality of latent subspace, multiview spectral embedding (MSE) [20] regard neighborhood of a sample as a whole and project them together. Basing on MCCA, graph regularized MCCA (GrMCC) [38] maximizes inter-view cumulative correlations, minimizes local intraclass scatter as well as maximizes local between-class scatter. Multiview laplacian least squares (MvLLS) [39] construct a iteratively updating global weighted graph for the multiple views to measure the distance between each two samples, and maintain the distance relation in the subspace. Multiview locality low-rank embedding (MvL²E) [29] aims to protect the correlations and manifold structure by adopting low-rank representations.

Co-regularization based on SVM. This line of methods are usually built upon support vector machine (SVM) [40]. The two-view learning method SVM-2K [22] thinks that SVM could be regarded as a 1-dimensional projection followed by a thresholding, and imposes the similarity between the two 1-dimensional projections got by SVMs constructed for the two views. Multiview twin SVMs (MvTSVMs) [24] is a multiview extension of TSVMs [41]. Similar to SVM-2K, MvTSVMs combines two views by imposing the similarity between the two 1-dimensional projections got by each two views. Multiview laplacian SVMs (MLSVM) [23] additionally considers manifold and multiview regularization in the framework of SVM. And multiview privileged SVM (PSVM-2V) [42] introduces privileged information to unleash the power of complementary principle.

Several generalized multiview frameworks were proposed for two-view subspace learning [43], [44]. And recently, many subspace learning researches concentrate on feature redundancy minimization [45], reducing noise [46] or margin consistency [25], [26].

III. MULTIVIEW NONLINEAR DISCRIMINANT STRUCTURE LEARNING

In this section, we present the preprocessing, formulation and optimization of multiview nonlinear discriminant structure

TABLE I
LIST OF IMPORTANT PARAMETERS

Notation	Description
$X_i \in \mathbb{R}^{d_i \times n}$	feature representations of samples in i_{th} view
$X_{ij} \in \mathbb{R}^{d_i}$	feature representation of j_{th} sample in X_i
d_i	dimensionality of samples in X_i
n, m	the number of samples
c_i	class label of i_{th} sample
c_{ij}	class label of X_{ij}
v	the number of views
d	dimensionality of the subspace
t	maximum number of iterations
$Y_i \in \mathbb{R}^{d \times n}$	projection of X_i in the subspace
$Y_{iu} \in \mathbb{R}^d$	projection of X_{iu} in the subspace
W, M	weight graph
$maxiter$	maximum number of iterations
I	identity matrix
$tr(\cdot)$	trace of a matrix
$(\cdot)^T$	transpose of a matrix or vector
$\ \cdot\ _{2,F}$	ℓ_2 -norm, Frobenius norm

learning (MNDSL), then introduce its out-of-sample extension. Consider a multiview supervised learning problem defined on feature representations of n samples from v views, TABLE I lists important notations in this paper. Note that dimensionality may varies across the views, while the labels are identical for all of them.

We nonlinearly reconstruct the feature representations of each view with local preserving and discriminant reconstruction (LPDR) to make full use of the information provided by each view. Then with the reconstructed views, we present multiview nonlinear discriminant structure learning (MNDSL) that aims to learn a discriminant structure and maintain the local neighborhood in the subspace. As dimensionality of each view owns its size, a metric based on global weights is introduced. After that, following the principle of maintaining distance structure, an out-of-sample extension that keeps the distance structure maintain is proposed for MNDSL to accommodate large-scale applications and get representations of new samples.

A. Local Preserving and Discriminant Reconstruction

Most of existing subspace learning methods directly use the provided feature sets. In this section, we reconstruct the features to get better category discriminant distribution with LPDR which could be considered as a preprocessing of subsequent work. LPDR aims to minimize the ℓ_2 -norm distances between every sample and its k nearest neighbors that have identical class labels, as well as maximize the ℓ_2 -norm distances between every sample and its k nearest neighbors that have different class labels. Suppose feature representation X_i is reconstructed as X'_i , we could get the following optimization problem:

$$\min_{X'_i} \sum_{j=1}^n \sum_{l=1}^k (\|X'_{ij} - X'_{ijl}^W\|_2^2 - \|X'_{ij} - X'_{ijl}^B\|_2^2) \quad (1)$$

where X'_{ijl}^W is one sample of the k nearest neighbors of X'_{ij} that have identical class labels, and X'_{ijl}^B is one sample of k

nearest neighbors of X'_{ij} that have different class labels. In this model, we not only protect the local neighborhood of within-class samples, but also separate between-class samples. This problem could also be expressed and simplified as follows:

$$\begin{aligned} & \min_{X'_i} \xi(X'_i) \text{ s.t. } X'_i X'^{T}_i = I \\ \xi(X'_i) &= \sum_{j=1}^n (\|(X'_{ij} e^T - X'_i) S_j^W\|_F^2 - \|(X'_{ij} e^T - X'_i) S_j^B\|_F^2) \\ &= tr(X'_i \tilde{S} X'^{T}_i) \\ \tilde{S} &= \sum_{j=1}^n (S_j^W S_j^{WT} - S_j^B S_j^{BT}) \end{aligned} \quad (2)$$

To avoid degenerate solutions, we constrain that $X'_i X'^{T}_i = I$. In (2), $e = [1, 1, \dots, 1]^T \in \mathbb{R}^n$ is a column vector, S_j^W and S_j^B are selection matrices to select the within-class neighbors and between-class neighbors for X'_{ij} . Specially, if $Neib_{ij}^W \in \mathbb{R}^{d_i \times k}$ denotes the k nearest samples that have the same labels with X'_{ij} , and $Neib_{ij}^B \in \mathbb{R}^{d_i \times k}$ denotes the k nearest neighbors that have different labels with X'_{ij} , the usages of selection matrices are formulated as follows:

$$Neib_{ij}^W = X'_i S_j^W, Neib_{ij}^B = X'_i S_j^B \quad (3)$$

The derivation of $\xi(X'_i)$ is given in APPENDIX A. Based on Ky-Fan theory [47], X'_i in (2) has a global optimal solution. Consequently, we get the presentations of features in each view after reconstruction.

B. Multiview Nonlinear Discriminant Structure Learning

After preprocessing with LPDR, we have fully exploited discriminant relationships and locality of each view. Then we get embeddings of the views $\{Y_1, Y_2, \dots, Y_v\}$ with MNDSL. In the proposed method MNDSL, we aim to find a common latent subspace for multiple views where within-class samples stay close and between-class samples stay separate with each other.

First, we discuss the relationships of samples within views in the subspace. For samples that have identical class labels, the closer they are distributed on original feature spaces, the closer they are expected to stay in the subspace. And for samples that have different class labels in each view, the more dispersed they are distributed on original feature spaces, the more dispersed they are expected to stay in the subspace. In MNDSL, we assign a larger connection weight to samples that are expected to stay close in subspace, and a smaller connection weight to samples that are expected to stay separate. So we could draw samples with large weights close and separate samples with small weights, by minimizing the embedding cost function below:

$$\min_{Y_1, \dots, Y_v} \sum_{i=1}^v \|Y_{iu} - Y_{iv}\|_2^2 W_{uv}^{ii} \quad (4)$$

where W_{uv}^{ii} weights the connection between X'_{iu} and X'_{iv} . If X'_{iu} and X'_{iv} have identical labels, W_{uv}^{ii} is defined as:

$$W_{uv}^{ii} = \exp(-\|X'_{iu} - X'_{iv}\|_2^2), \text{ if } c_{iu} = c_{iv} \quad (5)$$

While X'_{iu} and X'_{iv} have different labels, W_{uv}^{ii} is defined as:

$$W_{uv}^{ii} = \exp(-\|X'_{iu} - X'_{iv}\|_2^2), \text{ if } c_{iu} \neq c_{iv} \quad (6)$$

Besides, following the principle of complementarity, we replace within-class weights that smaller than hyper-parameter ε_1 with the maximum of corresponding weights of all views. This could be expressed as follows:

$$W_{uv}^{ii} = \max\{W_{uv}^{11}, W_{uv}^{22}, \dots, W_{uv}^{vv}\} \\ \text{if } W_{uv}^{ii} < \varepsilon_1 \text{ and } c_{iu} = c_{iv} \quad (7)$$

Within-class weights no smaller than ε_1 are remained to protect locality information. Similarly, between-class weights that larger than ε_2 are replaced by the minimum of corresponding weights of all views:

$$W_{uv}^{ii} = \min\{W_{uv}^{11}, W_{uv}^{22}, \dots, W_{uv}^{vv}\} \\ \text{if } W_{uv}^{ii} > \varepsilon_2 \text{ and } c_{iu} \neq c_{iv} \quad (8)$$

These equations are effective to describe the connections of samples in the subspace for all views respectively.

Second, we discuss relationships of samples between views in the subspace. For inter-view samples, we would also like to get their embeddings reasonably, referring to the connection relationships and category discriminant information:

$$\min_{Y_1, \dots, Y_v} \sum_{i,j=1, i \neq j}^v \|Y_{iu} - Y_{jv}\|_2^2 W_{uv}^{ij} \quad (9)$$

But dimensionality of feature set may varies across all views, which makes it difficult to characterize the relationship between inter-view samples. Besides, inter-view samples have no natural locality information, while they still need quantitative weights to measure their connections in the subspace, and they are encouraged to cluster according to their class labels. To address the two issues above, we propose a novel metric to measure the connection relationships of inter-view samples in the subspace.

For each pair of inter-view samples such as X'_{iu} and X'_{jv} , we construct two global weight vectors $GW_{iu} \in \mathbb{R}^n$ and $GW_{jv} \in \mathbb{R}^n$ which are the u_{th} row and v_{th} row of W^{ii} and W^{jj} respectively. Then we utilize GW_{iu} , GW_{jv} to measure the weight agreement of X'_{iu} and X'_{jv} in the subspace similarly:

$$W_{uv}^{ij} = \exp(-\|GW_{iu} - GW_{jv}\|_2^2) \\ \text{if } (c_{iu} = c_{jv}, i \neq j) \\ W_{uv}^{ij} = \exp(-\|GW_{iu} - GW_{jv}\|_2^2) \\ \text{if } (c_{iu} \neq c_{jv}, i \neq j) \quad (10)$$

where the global weight vectors state the contribution of X'_{iu} and X'_{jv} in the embedding of each view. The more similar GW_{iu} and GW_{jv} are, the closer X'_{iu} and X'_{jv} are expected to stay in the subspace. Combining (4) and (9), we could get:

$$\min_{Y_1, \dots, Y_v} \sum_{i,j=1}^v \|Y_{iu} - Y_{jv}\|_2^2 W_{uv}^{ij} \quad (11)$$

where W_{uv}^{ij} is defined accordingly. To avoid degenerate solutions, we impose the embedding vectors to have weighted

unit covariance. And we add a uncorrelated constrain that has been proved effective [21], [31] in many applications by make the projections have minimum redundancy. The uncorrelated principle could be regarded as another form of constrain that $Y_1 D_{11} Y_1^T = I, Y_2 D_{22} Y_2^T = I, \dots, Y_v D_{vv} Y_v^T = I$ in this paper. With simple algebraic formulation of (11), we get the following minimization problem for MNDL finally:

$$\min_{Y_1, \dots, Y_v} \sum_{i,j=1}^v \text{tr}(Y_i L_{ij} Y_j^T) \\ \text{s.t. } Y_i^k D_{ii} Y_i^{kT} = 1 \\ Y_i^k D_{ii} Y_i^{rT} = 0, r = 1, \dots, k-1 \quad (12)$$

where row vector $Y_i^k \in \mathbb{R}^n$ is the k_{th} projection of X'_i in the subspace. Suppose that W_{uv}^{ij} is in the u_{th} row and v_{th} column of the matrix W^{ij} , L_{ij} is then defined as the Laplacian matrix [48] that corresponds to W^{ij} , and D_{ii} is a diagonal matrix that meet $D_{ii} = L_{ii} + W^{ii}$. Note that L_{ij} , W^{ij} and D_{ii} are all symmetric matrices.

C. Optimization and Solution

When $k = 1$, for the main projection of each view $\{Y_1^1, Y_2^1, \dots, Y_v^1\}$, we could formulate that:

$$\min_{Y_1^1, \dots, Y_v^1} \sum_{i,j=1}^v Y_i^1 L_{ij} Y_j^{1T} \\ \text{s.t. } Y_i^1 D_{ii} Y_i^{1T} = 1 \quad (13)$$

where only the first projection of each view is considered. Using the Lagrangian multiplier method, we could get:

$$\begin{bmatrix} Y_1^1 & \dots & Y_v^1 \end{bmatrix} \begin{bmatrix} L_{11} & \dots & L_{1v} \\ \vdots & \ddots & \vdots \\ L_{v1} & \dots & L_{vv} \end{bmatrix} = \\ \begin{bmatrix} \lambda_1 Y_1^1 & \dots & \lambda_v Y_v^1 \end{bmatrix} \begin{bmatrix} D_{11} & & \\ & \ddots & \\ & & D_{vv} \end{bmatrix} \quad (14)$$

which is a generalized multivariate eigenvalue problem, $\{\lambda_1, \lambda_2, \dots, \lambda_v\}$ are the Lagrange multipliers that meet $\lambda_i \geq 0$. This problem could be simplified to:

$$[Y_1^1 \dots Y_v^1] L = [\lambda_1 Y_1^1 \dots \lambda_v Y_v^1] D \quad (15)$$

Horst's alternation algorithm [49] that is scalable and works well in practice is employed to solve this problem, as shown in Algorithm 1. Because (13) is a convex problem, Horst's algorithm is guaranteed to converge to globally optimal solution.

After getting the main projection, we introduce a method based on mathematical induction to get other projections of each view and approximate optimal solution. Suppose we have got the first $k-1$ ($k > 1$) projections for each view, for the k_{th} projection $\{Y_1^k, Y_2^k, \dots, Y_v^k\}$, we could formulate this equation according to (12)

$$\min_{Y_1^k, \dots, Y_v^k} \sum_{i,j=1}^v Y_i^k L_{ij} Y_j^{kT} \\ \text{s.t. } Y_i^k D_{ii} Y_i^{kT} = 1 \\ Y_i^k D_{ii} Y_i^{rT} = 0, r = 1, \dots, k-1 \quad (16)$$

Algorithm 1 Horst's algorithm to get main projections**Input:** $\{W^{11}, \dots, W^{ij}, \dots, W^{vv}\}$ **Output:** $Y^1 = [Y_1^1, Y_2^1, \dots, Y_v^1]$

- 1: Get L_{11}, \dots, L_{vv} and D_{11}, \dots, D_{vv} respectively.
- 2: $Y^1 \leftarrow$ transpose of eigenvector associated with the smallest eigenvalue of $D^{-1}L$.
- 3: **for** $iter = 1$ to t **do**
- 4: $Y^1 = [Y_1^1, Y_2^1, \dots, Y_v^1] = Y^1 L D^{-1}$
- 5: **for** $i = 1$ to v **do**
- 6: $Y_i^1 \leftarrow \frac{Y_i^1}{\|Y_i^1\|}$
- 7: **end for**
- 8: **end for**
- 9: **return** $Y^1 = [Y_1^1, Y_2^1, \dots, Y_v^1]$

In order to obtain a closed-form solution, the constraint that $Y_i^k D_{ii} Y_i^{kT} = 1$ is transformed into a single constraint $\sum_{i=1}^v Y_i^k D_{ii} Y_i^{kT} = 1$, which will be discussed later in this section. We could have

$$\begin{aligned} \min_{Y_1, \dots, Y_v} \quad & \sum_{i,j=1}^v Y_i^k L_{ij} Y_j^{kT} \\ \text{s.t.} \quad & \sum_{i=1}^v Y_i^k D_{ii} Y_i^{kT} = 1 \\ & Y_i^k D_{ii} Y_i^{rT} = 0, r = 1, \dots, k-1 \end{aligned} \quad (17)$$

We construct the Lagrangian function of (17) first:

$$\begin{aligned} L(Y_i^k, \lambda, \mu_{ir}) = & \sum_{i,j=1}^v Y_i^k L_{ij} Y_j^{kT} \\ & - \lambda \sum_{i=1}^v (Y_i^k D_{ii} Y_i^{kT} - 1) - \sum_{i=1}^v \sum_{r=1}^{k-1} \mu_{ir} Y_i^k D_{ii} Y_i^{rT} \end{aligned} \quad (18)$$

where λ, μ_{ir} are Lagrangian multipliers. Taking its derivative with respect to Y_i^k and making the derivative to be zero, we have

$$\begin{aligned} \frac{\partial L(Y_i^k, \lambda, \mu_{ir})}{\partial Y_i^k} = & 2 \left(\sum_{j=1}^v Y_j^k L_{ij} - \lambda Y_i^k D_{ii} - \sum_{r=1}^{k-1} \mu_{ir} Y_i^r D_{ii} \right) = 0 \\ \sum_{j=1}^v Y_j^k L_{ij} = & \lambda Y_i^k D_{ii} + \sum_{r=1}^{k-1} \mu_{ir} Y_i^r D_{ii} \end{aligned} \quad (19)$$

Then we multiply the right-hand of (19) with Y_i^{kT} . Considering that $Y_i^k D_{ii} Y_i^{kT} = 1$ and $Y_i^r D_{ii} Y_i^k = 0$ ($r = 1, \dots, k-1$), we could get

$$\sum_{j=1}^v Y_j^k L_{ij} Y_i^{kT} = \lambda Y_i^k D_{ii} Y_i^{kT} = \lambda \quad (20)$$

As i varies from 1 to v , we would have got v equations correspondingly. Accumulating the v equations, we could have

$$\sum_{i=1}^v \sum_{j=1}^v Y_j^k L_{ij} Y_i^{kT} = \sum_{i,j=1}^v Y_i^k L_{ij} Y_j^{kT} = v\lambda \quad (21)$$

which means $v\lambda$ is equal to the value of objective function in (17). Similarly, multiplying the right-hand of (19) with Y_i^{zT} , ($z = 1, \dots, k-1$), we could get

$$\sum_{j=1}^v Y_j^k L_{ij} Y_i^{zT} = \sum_{r=1}^{k-1} \mu_{ir} Y_i^r D_{ii} Y_i^{zT} \quad (z = 1, \dots, k-1) \quad (22)$$

which is transformed as follows

$$\begin{aligned} \sum_{j=1}^v Y_j^k L_{ij} \begin{bmatrix} Y_i^1 \\ Y_i^2 \\ \vdots \\ Y_i^{(k-1)} \end{bmatrix}^T = & \begin{bmatrix} \mu_{i1} \\ \mu_{i2} \\ \vdots \\ \mu_{i(k-1)} \end{bmatrix}^T \begin{bmatrix} Y_i^1 \\ Y_i^2 \\ \vdots \\ Y_i^{(k-1)} \end{bmatrix} D_{ii} \begin{bmatrix} Y_i^1 \\ Y_i^2 \\ \vdots \\ Y_i^{(k-1)} \end{bmatrix}^T \end{aligned} \quad (23)$$

Let

$$\begin{aligned} \mu_i = & [\mu_{i1}, \mu_{i2}, \dots, \mu_{i(k-1)}] \\ G_i = & [Y_i^{1T}, Y_i^{2T}, \dots, Y_i^{(k-1)T}] \end{aligned} \quad (24)$$

so (23) could be expressed as

$$\sum_{j=1}^v Y_j^k L_{ij} G_i = \mu_i G_i^T D_{ii} G_i \quad (25)$$

Then we obtain

$$\mu_i = \left(\sum_{j=1}^v Y_j^k L_{ij} G_i \right) (G_i^T D_{ii} G_i)^{-1} \quad (26)$$

With expressions of μ_i and G_i , (19) could be expressed as

$$\sum_{j=1}^v Y_j^k L_{ij} = \lambda Y_i^k D_{ii} + \mu_i G_i^T D_{ii} \quad (27)$$

Substituting (26) into this equation, we could get

$$\begin{aligned} \lambda Y_i^k D_{ii} = & \sum_{j=1}^v Y_j^k L_{ij} - \left(\sum_{j=1}^v Y_j^k L_{ij} G_i \right) (G_i^T D_{ii} G_i)^{-1} G_i^T D_{ii} \\ = & \sum_{j=1}^v Y_j^k L_{ij} - \left(\sum_{j=1}^v Y_j^k L_{ij} \right) G_i (G_i^T D_{ii} G_i)^{-1} G_i^T D_{ii} \\ = & \left(\sum_{j=1}^v Y_j^k L_{ij} \right) [I - G_i (G_i^T D_{ii} G_i)^{-1} G_i^T D_{ii}] \end{aligned} \quad (28)$$

Let

$$P_i = I - G_i (G_i^T D_{ii} G_i)^{-1} G_i^T D_{ii} \quad (29)$$

Similarly, substituting (29) into (28), we could get

$$\lambda Y_i^k D_{ii} = \left(\sum_{j=1}^v Y_j^k L_{ij} \right) P_i \quad (30)$$

Algorithm 2 MNDSL to get other projections

Input: $\{L_{11}, \dots, L_{vv}\}, \{D_{11}, \dots, D_{vv}\}$ and $\{Y_1^1, \dots, Y_v^1\}$.
Output: Y_1, \dots, Y_v

- 1: $G_1 = [Y_1^{1T}], \dots, G_v = [Y_v^{1T}]$
- 2: P_1, \dots, P_v are initialized using (29).
- 3: **for** $k = 2$ to d **do**
- 4: $\{Y_1^k, \dots, Y_v^k\}$ are computed using (31).
- 5: **for** $i = 1$ to v **do**
- 6: $Y_i^k \leftarrow \frac{Y_i^k}{\sqrt{Y_i^k D_{ii} Y_i^{kT}}}$
- 7: end for
- 8: Add Y_i^k to the last column of G_i .
- 9: P_i is updated using (29).
- 10: end for
- 11: $Y_1 = G_1^T, \dots, Y_v = G_v^T$
- 12: **return** Y_1, \dots, Y_v

Then we address the final generalized eigenvalue solution

$$\begin{bmatrix} Y_1^k & \dots & Y_v^k \end{bmatrix} \begin{bmatrix} L_{11} & \dots & L_{1v} \\ \vdots & \ddots & \vdots \\ L_{v1} & \dots & L_{vv} \end{bmatrix} \begin{bmatrix} P_1 \\ \ddots \\ P_v \end{bmatrix} = \lambda \begin{bmatrix} Y_1^k & \dots & Y_v^k \end{bmatrix} \begin{bmatrix} D_{11} & & \\ & \ddots & \\ & & D_{vv} \end{bmatrix} \quad (31)$$

In this equation, G_i that contains the first $k-1$ projections is given as assumed. Because $v\lambda$ is equal to the value of objective function in (17), to minimize the objective function, $\{Y_1^k, Y_2^k, \dots, Y_v^k\}$ should form the eigenvector that corresponds to minimum eigenvalue of (31). As we have relaxed the constraint earlier, Y_i^k is resized to meet $Y_i^k D_{ii} Y_i^{kT} = 1$:

$$Y_i^k \leftarrow \frac{Y_i^k}{\sqrt{Y_i^k D_{ii} Y_i^{kT}}} \quad (32)$$

Note that this is an approximate solution to (16). With new computed Y_i^k , G_i and P_i is updated to get Y_i^{k+1} again. This process is repeated until all projection vectors on the subspace are computed, as shown in Algorithm 2.

D. MNDSL with Iterative Optimization

As introduced above, MNDSL is an approximate solution but not the global optimal solution. To get the global optimal solution, we extend MNDSL to a more complex model MNDSL-It by using iterative optimization. If the constraints are not relaxed in (16), we could get the following equation after similar steps:

$$\sum_{j=1}^v Y_j^k L_{ij} Y_j^{kT} = \lambda_i Y_i^k D_{ii} Y_i^{kT} = \lambda_i \quad (33)$$

where λ_i is the Lagrangian multiplier. As i varies from 1 to v , accumulate the v equations:

$$\sum_{i,j=1}^v Y_i^k L_{ij} Y_j^{kT} = \sum_{i=1}^v \lambda_i \quad (34)$$

Algorithm 3 MNDSL-It to get other projections

Input: $\{L_{11}, \dots, L_{vv}\}, \{D_{11}, \dots, D_{vv}\}$ and $\{Y_1^1, \dots, Y_v^1\}$.
Output: Y_1, \dots, Y_v

- 1: $G_1 = [Y_1^{1T}], \dots, G_v = [Y_v^{1T}]$
- 2: P_1, \dots, P_v are initialized using (29).
- 3: **for** $k = 2$ to d **do**
- 4: $Y^k \leftarrow$ transpose of eigenvector associated with the smallest eigenvalue of $D^{-1} P^T L$.
- 5: **for** $t = 1$ to t **do**
- 6: $Y^k = [Y_1^k, Y_2^k, \dots, Y_v^k] = Y^1 L P D^{-1}$.
- 7: **for** $i = 1$ to v **do**
- 8: $Y_i^k \leftarrow \frac{Y_i^k}{\sqrt{Y_i^k D_{ii} Y_i^{kT}}}$
- 9: end for
- 10: end for
- 11: Add Y_i^k to the last column of G_i .
- 12: P_i is updated using (29).
- 13: end for
- 14: $Y_1 = G_1^T, \dots, Y_v = G_v^T$
- 15: **return** Y_1, \dots, Y_v

which means the sum of Lagrangian multipliers is equal to the value of objective function in (17). Then we could also get:

$$\begin{bmatrix} Y_1^k & \dots & Y_v^k \end{bmatrix} \begin{bmatrix} L_{11} & \dots & L_{1v} \\ \vdots & \ddots & \vdots \\ L_{v1} & \dots & L_{vv} \end{bmatrix} \begin{bmatrix} P_1 \\ \ddots \\ P_v \end{bmatrix} = \begin{bmatrix} \lambda_1 Y_1^k & \dots & \lambda_v Y_v^k \end{bmatrix} \begin{bmatrix} D_{11} & & \\ & \ddots & \\ & & D_{vv} \end{bmatrix} \quad (35)$$

This problem could be simplified to:

$$[Y_1^k \dots Y_v^k] L P = [\lambda_1 Y_1^k \dots \lambda_v Y_v^k] D \quad (36)$$

which is a generalized eigenvalue problem that could be solved by Horst's algorithm. We would have to deal with a generalized eigenvalue problem to get each projection of the views, as Algorithm 3 shows.

E. Out-of-Sample Extension

As a nonlinear and inexplicit method that utilizes class label information in the training process, MNDSL could only get the projections of training sets of the views. To adapt to large-scale applications, we introduce an out-of-sample extension for MNDSL to get projections of new samples in the subspace. The proposed method keeps the global nonlinearity of MNDSL from the local linear fits, along with the thinking of nonlinear learning.

Suppose that we have got $\{Y_1, Y_2, \dots, Y_v\}$ for $\{X_1, X_2, \dots, X_v\}$ using MNDSL, given feature representations of m new samples from v views $\{x_1, x_2, \dots, x_v\}$, out-of-sample extension aims to find projections in the subspace $\{y_1, y_2, \dots, y_v\}$ for these new

samples, where $x_i \in \mathbb{R}^{d_i \times m}$, $y_i \in \mathbb{R}^{d \times m}$. We introduce two basic ideas of our out-of-sample method.

On the one hand, following the principle of locality protection, we characterize the within-view locality information of the new samples. We find k -nearest neighbors for x_{ia} in X'_i at first, x_{ia} is the a_{th} sample in x_i . Then we aim to get a linear confidence vector $w_{ia} = [w_{ia1}, \dots, w_{iak}]$, and $\sum_{l=1}^k w_{ial} x_{ial}$ is expected to be close to weighted sum of its k -nearest neighbors, where w_{ial} indicates the confidence of the l_{th} neighbor x_{ial} :

$$\begin{aligned} & \min_{w_1, \dots, w_v} f_1(w) \\ f_1(w) &= \sum_{i=1}^v \sum_{a=1}^m \left\| \sum_{l=1}^k w_{ial} (x_{ia} - x_{ial}) \right\|_2^2 \\ &= \sum_{i=1}^v \sum_{a=1}^m w_{ia} \tilde{x}_{ia}^T \tilde{x}_{ia} w_{ia}^T \end{aligned} \quad (37)$$

where $\tilde{x}_{ia} = [x_{ia} - x_{ia1}, x_{ia} - x_{ia2}, \dots, x_{ia} - x_{iak}]$, $w_{ia} = [w_{ia1}, w_{ia2}, \dots, w_{iak}]$. Additionally, we continue to use the global weight vector introduced in III-B to measure the distance M_{ab}^{ij} between x_{ia} and x_{jb} . The closer x_{ia} and x_{jb} are, the closer they are expected to stay in the subspace:

$$\begin{aligned} M_{ab}^{ij} &= \exp(-\|GW_{ia} - GW_{jb}\|_2^2) \text{ if } (c_{ia} = c_{jb}) \\ M_{ab}^{ij} &= \exp(-\|GW_{ia} - GW_{jb}\|_2^2) \text{ if } (c_{ia} \neq c_{jb}) \end{aligned} \quad (38)$$

On the other hand, as MNDSL protects locality well, we maintain the between-view linear confidence relationship of new samples in the subspace. In other words, $\sum_{l=1}^k w_{ial} x_{ial}$ is considered equal to $\alpha_{ia} x_{ia}$, and $\alpha_{ia} = \sum_{l=1}^v w_{ial}$. We could formulate that:

$$\begin{aligned} & \min_{w_1, \dots, w_v} f_2(w) \\ f_2(w) &= \sum_{i \neq j}^v \sum_{a, b=1}^m \left\| \frac{1}{\alpha_{ia}} \sum_{l=1}^k w_{ial} x_{ial} - \frac{1}{\alpha_{jb}} \sum_{l=1}^k w_{jbl} x_{jbl} \right\|_2^2 M_{ab}^{ij} \\ &= \sum_{i \neq j}^v \sum_{a, b=1}^m \|x_{ia} - x_{jb}\|_2^2 M_{ab}^{ij} = \sum_{i \neq j}^v \text{tr}(x_i M^{ij} x_j^T) \\ &= \sum_{i \neq j}^v L_{ij} w_i \tilde{x}_i^T \tilde{x}_j w_j^T \end{aligned} \quad (39)$$

In (39), $\tilde{x}_i = [x_i - x_{i1}, x_i - x_{i2}, \dots, x_i - x_{ik}]$, M_{ij} and L_{ij} are the value on i_{th} row and j_{th} column of m and L respectively, L is the Laplacian matrix that corresponds to M . Combining (37) and (39), we could get

$$\begin{aligned} & \min_{w_1, \dots, w_v} f(w) \\ f(w) &= \sum_{i, j=1}^v L_{ij} w_i Q_i^T Q_j w_j^T \\ &= \sum_{i, j=1}^v L_{ij} A_i^T A_j = \text{tr}(ALA^T) \\ & \text{s.t. } AA^T = I \end{aligned} \quad (40)$$

When $i \neq j$, Q_i, Q_j are the same as \tilde{y}_i, \tilde{y}_j . When $i = j$, Q_i is got by Cholesky decomposition of $\tilde{y}_i^T \tilde{y}_i + (\tilde{x}_i^T \tilde{x}_i)/L_{ii}$,

which is a symmetric matrix. The i_{th} column of A is denoted by A_i that meets $A_i = Q_i w_i^T$. The constrain $AA^T = I$ is introduced to avoid degenerate solutions. Using Lagrange multiplier method, we could get A by solving the following generalized eigenvalue problem:

$$AL = \lambda A \quad (41)$$

where λ is a Lagrange multiplier. Then $\{w_1, \dots, w_v\}$ are computed after getting A , as $A_i = Q_i w_i^T$. Finally, y_i is give by

$$y_i = \frac{\sum_{l=1}^k w_{il} y_{il}}{\alpha_i} \quad (42)$$

F. Complexity Analysis

To preprocess feature sets of all views, LPDR needs a time complexity of $O((vn)^3)$. Time complexity of MNDSL is $O((t + d - 1)(vn)^3)$, dominated by complexity of Algorithm 1 and Algorithm 2. Time complexity of MNDSL-It is $O(td(vn)^3)$. Time complexity of the out-of-sample extension is $O(v^3)$.

IV. EXPERIMENTS

V. CONCLUSION

In this paper, we preprocess multiview data with local preserving and discriminant reconstruction to fully exploit the useful information, then get the latent subspace by the multiview nonlinear discriminant structure learning (MNDSL). The proposed nonlinear MSL method follows complementary and uncorrelated principle to protect local neighborhood information and exploit discriminant structure of multiple views. Weighted connections are introduced to characterize the distances of samples in the subspace. As dimensionality varies between views, a metric based on global vector is used to construct connection weights between inter-view samples. After that, an out-of-sample method is introduce to get projections of new samples. Extensive experiments verified the effectiveness and robustness of the proposed method. In the future, we will work to reduce computational complexity and apply MNDSL to more tasks.

APPENDIX A

THE DERIVATION OF (2)

The function of reconstructed feature representations $\xi(X'_i)$ is reformulated as follows

$$\begin{aligned} \xi(X'_i) &= \sum_{j=1}^n (\|(X'_{ij} e^T - X'_i) S_j^W\|_F^2 - \delta \|(X'_{ij} e^T - X'_i) S_j^B\|_F^2) \\ &= \sum_{j=1}^n (\|X'_i (t_j e^T - I) S_j^W\|_F^2 - \delta \|X'_i (t_j e^T - I) S_j^B\|_F^2) \\ &= \sum_{j=1}^n (\|X'_i \tilde{S}_j^W\|_F^2 - \delta \|X'_i \tilde{S}_j^B\|_F^2) \\ &= \sum_{j=1}^n \sum_{l=1}^{d_i} (X'_i{}^l \tilde{S}_j^W \tilde{S}_j^{WT} X'_i{}^{lT} - \delta X'_i{}^l \tilde{S}_j^B \tilde{S}_j^{BT} X'_i{}^{lT}) \end{aligned}$$

$$\begin{aligned}
&= \sum_{l=1}^{d_i} (X_i'^l \tilde{S}^B X_i'^{lT} - \delta X_i'^l \tilde{S}^W X_i'^{lT}) \\
&= \text{tr}(X_i'(\tilde{S}^B - \delta \tilde{S}^W)X_i'^T) \\
&= \text{tr}(X_i' \tilde{S} X_i'^T)
\end{aligned}$$

where $t_j \in \mathbb{R}^n$ is a column vector that selects the j_{th} sample in X_i' , $X_{ij}' = X_i' t_j$, $X_i'^l$ is the l_{th} row of X_i' which indicates features of the l_{th} dimension.

APPENDIX B

Appendix two text goes here.

ACKNOWLEDGMENT

The authors would like to thank...

REFERENCES

- [1] S. Bickel and T. Scheffer, "Multi-view clustering," in *ICDM*, vol. 4, 2004, pp. 19–26.
- [2] Y.-H. Yuan, Y. Li, X.-B. Shen, Q.-S. Sun, and J.-L. Yang, "Laplacian multiset canonical correlations for multiview feature extraction and image recognition," *Multimedia Tools and Applications*, vol. 76, no. 1, pp. 731–755, 2017.
- [3] C. Zhang, H. Fu, Q. Hu, P. Zhu, and X. Cao, "Flexible multi-view dimensionality co-reduction," *IEEE Transactions on Image Processing*, vol. 26, no. 2, pp. 648–659, 2016.
- [4] Z. Ding and Y. Fu, "Robust multi-view subspace learning through dual low-rank decompositions," in *Thirtieth AAAI conference on artificial intelligence*, 2016.
- [5] X. Zhu, X. Li, and S. Zhang, "Block-row sparse multiview multilabel learning for image classification," *IEEE transactions on cybernetics*, vol. 46, no. 2, pp. 450–461, 2015.
- [6] Y. Feng, J. Xiao, Y. Zhuang, and X. Liu, "Adaptive unsupervised multi-view feature selection for visual concept recognition," in *Asian conference on computer vision*. Springer, 2012, pp. 343–357.
- [7] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 24, no. 7, pp. 971–987, 2002.
- [8] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, vol. 1. IEEE, 2005, pp. 886–893.
- [9] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [10] C. Xu, D. Tao, and C. Xu, "Multi-view intact space learning," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 12, pp. 2531–2544, 2015.
- [11] —, "A survey on multi-view learning," *arXiv preprint arXiv:1304.5634*, 2013.
- [12] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, no. 3/4, pp. 321–377, 1936.
- [13] S. H. Lee and S. Choi, "Two-dimensional canonical correlation analysis," *IEEE Signal Processing Letters*, vol. 14, no. 10, pp. 735–738, 2007.
- [14] S. Akaho, "A kernel method for canonical correlation analysis," *arXiv preprint cs/0609071*, 2006.
- [15] J. Rupnik and J. Shawe-Taylor, "Multi-view canonical correlation analysis," in *Conference on Data Mining and Data Warehouses (SiKDD 2010)*, 2010, pp. 1–4.
- [16] Y. Ma, S. Lao, E. Takikawa, and M. Kawade, "Discriminant analysis in correlation similarity measure space," in *Proceedings of the 24th international conference on Machine learning*. ACM, 2007, pp. 577–584.
- [17] T.-K. Kim, J. Kittler, and R. Cipolla, "Learning discriminative canonical correlations for object recognition with image sets," in *European Conference on Computer Vision*. Springer, 2006, pp. 251–262.
- [18] M. Kan, S. Shan, H. Zhang, S. Lao, and X. Chen, "Multi-view discriminant analysis," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 1, pp. 188–194, 2016.
- [19] Y. H. Yuan, Y. Li, X. B. Shen, Q. S. Sun, and J. L. Yang, "Laplacian multiset canonical correlations for multiview feature extraction and image recognition," *Multimedia Tools & Applications*, vol. 76, no. 1, pp. 731–755, 2017.
- [20] T. Xia, D. Tao, T. Mei, and Y. Zhang, "Multiview spectral embedding," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 40, no. 6, pp. 1438–1446, 2010.
- [21] J. Yin and S. Sun, "Multiview uncorrelated locality preserving projection," *IEEE transactions on neural networks and learning systems*, 2019.
- [22] J. Farquhar, D. Hardoon, H. Meng, J. S. Shawe-taylor, and S. Szedmak, "Two view learning: Svm-2k, theory and practice," in *Advances in neural information processing systems*, 2006, pp. 355–362.
- [23] S. Sun, "Multi-view laplacian support vector machines," in *International Conference on Advanced Data Mining and Applications*. Springer, 2011, pp. 209–222.
- [24] X. Xie and S. Sun, "Multi-view twin support vector machines," *Intelligent Data Analysis*, vol. 19, no. 4, pp. 701–712, 2015.
- [25] S. Sun and G. Chao, "Multi-view maximum entropy discrimination," in *Twenty-third international joint conference on artificial intelligence*, 2013.
- [26] G. Chao and S. Sun, "Consensus and complementarity based maximum entropy discrimination for multi-view classification," *Information Sciences*, vol. 367, pp. 296–310, 2016.
- [27] —, "Alternative multiview maximum entropy discrimination," *IEEE transactions on neural networks and learning systems*, vol. 27, no. 7, pp. 1445–1456, 2015.
- [28] D. R. Hardoon and J. Shawe-Taylor, "Sparse canonical correlation analysis," *Machine Learning*, vol. 83, no. 3, pp. 331–353, 2011.
- [29] L. Feng, X. Meng, and H. Wang, "Multi-view locality low-rank embedding for dimension reduction," *Knowledge-Based Systems*, vol. 191, p. 105172, 2020.
- [30] Y. Zhao, X. You, S. Yu, C. Xu, W. Yuan, X.-Y. Jing, T. Zhang, and D. Tao, "Multi-view manifold learning with locality alignment," *Pattern Recognition*, vol. 78, pp. 154–166, 2018.
- [31] S. Sun, X. Xie, and M. Yang, "Multiview uncorrelated discriminant analysis," *IEEE transactions on cybernetics*, vol. 46, no. 12, pp. 3272–3284, 2015.
- [32] A. Sharma and D. W. Jacobs, "Bypassing synthesis: Pls for face recognition with pose, low-resolution and sketch," 2011.
- [33] R. Rosipal and L. J. Trejo, "Kernel partial least squares regression in reproducing kernel hilbert space," *Journal of machine learning research*, vol. 2, no. Dec, pp. 97–123, 2001.
- [34] Y. Luo, D. Tao, K. Ramamohanarao, C. Xu, and Y. Wen, "Tensor canonical correlation analysis for multi-view dimension reduction," *IEEE transactions on Knowledge and Data Engineering*, vol. 27, no. 11, pp. 3111–3124, 2015.
- [35] S. Sun, X. Xie, and M. Yang, "Multiview uncorrelated discriminant analysis," *IEEE transactions on cybernetics*, vol. 46, no. 12, pp. 3272–3284, 2015.
- [36] L. Han, X.-Y. Jing, and F. Wu, "Multi-view local discrimination and canonical correlation analysis for image classification," *Neurocomputing*, vol. 275, pp. 1087–1098, 2018.
- [37] T. Sun and S. Chen, "Locality preserving cca with applications to data visualization and pose estimation," *Image and Vision Computing*, vol. 25, no. 5, pp. 531–543, 2007.
- [38] Y. H. Yuan and Q. S. Sun, "Graph regularized multiset canonical correlations with applications to joint feature extraction," *Pattern Recognition*, vol. 47, no. 12, pp. 3907–3919, 2014.
- [39] S. Guo, L. Feng, Z.-B. Feng, Y.-H. Li, Y. Wang, S.-L. Liu, and H. Qiao, "Multi-view laplacian least squares for human emotion recognition," *Neurocomputing*, vol. 370, pp. 78–87, 2019.
- [40] J. A. Suykens and J. Vandewalle, "Least squares support vector machine classifiers," *Neural processing letters*, vol. 9, no. 3, pp. 293–300, 1999.
- [41] R. Khemchandani, S. Chandra et al., "Twin support vector machines for pattern classification," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 29, no. 5, pp. 905–910, 2007.
- [42] J. Tang, Y. Tian, P. Zhang, and X. Liu, "Multiview privileged support vector machines," *IEEE transactions on neural networks and learning systems*, vol. 29, no. 8, pp. 3463–3477, 2017.
- [43] A. Sharma, A. Kumar, H. Daume, and D. W. Jacobs, "Generalized multiview analysis: A discriminative latent space," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*. IEEE, 2012, pp. 2160–2167.

- [44] G. Cao, A. Iosifidis, K. Chen, and M. Gabbouj, “Generalized multi-view embedding for visual recognition and cross-modal retrieval,” *IEEE transactions on cybernetics*, vol. 48, no. 9, pp. 2542–2555, 2018.
- [45] T. Zhou, C. Zhang, C. Gong, H. Bhaskar, and J. Yang, “Multiview latent space learning with feature redundancy minimization,” *IEEE transactions on cybernetics*, 2018.
- [46] Z. Yue, H. Yong, D. Meng, Q. Zhao, Y. Leung, and L. Zhang, “Robust multiview subspace learning with nonindependently and nonidentically distributed complex noise,” *IEEE transactions on neural networks and learning systems*, 2019.
- [47] R. Bhatia, *Matrix analysis*. Springer Science & Business Media, 2013, vol. 169.
- [48] M. Belkin and P. Niyogi, “Laplacian eigenmaps for dimensionality reduction and data representation,” *Neural computation*, vol. 15, no. 6, pp. 1373–1396, 2003.
- [49] P. Horst, “Relations among sets of measures,” *Psychometrika*, vol. 26, no. 2, pp. 129–149, 1961.

PLACE
PHOTO
HERE

Michael Shell Biography text here.