Check for updates

A New Semi-supervised Multi-view Dimensionality Reduction Method

Huijie Guo, Yifei Chen, and Junyan Tan(✉)

College of Science, China Agricultural University, Beijing, China
tanjunyan0@126.com

Abstract. Among a lot of the multi-view dimension reduction methods, there are few methods for semi-supervised problem since it is difficult in obtaining the sample labels. In this paper, a new semi-supervised dimensionality reduction method which is called multi-view semi-supervised similarity projection (MSSP) is proposed. A new metric of similarity and correlations between views is firstly defined and then a regularization term based on the new metric is proposed. MSSP finds the low-dimensional projection matrix by maximizing a discriminant term and the new regularization term for semi-supervised data. Compared with other methods, MSSP makes full use of all sample information; MSSP emphasizes the samples that have different local structure from different views should be more similar. 1NN is finally used to test the effectiveness of MSSP. Experiments on YALE and ORL data sets show that MSSP is superior to other methods.

Keywords: Multi-View Learning · Dimensionality Reduction · Similarity · Correlation

1 Introduction

The data obtained in different ways is multi-view data [1]. For example, a news report can be showed in voice, text, video, image, etc. For the multi-view data, the traditional single-view methods are not effective in the learning tasks. Therefore, the multi-view learning [2–5] emerged. Wherein, dimensional reduction has attracted most researchers' attention since the dimension of multi-view data is usually very high.

At present, many effective dimensionality reduction (DR) [6] methods for multi-view data have been proposed. The most widely used method is canonical correlation analysis (CCA) [7–9], which finds one linear transformation for each view by maximizing the between-view correlation coefficient. Sun and Chen point out that CCA ignores the local the information of samples and propose the Locality preserving CCA (LPCCA) [11]. Similar to the locality preserving projection (LPP) [12], LPCCA finds the lower-dimensional projection by maximizing the correlation coefficient of the neighbors. However, CCA and LPCCA are suitable for the unsupervised problem. For the supervised problem, the discriminant CCA (DCCA) [10] and generalized multi-view analysis (GMA) [13] are proposed. DCCA combines the idea of discriminant analysis and CCA.

© ICST Institute for Computer Sciences, Social Informatics and Telecommunications Engineering 2025
Published by Springer Nature Switzerland AG 2025. All Rights Reserved
H.-H. Chen and W. Meng (Eds.): WiSATS 2024, LNICST 606, pp. 14–26, 2025.
https://doi.org/10.1007/978-3-031-86203-8_2

DCCA finds a common lower-dimensional space by maximizing the intra-class correlations and minimizing the inter-class correlations. GMA [13] learns the optimal projection matrix by solving a joint, relaxed QCQP (quadratic constrained quadratic program) over different feature spaces to obtain a single (non)linear subspace. Notice that, the above methods have good classification effect on two-view data and they do not perform well in multi-view problem (the number of views is greater than 2) [14, 15]. For the multi-view problem, multiset canonical correlation analysis (MCCA) [16, 17] is put out. MCCA learns the corresponding projection matrix for each view by maximizing the overall correlation coefficient between any pair of views. Although there are a lot of DR methods for multi-view data, the multi-view semi-supervised DR methods are less. Sparse regularized Discriminative Canonical Correlation Analysis (SrDCCA) [18] and MSDA [21] are two semi-supervised DR methods. Through constructing sparse weighted matrices in multiple views, SrDCCA incorporates the structure information into the original CCA framework to seek the projection matrix. The lower dimensional representations are the most correlated and they contain the important discriminative structure information. In MSDA, the labeled examples are used to infer the discriminant structure in each view. All the examples are used to discover the intrinsic geometrical structure in each view.

The existing multi-view DR methods consider only the correlation between views and the discrimination. So that, we define the similarity between views firstly, then a new regularization term is designed. This term contains both the similarity and the correlations between views. Finally, a new DR method is proposed by combining the new regularization term and the discrimination term.

The main contributions of this paper are following:

1. A new regularization term which can measure both the structural similarity and the correlations between views is constructed.
2. A new semi-supervised dimensional reduction method is proposed, which can make better uses of the information hidden in data than other traditional methods.
3. The experimental results on image dataset demonstrate the validity and advantage of the new method.

The rest of the paper is arranged as follows: The related works are introduced in Sect. 2. In Sect. 3, we propose multi-view semi-supervised dimensionality reduction method. In Sect. 4, the experiments on face data sets and object data set are conducted to verify the effectiveness of the proposed method. We have a conclusion of this paper in Sect. 5.

2 Related Work

In this section, we briefly introduce the related methods: Canonical Correlation Analysis (CCA).

2.1 Canonical Correlation Analysis (CCA)

CCA is a common unsupervised dimension reduction method for two-view data. Suppose that there are n examples from two views, $X_1 \in R^{m_1 \times n}$ is the matrix of the examples in the first view, $X_2 \in R^{m_2 \times n}$ is the matrix of the examples in the second view. Suppose that all examples have been centrally processed. CCA aims at learning two projection matrices by solving the following optimization problem:

$$\begin{aligned} \max_{P_1,P_2} & \frac{E[P_1^T X_1 X_2^T P_2]}{\sqrt{E[P_1^T X_1 X_1^T P_1]}\sqrt{E[P_2^T X_2 X_2^T P_2]}} \\ &= \frac{P_1^T X_1 X_2^T P_2}{\sqrt{P_1^T X_1 X_1^T P_1}\sqrt{P_2^T X_2 X_2^T P_2}} \\ &= \frac{P_1^T S_{12} P_2}{\sqrt{P_1^T S_{11} P_1}\sqrt{P_2^T S_{22} P_2}} \end{aligned} \tag{1}$$

where $S_{11} \in R^{m_1 \times m_1}$, $S_{22} \in R^{m_2 \times m_2}$ are the variance matrices of X_1, X_2, $S_{12} \in R^{m_1 \times m_2}$ is the covariance matrix between $X_1$and X_2. Problem (1) means that the examples from two views should be highly corelated in the lower dimensional space. In order to get the optimal solution, problem (1) is usually converted into the following form:

$$\begin{aligned} &\max_{P_1,P_2} tr\left(P_1^T S_{12} P_2\right) \\ &\text{s.t.} \begin{array}{l} P_1^T S_{11} P_1 = I \\ P_2^T S_{22} P_2 = I \end{array} \end{aligned} \tag{2}$$

By using Lagrange multipliers, problem (2) can be transformed into the following generalized eigenvalue problem:

$$\begin{bmatrix} 0 & S_{12} \\ S_{21} & 0 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} = \lambda \begin{bmatrix} S_{11} & 0 \\ 0 & S_{22} \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \end{bmatrix} \tag{3}$$

The eigenvectors corresponding to the first d maximum eigenvalues are selected from the projection matrix:

$$P = \begin{bmatrix} P_1 \\ P_2 \end{bmatrix}$$

3 The Multi-view Semi-supervised Projection

In this section, the Jaccard similarity coefficient is firstly introduced, then we define the structural similarity matrix and put out our new method.

3.1 Jaccard Similarity Coefficient

Jaccard similarity coefficient measures the similarity between two sets, it is defined as follows [19]:

$$J(\mathrm{A}, \mathrm{B}) = \frac{|\mathrm{A} \cap \mathrm{B}|}{|\mathrm{A} \cup \mathrm{B}|} \tag{4}$$

where A and B are two sets, $|\cdot|$ indicates the number of elements in the set. $0 \leq J(A, B) \leq 1$, it is the proportion of the intersection in the union of A and B. It can be seen that the larger $J(A, B)$ is, the more similar A and B are. $J(A, B) = 1$ means A is entirely equals to B; $J(A, B) = 0$ means the intersection of A and B is $\varnothing$, i.e. A is totally different from B.

3.2 Structural Similarity Matrix Between Views

Suppose that $X = \{x_1, \ldots, x_n\}$ and $Y = \{y_1, \ldots, y_n\}$ represent the sample set of view 1 and view 2 respectively, the i^{th} sample in two views are denoted as x_i and y_i. The structural similarity matrix between view 1 and view 2 is defined as:

$$W_{1,2} = \mathrm{diag}\left(W_{1,2}(1), W_{1,2}(2), \cdots, W_{1,2}(n)\right) \tag{5}$$

where $W_{1,2}(j) j = 1,2, \cdots, n$ is the Jaccard similarity coefficient of the j^{th} sample in two views, it is computed according to:

$$W_{1,2}(j) = J\left(N_k^1(j), N_k^2(j)\right) = \frac{\left|N_k^1(j) \cap N_k^2(j)\right|}{\left|N_k^1(j) \cup N_k^2(j)\right|} \tag{6}$$

$N_k^1(j)$ and $N_k^2(j)$ represent the sets that contains the k^{th} nearest neighbors in two views. Figure 1 is an example that can illustrate the similarity between two views clearly.

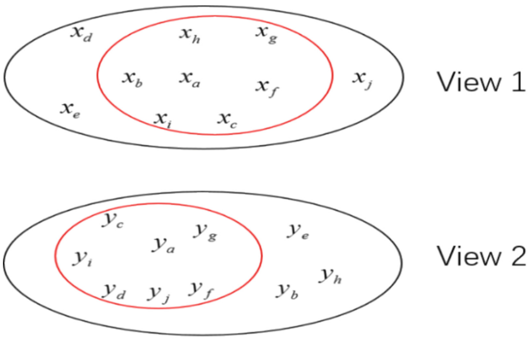


Fig. 1. 6 neighbors of sample a in different views

As showed in Fig. 1, the 6 neighbors of sample x_a in the view 1 is $N_k^1 = \{x_b, x_c, x_f, x_g, x_h, x_i\}$; the 6 neighbors of sample y_a in the view 2 is $N_k^2 = \{y_c, y_d, y_f, y_g, y_i, y_j\}$. The local structural similarity of the sample 'a' in the two views is measured by the Jaccard similarity coefficient:

$$W_{1,2}(a) = J\left(N_k^1(a), N_k^2(a)\right) = \frac{\left|N_k^1(a) \cap N_k^2(a)\right|}{\left|N_k^1(a) \cup N_k^2(a)\right|} \tag{7}$$

In this example,

$$\left|N_k^1(a) \cap N_k^2(a)\right| = |\{c, f, g, i\}| = 4$$
$$\left|N_k^1(a) \cup N_k^2(a)\right| = |\{b, c, d, f, g, h, i, j\}| = 8$$

Thus, the local structural similarity of sample a in the two views is 0.5. Follow the above method, we can get all samples' Jaccard similarity coefficients, then the structural similarity of all samples in two views can be computed.

3.3 Multi-view Semi-supervised Similarity Projection (MSSP)

Now, let us describe our notations. n is the number of samples, c is the number of classes, v is the number of views, m_k is the sample dimension in the k^{th} view. $X_k = \left[x_1^k, \ldots, x_i^k, \ldots, x_n^k\right] \in R^{m_k \times n}$ is the sample matrix of the k^{th} view. $\mathrm{X}_\mathrm{k}' \in \mathrm{R}^{\mathrm{m_k} \times \mathrm{nc}}$ is the labeled sample matrix in k^{th} view. $\mathrm{P}_1, \mathrm{P}_2, \cdots \mathrm{P}_\mathrm{v}$ are the projection matrices in v views, $\mathrm{P} = [\mathrm{P}_1, \ldots, \mathrm{P}_\mathrm{v}]^\mathrm{T}$, $\mathrm{z}_\mathrm{i}^\mathrm{k} = \mathrm{P}_\mathrm{k}\mathrm{x}_\mathrm{i}^\mathrm{k}$ is the low-dimension representation of the i^{th} sample in the k^{th} view.

MSSP learns a projection matrix for each view by using the discrimination information, the correlation between views and the local structural similarity between views. In the lower-dimensional space, MSSP hopes: 1) the intra-class scatter should be as small as possible, and the inter-class scatter should be as large as possible; 2) the between-view correlation and local structural similarity should be maximum. Based on the above motivation, the discrimination term $J_{dis}(P_1, ..., P_v)$ and the term that measures the correlation and similarity $J_{corr}(P_1, ..., P_v)$ are firstly proposed, then the optimization problem is put out. Now, let's introduce these two terms in detail.

3.4 The Formulation of MSSP

Discrimination Term $J_{dis}(P_1, ..., P_v)$. The intra-class scatter matrix in low-dimensional space is defined as follow:

$$S_w = \left[P_1^T, \cdots, P_v^T\right] \begin{bmatrix} S_{11} & \cdots & S_{1v} \\ \cdots & \cdots & \cdots \\ S_{v1} & \cdots & S_{vv} \end{bmatrix} = P^T SP \tag{8}$$

where S_{jr} is calculated as:

$$S_{jr} = \begin{cases} \sum\limits_{\mathrm{i}=1}^{\mathrm{c}} \left(\sum\limits_{\mathrm{k}=1}^{n_{ij}} x_{ik}^j \left(x_{ik}^j\right)^T - \frac{n_{ij} n_{ij}}{n_i} u_{ij} \left(u_{ij}\right)^T \right), \mathrm{j} = \mathrm{r} \\ -\sum\limits_{\mathrm{i}=1}^{\mathrm{c}} \frac{n_{ij} n_{ir}}{n_i} u_{ij} (u_{ir})^T, j \neq r \end{cases} \tag{9}$$

where $u_{ij} = \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} x_{ik}^j$ is the mean of samples belonging to i^{th} class in j^{th} view, n_{ij} is the number of samples belonging to i^{th} class in j^{th} view.

The inter-class scatter matric in low-dimensional space is defined as follow:

$$S_b = \left[P_1^T, \cdots, P_v^T\right] \begin{bmatrix} D_{11} & \cdots & D_{1v} \\ \cdots & \cdots & \cdots \\ D_{v1} & \cdots & D_{vv} \end{bmatrix} \begin{bmatrix} P_1 \\ \cdots \\ P_v \end{bmatrix} = P^T DP \tag{10}$$

where D_{jr} is calculated:

$$D_{jr} = \left(\sum_{\mathrm{i}=1}^{\mathrm{c}} \frac{n_{ij} n_{ir}}{n_i} u_{ij} u_{ir}^T \right) - \frac{1}{\mathrm{n}} \left(\sum_{\mathrm{i}=1}^{\mathrm{c}} n_{ij} u_{ij} \right) \left(\sum_{\mathrm{i}=1}^{\mathrm{c}} n_{ir} u_{ir} \right)^T \tag{11}$$

Then, the discrimination term is defined:

$$J_{dis}(\mathrm{P}) = tr\left(P^T DP - P^T SP\right) \tag{12}$$

Maximizing (12) means the within-class distance is small and the between-class distance is large.

Similarity Term. Notice that the same sample's local structural may be different between views since the data is obtained in different ways. While the traditional dimension reduction methods do not take this difference into condition. By incorporating the similarity, we define the new regularization term J_{corr} defined by (13). Maximizing the term (13) means the correlation coefficient between views with different local structural similarity should be larger.

$$\mathrm{J_{corr}} = tr\left(\sum\nolimits_{\mathrm{i}\neq\mathrm{j}}^{v} P_i^T X_i W_{i,j} X_j^T P_j\right) \tag{13}$$

$W_{i,j} = diag\left(W_{i,j}(x_1), \ldots, W_{i,j}(x_n)\right)$ is the similarity matrix between the i^{th} and ij^{th} view, and $W_{i,j}(x_k)$ represents the local structural similarity of sample x_k in the i^{th} and ij^{th} view computed according to Jaccard distance, $k = 1, 2, \cdots, n$.

For the convenience of solving, (13) can be rewritten as:

$$\begin{aligned} J_{corr} &= tr\left(\sum\nolimits_{\mathrm{i}\neq\mathrm{j}}^{v} P_i^T X_i W_{i,j} X_j^T P_j\right) \\ &= \left[P_1^T, \cdots, P_v^T\right] \begin{bmatrix} \mathrm{H}_{11} & \cdots & \mathrm{H}_{1v} \\ \cdots & \cdots & \cdots \\ \mathrm{H}_{v1} & \cdots & \mathrm{H}_{vv} \end{bmatrix} \begin{bmatrix} P_1 \\ \cdots \\ P_v \end{bmatrix} \\ &= \mathrm{tr}\left(P^T HP\right) \end{aligned} \tag{14}$$

where H_{jr} is:

$$H_{jr} = \begin{cases} 0, j = r \\ X_j W_{j,r} X_r^T, j \neq r \end{cases} \tag{15}$$

MSSP. Take the discrimination and similarity into consideration, the optimization problem of MSSP is constructed as:

$$\begin{aligned} &\max_{P_1,\cdots,P_v} J = J_{dis} + \mu J_{corr} \\ &= \mathrm{tr}\left(P^T DP - P^T SP + \mu P^T HP\right) \end{aligned} \tag{16}$$

In order to avoid the general solutions, a constraint is added to problem (16). Therefore, the final optimization problem of MSSP is:

$$\begin{aligned} &\max_{P_1,\cdots,P_v} J = J_{dis} + \mu J_{corr} \\ &= \mathrm{tr}\left(P^T DP - P^T SP + \mu P^T HP\right) \\ &= \mathrm{tr}\left(P^{T(\mathrm{D}-\mathrm{S}+\mu\mathrm{H})} P\right) \\ &= \mathrm{tr}\left(P^T \mathrm{L} P\right) \\ &s.t. P^T XX^T P = I \end{aligned} \tag{17}$$

where $P = [P_1, \ldots, P_v]^T, L = D - S + \mu H, X = diag(X_1X_1^T, \ldots, X_vX_v^T), I$ is the identity matrix.

Based on the above discussion, the algorithm of MSSP is summarized as follows:

Algorithm1. Multi-view semi-supervised similarity projection (MSSP) Input: samples matrix X_k and matrix of labeled samples $X_k', k = 1, \ldots, v$, the parameters μ, d.

Initialize:

1. Calculate the similarity term J_{corr} according to (13)
2. Calculate the discrimination term J_{dis} according to (12)
3. Solve the generalized eigenvalue problem: $(D - S + H)P = \lambda Xp$, where $p = \begin{bmatrix} p_1 \\ \cdots \\ p_v \end{bmatrix}$ and $X = diag(X_1X_1^T, \cdots, X_vX_v^T)$
4. The eigenvectors corresponding to the first d maximum eigenvalues of matrix P are selected and the projection matrix $P = \begin{bmatrix} P_1^* \\ \cdots \\ P_v^* \end{bmatrix}$, where P_k^* is the projection matrix of k^{th} view

Output: projection matrix P_k^*, $k = 1, \ldots, v$.

4 Experiment

4.1 Jaccard Similarity Coefficient

In this section, we compare MSSP and several other DR methods on three image datasets, including YALE, ORL face datasets and COIL-20 object dataset. ORL face dataset includes 400 images of 40 persons. The 10 images per person are collected under different time, lighting, facial expressions and facial details. Figure 2 is an example that contains 10 samples of one person.

Fig. 2. Sample images of one individual on ORL face datasets.

YALE face dataset includes 165 images of 15 persons. The 11 images per person are under different lighting conditions, facial expressions and glasses occlusion. Figure 3 shows 11 sample images of one person.

COIL-20 object dataset contains 1440 images of 20 objects. Each object has 72 samples in different poses. Several images of one object are shown in Fig. 4.

Fig. 3. Sample images of one individual on Yale face dataset.

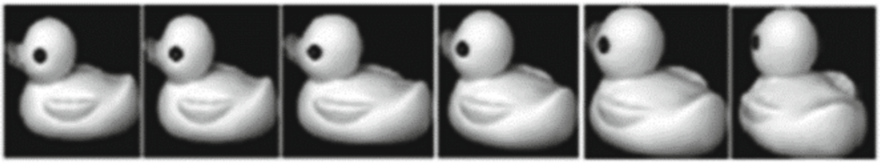


Fig. 4. Sample images of one individual on COIL-20 object.

4.2 The Experimental Settings

MSSP is compared with 7 popular multi-view dimensionality reduction methods: MCCA, LPCA, ALPCCA and MvPLS [20] (unsupervised methods), MvDA (supervised method) SrDccA [18] and MSDA [21].

The data is transformed into multi-view form through HOG, gray value and BLP, and the dimensionality is reduced to 100 by PCA for our experiments.

For the three datasets used in this paper, the parameter k is selected from the set {5,10,15,20}, the balance parameter μ is selected from the set $\{2^{-5}, 2^{-4}, \cdots, 2^{5}\}$.

In the experiment, we adopt the 1-nearest-neighborhood (1NN) classifier to evaluate the dimensional reduction effectiveness. For multi-view problem, the data from different views are projected to a common subspace, the 1NN classifier is used to classify multi-view data on this common subspace. The average classification accuracy of ten randomized experiments is regarded as our result.

4.3 The Experimental Settings

Comparison Experiment with Different Dimension. Tables 1, 2 and 3 show the best accuracies with different number of neighbors and training examples. In the tables, train, table and K represent the number of samples selected from each class, the number of labeled training samples, the number of k-nearest neighbors. According to the tables, the proposed method is obviously superior to other methods on each dataset because it has the best accuracy. Compared with the supervised and the unsupervised methods, MSSP is the best because it utilized all sample information including the labeled and unlabeled samples. Compared with semi-supervised methods, the maximum improvement of the best accuracy is 12.44%; the minimum improvement of the accuracy is 4.89%, MSSP is also the best one since it fully utilized the structural differences between views and discriminative information. In addition, the variance of the new method is also smaller than other methods that means that MSSP is more stable.

The best accuracies of the eight methods are shown in Fig. 5, it is obvious that the box plot of MSSP has the highest average accuracy and the smallest fluctuation range. So, we can conclude that our MSSP performs best on YALE, ORL and COIL-20.

Table 1. The optimal accuracies (%) for Yale

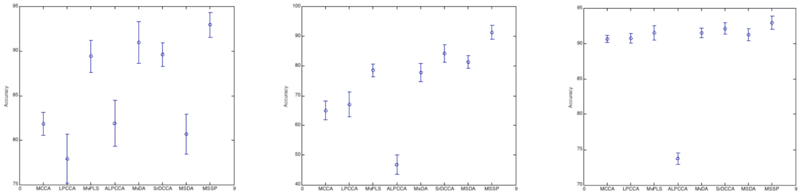
train		label	MCCA	LPCCA	MvPLS	ALPCCA	MvDA	SrDCCA	MSDA	MSSP
7	K = 1	5	76.11 ± 1.72	77.00 ± 1.52	80.44 ± 2.01	78.34 ± 2.87	76.44 ± 1.73	78.22 ± 2.38	80.56 ± 0.75	90.67 ± 1.34
	K = 3	5	76.11 ± 1.72	77.00 ± 1.52	80.44 ± 2.0	78.34 ± 2.87	76.44 ± 1.73	78.22 ± 2.38	76.89 ± 1.57	90.11 ± 1.21
	K = 5	5	76.11 ± 1.72	77.00 ± 1.52	80.44 ± 2.0	78.34 ± 2.87	76.44 ± 1.73	78.22 ± 2.38	78.56 ± 1.93	91.00 ± 1.06
8	K = 1	5	77.63 ± 2.07	77.48 ± 2.11	79.26 ± 1.87	76.23 ± 3.11	78.81 ± 1.83	79.26 ± 4.31	79.67 ± 1.21	88.15 ± 2.03
	K = 3	5	77.63 ± 2.07	77.48 ± 2.11	79.26 ± 1.87	76.23 ± 3.11	78.81 ± 1.83	79.26 ± 4.31	74.67 ± 1.37	89.19 ± 1.35
	K = 5	5	77.63 ± 2.07	77.48 ± 2.11	79.26 ± 1.87	76.23 ± 3.11	78.81 ± 1.83	79.26 ± 4.31	77.33 ± 1.04	88.00 ± 2.21
9	K = 1	5	80.44 ± 2.11	79.56 ± 2.23	79.78 ± 1.78	79.04 ± 3.21	84.44 ± 2.52	81.44 ± 3.36	82.44 ± 1.22	90.89 ± 1.56
	K = 3	5	80.44 ± 2.11	79.56 ± 2.23	79.78 ± 1.78	79.04 ± 3.21	84.44 ± 2.52	81.44 ± 3.36	79.78 ± 2.03	87.11 ± 1.63
	K = 5	5	80.44 ± 2.11	79.56 ± 2.23	79.78 ± 1.78	79.04 ± 3.21	84.44 ± 2.52	81.44 ± 3.36	82.22 ± 1.57	87.11 ± 0.22

Table 2. The optimal accuracies (%) for ORL

train		label	MCCA	LPCCA	MvPLS	ALPCCA	MvDA	SrDCCA	MSDA	MSSP
6	K = 1	5	72.38 ± 1.12	72.96 ± 0.68	88.17 ± 1.36	82.69 ± 2.98	80.25 ± 1.48	87.87 ± 1.88	88.42 ± 0.38	89.42 ± 1.65
	K = 3	5	72.38 ± 1.12	72.96 ± 0.68	88.17 ± 1.36	82.69 ± 2.98	80.25 ± 1.48	87.87 ± 1.88	86.83 ± 1.73	89.88 ± 1.22
	K = 5	5	72.38 ± 1.12	72.96 ± 0.68	88.17 ± 1.36	82.69 ± 2.98	80.25 ± 1.48	87.87 ± 1.88	86.54 ± 0.73	89.00 ± 1.26
7	K = 1	5	78.89 ± 1.94	79.78 ± 1.87	90.44 ± 2.06	85.76 ± 2.25	86.61 ± 1.61	88.72 ± 1.64	85.72 ± 1.37	92.50 ± 0.81
	K = 3	5	78.89 ± 1.94	79.78 ± 1.87	90.44 ± 2.06	85.76 ± 2.25	86.61 ± 1.61	88.72 ± 1.64	84.56 ± 1.75	92.67 ± 1.37
	K = 5	5	78.89 ± 1.94	79.78 ± 1.87	90.44 ± 2.06	85.76 ± 2.25	86.61 ± 1.61	88.72 ± 1.64	85.00 ± 0.93	91.94 ± 1.10
8	K = 1	5	83.58 ± 2.21	84.17 ± 2.01	90.75 ± 1.75	88.13 ± 2.37	89.67 ± 2.03	89.08 ± 2.13	88.00 ± 1.57	92.25 ± 1.05
	K = 3	5	83.58 ± 2.21	84.17 ± 2.01	90.75 ± 1.75	88.13 ± 2.37	89.67 ± 2.03	89.08 ± 2.13	84.67 ± 1.23	91.33 ± 0.96
	K = 5	5	83.58 ± 2.21	84.17 ± 2.01	90.75 ± 1.75	88.13 ± 2.37	89.67 ± 2.03	89.08 ± 2.13	84.83 ± 1.57	92.33 ± 1.25

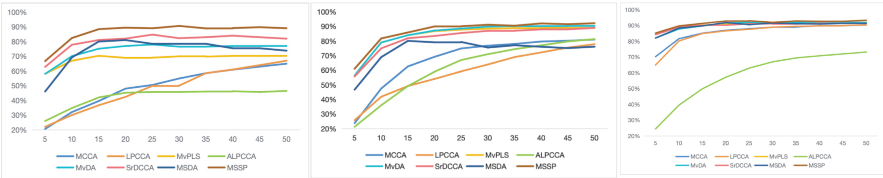
Table 3. The optimal accuracies (%) for COLL-20

train		label	MCCA	LPCCA	MvPLS	ALPCCA	MvDA	SrDCCA	MSDA	MSSP
15	K = 5	10	88.82 ± 0.8	90.46 ± 0.77	92.16 ± 0.90	86.82 ± 1.53	88.38 ± 1.04	89.44 ± 0.91	89.60 ± 0.84	93.78 ± 0.18
	K = 10	10	88.82 ± 0.8	90.46 ± 0.77	92.16 ± 0.90	86.82 ± 1.53	88.38 ± 1.04	89.44 ± 0.91	89.34 ± 0.94	92.34 ± 0.32
20	K = 5	15	92.35 ± 0.55	93.01 ± 0.61	93.01 ± 0.55	89.16 ± 0.97	92.37 ± 0.57	92.01 ± 1.03	89.29 ± 0.52	93.43 ± 0.48
	K = 10	15	92.35 ± 0.55	93.01 ± 0.61	93.01 ± 0.55	89.16 ± 0.97	92.37 ± 0.57	92.01 ± 1.03	88.39 ± 1.14	93.68 ± 0.73
	K = 15	15	92.35 ± 0.55	93.01 ± 0.61	93.01 ± 0.55	89.16 ± 0.97	92.37 ± 0.57	92.01 ± 1.03	88.85 ± 0.84	93.48 ± 1.03
25	K = 5	15	94.61 ± 0.66	94.78 ± 0.83	94.53 ± 0.60	92.34 ± 1.35	94.84 ± 0.71	92.68 ± 0.53	92.30 ± 0.82	95.01 ± 0.67
	K = 10	15	94.61 ± 0.66	94.78 ± 0.83	94.53 ± 0.60	92.34 ± 1.35	94.84 ± 0.71	92.68 ± 0.53	92.74 ± 0.95	95.34 ± 0.31
	K = 15	15	94.61 ± 0.66	94.78 ± 0.83	94.53 ± 0.60	92.34 ± 1.35	94.84 ± 0.71	92.68 ± 0.53	92.08 ± 0.82	95.24 ± 0.39
30	K = 5	15	95.73 ± 0.95	95.72 ± 0.44	95.60 ± 0.44	94.79 ± 2.03	95.04 ± 0.20	93.40 ± 1.14	92.92 ± 1.01	96.19 ± 0.73
	K = 10	15	95.73 ± 0.95	95.72 ± 0.44	95.60 ± 0.44	94.79 ± 2.03	95.04 ± 0.20	93.40 ± 1.14	91.97 ± 0.90	96.31 ± 0.41
	K = 15	15	95.73 ± 0.95	95.72 ± 0.44	95.60 ± 0.44	94.79 ± 2.03	95.04 ± 0.20	93.40 ± 1.14	92.28 ± 0.62	96.49 ± 0.56



(a) ORL (b)YALE (c)COIL-20

Fig. 5. The best accuracies of the eight methods on three datasets



(a) YALE database. (b) ORL database. (c) COIL-20 database.

Fig. 6. The classification accuracies by different methods on three datasets

Figure 6 compares the classification accuracies on three datasets when the samples are projected to different dimensions. On the data set YALE and OPL, MSSP shows a great improvement on other methods. On COLL-20, MSSP has a weak advantage over other methods. Furthermore, we can see that all methods have the best accuracies when the dimensions are reduced to 50.

5 Conclusion

This paper proposes a way to measure the structural similarity between views firstly, then a multi-view dimensionality reduction is proposed, which not only considers the discrimination of the samples in all views, but also includes the structural similarity between views. MSSP emphasizes the samples that have different local structure from different views should be more similar. 1NN classifier is used to test the effectiveness of MSSP and other dimensionality reduction methods. The experiments are conducted on three real face datasets, YALE, ORL and COIL-20. The results show that MSSP is an effective semi-supervised dimensionality reduction method because it has the best accuracies and the best stability.

References

1. Sargin, M.E., et al.: Audiovisual synchronization and fusion using canonical correlation analysis. IEEE Trans. Multimedia 1396–1403 (2007)
2. Sun, S.: A survey of multi-view machine learning. Neural Comput. Appl. 2031–2038 (2013)
3. Blum, A., Mitchell, T.: Combining labeled and unlabeled data with co-training. In: Conference on Computational Learning Theory (1998)

4. Gönen, M., Alpaydın, E.: Multiple kernel learning algorithms. J. Mach. Learn. Res. 2211–2268 (2011)
5. Xu, C., Tao, D., Xu, C.: A survey on multi-view learning, arXiv preprint arXiv:1304.5634 (2013)
6. Liu, H., Motoda, H.: Feature transformation and dimensionality reduction (1998)
7. Hardoon, D.R., Szedmak, S., Shawe-Taylor, J.: Canonical correlation analysis: an overview with application to learning methods. Neural Comput. 2639–2664 (2004)
8. Kakade, S.M., Foster, D.P.: Multi-view regression via canonical correlation analysis. Learn. Theory (2007)
9. Sun, L., Ji, S., Ye, J.: Canonical correlation analysis for multilabel classification: a least-squares formulation, extensions, and analysis. IEEE Trans. Pattern Anal. Mach. Intell. 194–200 (2011)
10. Sun, T., Chen, S., Yang, J., et al.: A novel method of combined feature extraction for recognition. In: Eighth IEEE International Conference on Data Mining, IEEE Computer Society (2008)
11. Sun, T., Chen, S.: Locality preserving CCA with applications to data visualization and pose estimation. Image Vis. Comput. 531–543 (2007)
12. He, X.F., Niyogi, P.: Locality preserving projections. In: Proceedings of Neural Information Processing Systems (NIPS) (2003)
13. Jacobs, D.W., Daume, H., Kumar, A., et al.: Generalized multiview analysis: a discriminative latent space. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition, IEEE Computer Society (2012)
14. Nielsen, A.: Multiset canonical correlations analysis and multispectral, truly multitemporal remote sensing data. IEEE Trans. Image Process. 293–305 (2002)
15. Gong, Y., Ke, Q., Isard, M., et al.: A multi-view embedding space for modeling internet images, tags, and their semantics. Int. J. Comput. Vis. 210–233 (2012)
16. Kettering, J.R.: Canonical analysis of several sets of variables. Biometrika 433–451 (1971)
17. Rupnik, J., Shawe-Taylor, J.: Multi-view canonical correlation analysis. In: Conference on Data Mining and Data Warehouses, pp.1–4 (2010)
18. Hou, S., Liu, H., Sun, Q.: Sparse regularized discriminative canonical correlation analysis for multi-view semi-supervised learning. Neural Comput. Appl. (2018)
19. Jaccard, P.: The distribution of flora in the alpine zone. New Phytologist 37–50 (2010)
20. Cao, G., Iosifidis, A., Chen, K., et al.: Generalized multi-view embedding for visual recognition and cross-modal retrieval. IEEE Trans. Cybern. 1–14 (2017)
21. Zhao, X., Evans, N., Dugelay, J.L.: Multi-view semi-supervised discriminant analysis: a new approach to audio-visual person recognition. In: Signal Processing Conference, IEEE (2012)