Multi-view Sparse Embedding Analysis Based Image Feature Extraction and Classification

Yangping Zhu[1], Xiaoyuan Jing[1,2](✉), Qing Wang[3], Fei Wu[4], Hui Feng[2], and Shanshan Wu[1]

[1] College of Automation, Nanjing University of Posts and Telecommunications, Nanjing 210023, China
zhuyangping_1991@163.com, jingxy_2000@126.com
[2] State Key Laboratory of Software Engineering, Wuhan University, Wuhan 430079, China
[3] College of Computer, Nanjing University of Posts and Telecommunications, Nanjing 210003, China
happywq2009@126.com
[4] College of Communication and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China
wufei_8888g@126.com

Abstract. Multi-view feature extraction is an attractive research topic in computer vision domain, since it can well reveal the inherent property of images. Most existing multi-view feature extraction methods focus on investigating the intra-view or inter-view correlation. However, they fail to consider the sparse reconstruction relationship and the discriminant correlation in multi-view data, simultaneously. In this paper, we propose a novel multi-view feature extraction approach named Multi-view Sparse Embedding Analysis (MSEA). MSEA not only explores the sparse reconstruction relationship that hides in multi-view data, but also considers discriminant correlation by maximizing the within-class correlation and simultaneously minimizing the between-class correlation from intra-view. Moreover, we add orthogonal constraints of embedding matrices to remove the redundancy among views. To tackle the linearly inseparable problem in original feature space, we further provide a kernelized extension of MSEA called KMSEA. The experimental results on two datasets demonstrate the proposed approaches outperform several state-of-the-art related methods.

Keywords: Sparse embedding analysis · Multi-view · Discriminant correlation · Orthogonal constraints

1 Introduction

In computer vision domain, many applications are usually involved with different views of data. With respect to feature extraction, multi-view features can

Y. Zhu and X. Jing—The work described in this paper was fully supported by the National Natural Science Foundation of China under Project No. 61272273, the Research Project of Nanjing University of Posts and Telecommunications under Project No. XJKY14016, and the Postgraduate Scientific Research and Innovation Plan of Jiangsu Province Universities under Project No. CXLX13-465.

H. Zha et al. (Eds.): CCCV 2015, Part II, CCIS 547, pp. 51–60, 2015.
DOI: 10.1007/978-3-662-48570-5_6

well reveal the inherent property of data. Multi-view feature extraction aims to exploit different characteristics or views of data, which is an attractive and important research direction [1,2].

Existing supervised multi-view extraction methods can be roughly categorized into two types. ***(1) Shared subspace learning based methods.*** They focus on learning a common shared subspace, in which the correlation among multiple views can be well revealed. Mostly they are based on canonical correlation analysis (CCA) [3], which is a vital multi-view extraction technique, since it can well utilize the inter-view correlation. Other shared subspace learning based methods include discriminant analysis of canonical correlations (DCC) [4], multiple discriminant CCA (MDCCA) [5], multi-view discriminant analysis (MvDA) [6], intra-view and inter-view supervised correlation analysis (I^2SCA) [7], etc. ***(2) Transfer learning and dictionary learning based methods.*** They focus on incorporating the transfer learning or dictionary learning techniques into the multi-view feature extraction process. Transfer learning can alleviate the distribution differences among different views. And dictionary learning holds favorable reconstruction capability for multi-view features. Based on them, transfer component analysis (TCA) [8] and uncorrelated multi-view fisher discrimination dictionary learning (UMDDL) [9] are presented.

Although there exist much effort on multi-view extraction, existing methods almost fail to preserve the sparse reconstruction relationship and simultaneously consider the discriminant correlation in multi-view data. In this paper, we propose a novel multi-view feature extraction approach named Multi-view Sparse Embedding Analysis (MSEA). The contributions are summarized as follows:

1. We incorporate the sparse embedding analysis and learn a shared dictionary for multiple views, such that the sparse reconstruction relationship in multi-view data can be well preserved. Moreover, we consider the discriminative correlation by maximizing the within-class correlation and simultaneously minimizing the between-class correlation from intra-view. Since there exist much redundancy in multi-view features, we add the orthogonal constraints into the objective function, such that the redundant information among views can be effectively reduced.
2. We further provide a kernelized extension of MSEA, that is, KMSEA, to tackle the linearly inseparable problem in the original feature space.

The rest of this paper is organized as follows. In Section 2, we briefly review the related work. In Section 3, we describe the proposed MSEA approach and its kernelized extension KMSEA. Experimental results and analysis are provided in Section 4, and conclusion is drawn in Section 5.

2 Related Work

In this section, we briefly review the related methods, which are generally divided into following two types.

Shared subspace learning based methods mainly try to learn a common shared subspace for multiple views. Discriminant analysis of canonical correlations (DCC) [4] maximizes the within-class correlation and minimizes the between-class correlation for two sets of variables. Multiple discriminant CCA (MDCCA) [5] was designed for multiple views in the comparison with DCC. Kan et al. [6] presented a Multi-view discriminant analysis (MvDA) method, which maximizes between-class variations and minimizes within-class variations of the learning common space from both intra-view and inter-view. Intra-view and inter-view supervised correlation analysis (I^{2}SCA) [7] simultaneously extracts the discriminatingly correlated features from both inter-view and intra-view.

Transfer learning and dictionary learning based methods are mainly based on the transfer learning and dictionary learning techniques. Transfer component analysis (TCA) [8] attempts towards learning a few transfer components across domains by using maximum mean miscrepancy strategy. In the subspace spanned by these transfer components, data properties are preserved and data distributions in different domains are close to each other. Uncorrelated Multi-view Fisher Discrimination Dictionary Learning (UMDDL) [9] learns the multiple structural and discriminant dictionaries, which can well reconstruct the multi-view data.

3 Proposed Approach

3.1 Multi-view Sparse Embedding Analysis (MSEA)

Multi-view features can reveal the inherent property of data. Although these features come from different views, there exist some useful latent shared information, e.g., sparse structure, in the multi-view data [9]. How to effectively exploit this kind of latent sparse structure is vital for improving the performance of multi-view feature extraction. In this paper, we attempt towards incorporating the sparse embedding analysis into multi-view feature extraction. We learn a shared dictionary and multiple embedding matrices, which can make inherent sparse structure still be preserved in the projected multi-view features. The scheme of our MSEA is illustrated in Fig. 1.

The entire objective function of our MSEA contains three parts: sparse embedding analysis, intra-view discriminant correlation, and orthogonal constraints of embedding matrices. Then we describe these three parts in detail.

(1) Sparse Embedding Analysis. Let Y_i denote the i^{th} view of samples, and assume that they have been normalized, that is, $\hat{Y}_i\hat{Y}_i^T = 1, i = 1, 2, ..., N$, where N is the number of view. We try to learn multiple embedding matrices, with each corresponding to one view. The target is to project the original feature of multi-view samples into a shared subspace and help learn the shared dictionary. Then, this part of the objective function is defined as follows:

$$\langle D, W, X\rangle \arg \min_{D,W,X} \sum_{i=1}^{N} \left\| W_i\hat{Y}_i - DX \right\|_F^2, \quad (1)$$
$$s.t. \ \|x_j\|_0 \leq T_0, \ \forall j$$

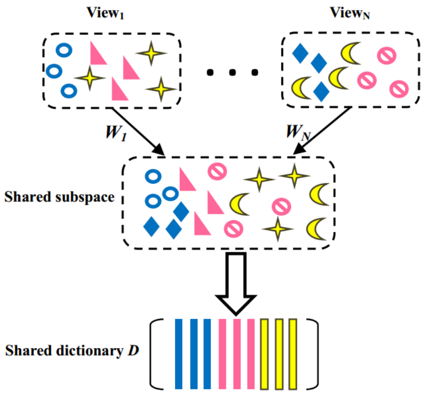

Fig. 1. Illustration of the Scheme of MSEA.

where W_i is the i^{th} embedding corresponding to the i^{th} view. The dimensionality of W_i is $p \times n$, where n is the dimensionality of original samples feature and p is the dimensionality of feature after embedding. D is the shared dictionary, and X is the sparse representation coefficients.

(2) Intra-view Discriminant Correlation Maximization. Recently, some study shows that the correlation information in views is significant in the feature extraction [7],[9]. To make the extracted features hold favorable discriminability, our MSEA tries to incorporate the intra-view discriminant correlation into the objective function. This target is to maximize the within-class correlation and minimize the between-class correlation from intra-view, simultaneously, that is,

$$\langle W \rangle = \arg\max_{W_i} \sum_{i=1}^{N} \left(C^i \right). \tag{2}$$

The discriminant correlation mentioned above can be defined as $C^i = C_w^i - \beta C_b^i$, where C_w^i is the intra-view within-class correlation and C_b^i is the intra-view between-class correlation of the i^{th} view. $\beta > 0$ is a tunable parameter that indicates the relative significance of C_w^i versus C_b^i. Specifically, C_w^i and C_b^i are defined as

$$C_w^i = \frac{\left[1/\sum\limits_{p=1}^{c} n_p^2 \right] \sum\limits_{p=1}^{c} \sum\limits_{r=1}^{n_p} \sum\limits_{t=1}^{n_p} \hat{y}_{pr}^{i}{}^{T} W_i^T W_i \hat{y}_{pt}^i}{\sqrt{\frac{1}{n} \sum\limits_{p=1}^{c} \sum\limits_{r=1}^{n_p} \left(y_{pr}^i - \bar{y}^i \right)^T W_i^T \left(y_{pr}^i - \bar{y}^i \right) W_i} \sqrt{\frac{1}{n} \sum\limits_{p=1}^{c} \sum\limits_{t=1}^{n_p} \left(y_{pt}^i - \bar{y}^i \right)^T W_i^T \left(y_{pt}^i - \bar{y}^i \right) W_i}},$$

$$= \frac{n \bullet tr \left\{ \sum\limits_{p=1}^{c} \sum\limits_{r=1}^{n_p} \sum\limits_{t=1}^{n_p} W_i \hat{y}_{pt}^i \hat{y}_{pr}^{iT} W_i^T \right\}}{\left(\sum\limits_{p=1}^{c} n_p^2 \right) \sqrt{\hat{Y}_i^T W_i^T W_i \hat{Y}_i} \sqrt{\hat{Y}_i^T W_i^T W_i \hat{Y}_i}} = \frac{n \bullet tr \left\{ W_i \hat{Y}_i A \hat{Y}_i^T W_i^T \right\}}{\left(\sum\limits_{p=1}^{c} n_p^2 \right) \hat{Y}_i^T W_i^T W_i \hat{Y}_i}$$

$$C_b^i = \frac{\left[1/\left(n^2 - \sum\limits_{p=1}^{c} n_p^2\right)\right] \sum\limits_{p=1}^{c} \sum\limits_{\substack{q=1 \\ q \neq p}}^{c} \sum\limits_{r=1}^{n_p} \sum\limits_{t=1}^{n_q} {\hat{y}_{pr}^i}^T W_i^T W_i \hat{y}_{qt}^i}{\sqrt{\frac{1}{n} \sum\limits_{p=1}^{c} \sum\limits_{r=1}^{n_p} \left(y_{pr}^i - \bar{y}^i\right)^T W_i^T \left(y_{pr}^i - \bar{y}^i\right) W_i} \sqrt{\frac{1}{n} \sum\limits_{q=1}^{c} \sum\limits_{t=1}^{n_p} \left(y_{qt}^i - \bar{y}^i\right)^T W_i^T \left(y_{qt}^i - \bar{y}^i\right) W_i}},$$

$$= \frac{n \bullet tr\left\{\sum\limits_{p=1}^{c} \sum\limits_{r=1}^{n_p} \sum\limits_{t=1}^{n_p} W_i \hat{y}_{pt}^i \hat{y}_{pr}^{iT} W_i^T\right\}}{\left(n^2 - \sum\limits_{p=1}^{c} n_p^2\right) \sqrt{\hat{Y}_i^T W_i^T W_i \hat{Y}_i} \sqrt{\hat{Y}_i^T W_i^T W_i \hat{Y}_i}} = -\frac{n \bullet tr\left\{W_i \hat{Y}_i A \hat{Y}_i^T W_i^T\right\}}{\left(n^2 - \sum\limits_{p=1}^{c} n_p^2\right) \hat{Y}_i^T W_i^T W_i \hat{Y}_i}$$

where $A = diag\left(E_{n_1}, E_{n_2}, ..., E_{n_c}\right)$ denotes a $n \times n$ symmetric, positive semi-definite, blocked diagonal matrix. E_{n_k} is a $n_k \times n_k$ matrix with all elements equalling to 1. Since A is a positive semi-definite matrix, we let $A = HH^T$ and obtain a more brief representation of this part in objective function:

$$\langle W \rangle = \arg\min_{\tilde{W}} \gamma \left\| \tilde{W} \tilde{Y} H \right\|_F^2, \tag{3}$$

where $\gamma = \left(\frac{n}{\sum\limits_{p=1}^{c} n_p^2} + \frac{n\beta}{n^2 - \sum\limits_{p=1}^{c} n_p^2}\right)$, $\tilde{W} = [W_1, ..., W_N]$, $\tilde{Y} = \begin{pmatrix} \hat{Y}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \hat{Y}_N \end{pmatrix}$.

(3) Orthogonal Constraints of Embedding Matrices. Although multi-view data reveal different characteristics of data, there also exist some redundant information among those views describing the same object. Therefore, to remove this kind of redundant information, we add the orthogonal constraints of above learned embedding matrices, that is,

$$W_i {W_i}^T = I, i = 1, 2, ..., N. \tag{4}$$

By combining the Formula (1), (3) and (4), the entire objective function of our MSEA is defined as follows:

$$\begin{aligned} &\langle D, W, X \rangle = \arg \min_{D, \tilde{W}, X} \left\| \tilde{W} \tilde{Y} - DX \right\|_F^2 - \gamma \left\| \tilde{W} \tilde{Y} H \right\|_F^2 . \\ &s.t. W_i {W_i}^T = I, i = 1, 2, ..., N, and \left\| x_j \right\|_0 \leq T_0, \forall j \end{aligned} \tag{5}$$

3.2 The Optimization of MSEA

There is no theoretical guarantee that our objective function in Formula (5) is jointly convex to $(D,\ W,\ X)$. However, it is convex with respect to each of $D,\ W,\ X$ when the others are fixed. Hence, this objective function can be solved based on the idea of divide-and-conquer. Before we conduct the iterative solution for MSEA, we try to simplify the optimization problem in Formula (5) by using a optimization trick in the literature [15]. We introduce two matrices, $Q \in n \times p$

and $B \in n \times p$, where n is the size of original samples feature and p is the size of feature after embedding. Then, the embedding matrices W_i can be represented by $W_i = {Q_i}^T {Y_i}^T$ and the shared dictionary D can be represented by $D = \tilde{W}\tilde{Y}B$. With this optimization trick and after some manipulations, the original optimization problem in Formula (5) can be simplified as follows:

$$\begin{aligned} &\arg\min_{\tilde{Q},X,B} \left\| \tilde{Q}^T K - \tilde{Q}^T KBX \right\|_F^2 - \gamma \left\| \tilde{Q}^T KH \right\|_F^2, \\ &s.t. {Q_i}^T K_i Q_i = I, i = 1, 2, ..., N, \left\| x_j \right\|_0 \leq T_0, \forall j \end{aligned} \tag{6}$$

where $K_i = \hat{Y}_i^T \hat{Y}_i$, $\tilde{Q} = [Q_1, ..., Q_N]$, and $K = \begin{pmatrix} K_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & K_N \end{pmatrix}$. The above Formula also can be tackled by the divide-and-conquer strategy. We divide the objective function in Formula (6) into three sub-problems:

(1) Updating X***.*** We update the sparse representation coefficients X by fixing the matrix B and $\tilde{Q}$. The objective function can be simplified as follows:

$$\begin{aligned} &\langle X \rangle = \arg\min_{X} \left\| \tilde{Q}^T K - \tilde{Q}^T KBX \right\|_F^2. \\ &s.t. \left\| x_j \right\|_0 \leq T_0, \forall j \end{aligned} \tag{7}$$

This is a typical sparse representation problem, which has been effectively solved by method of optimal directions (MOD) [10]. We directly utilize the MOD algorithm to update X.

(2) Updating B***.*** We update the matrix B by fixing X and $\tilde{Q}$. The objective function can be simplified as follows:

$$\langle B \rangle = \arg\min_{B} \left\| \tilde{Q}^T K - \tilde{Q}^T KBX \right\|_F^2. \tag{8}$$

We let $L(B) = \left\| \tilde{Q}^T K - \tilde{Q}^T KBX \right\|_F^2$, and take the derivative of $L(B)$ with respect to B. By setting the derivative result being equal to zero, we obtain:

$$B = X^T \left(X X^T \right)^{-1}. \tag{9}$$

(3) Updating $\tilde{Q}$***.*** We update the matrix $\tilde{Q}$ by fixing X and B. First, we conduct singular value decomposition (SVD) on K, and further let $U = S^{\frac{1}{2}} V^T ((I - BX)(I - BX)^T - A) V S^{\frac{1}{2}}$, $G_i = S^{\frac{1}{2}} V^T Q_i$. Then the objective function with respect to matrix $\tilde{Q}$ can be reformulated as follows:

$$\begin{aligned} &\langle G_i \rangle = \arg\min_{G} \; tr\left({G_i}^T U G_i \right) \\ &s.t. \;\; {G_i}^T G_i = I \end{aligned} . \tag{10}$$

The matrices G_i can be solved by using the generalized eigen-value decomposition. The eigen-vectors of U are made up of the matrices G_i. Once we obtain G_i, we can calculate the matrices Q based on $G_i = S^{\frac{1}{2}} V^T Q_i$.

The entire optimization of our approach is summarized in Algorithm 1.

Algorithm 1. MSEA

Step 1: Randomly initialize the $\tilde{Q}$, B and X;
Step 2: while $j < m$ (max iteration number) do:
  2.1 Updating the matrix X in Formula (7)with MOD algorithm;
  2.2 Updating the matrix B , by using the Formula (9);
  2.3 Updating the matrix $\tilde{Q}$, by using $Q_i = VS^{-\frac{1}{2}} G_i$,
  where G_i is the eigen-vectors of U in Formula (10);
Step 3: Output the sparse representation coefficients X, the embedding matrices $W_i = {Q_i}^T {Y_i}^T$ and the shared dictionary $D = \tilde{W}\tilde{Y}B$.

3.3 Classification Strategy of MSEA

Given the learned D, W_i, and X, we design an effective classification strategy for MSEA. Specifically, we first project the testing samples into a novel feature space by using the multi-view embedding matrices W_i. Then, we employ the shared dictionary D to represent the projected features of samples, that is,

$$x = \arg\min_x \left\{ \|y - Dx\|_2^2 + \gamma \|x\|_1 \right\},$$

where x is the sparse representation coefficients. We classify the testing samples according to identity $(y) = \arg\min_i \{e_i\}$, where $e_i = \|y - D_i \alpha_i\|_2$ is the representation error of each class, and $\alpha_i = [\alpha_1, \alpha_2, \cdots, \alpha_c]^T$ is the sparse representation coefficients of each class. We classify the testing samples into the class with the smallest reconstruction error.

3.4 Kernelized MSEA

To tackle the linearly inseparable problem in the original feature space, we extend a kernelized extension of MSEA called KMSEA by using kernel trick. Kernel trick has shown its effectiveness in some methods [11, 12]. We first perform the kernel mapping for samples and then realize the MSEA in the mapped space.

Assume that $\phi : R^d \to F$ denotes a nonlinear mapping from the low-dimensional feature space to high-dimensional feature space. Then the mapping process from the sample set Y to space F can be represented as $Y \to \phi(Y)$. The objective function of KMSEA is defined as

$$\langle D, W, X \rangle = \arg\min_{D,\tilde{W},X} \left\| \tilde{W}\phi\left(\tilde{Y}\right) - DX \right\|_F^2 - \gamma \left\| \tilde{W}\phi\left(\tilde{Y}\right) H \right\|_F^2. \quad (11)$$
$$s.t. W_i {W_i}^T = I, i = 1, 2, ..., N, and \|x_j\|_0 \le T_0, \forall j$$

The optimization of KMSEA is similar to that of MSEA. We similarly introduce two matrices, $Q \in n \times p$ and $B \in n \times p$, and then the embedding matrices W_i and shared dictionary D can be represented by $W_i = {Q_i}^T \phi\left({Y_i}^T\right)$ and $D = \tilde{W} \phi\left(\tilde{Y}\right) B$, respectively. Substituting W_i and D into the Formula (11), we employ the kernel trick and then the objective function of KMSEA can be reformulated as

$$\begin{aligned} &\arg \min_{D, \tilde{W}, X} \left\| \tilde{Q}^T \tilde{K} - \tilde{Q}^T \tilde{K} B X \right\|_F^2 - \gamma \left\| \tilde{Q}^T \tilde{K} H \right\|_F^2 , \\ &s.t. {Q_i}^T \tilde{K}_i Q_i = I, i = 1, 2, ..., N, and \|x_j\|_0 \leq T_0, \ \forall j \end{aligned} \tag{12}$$

where $\tilde{K}_i = \phi\left(\tilde{Y}_i\right)^T \phi\left(\tilde{Y}_i\right)$ is the RBF kernel trick. $\tilde{K} = \begin{pmatrix} \tilde{K}_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \tilde{K}_N \end{pmatrix}$, $\tilde{Q} = [Q_1, ..., Q_N]$. Similar to the MSEA, the KMSEA also can be solved by using the divide-and-conquer strategy. Its optimization process is similar to Algorithm 1.

4 Experiments

In this section, we evaluate our two approaches MSEA and KMSEA. We choose three state-of-the-art multi-view feature extraction methods, including the**TCA**[8], **UMDDL**[9], and **I^{2}SCA** [7], as the compared methods. We validate the effectiveness of our approaches through two aspects: the mean recognition rate and the sample distribution figure.

The experiments are conducted on two widely-used multi-view datasets. Multiple feature dataset (MFD) [13] contains 10 classes of handwritten numerals. These digit characters are represented in terms of six views of feature sets. In the experiment, we randomly choose 100 samples per class as the training set and the remaining 100 samples as the testing set. Multi-PIE dataset [14] contains various views, illumination and expressions variations. We choose its subset containing 1632 samples from 68 classes in 5 poses (C05, C07, C09, C27, C29). We randomly select 5 samples per class as the training samples and the remaining as the testing set.

Table 1 shows the average recognition rates and the standard deviation of 20 random runs for all methods on MFD and Multi-PIE datasets. We can observe that both MSEA and KMSEA outperform the compared methods on two datasets. Moreover, KMSEA obtains better performance than MSEA.

Table 1. Average recognition rate (± standard deviation) on two datasets.

Datasets	TCA	UMMDL	I^{2}SCA	MSEA	KMSEA
MFD	91.87±3.67	92.07±4.67	92.11 ±3.97	92.22± 4.23	**92.84±4.93**
Multi-PIE	91.87±4.56	92.53±4.02	92.91±3.46	93.51±3.38	**93.92±3.27**

In order to analyze the separabilities of all methods, we provide the distribution of samples with two principal features extracted from 5 different views by using all related methods on Multi-PIE dataset. Here, we employ the PCA transform to obtain two principal features. Note that since UMDDL is a dictionary learning method, not a feature extraction method, we cannot provide its sample distribution figure.

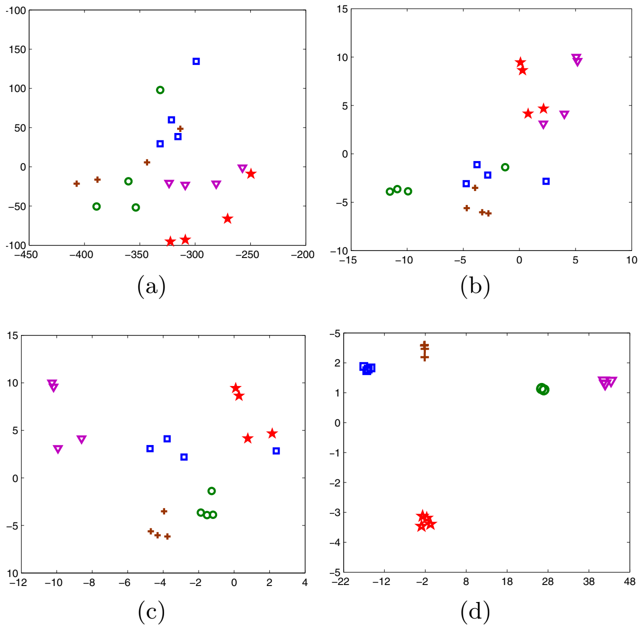

Fig. 2. Sample distributions of methods on Multi-PIE dataset of 20 samples in the feature space. (a): TCA; (b): I^2SCA; (c): MSEA; (d): KMSEA.

Fig. 2 shows the distribution of two principal features of 20 samples (from 5 different persons and 4 samples per person) extracted on the Multi-PIE dataset. The markers with different shapes and colors stand for 5 different persons. It shows that the proposed approaches achieve preferable separabilities in comparison with other methods. As for the MFD dataset, we obtain the similar results. Due to the limited space, we don't provide the results in detail here.

5 Conclusion

In this paper, we propose a novel multi-view feature extraction approach named MSEA. It not only can preserve the sparse reconstruction information, but also can consider the discriminative correlation in multi-view data. To remove the redundancy among views, we add orthogonal constraints of embedding matrices.

Furthermore, we provide a kernelized extension KMSEA to tackle the linearly inseparable problem. Experiments demonstrate that the MSEA and KMSEA outperform several state-of-the-art related methods with respect to the recognition rate and separabilities on sample distribution figures.

References

1. Guo, Y. H.: Convex subspace representation learning from multi-view data. In: AAAI Conference on Artificial Intelligence, pp. 387–393 (2013)
2. Long, B., Yu, P.S., Zhang, Z.M.: A general model for multiple view unsupervised learning. In: Proceedings of the SIAM International Conference on Data Mining, pp. 822–833 (2008)
3. Hardoon, D.R., Szedmak, S., Shawe-Taylor, J.: Canonical Correlation analysis: An Overview with Application to Learning Methods. Neural Computation **16**(12), 2639–2664 (2004)
4. Kim, T.K., Kittler, J., Cipolla, R.: Discriminative Learning and Recognition of Image Set Classes Using Canonical Correlations. IEEE Transactions on Pattern Analysis and Machine Intelligence **29**(6), 1005–1018 (2007)
5. Gao, L., Qi, L., Chen, E.Q., Guan, L.: Discriminative multiple canonical correlation analysis for multi-feature information fusion. In: IEEE International Symposium on Multimedia, pp. 36–43 (2012)
6. Kan, M., Shan, S., Zhang, H., Lao, S., Chen, X.: Multi-view discriminant analysis. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part I. LNCS, vol. 7572, pp. 808–821. Springer, Heidelberg (2012)
7. Jing, X.Y., Hu, R.M., Zhu, Y.P., Wu, S.S., Liang, C., Yang, J.Y.: Intra-view and inter-view supervised correlation analysis for multi-view feature learning. In: AAAI Conference on Artificial Intelligence, pp. 589–602 (2014)
8. Pan, S.J., Tsang, I.W., Kwok, J.T., Yang, Q.: Domain Adaptation via Transfer Component Analysis. IEEE Transactions on Neural Networks **22**(2), 199–210 (2011)
9. Jing, X.Y., Hu, R.M., Wu, F., Liang, C., Yang, J.Y.: Uncorrelated multi-view Fisher discrimination dictionary learning for recognition. In: AAAI Conference on Artificial Intelligence, pp. 470–474 (2014)
10. Engan, K., Aase, S.O., Husoy, J.H.: Method of Optimal Directions for Frame Design. IEEE International Conference on Acoustics, Speech and Signal Process **5**(1), 2443–2446 (1999)
11. Sun, T.K., Chen, S.C., Jin, Z., Yang, J.Y.: Kernelized discriminative canonical correlation analysis. In: International Conference on Wavelet Analysis and Pattern Recognition, pp. 1283–1287 (2007)
12. Bach, F.R., Jordan, M.I.: Kernel Independent Component Analysis. Journal of Machine Learning Research **3**, 1–48 (2002)
13. Yuan, Y.H., Sun, Q.S., Zhou, Q., Xia, D.S.: A Novel Multiset Integrated Canonical Correlation Analysis Framework and Its Application in Feature Fusion. Pattern Recognition **44**(5), 1031–1040 (2010)
14. Cai, D., He, X., Han, J., Zhang, H.J.: Orthogonal Laplacianfaces for Face Recognition. IEEE Transactions on Image Processing **15**(11), 3608–3614 (2006)
15. Nguyen, H.V., Patel, V.M., Nasrabadi, N.M., Chellappa, R.: Sparse embedding: a framework for sparsity promoting dimensionality reduction. In: Fitzgibbon, A., Lazebnik, S., Perona, P., Sato, Y., Schmid, C. (eds.) ECCV 2012, Part VI. LNCS, vol. 7577, pp. 414–427. Springer, Heidelberg (2012)