

Robust Multi-view Common Component Learning

Jiamiao Xu, Xinge You*, Shi Yin, Peng Zhang, and Wei Yuan

*School of Electronics and Information Engineering,
Huazhong University of Science and Technology, Wuhan, China
youxg@mail.hust.edu.cn

Abstract. In many computer vision applications, one object usually exists more than one data representation. This paper focuses on the specific problem of cross-view recognition, which aims to recognize objects from different views. A majority of representative works mainly attempt to seek a common subspace, in which the Euclidean distance of within-class data is short. Intuitively, the recognition performance will be better if the various data from the same object have completely same representation in the common space. Therefore, this paper proposes robust multi-view common component learning (RMCCL) algorithm, which learns multiple linear transforms to extract the common component of multi-view data from the same instance. To enhance the discriminant ability and robustness of subspace, we introduce binary label matrix technology and serve Cauchy loss as our error measurement. RMCCL can be decomposed into two subproblems by Alternating optimization method, and each subproblem can be optimized by Iteratively Reweight Residuals (IRR) technique. Extensive experiments in both two-view and multi-view datasets demonstrate that the our method outperforms other state-of-the-art approaches.

Keywords: Cross-view recognition, multi-view learning, common space.

1 Introduction

In many computer vision applications, one object can be captured by different sensors or observed at diverse viewpoints, e.g., [15]. Thus, one object often exists multiple data representation, usually denoted as multi-view data. Recently, cross-view recognition, which aims to recognize the samples from completely heterogeneous views, becomes more and more significant. However, due to big gap possibly existing between views, the recognition performance will be poor by directly recognizing samples from diverse views.

* J. Xu, X. You, S. Yin, P. Zhang and W. Yuan are with the School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan 430074, China.(Corresponding Author: Xinge You.)

In order to better handle the cross-view recognition problems, many excellent and effective works have been done. Most of approaches can be attributed to metric learning methods or subspace learning approaches. Owing to the large difference between views, Euclidean distance metric are not applicable. To handle this problem, metric learning methods aim to learn a kind of new similarity measurement manners. Nevertheless, subspace learning try to find a common space shared by multiple views, in which Euclidean distance metric can work. Certainly, there are also many works can be utilized to handle the cross-view recognition problems, such as transfer learning [6, 30]. Because our proposed method belongs to subspace learning, we will introduce this kind of methods in detail. Regarding subspace learning, according to the number of views, this kind of approaches can be grouped into two-view and multi-view approaches.

1.1 Two-view Subspace Learning Approaches

The canonical correlation analysis (CCA) [9] is proposed in 1936, and is the earliest method of cross-view subspace learning. CCA attempts to find two linear transformations, one for each view, such that the correlation coefficient between two views is maximized. To handle information dissipation caused by CCA, Chen et al. [2] propose continuum regression (CR) model. Experiments on cross-modal retrieval demonstrates that CR model can better address the cross-view problems. Heterogeneous face recognition is a specific but important subproblem in the cross-view recognition problems. To better address this subproblem, in [19, 21], PLS is employed to conduct effective feature selection. CCA is only a unsupervised subspace learning approach. To take full advantage of label information, multi-view fisher discriminant analysis (MFDA) [3, 4] effectively utilizes labels information to learn informative projections. Experimental results show that MFDA works better than previous unsupervised methods. For heterogeneous face recognition, Conventional approaches, which bring in a middle conversion stage, are easy to cause performance degradation. In order to better handle this problem, Lin and Tang [17] propose common discriminant feature extraction (CDFE) to simultaneously learn two transformation, which transform two-view samples to common subspace. Certainly, in addition to previously mentioned approaches, there are some excellent works [1, 16, 23, 25].

1.2 Multi-view Subspace Learning Approaches

No doubt that we can use two-view subspace learning approaches to solve multi-view recognition problem by one-versus-one strategy. To efficiently handle this problem, we tend to design multi-view subspace learning methods, which can transform multiple view samples into a latent subspace, simultaneously. In [18, 20], Multi-view Canonical Correlation Analysis (MCCA) is designed as a multiple versions of the CCA. In addition, Sim et al. [22] present Multimodal Discriminant Analysis(MMDA) to decompose samples into independent modes. Recently, In [11], kan et al. propose a novel method named multi-view discriminant

analysis (MvDA), which tries to project the multi-view data into a common space, in which within-class samples are close to each other and between-class samples are far away. Kan et al. believe that there exists some relations between different transformations. To exploit these relations, Kan et al. [12] develop view-consistency version of MvDA in 2016, which is called MvDA-VC. Besides, low-rank representation (LRR) is good at digging relations between views. Therefore, recently, LRR is also utilized to handle the cross-view problems, such as SRRS [14], LRCS [5] and LRDE [13].

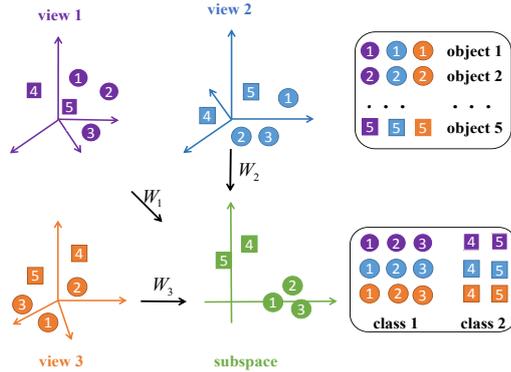


Fig. 1. Illustration of our idea. There are five objects in the Figure, and each object has three views. Each number denotes one object, and each color stands for one view. The first three objects from circle class and the last two from rectangle class. As we have seen, three view data are mapped to a unified space by W_1, W_2, W_3 projection matrices, in which various data from one object have unique representation. Furthermore, circle class is far from rectangle class and within-class objects are close to each other.

As mentioned above, one object can generate multiple view data. However, existing subspace learning approaches can only minimize the difference of within-class samples, i.e., the distance of data from the same instance is close in the unified subspace. Intuitively, we can get a better performance for cross-view recognition problem if the heterogeneous data from one object have completely same representation in the latent common subspace. For this reason, this paper learns multiple linear projection matrices, one for each view, to extract the common component of the heterogeneous data. In this paper, to strengthen the robustness to noise, we employ Cauchy loss [27] as the error measurement in our algorithm. Through the above two thoughts, we still can't learn a discriminative common space. Therefore, we introduce a non-negative label matrix [28] to enlarge the margins between different classes and diminish the discrepancy of samples in the same class. Finally, the proposed method is applied to several cross-view recognition tasks. Experimental results on two heterogeneous face databases demonstrate that the proposed method outperforms previous cross-view recognitions approaches. The fundamental idea of our algorithm is illustrated in Fig. 1.

2 Robust Multi-view Common Component Learning

In this section, we first present the problem formulation, and then bring in label matrix and Cauchy loss to enhance discriminative ability and robustness of algorithm.

2.1 Problem Formulation

Let $\mathcal{X} = \{x_i^1, x_i^2, \dots, x_i^n\}_{i=1}^m$ stands for a dataset containing m objects, and each object exists n samples, usually denoted as n views in this area. $x_i^v \in \mathcal{R}^{d_v}$ denotes the v th view of the i th object embedded in d_v dimension space. Generally, there are a large gap between different views, i.e., the between-class samples in a view might have higher similarity than within-class samples from different views. That will lead to poor performance. To better solve this problem, we expect that multiple samples from one object have completely uniform feature representation. Certainly, we can also consider that we extract component features in the process. We denote the linear transformations as $W = \{W_v\}_{v=1}^n$, where $W_v \in \mathcal{R}^{d_i \times d}$ is the transformation matrix of i th view, and projection results are $\mathcal{Z} = \{z_i\}_{i=1}^m$, where $z_i \in \mathcal{R}^d$ is the mapping result of i th object embedded in d dimension common space.

It is obvious that our objective is to learn a series of view projection matrices $\{W_v\}_{v=1}^n$ to extract common component z . A straightforward method is minimizing the empirical risk $z_i - W_v^T x_i^v$ in the whole dataset:

$$\min_W \frac{1}{mn} \sum_{i=1}^m \sum_{v=1}^n \|z_i - W_v^T x_i^v\|_2^2 + c_1 \sum_{v=1}^n \|W_v\|_F^2 + c_2 \sum_{i=1}^m \|z_i\|_2^2 \quad (1)$$

As can be seen in Eq. (1), least square error are used to extract common features of multiple view data, and regularization items are employed to penalize view projection matrices W and generating samples z_i , where $c_1 > 0$, $c_2 > 0$ are the balance parameters, which are presented to prevent overfitting.

No doubt that Eq. (1) is able to extract the common component of multiple views from one object. However, the constraint of Eq. (1) is too weak, so that it is not enough to obtain a discriminant subspace. To improve the performance of algorithm, like most previous works, we expect to unite within-class samples and separate between-class samples in the common space. Certainly, it is usual that manifold learning methods [7, 8] can be use to improve the discriminant ability of algorithms. However, we adopt a simple but effective methods in this paper. we introduce binary label matrix to enhance the discriminant ability of our algorithm.

2.2 Least square regression for multi-class classification

Given a dataset $\mathcal{X} = \{(x_i, y_i)\}_{i=1}^m$, where x_i is the i th sample of dataset \mathcal{X} , and each input feature x_i corresponds to a target vector y_i . As usual, the regression

problem can be expressed as

$$\min_{W,b} \sum_{i=1}^m \|W^T x_i + b - y_i\|_2^2 + \frac{\lambda}{2} \|W\|_F^2 \quad (2)$$

where $W \in \mathcal{R}^{d \times p}$ is weight matrix, $b \in \mathcal{R}^p$ is the bias vector and λ is the regularized parameter, which is used to prevent overfitting. In the regression problem, the target y_i in Eq. (2) is continuous. Certainly, by setting target to 0 and 1, Eq. (2) can be extended to two-class classification. Regarding multi-class classification, we can adopt one-versus-one policy to generalize Eq. (2). However, we will have to address multiple subproblems generated by this policy.

In order to use regression model to better handle multi-class classification problem, we introduce label matrix in this paper. Supposing \mathcal{X} is from c classes, and x_i is from class j . Then, for sample x_i , its target vector is $y_i = [0, \dots, 0, 1, 0, \dots, 0]^T$ with only the j th element equal to one, where $y_i \in \mathcal{R}^c$. The target vector constructed by this rule is called as binary label matrix. Then, Eq.(2) can be generalized to multi-class classification easily.

As a matter of fact, numerous works utilize linear regression model to address multi-class classification problem by bringing in binary label matrix [28]. In this paper, we utilize it to restraint common space z learning in Eq. (1). Thus, Eq. (1) can be rewritten as

$$\min_W \frac{1}{mn} \sum_{i=1}^m \sum_{v=1}^n \|z_i - W_v^T x_i^v\|_2^2 + c_1 \sum_{v=1}^n \|W_v\|_F^2 + c_2 \sum_{i=1}^m \|z_i - y_i\|_2^2 \quad (3)$$

where y_i is the binary label vector of x_i . Supposed that the training set is from c classes, we can regard y_i as a point embedded in c dimension space. In this space, within-class samples correspond to the same point and between-class samples are forced to be separated. Therefore, we can obtain a discriminant space by using label matrix. We might as well call Eq. (3) as MCCL-L2 to distinguish what we will propose below.

2.3 Cauchy loss

Least square loss is generally used as error measurement, as we have done in Eq. (3). However, thoroughly theory studies [10] show that least square estimator is not robust to noise, thus the performance of cross-view recognition will be seriously degraded when the data is not clean. Xu et al. [27] indicates that Cauchy loss is more robust than L_2 loss. In order to strength the robustness of algorithm, we employ Cauchy loss to instead of L_2 loss, then the objective function can be rewritten as

$$\min_W \frac{1}{mn} \sum_{i=1}^m \sum_{v=1}^n \log \left(1 + \frac{\|z_i - W_v^T x_i^v\|_2^2}{c^2} \right) + c_1 \sum_{v=1}^n \|W_v\|_F^2 + c_2 \sum_{i=1}^m \|z_i - y_i\|_2^2 \quad (4)$$

where c is a constant. The Eq. (4) jointly models the relationships between each view space \mathcal{X}^v and the common subspace Z by a robust algorithm. It can be expected that we can acquire multiple linear projections by solving this problem with an input of multiple view data. Eq. (4) is our method in this paper, named Robust Multi-view Common Component Learning.

3 Optimization

Through alternating optimization method, problem (4) can be decomposed into two subproblems. Inspired by generalized Weiszfeld's method [24], Xu et al. [27] develop Iteratively Reweight Residuals (IRR) technology. In this paper, we utilize IRR to efficiently optimize the subproblems, respectively.

Firstly, fix view projection matrices $\{W_v\}_{v=1}^n$ and data representation in common subspace $\{z_j\}_{j=1}^n, j \neq i$, then Eq. (4) is decomposed into subproblem over z_i , which can be written as

$$\min_{z_i} \mathcal{J} = \frac{1}{n} \sum_{v=1}^n \log \left(1 + \frac{\|z_i - W_v^T x_i^v\|_2^2}{c^2} \right) + c_2 \|z_i - y_i\|_2^2 \quad (5)$$

setting the gradient of \mathcal{J} with respect to z_i to zero, we have

$$\sum_{v=1}^n \frac{z_i - W_v^T x_i^v}{c^2 + \|z_i - W_v^T x_i^v\|_2^2} + nc_2 (z_i - y_i) = 0 \quad (6)$$

which can also be rewritten as

$$\left(\sum_{v=1}^n \frac{1}{c^2 + \|z_i - W_v^T x_i^v\|_2^2} + nc_2 \right) z_i = \left(\sum_{v=1}^n \frac{W_v^T x_i^v}{c^2 + \|z_i - W_v^T x_i^v\|_2^2} + nc_2 y_i \right) \quad (7)$$

where $r^v = z_i - W_v^T x_i^v$ is referred to as the residual. A weight function is then defined as

$$Q = \left[\frac{1}{c^2 + \|r^1\|_2^2}, \dots, \frac{1}{c^2 + \|r^n\|_2^2} \right] \quad (8)$$

Then, put together Eq. (7) and (8)

$$z_i = \left(\left(\sum_{v=1}^n Q_v \right) + nc_2 \right)^{-1} \left(\left(\sum_{v=1}^n W_v^T x_i^v \right) + nc_2 y_i \right) \quad (9)$$

It is obvious that Q is the function of z_i , we thus utilize IRR technology to iteratively update z_i using Eq. (9) with an initial value until convergence. The procedure is described in Algorithm 1.

By fixing all common subspace data points $\{z_i\}_{i=1}^m$ and view projection matrix $\{W_k\}_{k=1}^n, k \neq v$, Eq. (4) is reduced to the following subproblem over projection matrix W_v :

Algorithm 1 solving Eq. (9) by IRR Method

Input: The view data of object i : $x_i^1, x_i^2, \dots, x_i^n$; projection matrices: $\{W_v\}_{v=1}^n$; the binary label vector of object i : y_i ; and the initial value: z_i^0

Output: z_i

- 1: **for** $k = 1, \dots, itermax$ **do**
 - 2: 1. calculate residuals $\{r^v\}_{v=1}^n$ using z_i^{k-1} ;
 - 3: 2. update the weight function Q through Eq. (11);
 - 4: 3. using Eq.(12) to update z_i^k ;
 - 5: **if** the estimates of z_i converge **then**
 - 6: break;
 - 7: **end if**;
 - 8: **end for**;
 - 9: **Return:** z_i
-

$$\min_{W_v} \mathcal{J} = \frac{1}{m} \sum_{i=1}^m \log \left(1 + \frac{\|z_i - W_v^T x_i^v\|_2^2}{c^2} \right) + c_1 \|W_v\|_F^2 \quad (10)$$

Setting the derivation of \mathcal{J} with respect to W_v to zero. Finally, the projection matrix W_v can be updated by

$$W_v = \left(\left(\sum_{i=1}^m x_i^v Q_i (x_i^v)^T \right) + m c_1 \right)^{-1} \sum_{i=1}^m x_i^v Q_i z_i^T \quad (11)$$

where Q_i is the i th element of Q , and Q can be written as

$$Q = \left[\frac{1}{c^2 + \|r_1\|_2^2}, \dots, \frac{1}{c^2 + \|r_m\|_2^2} \right] \quad (12)$$

where $r_i = z_i - W_v^T x_i^v$. Considering Q_i depends on W_v , we thus iteratively update W_v using Eq. (12). The iterative procedure is similar to Algorithm 1.

4 EXPERIMENTS

In this section, the proposed method RMCCCL is evaluated on two heterogeneous face databases, including the Heterogeneous Face Biometrics (HFB) database and the CMU Pose, Illumination, and Expression (PIE) database (CMU PIE). The representative methods CCA [9], MCCA [20], LRCS [5], MvDA [11] and MvDA-VC [12] and MCCL-L2 are brought in for comparison and the traditional single-view methods PCA and LDA serve as baselines. It is worth noting that MCCL-L2 algorithm is proposed by replacing Cauchy loss in Eq. (4) with L_2 loss. It can be regard as a middle algorithm of RMCCCL. We expect to prove Cauchy loss works better than L_2 loss.



Fig. 2. Selected subjects from the HFB dataset. The first row is visual light images, and the second row is near infrared images

4.1 Experimental Settings

One object can be captured by different sensors, such as visual light camera and near-infrared camera. There are 100 individuals on the HFB dataset, in which each subject has four visual light photos and four near infrared photos. We use 70 individuals to train the RMCCCL model, and the rest are utilized to evaluate the performance of our algorithm. Some selected faces from the HFB database can be seen in Fig. 2. Regarding the CMU PIE database, five poses, i.e., C11, C29, C27, C05 and C37, are chosen as the multiple view data, and each subject has four face images under every pose. Some selected faces under five poses can be seen in Fig. 3. 45 people are chosen to train the model, and the rest are employed to evaluate algorithm.

The HFB and the CMU PIE heterogeneous face datasets are cropped according standard protocol and are resized to 32×32 and 64×64 , respectively. It is worth noting that, in the HFB or the CMU PIE dataset, there are multiple schemes to divide datasets into training set and testing set. To make the results more convincing, all experiments are repeated ten times by randomly dividing these into two parts. The average results are regarded as the final accuracies (mACC). Furthermore, all methods employ principle component analysis to realize dimensionality reduction, and each approach sets dimension to get the best accuracy.

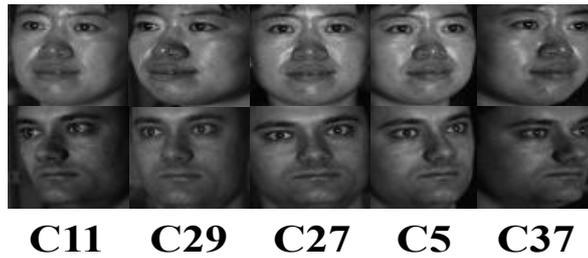


Fig. 3. Selected subjects from the CMU PIE dataset. C11, C29, C27, C05 and C37 poses are selected as the views.

4.2 Face Recognition between Visual Light and Near Infrared Images

In this experiment, we evaluate the performance of RMCCCL on the HFB, in which objects are captured by different sensors. The proposed methods RMCCCL and MCCL-L2 are compared to classic works including PCA, LDA, CCA [9], MCCA [20], LRCS [5], MvDA [11], MvDA-VC [12]. The experimental results are shown in Table 1.

As can be seen, MvDA-VC is the best competitive method with an high recognition accuracy of 55.88 percent. However, our proposed method RMCCCL achieves a highest recognition rate of 61.46%. Note that, because low rank representation methods usually select train set as the over-complete dictionary, LRCS fails to work when dataset is small. Therefore, it obtains a recognition rate of 33.21%. It is just a bit better than MCCA. What’s more, our method RMCCCL has better performance than MCCL-L2. Obviously, it accords with the theoretical analysis. Through the experimental results on the HFB, it shows that RMCCCL outperforms other subspace learning methods in two-view case.

Table 1. The Average Recognition Results (%) on the HFB Dataset

	PCA	LDA	CCA	LRCS	MvDA	MvDA-VC	MCCL-L2	RMCCCL
NIR-VIS	9.17	13.00	30.83	31.42	41.75	56.42	59.42	61.67
VIS-NIR	8.17	16.33	31.42	35.00	45.00	55.33	60.50	61.25
Average	8.67	14.67	31.13	33.21	43.38	55.88	59.96	61.46
Std	2.10	1.75	4.41	7.00	8.87	5.34	3.80	3.36

4.3 Face Recognition across poses

In the experiments, the CMU PIE dataset is utilized to evaluate face recognition across poses, and pair-wise manner is employed to conduct the experiment. Since each object owns five view data, it generates $5 \times 4 = 20$ recognition accuracies. Then, we utilize the mACC as final evaluation index. The Experimental results on the CMU PIE dataset can be seen in Table 2.

As can be seen, due to unite the within-class samples and separate between-class samples in the latent unified space, the MvDA outperforms PCA, LDA, MCCA on the CMU PIE database. As we have explained in the HFB dataset experiments, since the train set fails to compose a over-complete basic, LRCS performs even worse than MvDA. Through considering the relation of multiple view projections, MvDA-VC further improves performance by 11.7%. It is worth noting that, proposed method RMCCCL outperforms MvDA-VC with an absolute improvement by 6.1%. In addition, RMCCCL also performs better than MCCL-L2 with an improvement by 2.8%. It demonstrates that Cauchy loss has better robustness than L_2 loss. Experimental results show that our method RMCCCL performs better than several state-of-the-art algorithms on the CMU PIE

Table 2. The Average Recognition Results (%) on the CMU PIE Dataset

Gallery	Probe	PCA	LDA	MCCA	LRCS	MvDA	MvDA-VC	MCCL-L2	RMCCCL
C11	C29	67.1	20.2	48.0	72.3	69.3	79.7	79.1	81.0
	C27	29.6	8.3	26.5	33.2	46.7	58.3	61.3	64.2
	C05	13.9	6.1	22.3	19.5	37.0	48.2	51.2	58.0
	C37	12.7	5.1	20.5	14.7	37.4	49.6	46.3	51.6
C29	C11	69.2	19.8	44.2	69.8	67.7	80.4	77.4	79.5
	C27	42.3	14.6	43.2	51.6	57.6	77.0	78.5	80.0
	C05	21.5	5.8	33.6	29.0	47.8	62.7	69.9	73.4
	C37	17.6	6.0	24.9	22.0	42.9	55.8	54.6	59.1
C27	C11	26.5	9.2	29.6	33.0	47.7	57.0	62.2	63.2
	C29	40.1	15.7	46.2	51.2	59.2	75.0	81.6	82.3
	C05	39.9	11.5	41.6	48.7	58.8	69.5	77.2	82.2
	C37	26.6	8.8	27.4	35.1	44.0	57.8	58.2	63.7
C05	C11	11.2	5.9	23.2	19.8	37.4	48.2	52.4	54.6
	C29	20.4	6.1	28.8	25.4	47.1	61.7	69.6	71.5
	C27	34.8	13.4	35.0	46.3	60.4	70.0	75.1	78.2
	C37	65.7	29.0	48.8	73.7	65.7	74.8	78.2	81.5
C37	C11	12.2	5.4	17.2	15.1	37.8	47.9	48.3	51.1
	C29	20.0	5.9	19.6	22.1	40.1	52.4	53.0	59.0
	C27	24.1	8.2	22.9	35.1	47.8	56.2	57.9	61.8
	C37	72.5	26.3	45.3	76.7	68.5	74.1	83.0	82.6
Average		33.4	11.6	32.4	39.7	51.1	62.8	66.0	68.9
Std		2.7	1.7	2.3	3.0	2.1	2.7	1.7	1.5

dataset. Furthermore, it demonstrates that our approach is also suitable for multi-view data.

5 Conclusion

In this paper, we propose RMCCCL algorithm, which learns multiple linear transforms to extract the common component of multi-view data from the same instance. To enhance the discriminant ability and robustness of subspace, we introduce binary label matrix technology and serve Cauchy loss as our error measurement. Experimental results show that our approach outperforms several state-of-the-art algorithms. In future, we'll further analyze the robust and convergence of algorithm. Certainly, in order to better address various data, we can also bring kernel technology [26, 29] to develop the kernel version of RMCCCL.

6 Acknowledgment

This work was supported partially by National Key Technology Research and Development Program of the Ministry of Science and Technology of China (No. 2015BAK36B00), in part by the Key Science and Technology of Shen zhen(No.

CXZZ20150814155434903), in part by the Key Program for International S&T Cooperation Projects of China (No. 2016YFE0121200), in part by the National Natural Science Foundation of China (No. 61571205), in part by the National Natural Science Foundation of China (No. 61772220).

References

1. Chen, N., Zhu, J., Xing, E.P.: Predictive subspace learning for multi-view data: a large margin approach. In: *Advances in neural information processing systems*. pp. 361–369 (2010)
2. Chen, Y., Wang, L., Wang, W., Zhang, Z.: Continuum regression for cross-modal multimedia retrieval. In: *Image Processing (ICIP), 2012 19th IEEE International Conference on*. pp. 1949–1952. IEEE (2012)
3. Diethe, T., Hardoon, D., Shawe-Taylor, J.: Constructing nonlinear discriminants from multiple data views. *Machine learning and knowledge discovery in databases* pp. 328–343 (2010)
4. Diethe, T., Hardoon, D.R., Shawe-Taylor, J.: Multiview fisher discriminant analysis. In: *NIPS workshop on learning from multiple sources* (2008)
5. Ding, Z., Fu, Y.: Low-rank common subspace for multi-view learning. In: *Data Mining (ICDM), 2014 IEEE International Conference on*. pp. 110–119. IEEE (2014)
6. Gopalan, R., Li, R., Chellappa, R.: Domain adaptation for object recognition: An unsupervised approach. In: *Computer Vision (ICCV), 2011 IEEE International Conference on*. pp. 999–1006. IEEE (2011)
7. Gui, J., Jia, W., Zhu, L., Wang, S.L., Huang, D.S.: Locality preserving discriminant projections for face and palmprint recognition. *Neurocomputing* 73(13), 2696–2707 (2010)
8. Gui, J., Sun, Z., Jia, W., Hu, R., Lei, Y., Ji, S.: Discriminant sparse neighborhood preserving embedding for face recognition. *Pattern Recognition* 45(8), 2884–2893 (2012)
9. Hotelling, H.: Relations between two sets of variates. *Biometrika* 28(3/4), 321–377 (1936)
10. Huber, P.J.: *Robust statistics*. Springer (2011)
11. Kan, M., Shan, S., Zhang, H., Lao, S., Chen, X.: *Multi-view Discriminant Analysis*, pp. 808–821. Springer Berlin Heidelberg, Berlin, Heidelberg (2012)
12. Kan, M., Shan, S., Zhang, H., Lao, S., Chen, X.: Multi-view discriminant analysis. *IEEE transactions on pattern analysis and machine intelligence* 38(1), 188–194 (2016)
13. Li, J., Wu, Y., Zhao, J., Lu, K.: Low-rank discriminant embedding for multiview learning. *IEEE transactions on cybernetics* (2016)
14. Li, S., Fu, Y.: Robust subspace discovery through supervised low-rank constraints. In: *Proceedings of the 2014 SIAM International Conference on Data Mining*. pp. 163–171. SIAM (2014)
15. Li, S.Z., Lei, Z., Ao, M.: The hfb face database for heterogeneous face biometrics research. In: *Computer Vision and Pattern Recognition Workshops, 2009. CVPR Workshops 2009. IEEE Computer Society Conference on*. pp. 1–8. IEEE (2009)
16. Li, W., Wang, X.: Locally aligned feature transforms across views. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. pp. 3594–3601 (2013)

17. Lin, D., Tang, X.: Inter-modality face recognition. *Computer Vision–ECCV 2006* pp. 13–26 (2006)
18. Nielsen, A.A.: Multiset canonical correlations analysis and multispectral, truly multitemporal remote sensing data. *IEEE transactions on image processing* 11(3), 293–305 (2002)
19. Rosipal, R., Krämer, N.: Overview and recent advances in partial least squares. In: *Subspace, latent structure and feature selection*, pp. 34–51. Springer (2006)
20. Rupnik, J., Shawe-Taylor, J.: Multi-view canonical correlation analysis. In: *Conference on Data Mining and Data Warehouses (SIKDD 2010)*. pp. 1–4 (2010)
21. Sharma, A., Jacobs, D.W.: Bypassing synthesis: Pls for face recognition with pose, low-resolution and sketch. In: *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on*. pp. 593–600. IEEE (2011)
22. Sim, T., Zhang, S., Li, J., Chen, Y.: Simultaneous and orthogonal decomposition of data using multimodal discriminant analysis. In: *Computer Vision, 2009 IEEE 12th International Conference on*. pp. 452–459. IEEE (2009)
23. Tzimiropoulos, G., Zafeiriou, S., Pantic, M.: Subspace learning from image gradient orientations. *IEEE transactions on pattern analysis and machine intelligence* 34(12), 2454–2466 (2012)
24. Voß, H., Eckhardt, U.: Linear convergence of generalized weiszfeld’s method. *Computing* 25(3), 243–251 (1980)
25. Wang, K., He, R., Wang, W., Wang, L., Tan, T.: Learning coupled feature spaces for cross-modal matching. In: *Proceedings of the IEEE International Conference on Computer Vision*. pp. 2088–2095 (2013)
26. Xiao, M., Guo, Y.: Feature space independent semi-supervised domain adaptation via kernel matching. *IEEE transactions on pattern analysis and machine intelligence* 37(1), 54–66 (2015)
27. Xu, C., Tao, D., Xu, C.: Multi-view intact space learning. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 37(12), 2531–2544 (2015)
28. Xu, Y., Fang, X., Wu, J., Li, X., Zhang, D.: Discriminative transfer subspace learning via low-rank and sparse representation. *IEEE Transactions on Image Processing* 25(2), 850–863 (2016)
29. You, X., Guo, W., Yu, S., Li, K., Príncipe, J.C., Tao, D.: Kernel learning for dynamic texture synthesis. *IEEE Transactions on Image Processing* 25(10), 4782–4795 (2016)
30. Yu, S., Abraham, Z.: Concept drift detection with hierarchical hypothesis testing. In: *Proceedings of the 2017 SIAM International Conference on Data Mining*. pp. 768–776. SIAM (2017)