Multi-view Discriminant Analysis

Meina Kan[1], Shiguang Shan[1], Haihong Zhang[2],
Shihong Lao[2], and Xilin Chen[1]

[1] Key Lab of Intelligent Information Processing of Chinese Academy of Sciences (CAS), Institute of Computing Technology, CAS, Beijing, 100190, China
[2] Omron Social Solutions Co., LTD., Kyoto, Japan
{meina.kan,shiguang.shan,xilin.chen}@vipl.ict.ac.cn,
angelazhang@ssb.kusatsu.omron.co.jp, lao@ari.ncl.omron.co.jp

Abstract. The same object can be observed at different viewpoints or even by different sensors, thus generating multiple distinct even heterogeneous samples. Nowadays, more and more applications need to recognize object from distinct views. Some seminal works have been proposed for object recognition across two views and applied to multiple views in some inefficient pairwise manner. In this paper, we propose a Multi-view Discriminant Analysis (MvDA) method, which seeks for a discriminant common space by jointly learning multiple view-specific linear transforms for robust object recognition from multiple views, in a non-pairwise manner. Specifically, our MvDA is formulated to jointly solve the multiple linear transforms by optimizing a generalized Rayleigh quotient, i.e., maximizing the between-class variations and minimizing the within-class variations of the low-dimensional embeddings from both intra-view and inter-view in the common space. By reformulating this problem as a ratio trace problem, an analytical solution can be achieved by using the generalized eigenvalue decomposition. The proposed method is applied to three multi-view face recognition problems: face recognition across poses, photo-sketch face recognition, and Visual (VIS) image vs. Near Infrared (NIR) image face recognition. Evaluations are conducted respectively on Multi-PIE, CUFSF and HFB databases. Intensive experiments show that MvDA can achieve a more discriminant common space, with up to 13% improvement compared with the best known results.

Keywords: Multi-view Discriminant Analysis, Multi-view Face Recognition, Common space for Multi-view.

1 Introduction

In many computer vision applications, the same object can be observed at different viewpoints or even by different sensors, thus generating multiple distinct even heterogeneous samples. For example, given a face, photos can be taken from different viewpoints, resulting multi-pose face images [1]; a face can be also illuminated by visible lighting or near infrared lighting to capture visual images or near infrared images respectively [2]. Recently, more and more applications need

A. Fitzgibbon et al. (Eds.): ECCV 2012, Part I, LNCS 7572, pp. 808–821, 2012.

to perform classification from both intra-view and inter-view. However, they generally cannot be conducted directly since the samples from different views may lie in quite different spaces and cannot be compared. Therefore, most of previous works addressing this problem endeavored to learn a common space shared by the multiple views, in which samples from multiple views can be compared.

The most typical approach to obtain a common space for multiple views should be the Canonical Correlation Analysis (CCA) [3] and its kernelized variant [4]. CCA learned two transforms, one for each view, to respectively project samples to a common space. Both transforms were obtained by maximizing the cross correlation between the two views. In [5][6], a pseudo-photo in the target view was synthesized for the sample from the query view. In these methods, the target space is actually used as the common space. Further, a method dealing with face recognition with pose, low-resolution and sketch was proposed by employing Partial Least Square regression [7] to project samples from two views to a common latent subspace, in which samples from one view as regressor and samples from another view as response. For specific photo-sketch face recognition, a Coupled Information-theoretic projection tree [8] was proposed to reduce the modality gap between photo and sketch. In [9], a pair of semi-coupled dictionaries are was proposed to characterize both views with a mapping function modeling the relationship between the two dictionaries. Although the gap between two views was minimized by these methods, the discriminant information, e.g., label information, was not explicitly taken into account.

To learn a discriminant common space for two views, Correlation Discriminant Analysis (CDA) [10] and Discriminative Canonical Correlation Analysis (DCCA) [11][12] were proposed to extend CCA by maximizing the difference of within-class and between-class variations across two views. In [13][14], Multiview Fisher Discriminant Analysis (MFDA) was proposed to employ the label information for binary classification. Then in [15], Common Discriminant Feature Extraction (CDFE) was proposed to minimize the intra-class scatter and meanwhile maximize the inter-class separability, with the local consistency as a regularizer to reduce the risk of over-fitting. As a result, CDFE achieved very encouraging performance. However CDFE is sensitive to the parameter β which controls the local consistency. In [16], a large margin approach was proposed to learn the supervised multi-view latent space Markov Networks by using the maximum likelihood estimation. Additionally, Coupled Spectral Regression (CSR) [17] was proposed by employing the label as the common space for two views. First, a common low-dimensional embedding was calculated only according to the label information which was the same for samples of one class but from multiple views, and then a projection matrix between the observation space and the low-dimensional embedding was learned through least square regression for each view. In [18], a local feature based discriminant analysis was proposed by employing the Scale-invariant feature transform (SIFT) and multi-scale Local Binary Patterns (LBP) feature to match a forensic sketch and a mug shot photo.

All above methods work well only in the scenario of two views. In case of multiple views, the pairwise (i.e., one-versus-one) strategy is generally exploited

to convert one common space for v views problem to C_v^2 common spaces problem. However, such a pairwise manner is neither efficient nor optimal for classification across different views. What we need is a unified semantic common space, which should embody invariant features or attributes that can identify the underlying object, commonly shared by all the views rather than only two views. For this purpose, the Multiview CCA (MCCA) [19] was proposed to obtain one common space for v views. In MCCA, the v view-specific transforms, one for each view, were obtained by maximizing the total correlations between any two views.

Although MCCA can obtain a common space for multiple views, it does not take discriminant information into account. In other words, the resulting common space is not discriminative, thus not good for classification across views. Moreover, in MCCA and all existing pairwise methods, only the inter-view correlation is considered, but ignoring the intra-view correlations which implies unstable classification within view. To deal with these problems, this paper proposes a Multi-view Discriminant Analysis (MvDA) method that can learn single unified discriminant common space for v views by jointly optimizing v view-specific transforms, one for each view. In this common space, the between-class variations from both inter-view and intra-view are maximized, while the within-class variations from both inter-view and intra-view are minimized. In our implementation, the between-class and within-class variations are combined to form a generalized Rayleigh quotient, which can be solved analytically by using the generalized eigenvalue decomposition.

Compared with previous works, our method has several advantages: 1) one discriminant common space is obtained for multiple views by jointly optimizing v view-specific transforms, which is efficient and leads to better generalization ability for classification from multiple views. 2) variations from both inter-view and intra-view are considered in a generalized Rayleigh quotient, leading to a more discriminant common space. 3) the problem is solved analytically by using the generalized eigenvalue decomposition.

The rest of the paper is organized as follows. Section 2 gives a detailed description of some related works. Section 3 presents the formulation and solution of the proposed Multi-view Discriminant Analysis. Section 4 evaluates the Multi-view Discriminant Analysis on three databases, followed by a conclusion.

2 Related Works

2.1 Canonical Correlation Analysis (CCA) [3]

CCA was proposed to find a common subspace in which the low dimensional embedding of samples from two views are most correlated. Formally, let S represent the set of samples from two views: $S = \{(x_{11}, x_{12}), (x_{21}, x_{22}), \cdots, (x_{n1}, x_{n2})\}$, where $x_{ik} \in R^{p_k}, k = 1, 2$ represents the i^{th} sample from the k^{th} view in p_k dimensionality. Two matrices $X_1 = [x_{11}, x_{21}, \cdots, x_{n1}]$ and $X_2 = [x_{12}, x_{22}, \cdots, x_{n2}]$ are defined for representing the data matrix from the two views. Two linear transforms w_1 and w_2 can be obtained to respectively project the samples from

two views into the common space, by maximizing the correlation between the low-dimensional embeddings $w_1^T X_1$ and $w_2^T X_2$:

$$\begin{aligned} &\max_{w_1, w_2} w_1^T X_1 X_2^T w_2 \\ &s.t.\ w_1^T X_1 X_1^T w_1 = 1,\ w_2^T X_2 X_2^T w_2 = 1. \end{aligned} \tag{1}$$

Employing the Lagrange multiplier, this problem can be solved by resorting to the eigenvalue decomposition. However, the data in CCA must be given in pairwise, i.e., the number of samples from both views should be same to make $X_1 X_2^T$ computable. Having w_1 and w_2, samples from two views can be compared by being projected to the common space. Actually, CCA is a two-view extension of PCA [20], i.e., a unsupervised approach. When applied in the multi-view scenario, the one-versus-one strategy can be employed.

2.2 Multi-view Canonical Correlation Analysis (MCCA) [19]

In [19], a generalization of CCA for multi-view scenario is proposed called Multi-view Canonical Correlation Analysis. The goal of MCCA is to find a set of linear transforms $w_i \in R^{p_i}, i = 1, 2, \cdots, v$, to project samples from v views $\{X_1, \cdots, X_v\}$ to one common space where the total correlation of the low-dimensional embeddings $\{w_1^T X_1, \cdots, w_v^T X_v\}$ from any two views is maximized:

$$\begin{aligned} &\max_{w_1, w_2, \cdots, w_v} \textstyle\sum_{i<j} w_i^T X_i X_j^T w_j \\ &s.t.\ w_i^T X_i X_i^T w_i = 1, i = 1, 2, \cdots, v, \end{aligned} \tag{2}$$

where $X_i \in R^{p_i \times n}$ is the data matrix from the i^{th} view containing n samples in p_i dimensionality. This problem can be reformulated as a generalized multivariate eigenvalue problem by using the Lagrange multiplier:

$$\begin{bmatrix} A_{11} & \cdots & A_{1v} \\ \vdots & \ddots & \vdots \\ A_{v1} & \cdots & A_{vv} \end{bmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_v \end{pmatrix} = \begin{pmatrix} \lambda_1 \beta_1 \\ \lambda_2 \beta_2 \\ \vdots \\ \lambda_v \beta_v \end{pmatrix}, \tag{3}$$

where $w_i^T X_i$ is expressed in the dual form $\beta_i^T K_i$, K_i is the kernel matrix and $A_{ij} = K_i K_j^T$. This problem can be solved by an alternation method [21]. Same as CCA, the number of samples in each view should be the same.

3 Multi-view Discriminant Analysis (MvDA)

In this section, we first introduce the basic idea and formulation of the MvDA. Then, describe how to solve it analytically. Finally, we discuss more on the differences and advantages of MvDA over previous methods. Please note that, for the sake of clarity, some of the detailed reformulations are put in the appendix.

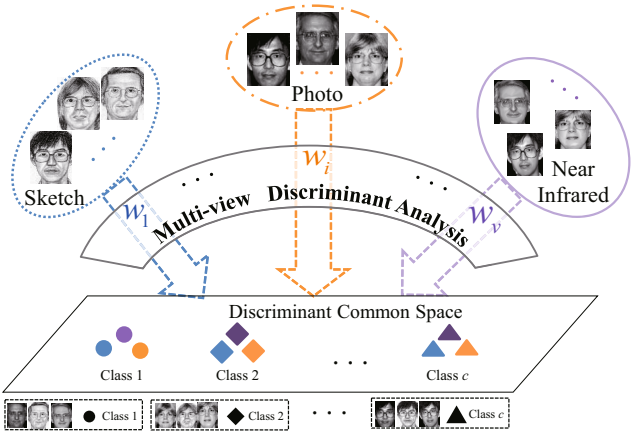

Fig. 1. The overview of MvDA. Samples from different views are projected into a discriminant common space by using v transforms, one for each view. In this common space, samples in one class from multiple views are close to each other, while samples in different classes from multiple views are far away from each other. Here, images from distinct views, such as photo, sketch, NIR are denoted in different colors and images from different classes are denoted in different shapes.

3.1 MvDA: Overview and Formulation

As shown in Fig. 1, our MvDA attempts to find v linear transforms $w_1, w_2, \cdots, w_v$ that can respectively project the samples from v views to one discriminant common space, in which the between-class variation is maximized while the within-class variation is minimized. For this purpose, formally, let us first define $\mathcal{X}^{(j)} = \{x_{ijk} | i = 1, \cdots, c; k = 1, \cdots, n_{ij}\}$ as the samples from the j^{th} view, where $x_{ijk} \in R^{d_j}$ is the k^{th} sample from the j^{th} view of the i^{th} class in d_j dimensionality, c is the number of classes and n_{ij} is the number of samples from the j^{th} view of the i^{th} class.

The samples from v views can be projected to the common space by using the v linear transforms denoted as $\mathcal{Y} = \{y_{ijk} = w_j^T x_{ijk} | i = 1, \cdots, c; j = 1, \cdots, v; k = 1, \cdots, n_{ij}\}$. In this common space, the between-class variation S_B^y from all views is maximized while the within-class variation S_W^y from all views is minimized. We formulate this objective as a generalized Rayleigh quotient:

$$(w_1^*, w_2^*, \cdots, w_v^*) = \arg \max_{w_1, \cdots, w_v} \frac{Tr\left(S_B^y\right)}{Tr\left(S_W^y\right)}. \tag{4}$$

Here, the within-class scatter matrix S_W^y of the low-dimensional embeddings in the common space is calculated as:

$$S_W^y = \sum_{i=1}^{c} \sum_{j=1}^{v} \sum_{k=1}^{n_{ij}} \left(y_{ijk} - \mu_i\right) \left(y_{ijk} - \mu_i\right)^T, \tag{5}$$

where $u_i = \frac{1}{n_i} \sum_{j=1}^{v} \sum_{k=1}^{n_{ij}} y_{ijk}$ is the mean of the low-dimensional embeddings from the i^{th} class and n_i is the number of samples in the i^{th} class.

Similarly, the between-class scatter matrix S_B^y of the low-dimensional embeddings in the common space is calculated as:

$$S_B^y = \sum\nolimits_{i=1}^{c} n_i \left(\mu_i - \mu\right)\left(\mu_i - \mu\right)^T, \tag{6}$$

where $\mu = \frac{1}{n}\sum_{i=1}^{c}\sum_{j=1}^{v}\sum_{k=1}^{n_{ij}} y_{ijk}$ is the mean of all low-dimensional embeddings with n is the number of all samples.

In the above formulation, it is clear that the within-class and between-class variations are calculated from the samples from all views, not only those from the distinct views. In other words, the variations from the intra-view are also considered in addition to that from the inter-view. After obtaining $w_1, w_2, \cdots, w_v$ by (4), the samples from v views can be compared by being respectively projected to the discriminant common space.

3.2 Analytical Solution of MvDA

Equ. (4) seems like traditional Fisher Linear Discriminant (FLD) analysis, however it is actually much more complicated than FLD, as we need to jointly optimize v distinct linear transforms. Fortunately, we work out an analytic solution by re-formulate it into a ratio trace problem. Formally, the within-class scatter matrix (5) of the low-dimensional embeddings in the common space can be reformulated as follows (please refer to the appendix for the detailed derivation):

$$S_W^y = \begin{bmatrix} w_1^T & w_2^T & \cdots & w_v^T \end{bmatrix} \begin{pmatrix} S_{11} & S_{12} & \cdots & S_{1v} \\ S_{21} & S_{22} & \cdots & S_{2v} \\ \vdots & \vdots & \vdots & \vdots \\ S_{v1} & S_{v2} & \cdots & S_{vv} \end{pmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_v \end{bmatrix} = W^T S W, \tag{7}$$

with $W = \begin{bmatrix} w_1^T, w_2^T, \cdots, w_v^T \end{bmatrix}^T$ and S_{jr} defined as bellow with $\mu_{ij}^{(x)} = \frac{1}{n_{ij}}\sum_{k=1}^{n_{ij}} x_{ijk}$:

$$S_{jr} = \begin{cases} \sum_{i=1}^{c}\left(\sum_{k=1}^{n_{ij}} x_{ijk}x_{ijk}^T - \frac{n_{ij}n_{ij}}{n_i}\mu_{ij}^{(x)}\mu_{ij}^{(x)T}\right) & j = r \\ -\sum_{i=1}^{c}\frac{n_{ij}n_{ir}}{n_i}\mu_{ij}^{(x)}\mu_{ir}^{(x)T} & \text{otherwise} \end{cases} \tag{8}$$

Similarly, the between-class scatter matrix (6) can be further reformulated as follows (please also refer to the appendix for more details):

$$S_B^y = \begin{bmatrix} w_1^T & w_2^T & \cdots & w_v^T \end{bmatrix} \begin{pmatrix} D_{11} & D_{12} & \cdots & D_{1v} \\ D_{21} & D_{22} & \cdots & D_{2v} \\ \vdots & \vdots & \vdots & \vdots \\ D_{v1} & D_{v2} & \cdots & D_{vv} \end{pmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_v \end{bmatrix} = W^T D W, \tag{9}$$

with W defined as above and D_{jr} is defined as:

$$D_{jr} = \left(\sum_{i=1}^{c}\frac{n_{ij}n_{ir}}{n_i}\mu_{ij}^{(x)}\mu_{ir}^{(x)T}\right) - \frac{1}{n}\left(\sum_{i=1}^{c} n_{ij}\mu_{ij}^{(x)}\right)\left(\sum_{i=1}^{c} n_{ir}\mu_{ir}^{(x)}\right)^T, \tag{10}$$

By taking (7) and (9), (4) can be reformulated as:

$$(w_1^*, w_2^*, \cdots, w_v^*) = \arg \max_{w_1,\cdots,w_v} \frac{Tr\left(W^T D W\right)}{Tr\left(W^T S W\right)}. \tag{11}$$

According to [22], the objective function in (11) is in the form of trace ratio, which implies the closed form solution does not exist. We therefore reformulate it into a more tractable one in the form of ratio trace as bellow:

$$(w_1^*, w_2^*, \cdots, w_v^*) = \arg \max_{w_1,\cdots,w_v} Tr\left(\frac{W^T D W}{W^T S W}\right) \tag{12}$$

which can be solved analytically through generalized eigenvalue decomposition.

3.3 Discussion

Difference with Other Inter-View Methods. In our formulation, both intra-view and inter-view correlations are considered when calculating the within-class and between-class variations. To show this, we can reformulate (5) as bellow :

$$\begin{aligned} S_W^y = & \sum_{i=1}^{c} \Big(\sum_{j=1}^{v} \sum_{k=1}^{n_{ij}} \sum_{l=1}^{n_{ij}} \left(y_{ijk} - y_{ijl}\right) \left(y_{ijk} - y_{ijl}\right)^T \\ & + \sum_{j=1}^{v} \sum_{r=1, r\neq j}^{v} \sum_{k=1}^{n_{ij}} \sum_{l=l}^{n_{ij}} \left(y_{ijk} - y_{irl}\right) \left(y_{ijk} - y_{irl}\right)^T \Big). \end{aligned} \tag{13}$$

It can be easily seen that the first term captures the variations within one view and the second term represents the variations across different views. Another key difference is that our MvDA projects the samples from v views to one common space, not taking the one-versus-one manner to project the samples into C_v^2 common spaces as most previous methods did.

Difference with Metric Learning Methods. Generally, the metric learning methods convert the multi-view problem to multiple two-view problems leading to multiple metrics, one for each pair of views, while MvDA directly deals with the multi-view problem by seeking for one common space for multiple views.

Difference with MCCA [19]. Both MCCA and MvDA obtain one common space for multiple views. MCCA only obtain a common space in which the correlation between multiple views is maximized, neither intra-view correlation nor label information is considered, while MvDA endeavors to obtain a discriminant common space, in which the within-class variations from multiple views is minimized and the between-class variations from multiple views is maximized.

Difference with MFDA [13]. MFDA is only applicable for binary classification. Although the one-versus-all or hierarchical strategy can be employed for multi-class scenarios, it's still not applicable when the training classes not overlap with the testing classes, which is mainly considered in this work.

Difference with GDA [23]. 1) In GDA, the discriminability is obtained within each view, while in MvDA it is achieved across all views . 2) In GDA, the cross-view correlation is obtained only from those observations corresponding to the

same underlying sample, while in MvDA it is obtained from all observations from different views. 3) GDA has many parameters especially for large number of views, while our MvDA has no parameter.

4 Experiments

In this section, MvDA is evaluated on face recognition across pose, photo-sketch face recognition and Near Visual vs. Infrared image heterogeneous face recognition on three datasets. Intensive results demonstrate the effectiveness of MvDA.

4.1 Datasets

Multi-PIE dataset [1] is employed to evaluate face recognition across pose. It contains more than 750,000 images of 337 people under various view points, illumination and expressions. In this work, a subset about 14,450 images from the 337 subjects in 7 poses (-45°, -30°, -15°, 0°, 15°, 30°, 45°), 3 expression (Neutral, Smile, Disgust), no flush illumination from 4 sessions are selected as the evaluation dataset. This subset is divided into two parts: images from 231 subjects with 4 randomly selected images under each pose of each subject ($231\times7\times4$=6,468) are used for training and images from the rest subjects (2,289) for testing.

CUHK Face Sketch FERET (CUFSF) dataset [24][8] is used to evaluate photo-sketch face recognition. It contains 1,194 subjects with lighting variations from FERET dataset [25]. For each subject, a sketch is drawn with shape exaggeration. On this dataset, images from the first 700 subjects are used as the training data and images from the rest 494 subjects are used as the testing data.

Heterogeneous Face Biometrics (HFB) dataset [2] is used to evaluate Visual (VIS) image vs. Near Infrared (NIR) image heterogeneous face recognition. This dataset contains images from 100 subjects, with 4 NIR and 4 VIS images per subject. The evaluation follows the Protocol II, i.e., images from 70 and the rest 30 subjects are respectively employed for training and testing.

4.2 Experimental Settings

All images from Multi-PIE and CUFSF datasets are cropped into 64x80 pixels without any further preprocess and images from HFB dataset are cropped into 32x32 according to the standard protocol. The proposed MvDA is compared to Pairwise CCA (PW-CCA) [3], Pairwise FDA (PW-FDA) [20], CDFE [15], CSR [17], PLS [7], Unified FDA [20] (U-FDA) and MCCA [19]. Among them, PW-CCA, PW-FDA, CDFE, CSR and PLS are pairwise methods for multi-view classification; the so called U-FDA is the directly applied Fisherface [20] regardless of the view and MCCA is a multi-view method. In all our experiments, the alternation solution for MCCA performs very poorly. So instead, we constrain $\lambda_1 = \lambda_2 = \cdots = \lambda_v$ leading to an analytical solution to (3) and this analytical solution performs much better than the alternation one. Therefore, all results of MCCA reported in this paper are obtained by using this analytical solution.

Table 1. Evluation on Multi-PIE dataset in terms of mean accuracy (mACC)

PW-CCA[3]	PW-FDA[20]	CDFE[15]	CSR[17]	PLS[7]	U-FDA[20]	MCCA[19]	**MvDA**
82.1%	89.0%	88.8%	72.0%	77.4%	84.3%	91.6%	**95.0%**

Table 2. Results of CDFE [15] on MultiPIE dataset in terms of rank-1 recognition rate

Probe \ Gallery	-45°	-30°	-15°	0°	15°	30°	45°
-45°	100.0%	99.1%	88.7%	70.3%	**71.9%**	**71.6%**	**78.0%**
-30°	98.8%	100.0%	99.7%	89.6%	89.6%	**89.0%**	**82.6%**
-15°	89.6%	99.4%	100.0%	98.8%	96.9%	92.4%	**79.8%**
0°	70.6%	89.9%	98.8%	100.0%	99.1%	94.2%	77.7%
15°	**71.9%**	89.3%	98.2%	99.1%	100.0%	100.0%	92.7%
30°	**73.4%**	**90.2%**	91.4%	93.0%	99.7%	100.0%	99.4%
45°	**80.4%**	**86.2%**	**79.8%**	77.4%	91.1%	99.4%	100.0%

The Principal Component Analysis (PCA) [26] is applied for dimension reduction. The dimensionality is set to 100, 100 and 78 to preserve more than 95% energy on Multi-PIE, CUFSF and HFB datasets respectively for all methods except CDFE, CSR and PLS, for which the dimensionality is set to the value that can obtain the best performance. For CDFE, the parameter α and β are traversed in [0.01 1] and [0.0001 1] respectively to report the best result and for CSR, the parameter λ and η are traversed in [0.01 10] to obtain a best result.

4.3 Face Recognition across Pose

Face recognition across pose is evaluated in both pairwise manner and multi-view manner on Multi-PIE dataset. The pairwise manner means images from one view are used as gallery set while images from another view used as probe set. Samples in Multi-PIE dataset are from 7 views, thus leading to 7×6=42 evaluations in terms of rank-1 recognition rate (as shown in Table 2∼4). Then, all 42 results are averaged as the mean accuracy (mACC) as shown in Table 1.

From Table 1, we can see that the multi-view method MCCA and MvDA perform better than other pairwise methods significantly. Compared with MCCA, the proposed MvDA can perform even better with 3.4% improvement which is very impressive since it is the average of 42 results. Seen from Table 2 ∼ Table 4, the improvement between samples across large view (e.g., the results in bold) can be up to 18.3% and 10.4% compared with CDFE and MCCA. Furthermore, the proposed MvDA is also evaluated in the multi-view scenario that samples in the gallery are from multiple views. In this scenario, samples from each view are used as probe data and samples from the rest views that have at least 30° away from the probe view are employed as the gallery data. The results are displayed in Table 5. As seen, MCCA performs bettern than all the other methods and our MvDA works better than MCCA up to 5.2%, which demonstrates our MvDA can obtain a more discriminant common space.

Table 3. Results of MCCA [19] on MultiPIE in terms of rank-1 recognition rate

Probe \ Gallery	-45°	-30°	-15°	0°	15°	30°	45°
-45°	100.0%	95.4%	86.5%	77.4%	**76.8%**	**79.5%**	**86.0%**
-30°	99.1%	100.0%	98.5%	93.0%	90.0%	**91.7%**	**88.4%**
-15°	93.0%	99.7%	100.0%	100.0%	98.8%	94.5%	**88.1%**
0°	78.9%	92.1%	99.4%	100.0%	99.4%	94.5%	84.4%
15°	**82.3%**	91.1%	98.2%	99.7%	100.0%	99.4%	93.0%
30°	**85.6%**	**93.0%**	93.0%	97.6%	99.4%	100.0%	99.0%
45°	**84.1%**	**87.2%**	**84.1%**	85.0%	92.0%	98.2%	100.0%

Table 4. Results of our MvDA on MultiPIE dataset in terms of rank-1 recognition rate

Probe \ Gallery	-45°	-30°	-15°	0°	15°	30°	45°
-45°	100.0%	100.0%	92.7%	80.0%	**87.2%**	**88.4%**	**91.4%**
-30°	99.4%	100.0%	100.0%	94.8%	93.6%	**96.3%**	**94.2%**
-15°	95.4%	100.0%	100.0%	99.7%	98.8%	98.5%	**91.7%**
0°	80.1%	94.5%	100.0%	100.0%	100.0%	96.6%	87.8%
15°	**90.2%**	96.0%	100.0%	99.4%	100.0%	100.0%	95.7%
30°	**91.4%**	**98.5%**	98.2%	98.5%	100.0%	100.0%	100.0%
45°	**91.4%**	**93.3%**	**92.7%**	87.2%	95.1%	100.0%	100.0%

4.4 Photo-Sketch Recognition

Photo-Sketch recognition is evaluated on CUFSF dataset. Samples in this dataset come from only two views, photo and sketch. In this case, MCCA degenerates to the PW-CCA, and U-FDA degenerates to PW-FDA. The comparison results are shown in Table 6. As seen, MCCA performs much worse than our MvDA and this may be due to the disappearance of the generalization benefited from multiple views, while our MvDA can still benefit from the intra-view variations.

4.5 Near Infrared vs. Visual Image Heterogeneous Face Recognition

We also test MvDA for heterogeneous face recognition on HFB dataset. Same as photo-sketch recognition, samples are only from two views, Visual image and Near infrared image. Seen from the comparison in Table 6, MvDA can achieve a significant improvement up to 13% compared to the best performer. Since there are more (i.e., 4) images for each view per subject on HFB than that (i.e., 1) on CUFSF, more information from intra-view can be exploited by MvDA on HFB leading to a larger improvement. From the above intensive evaluations, it can be seen that the common space obtained by the multi-view methods MCCA and our MvDA is more suitable for multi-view classification. Furthermore, benefited from the variations within the same view besides that across view, MvDA can obtain a more discriminant common space for multiple views.

Table 5. Evaluation results on MultiPIE dataset in terms of rank-1 recognition rate. Testing follows the multi-view manner, i.e., samples in gallery are from multiple view.

Gallery View	-15° ∼45°	0° ∼45°	-45° 15° ∼45°	-45°,-30° 30°,45°	-45° ∼-15° 45°	-45° ∼0°	-45° ∼15°
Probe View	-45°	-15°	-30°	0°	15°	30°	45°
CCA[3]	49.9%	63.9%	66.9%	62.1%	66.9%	61.8%	52.0%
PW-FDA[20]	58.7%	62.7%	68.5%	64.2%	66.7%	66.9%	63.0%
CDFE[15]	31.5%	27.8%	33.9%	18.4%	35.2%	31.8%	32.1%
CSR[17]	42.8%	53.5%	58.7%	54.1%	59.3%	54.7%	47.7%
PLS[7]	47.7%	54.7%	58.1%	56.9%	61.5%	58.1%	50.5%
U-FDA[20]	53.8%	61.5%	66.7%	61.2%	64.8%	65.4%	58.4%
MCCA[19]	61.7%	63.9%	67.9%	65.8%	70.3%	64.5%	63.0%
MvDA	**63.9%**	**69.1%**	**70.1%**	**68.2%**	**70.6%**	**69.4%**	**67.3%**

Table 6. Evaluation on CUFSF and HFB datasets in terms of rank-1 recognition rate

		PW-CCA[3]	CDFE[15]	CSR[17]	PLS[7]	U-FDA[20]	**MvDA**
CUFSF	**Photo-Sketch**	38.7%	45.6%	50.2%	48.6%	46.8%	**53.4%**
	Sketch-Photo	47.5%	47.6%	49.0%	51.0%	53.4%	**55.5%**
HFB	**VIS-NIR**	29.2%	40.8%	26.7%	38.3%	39.1%	**53.3%**
	NIR-VIS	30.8%	36.7%	32.5%	40.8%	40.0%	**50.0%**

5 Conclusions

In this paper, we have developed a multi-view discriminant analysis method that can obtain one discriminant common space for object recognition from multiple views. Our method not only exploits the correlations from inter-view, but also that from the intra-view to obtain better discriminability and generalizability. The problem is formulated to optimize a generalized Rayleigh quotient and solved analytically by reformulating it to a ratio trace problem. Experiments on three multi-view datasets demonstrate the superior of our method over other state-of-the-art techniques. Obviously, our work can be easily kernelized which will be our future work. Furthermore, we will attempt to directly optimize (11).

Acknowledgments. This work is partially supported by National Basic Research Program of China (973 Program) under contract 2009CB320902, Natural Science Foundation of China under contracts Nos. 61025010, 61173065, 60832004, and the FiDiPro program of Tekes. Haihong Zhang and Shihong Lao are partially supported by "R&D Program for Implementation of Anti-Crime and Anti-Terrorism Technologies for a Safe and Secure Society", Special Coordination Fund for Promoting Science and Technology of MEXT, the Japanese Government.

References

1. Gross, R., Matthews, I., Cohn, J., kanada, T., Baker, S.: The cmu multi-pose, illumination, and expression (multi-pie) face database. Technical report, Carnegie Mellon University Robotics Institute. TR-07-08 (2007)

2. Li, S.Z., Lei, Z., Ao, M.: The hfb face database for heterogeneous face biometrics research. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 1–8 (2009)
3. Hotelling, H.: Relations between two sets of variates. Biometrika 28, 321–377 (1936)
4. Akaho, S.: A kernel method for canonical correlation analysis (2006)
5. Tang, X., Wang, X.: Face sketch recognition. IEEE Transactions on Circuits and Systems for Video Technology 14, 50–57 (2004)
6. Tang, X., Jin, H., Lu, H., Ma, S.: A nonlinear approach for face sketch synthesis and recognition. In: IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, pp. 1005–1010 (2005)
7. Sharma, A., Jacobs, D.W.: Bypassing synthesis: Pls for face recognition with pose, low-resolution and sketch. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 593–600 (2011)
8. Zhang, W., Wang, X., Tang, X.: Coupled information-theoretic encoding for face photo-sketch recognition. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 513–520 (2011)
9. Wang, S., Zhang, L., Liang, Y., Pan, Q.: Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch image synthesis. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR (2012)
10. Ma, Y., Lao, S., Takikawa, E., Kawade, M.: Discriminant analysis in correlation similarity measure space. In: Proceedings of the International Conference on Machine Learning (ICML), pp. 577–584 (2007)
11. Sun, T., Chen, S., Yang, J., Shi, P.: A novel method of combined feature extraction for recognition. In: IEEE Intl. Conf. on Data Mining (ICDM), pp. 1043–1048 (2008)
12. Kim, T.-K., Kittler, J., Cipolla, R.: Learning Discriminative Canonical Correlations for Object Recognition with Image Sets. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3953, pp. 251–262. Springer, Heidelberg (2006)
13. Diethe, T., Hardoon, D.R., Shawe-Taylor, J.: Multiview fisher discriminant analysis. In: NIPS Workshop on Learning from Multiple Sources (2008)
14. Diethe, T., Hardoon, D.R., Shawe-Taylor, J.: Constructing Nonlinear Discriminants from Multiple Data Views. In: Balcázar, J.L., Bonchi, F., Gionis, A., Sebag, M. (eds.) ECML PKDD 2010, Part I. LNCS, vol. 6321, pp. 328–343. Springer, Heidelberg (2010)
15. Lin, D., Tang, X.: Inter-modality Face Recognition. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3954, pp. 13–26. Springer, Heidelberg (2006)
16. 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 (NIPS), vol. 23 (2010)
17. Lei, Z., Li, S.Z.: Coupled spectral regression for matching heterogeneous faces. In: IEEE Conference on Computer Vision and Pattern Recognition, pp. 1123–1128 (2009)
18. Klare, B.F., Li, Z., Jain, A.K.: Matching forensic sketches to mug shot photos. IEEE Trans. on Pattern Analysis And Machine Intelligence 33, 39–646 (2011)
19. Rupnik, J., Shawe-Taylor, J.: Multi-view canonical correlation analysis. In: SiKDD (2010)
20. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI) 19, 711–720 (1997)
21. Horst, P.: Relations among m sets of measures. Journal of Foo 26, 129–149 (1961)
22. Wang, H., Yan, S., Xu, D., Tang, X., Huang, T.: Trace ratio vs. ratio trace for dimensionality reduction. In: IEEE Conf. on Computer Vision and Pattern Recognition, CVPR (2007)

23. A. Sharma, A. Kumar, H. Daume III, D.W. Jacobs: Generalized multiview analysis: A discriminative latent space. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR (2012)
24. Wang, X., Tang, X.: Face photo-sketch synthesis and recognition. IEEE Transactions on Pattern Analysis And Machine Intelligence (TPAMI) 31, 1955–1967 (2009)
25. Phillips, P., Wechslerb, H., Huangb, J., Rauss, P.J.: The feret database and evaluation procedure for face-recognition algorithms. Image and Vision Computing 16, 295–306 (1998)
26. Turk, M.A., Pentland, A.P.: Face recognition using eigenfaces. In: IEEE Conference on Computer Vision and Pattern Recognition (CVPR), vol. 591, pp. 586–591 (1991)

Appendix: The derivation of S_W^y and S_B^y

The within-class scatter matrix S_W^y of the low-dimensional embeddings from multiple views are reformulated as follows:

$$
\begin{aligned}
S_W^y &= \sum\nolimits_{i=1}^{c} \sum\nolimits_{j=1}^{v} \sum\nolimits_{k=1}^{n_{ij}} \left(y_{ijk} - \mu_i\right)\left(y_{ijk} - \mu_i\right)^T \\
&= \sum_{i=1}^{c} \left(\sum_{j=1}^{v} \sum_{k=1}^{n_{ij}} y_{ijk} y_{ijk}^T - \sum_{j=1}^{v} \sum_{k=1}^{n_{ij}} y_{ijk} \mu_i^T - \sum_{j=1}^{v} \sum_{k=1}^{n_{ij}} \mu_i y_{ijk}^T + \sum_{j=1}^{v} \sum_{k=1}^{n_{ij}} \mu_i \mu_i^T \right) \\
&= \sum\nolimits_{i=1}^{c} \left(\sum\nolimits_{j=1}^{v} \sum\nolimits_{k=1}^{n_{ij}} y_{ijk} y_{ijk}^T - n_i \mu_i \mu_i^T \right) \\
&= \sum\nolimits_{i=1}^{c} \left(\sum_{j=1}^{v} \sum_{k=1}^{n_{ij}} y_{ijk} y_{ijk}^T - \frac{1}{n_i} \left(\sum\nolimits_{j=1}^{v} \sum\nolimits_{k=1}^{n_{ij}} y_{ijk} \right) \left(\sum\nolimits_{j=1}^{v} \sum\nolimits_{k=1}^{n_{ij}} y_{ijk} \right)^T \right) \\
&= \sum\nolimits_{i=1}^{c} \left(\sum\nolimits_{j=1}^{v} \sum\nolimits_{k=1}^{n_{ij}} y_{ijk} y_{ijk}^T - \frac{1}{n_i} \left(\sum\nolimits_{j=1}^{v} n_{ij} \mu_{ij} \right) \left(\sum\nolimits_{j=1}^{v} n_{ij} \mu_{ij} \right)^T \right) \\
&= \sum\nolimits_{i=1}^{c} \left(\sum\nolimits_{j=1}^{v} \sum\nolimits_{k=1}^{n_{ij}} y_{ijk} y_{ijk}^T - \frac{1}{n_i} \sum\nolimits_{j=1}^{v} \sum\nolimits_{r=1}^{v} n_{ij} n_{ir} \mu_{ij} \mu_{ir}^T \right) \\
&= \sum_{i=1}^{c} \left(\sum_{j=1}^{v} w_j^T \left(\sum_{k=1}^{n_{ij}} x_{ijk} x_{ijk}^T \right) w_j - \frac{1}{n_i} \sum_{j=1}^{v} \sum_{r=1}^{v} w_j^T \left(n_{ij} n_{ir} \mu_{ij}^{(x)} \mu_{ir}^{(x)T} \right) w_r \right) \\
&= \sum_{j=1}^{v} w_j^T \left(\sum_{i=1}^{c} \sum_{k=1}^{n_{ij}} x_{ijk} x_{ijk}^T \right) w_j - \sum_{j=1}^{v} \sum_{r=1}^{v} w_j^T \left(\sum_{i=1}^{c} \frac{n_{ij} n_{ir}}{n_i} \mu_{ij}^{(x)} \mu_{ir}^{(x)T} \right) w_r \\
&= \sum\nolimits_{j=1}^{v} \sum\nolimits_{r=1}^{v} w_j^T S_{jr} w_r \\
&= \begin{bmatrix} w_1^T & w_2^T & \cdots & w_v^T \end{bmatrix} \begin{pmatrix} S_{11} & S_{12} & \cdots & S_{1v} \\ S_{21} & S_{22} & \cdots & S_{2v} \\ \vdots & \vdots & \vdots & \vdots \\ S_{v1} & S_{v2} & \cdots & S_{vv} \end{pmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_v \end{bmatrix},
\end{aligned}
\tag{14}
$$

$$\text{with} \quad S_{jr} = \begin{cases} \sum_{i=1}^{c} \left(\sum_{k=1}^{n_{ij}} x_{ijk} x_{ijk}^T - \frac{n_{ij} n_{ij}}{n_i} \mu_{ij}^{(x)} \mu_{ij}^{(x)T} \right) & j = r \\ -\sum_{i1}^{c} \frac{n_{ij} n_{ir}}{n_i} \mu_{ij}^{(x)} \mu_{ir}^{(x)T} & \text{otherwise} \end{cases} \tag{15}$$

where c is the number of classes, v is the number of views, n_{ij} is the number of samples from the i^{th} view of the c^{th} class, n_i is the number of samples from the i^{th} class, $\mu_i = \frac{1}{n_i} \sum_{j=1}^{v} \sum_{k=1}^{n_{ij}} y_{ijk}$ is the mean of the i^{th} class across all view, $\mu_{ij} = \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} y_{ijk}$ is the mean of the j^{th} view of the i^{th} class and $\mu_{ij}^{(x)} = \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} x_{ijk}$.

Similarly, the between-class scatter matrix S_B^y of low-dimensional embedding from multi-views are reformulated as bellow with $\mu = \frac{1}{n} \sum_{i=1}^{c} \sum_{j=1}^{v} \sum_{k=1}^{n_{ij}} y_{ijk}$:

$$\begin{aligned}
S_B^y &= \sum_{i=1}^{c} n_i (\mu_i - \mu)(\mu_i - \mu)^T \\
&= \sum_{i=1}^{c} n_i \mu_i \mu_i^T - \sum_{i=1}^{c} n_i \mu_i \mu^T - \sum_{i=1}^{c} n_i \mu \mu_i^T + \sum_{i=1}^{c} n_i \mu \mu^T \\
&= \sum_{i=1}^{c} n_i \mu_i \mu_i^T - n \mu \mu^T \\
&= \sum_{i=1}^{c} n_i \left(\frac{1}{n_i} \sum_{j=1}^{v} \sum_{k=1}^{n_{ij}} y_{ijk} \right) \left(\frac{1}{n_i} \sum_{j=1}^{v} \sum_{k=1}^{n_{ij}} y_{ijk} \right)^T \\
&\quad - n \left(\frac{1}{n} \sum_{i=1}^{c} \sum_{j=1}^{v} \sum_{k=1}^{n_{ij}} y_{ijk} \right) \left(\frac{1}{n} \sum_{i=1}^{c} \sum_{j=1}^{v} \sum_{k=1}^{n_{ij}} y_{ijk} \right)^T \\
&= \sum_{i=1}^{c} \frac{1}{n_i} \left(\sum_{j=1}^{v} n_{ij} \mu_{ij} \right) \left(\sum_{j=1}^{v} n_{ij} \mu_{ij} \right)^T \\
&\quad - \frac{1}{n} \left(\sum_{j=1}^{v} \sum_{i=1}^{c} n_{ij} \mu_{ij} \right) \left(\sum_{j=1}^{v} \sum_{i=1}^{c} n_{ij} \mu_{ij} \right)^T \\
&= \sum_{i=1}^{c} \frac{1}{n_i} \left(\sum_{j=1}^{v} n_{ij} w_j^T \mu_{ij}^{(x)} \right) \left(\sum_{j=1}^{v} n_{ij} w_j^T \mu_{ij}^{(x)} \right)^T \\
&\quad - \frac{1}{n} \left(\sum_{j=1}^{v} \sum_{i=1}^{c} n_{ij} w_j^T \mu_{ij}^{(x)} \right) \left(\sum_{j=1}^{v} \sum_{i=1}^{c} n_{ij} w_j^T \mu_{ij}^{(x)} \right)^T \\
&= \sum_{j=1}^{v} \sum_{r=1}^{v} w_j^T \left(\sum_{i=1}^{c} \frac{n_{ij} n_{ir}}{n_i} \mu_{ij}^{(x)} \mu_{ir}^{(x)T} \right) w_r \\
&\quad - \frac{1}{n} \sum_{j=1}^{v} \sum_{r=1}^{v} w_j^T \left(\left(\sum_{i=1}^{c} n_{ij} \mu_{ij}^{(x)} \right) \left(\sum_{i=1}^{c} n_{ir} \mu_{ir}^{(x)} \right)^T \right) w_r \\
&= \sum_{j=1}^{v} \sum_{r=1}^{v} w_j^T D_{jr} w_r \\
&= \begin{bmatrix} w_1^T & w_2^T & \cdots & w_v^T \end{bmatrix} \begin{pmatrix} D_{11} & D_{12} & \cdots & D_{1v} \\ D_{21} & D_{22} & \cdots & D_{2v} \\ \vdots & \vdots & \vdots & \vdots \\ D_{v1} & D_{v2} & \cdots & D_{vv} \end{pmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_v \end{bmatrix},
\end{aligned} \tag{16}$$

$$D_{jr} = \left(\sum_{i=1}^{c} \frac{n_{ij} n_{ir}}{n_i} \mu_{ij}^{(x)} \mu_{ir}^{(x)T} \right) - \frac{1}{n} \left(\sum_{i=1}^{c} n_{ij} \mu_{ij}^{(x)} \right) \left(\sum_{i=1}^{c} n_{ir} \mu_{ir}^{(x)} \right)^T. \tag{17}$$