Short Papers

Multi-View Discriminant Analysis

Meina Kan, Shiguang Shan, *Senior Member, IEEE*, Haihong Zhang, Shihong Lao, and Xilin Chen, *Senior Member, IEEE*

Abstract—In many computer vision systems, the same object can be observed at varying viewpoints or even by different sensors, which brings in the challenging demand for recognizing objects from distinct even heterogeneous views. In this work we propose a Multi-view Discriminant Analysis (MvDA) approach, which seeks for a single discriminant common space for multiple views in a non-pairwise manner by jointly learning multiple view-specific linear transforms. 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 from both intra-view and inter-view in the common space. By reformulating this problem as a ratio trace problem, the multiple linear transforms are achieved analytically and simultaneously through generalized eigenvalue decomposition. Furthermore, inspired by the observation that different views share similar data structures, a constraint is introduced to enforce the view-consistency of the multiple linear transforms. The proposed method is evaluated on three tasks: face recognition across pose, photo versus. sketch face recognition, and visual light image versus near infrared image face recognition on Multi-PIE, CUFSF and HFB databases respectively. Extensive experiments show that our MvDA achieves significant improvements compared with the best known results.

Index Terms—Multi-view discriminant analysis, cross-view recognition, heterogeneous recognition, common space

1 Introduction

In many computer vision applications, the same object can be observed at various viewpoints or even by heterogeneous sensors, thus generating multiple distinct even heterogeneous images, e.g., [1], [2]. Recently, more and more applications need to match images from different viewpoints or different sensors, usually denoted as heterogeneous recognition or cross-view recognition. Due to the large gap between views, the samples from different views might lie in completely different spaces. Therefore, directly matching the samples from different views is no longer applicable.

To address the above mentioned heterogeneous recognition (or cross-view matching) problem, one need either transform samples of different views into a common space or learn distance metrics that can match heterogeneous samples of various views. As these two methodologies can be equivalently converted in some cases [3], this work focuses on the former, i.e., learning a common subspace shared by various views. This line of methods can be further grouped into two categories: two-view methods and multi-view methods. The multi-view methods attempt to seek for a single unified common space shared by all views. In contrast, the two-view methods essentially can only obtain a common space for two views, but can also be extended to address multiple views problem by using pairwise (i.e., one-versus-one) strategy, i.e., converting a v-view problem to C_v^2 two-view problems. However, such a pairwise manner is neither efficient nor optimal for recognition across multiple views. In addition, according to whether the class label information is exploited, the methods in each category can run in either supervised mode or unsupervised mode.

Two-view methods in unsupervised mode. The most typical approach to obtain a common space for two views should be the canonical correlation analysis (CCA) [4], [5]. CCA attempted to learn two transforms, one for each view, to respectively project the samples from the two views into a common subspace, by maximizing the cross correlation between two views. In [6], [7], to recognize faces with variations in pose, resolution and imaging source, partial least squares (PLS) regression was employed to regress the samples from one view to another. For photo versus sketch face recognition, a coupled information-theoretic projection tree [8] was proposed to reduce the modality gap by maximizing the mutual information between photos and sketches in the quantized feature spaces. In [9], a pair of semi-coupled dictionaries were proposed to characterize both views with a mapping function modeling the intrinsic relationship between the two views, and this work was further extended by using a unified model for coupled dictionary and feature space learning in [10]. Besides, some methods employed either view as the common space, e.g., a pseudo-sketch of photo was synthesized for photo-sketch recognition [11], [12]. Although the gap between two views was minimized by these methods, the discriminant information, e.g., class label, was not explicitly taken into account.

Two-view methods in supervised mode. To learn a discriminant common space for two views, the class label information is generally incorporated. In [13], [14], [15], CCA was extended to correlation discriminant analysis (CDA) and discriminative canonical correlation analysis (DCCA) by maximizing the within-class correlation and minimizing between-class correlation across two-view. In [16], [17], multiview Fisher discriminant analysis (MFDA) was proposed to employ the label information for binary classification. In [18], the Fisher linear discriminant analysis is interpreted as CCA between appropriately defined vectors. In [19], common discriminant feature extraction (CDFE) was proposed to minimize the intra-class scatter and meanwhile maximize the inter-class separability, resulting in very encouraging performance. In [20], a large margin approach was proposed to discover a predictive latent subspace representation shared by two views based on an undirected latent space Markov network. In [21], coupled spectral regression (CSR) learnt a projection from the observation to the common low-dimensional embedding of the class label through least squares regression. Similarly, in [22], two coupled linear regression models were used to project data from different modalities into a common subspace that is directly defined by the class label. In [23], a local feature-based discriminant analysis method was proposed to match a forensic sketch and a mug shot photo, and also other effective features can be used such as [24]. Besides, some other methods proposed to apply discriminant classifier in the common space achieved from some unsupervised method, like [25]. Benefitted from the supervised information, these discriminant common space methods usually outperform the unsupervised ones.

Multi-view methods. As mentioned, the two-view methods essentially are only applicable to two-view scenario. To deal with multi-view cases, the pairwise strategy is usually exploited, resulting in multiple two-view models. However, in scenario of multiple views, a more efficient and robust solution is to learn a unified common space shared by all views rather than only two views. For

- *M. Kan, S. Shan, and X. Chen are with the Key Lab of Intelligent Information Processing, Chinese Academy of Sciences (CAS), Institute of Computing Technology (ICT), CAS, Beijing 100190, China. E-mail: {kanmeina, sgshan, xlchen}@ict.ac.cn.*
- *H. Zhang and S. Lao are with Omron Social Solutions Co., LTD., Kyoto, Japan. E-mail: haihong_zhang@oss.omron.co.jp, lao@ari.ncl.omron.co.jp.*

Manuscript received 23 May 2014; revised 3 Mar. 2015; accepted 4 May 2015. Date of publication 19 May 2015; date of current version 9 Dec. 2015.
Recommended for acceptance by M. Pantic.
For information on obtaining reprints of this article, please send e-mail to: reprints@ieee.org, and reference the Digital Object Identifier below.
Digital Object Identifier no. 10.1109/TPAMI.2015.2435740

0162-8828 © 2015 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

this purpose, the Multiview CCA (MCCA) [26], [27] was proposed to obtain one common space for v views. In MCCA, v view-specific transforms, one for each view, were obtained by maximizing the total correlations between any two views. However, MCCA did not take discriminant information into account, which may be not good for classification across views. Recently, a generalized multi-view analysis (GMA) framework was proposed in [28], in which the supervised information was incorporated, leading to a discriminant common subspace. Although GMA can obtain a discriminant common subspace, it only considered the intra-view discriminant information, but ignored the inter-view discriminant information. Some other methods attempted to decompose the variations of each view. In [29], a multilinear analysis method named Tensorfaces was proposed to decompose the modes due to identity, pose, and illumination. Furthermore, a multimodal discriminant analysis (MMDA) [30] method was proposed for discriminative multimodal decomposition based on the Fisher Criterion, thus favorable for multimodal pattern recognition.

Following the multi-view strategy for cross-view recognition, 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. Moreover, inspired by the observation that different views share similar structures, a constraint enforcing the consistency of the multiple linear transforms is introduced to achieve a more robust common space. Specifically, the between-class and within-class variations are formulated into a Rayleigh quotient, with which the v view-specific transforms can be solved analytically and simultaneously through generalized eigenvalue decomposition. Overal speaking, MvDA is a multi-view method, rather than pairwise two-view method; MvDA considers both inter-view and intra-view variations leading to a more discriminative common space; and MvDA can be solved analytically.

In the following, Section 2 introduces the related works, Section 3 presents the formulation of MvDA with some discussions on difference from previous works, and Section 4 evaluates the MvDA on three databases, followed by a conclusion.

2 Related Works

2.1 Canonical Correlation Analysis [4]

CCA attempts to find a common subspace where the samples from two views are most correlated. Formally, let $\mathcal{S}$ represent the samples from two views: $\mathcal{S} = \{(\mathbf{x}_{11}, \mathbf{x}_{12}), \ldots, (\mathbf{x}_{n1}, \mathbf{x}_{n2})\}$, where $\mathbf{x}_{ij} \in \mathbb{R}^{p_j}, j = 1, 2$, represents the ith sample from the jth view of p_j dimension. Two matrices $\mathbf{X}_1 = [\mathbf{x}_{11}, \ldots, \mathbf{x}_{n1}]$ and $\mathbf{X}_2 = [\mathbf{x}_{12}, \ldots, \mathbf{x}_{n2}]$ are defined to represent the data from the two views. Two linear transforms $\mathbf{w}_1$ and $\mathbf{w}_2$ are obtained to respectively project the samples from two views into the common subspace, by maximizing the correlation between $\mathbf{w}_1^T\mathbf{X}_1$ and $\mathbf{w}_2^T\mathbf{X}_2$ as below:

$$\begin{aligned} &\max_{\mathbf{w}_1,\mathbf{w}_2} \mathbf{w}_1^T\mathbf{X}_1\mathbf{X}_2^T\mathbf{w}_2 \\ &s.t.\ \mathbf{w}_1^T\mathbf{X}_1\mathbf{X}_1^T\mathbf{w}_1 = 1,\ \mathbf{w}_2^T\mathbf{X}_2\mathbf{X}_2^T\mathbf{w}_2 = 1. \end{aligned} \tag{1}$$

With the Lagrange multiplier, Eq. (1) can be solved by resorting to the eigenvalue decomposition.

With $\mathbf{w}_1$ and $\mathbf{w}_2$, the samples from two views can be compared after projecting to the common space. As an unsupervised approach, CCA can be considered as a two-view extension of PCA [31]. CCA is only designed for two-view case, and thus the pairwise strategy is needed when applied to the multi-view scenario. Another limitation of CCA is that the training data for CCA must be given in view-pair mode, i.e., the number of samples from both views should be the same to make $\mathbf{X}_1\mathbf{X}_2^T$ computable.

2.2 Multi-view CCA [26], [27]

In [27], CCA is further generalized for multi-view scenario termed as multi-view canonical correlation analysis (MCCA). The goal of MCCA is to find a set of linear transforms $\mathbf{w}_i|_{i=1}^v$, to respectively project the samples of v views $\{\mathbf{X}_1, \ldots, \mathbf{X}_v\}$ to one common space, i.e., $\{\mathbf{w}_1^T\mathbf{X}_1, \ldots, \mathbf{w}_v^T\mathbf{X}_v\}$. The total correlation in the common space is maximized as below:

$$\begin{aligned} &\max_{\mathbf{w}_1,\mathbf{w}_2,\ldots,\mathbf{w}_v} \sum_{i<j} \mathbf{w}_i^T\mathbf{X}_i\mathbf{X}_j^T\mathbf{w}_j \\ &s.t.\ \mathbf{w}_i^T\mathbf{X}_i\mathbf{X}_i^T\mathbf{w}_i = 1, i = 1, 2, \ldots, v, \end{aligned} \tag{2}$$

where $\mathbf{X}_i \in \mathbb{R}^{p_i \times n}$ is the data matrix of the ith view with n samples of p_i dimension. Like CCA, the number of samples in each view should be the same. Similarly, MCCA is also an unsupervised method.

In [26], CCA is also generalized to multi-view CCA, and several kinds of the characteristics calculated from the transformed variables are investigated. The method in [27] can be considered as a special case of [26] under the denoted constraint 3.

2.3 Generalized Multiview Analysis (GMA) [28]

In [28], a general framework for multiview analysis is proposed to achieve a discriminative common subspace for all views. The GMA aims at preserving the supervised structure of each view and meanwhile keeping the projections of different views close to each other in the latent common space as follows:

$$\begin{aligned} \max_{\mathbf{w}_1,\cdots,\mathbf{w}_v} \sum_{i=1}^{v} \mu_i\mathbf{w}_i^T\mathbf{A}_i\mathbf{w}_i + \sum_{i<j} 2\lambda_{ij}\mathbf{w}_i^T\mathbf{X}_i\mathbf{X}_j^T\mathbf{w}_j, \\ s.t.\ \textstyle\sum_i \gamma_i\mathbf{w}_i^T\mathbf{B}_i\mathbf{w}_i = 1, \end{aligned} \tag{3}$$

where μ_i, λ_{ij}, γ_i are balance parameters, $\mathbf{A}_i$ and $\mathbf{B}_i$ are the between-class and within-class scatter matrices from the ith view respectively. GMA can be regarded as an extension of Fisher Discriminant analysis [31] for cross-view matching.

GMA considers class label information in a multi-view manner which makes it efficient and discriminative for recognition across multiple views. However, GMA only employs the discriminant information within each individual view but without considering the discriminant information from the inter-view, and this may degenerate the performance of cross-view matching. Besides, GMA has about $\frac{v\times(v+3)-4}{2}$ parameters (i.e., those λ_{ij}, μ_i and γ_i) in case of v views, which means a lot of tedious parameter tuning.

3 Multi-View Discriminant Analysis

In this section, we firstly introduce the basic idea and formulation of MvDA and then present its analytic solution, followed by extended MvDA with view consistency (MvDA-VC). Finally, we discuss the differences of MvDA from previous methods. Please note that, for the sake of clarity, some of the detailed inferences are put in the supplemental material, which can be found on the Computer Society Digital Library at http://doi.ieeecomputersociety.org/10.1109/TPAMI.2015.2435740.

3.1 MvDA: Overview and Formulation

As shown in Fig. 1, our MvDA attempts to find v linear transforms $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_v$ that can respectively project the samples from v views to one discriminant common space, where the between-class variation is maximized while the within-class variation is

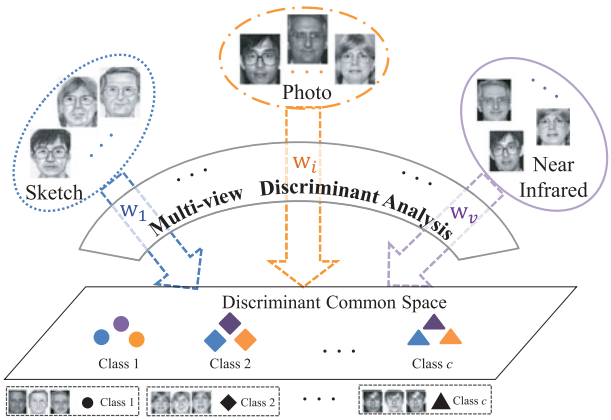

Fig. 1. The overview of Multi-view Discriminant Analysis. The samples from distinct views are projected into a discriminant common space by using v transforms, one for each view. Here, images from distinct views, e.g., photo, sketch, NIR are denoted in distinct colors and images from distinct classes are denoted in distinct shapes. (Best viewed in color)

minimized. For this purpose, formally, let us define $\mathcal{X}^{(j)} = \{\mathbf{x}_{ijk} | i = 1, \ldots, c; k = 1, \ldots, n_{ij}\}$ as the samples from the jth view ($j = 1, \ldots, v$), where $\mathbf{x}_{ijk} \in \mathbb{R}^{d_j}$ is the kth sample from the jth view of the ith class of d_j dimension, c is the number of classes and n_{ij} is the number of samples from the jth view of ith class.

The samples from v views are then projected to the same common space by using v view-specific linear transforms. We denote the projection results as $\mathcal{Y} = \{\mathbf{y}_{ijk} = \mathbf{w}_j^T \mathbf{x}_{ijk} | i = 1, \ldots, c; j = 1, \ldots, v; k = 1, \ldots, n_{ij}\}$. In the common space, according to our goal, the between-class variation $\mathbf{S}_B^y$ from all views should be maximized while the within-class variation $\mathbf{S}_W^y$ from all views should be minimized. We formulate this objective as a generalized Rayleigh quotient:

$$(\mathbf{w}_1^*, \mathbf{w}_2^*, \ldots, \mathbf{w}_v^*) = \arg \max_{\mathbf{w}_1, \cdots, \mathbf{w}_v} \frac{Tr(\mathbf{S}_B^y)}{Tr(\mathbf{S}_W^y)}. \tag{4}$$

Here, the within-class scatter matrix $\mathbf{S}_W^y$ and the between-class scatter matrix $\mathbf{S}_B^y$ of the projected samples in the common space are computed as below:

$$\mathbf{S}_W^y = \sum_{i=1}^{c} \sum_{j=1}^{v} \sum_{k=1}^{n_{ij}} (\mathbf{y}_{ijk} - \boldsymbol{\mu}_i)(\mathbf{y}_{ijk} - \boldsymbol{\mu}_i)^T, \tag{5}$$

$$\mathbf{S}_B^y = \sum_{i=1}^{c} n_i (\boldsymbol{\mu}_i - \boldsymbol{\mu})(\boldsymbol{\mu}_i - \boldsymbol{\mu})^T, \tag{6}$$

where $\boldsymbol{\mu}_i = \frac{1}{n_i} \sum_{j=1}^{v} \sum_{k=1}^{n_{ij}} \mathbf{y}_{ijk}$ is the mean of all the samples of the ith class over all views in the common space, $\boldsymbol{\mu} = \frac{1}{n} \sum_{i=1}^{c} \sum_{j=1}^{v} \sum_{k=1}^{n_{ij}} \mathbf{y}_{ijk}$ is the mean of all samples over all views in the common space, $n_i = \sum_{j=1}^{v} n_{ij}$ the number of samples of the ith class in all views, and $n = \sum_{i=1}^{c} n_i$ is the number of samples from all classes and all views.

From Eq. (5) and Eq. (6), it is clear that the within-class and between-class variations are computed from the samples of all views, not only intra-view ones but also inter-view ones. In other words, not only the discriminant information from the intra-view but also that from the inter-view are considered. After obtaining $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_v$ from Eq. (4), the samples from v views can be compared after respectively projected to the discriminant common space.

3.2 Analytical Solution of MvDA

Although Eq. (4) seems like LDA [31], it is much more complicated, as it needs to jointly optimize v distinct linear transforms. Fortunately, we work out an analytic solution by reformulating the trace ratio problem in Eq. (4) into a ratio trace problem.

Formally, the within-class scatter matrix in the common space in Eq. (5) can be reformulated as follows:

$$\mathbf{S}_W^y = \begin{bmatrix} \mathbf{w}_1^T \mathbf{w}_2^T \cdots \mathbf{w}_v^T \end{bmatrix} \begin{pmatrix} \mathbf{S}_{11} & \cdots & \mathbf{S}_{1v} \\ \vdots & \vdots & \vdots \\ \mathbf{S}_{v1} & \cdots & \mathbf{S}_{vv} \end{pmatrix} \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \\ \vdots \\ \mathbf{w}_v \end{bmatrix} = \mathbf{W}^T \mathbf{S} \mathbf{W}, \tag{7}$$

with $\mathbf{W} = \left[\mathbf{w}_1^T, \mathbf{w}_2^T, \ldots, \mathbf{w}_v^T\right]^T$ and $\mathbf{S}_{jr}$ is defined as below with $\boldsymbol{\mu}_{ij}^{(x)} = \frac{1}{n_{ij}} \sum_{k=1}^{n_{ij}} \mathbf{x}_{ijk}$:

$$\mathbf{S}_{jr} = \begin{cases} \sum_{i=1}^{c} \left(\sum_{k=1}^{n_{ij}} \mathbf{x}_{ijk} \mathbf{x}_{ijk}^T - \frac{n_{ij} n_{ij}}{n_i} \boldsymbol{\mu}_{ij}^{(x)} \boldsymbol{\mu}_{ij}^{(x)T} \right), & j = r \\ -\sum_{i=1}^{c} \frac{n_{ij} n_{ir}}{n_i} \boldsymbol{\mu}_{ij}^{(x)} \boldsymbol{\mu}_{ir}^{(x)T}, & \text{otherwise} \end{cases}. \tag{8}$$

Similarly, the between-class scatter matrix in Eq. (6) can be further reformulated as follows:

$$\mathbf{S}_B^y = \begin{bmatrix} \mathbf{w}_1^T \mathbf{w}_2^T \cdots \mathbf{w}_v^T \end{bmatrix} \begin{pmatrix} \mathbf{D}_{11} & \cdots & \mathbf{D}_{1v} \\ \vdots & \vdots & \vdots \\ \mathbf{D}_{v1} & \cdots & \mathbf{D}_{vv} \end{pmatrix} \begin{bmatrix} \mathbf{w}_1 \\ \mathbf{w}_2 \\ \vdots \\ \mathbf{w}_v \end{bmatrix} = \mathbf{W}^T \mathbf{D} \mathbf{W}, \tag{9}$$

where $\mathbf{W}$ is the same as above and $\mathbf{D}_{jr}$ is defined as:

$$\mathbf{D}_{jr} = \left(\sum_{i=1}^{c} \frac{n_{ij} n_{ir}}{n_i} \boldsymbol{\mu}_{ij}^{(x)} \boldsymbol{\mu}_{ir}^{(x)T} \right) - \frac{1}{n} \left(\sum_{i=1}^{c} n_{ij} \boldsymbol{\mu}_{ij}^{(x)} \right) \left(\sum_{i=1}^{c} n_{ir} \boldsymbol{\mu}_{ir}^{(x)} \right)^T, \tag{10}$$

With this, Eq. (4) can be reformulated as:

$$(\mathbf{w}_1^*, \mathbf{w}_2^*, \ldots, \mathbf{w}_v^*) = \arg \max_{\mathbf{w}_1, \cdots, \mathbf{w}_v} \frac{Tr(\mathbf{W}^T \mathbf{D} \mathbf{W})}{Tr(\mathbf{W}^T \mathbf{S} \mathbf{W})}. \tag{11}$$

According to [32], the objective in Eq. (11) is in the form of trace ratio, which implies the closed form solution does not exist. We therefore relax it into a more tractable one in the form of ratio trace:

$$(\mathbf{w}_1^*, \mathbf{w}_2^*, \cdots, \mathbf{w}_v^*) = \arg \max_{\mathbf{w}_1, \ldots, \mathbf{w}_v} Tr\left(\frac{\mathbf{W}^T \mathbf{D} \mathbf{W}}{\mathbf{W}^T \mathbf{S} \mathbf{W}} \right), \tag{12}$$

which can be solved analytically through generalized eigenvalue decomposition.

3.3 MvDA with View-Consistency

As multiple views actually correspond to the same objects, there should be some correspondence between multiple views. For example, if two views of human faces are taken from left 45° and right 45° (yaw), denoted as $\mathbf{X}_1$ and $\mathbf{X}_2$ respectively, each view should be the flipping of the other one, i.e., $\mathbf{X}_1 = \mathbf{R}\mathbf{X}_2$, where $\mathbf{R}$ is the transform matrix that can flip the image from left 45 to right 45° or vice versa. As a result, the transforms obtained from Eq. (12) for these two views should also have similar relationship, i.e., $\mathbf{w}_1 = \mathbf{R}\mathbf{w}_2$. Following Representer Theorem, the transform $\mathbf{w}_i$ for ith view can be equivalently formulated as follows:

$$\mathbf{w}_i = \mathbf{X}_i \boldsymbol{\beta}_i, \tag{13}$$

where $\boldsymbol{\beta}_i$ captures the structure of each $\mathbf{w}_i$.

Then, we can reach the following equivalence:

$$\mathbf{X}_1 \boldsymbol{\beta}_1 = \mathbf{R}\mathbf{X}_2 \boldsymbol{\beta}_2 = \mathbf{X}_1 \boldsymbol{\beta}_2, \tag{14}$$

which demonstrates that $\boldsymbol{\beta}_1 = \boldsymbol{\beta}_2$. In other words, the structure of each transform $\mathbf{w}_i$ captured by $\boldsymbol{\beta}_i$ is the same for different views.

Without loss of generality, if $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_v$ from v views correspond to the same underlying objects, they should have similar

structures, which implies that the structures of the transforms of different views should be also similar. This observation further implies that $\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_v$, which depict the transforms, should resemble mutually. In this work, we call this resemblance as *view-consistency*, modeled by the following term:

$$\sum\nolimits_{i,j=1}^{v} \|\boldsymbol{\beta}_i - \boldsymbol{\beta}_j\|_2^2. \tag{15}$$

We then minimize this term by adding it into the original denominator of Eq. (11) and reach the following new objective:

$$\begin{aligned}&(\mathbf{w}_1^*, \mathbf{w}_2^*, \ldots, \mathbf{w}_v^*)\\ &= \arg\max\nolimits_{\mathbf{w}_1,\ldots,\mathbf{w}_v} \frac{Tr\big(\mathbf{W}^T\mathbf{D}\mathbf{W}\big)}{Tr\big(\mathbf{W}^T\mathbf{S}\mathbf{W}\big) + \lambda\sum_{i,j=1}^{v}\|\boldsymbol{\beta}_i - \boldsymbol{\beta}_j\|_2^2},\end{aligned} \tag{16}$$

where λ is the balance parameter. We denote this extended MvDA as MvDA-VC.

At first glance, this new objective might make the optimization very complicated, but fortunately, it still has analytical solution as illustrated below.

From Eq. (13), $\boldsymbol{\beta}_i$ can be equivalently represented as:

$$\boldsymbol{\beta}_i = \big(\mathbf{X}_i^T\mathbf{X}_i\big)^{-1}\mathbf{X}_i^T\mathbf{w}_i \triangleq \mathbf{P}_i\mathbf{w}_i, \tag{17}$$

with $\mathbf{P}_i \triangleq \big(\mathbf{X}_i^T\mathbf{X}_i\big)^{-1}\mathbf{X}_i^T$. Then, Eq. (15) can be reformulated as follows:

$$\sum\nolimits_{i,j=1}^{v} \|\boldsymbol{\beta}_i - \boldsymbol{\beta}_j\|^2 = Tr\big(\mathbf{W}^T\mathbf{M}\mathbf{W}\big), \tag{18}$$

$$\mathbf{M} = \begin{pmatrix} \mathbf{M}_{11} & \cdots & \mathbf{M}_{1v} \\ \vdots & \vdots & \vdots \\ \mathbf{M}_{v1} & \cdots & \mathbf{M}_{vv} \end{pmatrix}, \mathbf{M}_{ij} = \begin{cases} 2(v-1)\mathbf{P}_i^T\mathbf{P}_i, & i = j \\ -2\mathbf{P}_i^T\mathbf{P}_j, & i \neq j \end{cases} \tag{19}$$

With Eq. (18), the objective of MvDA-VC in Eq. (16) can be rewritten as a trace ratio form:

$$(\mathbf{w}_1^*, \mathbf{w}_2^*, \ldots, \mathbf{w}_v^*) = \arg\max_{\mathbf{w}_1,\ldots,\mathbf{w}_v} \frac{Tr\big(\mathbf{W}^T\mathbf{D}\mathbf{W}\big)}{Tr\big(\mathbf{W}^T(\mathbf{S} + \lambda\mathbf{M})\mathbf{W}\big)}. \tag{20}$$

Evidently, Eq. (20) can also be solved analytically after relaxing to the ratio trace problem as Eq. (11).

3.4 Discussions

To further clarify the significance of our method, this section will discuss in details the difference between our MvDA and previous closely related methods.

Difference from other inter-view methods. In the proposed MvDA, both intra-view and inter-view variations are considered when calculating the within-class and between-class scatter matrices. To show this, we reformulate the within-class scatter in Eq. (5) as :

$$\begin{aligned}\mathbf{S}_W^y = &\sum\nolimits_{i=1}^{c}\Big(\sum\nolimits_{j=1}^{v}\sum\nolimits_{k=1}^{n_{ij}}\sum\nolimits_{l=1}^{n_{ij}}\big(\mathbf{y}_{ijk} - \mathbf{y}_{ijl}\big)\big(\mathbf{y}_{ijk} - \mathbf{y}_{ijl}\big)^T\\ &+ \sum\nolimits_{j=1}^{v}\sum\nolimits_{r=1,r\neq j}^{v}\sum\nolimits_{k=1}^{n_{ij}}\sum\nolimits_{l=l}^{n_{ij}}\big(\mathbf{y}_{ijk} - \mathbf{y}_{irl}\big)\big(\mathbf{y}_{ijk} - \mathbf{y}_{irl}\big)^T\Big).\end{aligned} \tag{21}$$

It can be easily seen that the first term models the variations within each view and the second term models the variations across view. Similarly, in our MvDA both intra-view and inter-view variations are considered in the between-class scatter matrix. In contrast, in most existing inter-view methods, e.g., CDFE and GMA, only part of the discriminative information in either intra-view or inter-view variations is considered. Another key difference is that our MvDA projects the samples from v views to a single common space, rather than C_v^2 common spaces as most previous two-view methods do.

Difference from metric learning methods. Many researchers argue that metric learning and dimension reduction are equivalent in some sense but with quite different primary objectives [3]. Our MvDA can be also considered to learn a metric computed as $d_{MvDA} = \sqrt{\big(\mathbf{w}_1^T\mathbf{X}_1 - \mathbf{w}_2^T\mathbf{X}_2\big)^T\big(\mathbf{w}_1^T\mathbf{X}_1 - \mathbf{w}_2^T\mathbf{X}_2\big)}$. However, for multi-view problem, our MvDA is superior to the cross-view metric learning methods. Specifically, for v view problem, the cross-view or heterogenous metric learning methods have to learn C_v^2 view-pair metrics, while MvDA needs only v transforms. In addition, in case of multi-class scenario, the cross-view or heterogenous metric learning methods usually need convert multi-class problem to two-class problem, while MvDA can naturally handle the multi-class problem more efficiently, benefited from the objective in the form of Rayleigh quotient.

Difference from MCCA [26], [27]. Both MCCA and MvDA obtain only one common space for multiple views. MCCA obtains a common space where only the correlation between views is maximized, but neither intra-view correlation nor class label is considered. Differently, MvDA endeavors to obtain a discriminant common space, which considers the discriminative information from both intra-view and inter-view.

*Difference from MFDA [16] and MMDA [30].*MFDA is originally designed for binary classification, and one-versus-all or hierarchical strategy are needed for multi-class scenarios. For the same reason, MFDA also requires the classes to classify should be the same as those in the training set. On the contrary, as a feature extraction method, our MvDA is originally designed for multi-view and multi-class scenario, and can even be applied to classes not presented in the training set. Both MMDA and MvDA are discriminant methods for multi-view or multimodal problem. MMDA individually decomposes each mode, which implies the between-class variation in one mode (e.g., pose) is contained in the within-class variations of another mode (e.g., expression), leading to a better performance than LDA. But also attributed to this property, we argue that it becomes difficult to eliminate all the factors irrelevant to identity. In contrast, our MvDA can remove all the identity-irrelevant factors and induce a more discriminant model.

Difference from GMA [28]. Both MvDA and GMA are discriminant multi-view methods for recognition across multiple views. However, MvDA is quite different from GMA in the following aspects: 1) In GMA, only the intra-view discriminant information is considered, while in MvDA both intra-view and inter-view discriminant information is considered. The inter-view discriminant information is especially important since object recognition across views is about the inter-view distinguishing. 2) In case of only single sample per class per view, the supervised GMA will fail to work because it is impossible to compute the within-class variations with single sample per class. On the contrary, in this case our MvDA can still work since the within-class variations are computed from all views rather than single view. 3) GMA has many (about $\frac{v\times(v+3)-4}{2}$) free parameters to tune, which can be very tedious especially in case of large number of views. In contrast, our MvDA has no parameter to tune, while MvDA-VC has only one balance parameter, thus much easier to use in practice.

4 Experiments

In this section, MvDA is evaluated on three heterogeneous face recognition tasks, i.e., face recognition across pose, photo vs. sketch recognition and visual light (VIS) image versus near infrared (NIR) image recognition respectively on three datasets.

TABLE 1
The Performance of MvDA-VC w.r.t. λ in Terms of Mean Accuracy (mACC) on Multi-PIE Dataset

λ	0	0.001	0.01	0.02	0.03	0.1	0.2	0.3
mACC	95.0%	95.3%	96.0%	96.2%	96.3%	95.9%	95.5%	95.0%

TABLE 2
Evluation on Multi-PIE Dataset in Terms of Mean Accuracy (mACC)

Pairwise Methods				Multi-view Methods						
PW-CCA [4]*	CDFE [19]	CSR [21]	PLS [6]	U-LDA [31]	MMDA [30]	MCCA [26]	MCCA[26]+LDA	GMA [28]	**MvDA**	**MvDA-VC**
83.7%	88.8%	72.0%	77.4%	84.3%	86.9%	91.6%	92.6%	92.0%	**95.0%**	**96.3%**

In [36], the constraints in Eq. (1) were not enforced to satisfy. Here, it is corrected and thus the results are slightly different.

4.1 Datasets

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

CUHK Face Sketch FERET (CUFSF) dataset [8], [33] is used to evaluate photo versus sketch face recognition. CUFSF consists of the images from 1,194 subjects from FERET dataset [34] with lighting variations. For each subject, only one photo is available and a sketch is drawn with shape exaggeration according to each photo. On this dataset, the images from the first 700 subjects are used for training and images from the rest 494 subjects are used for testing.

Heterogeneous Face Biometrics (HFB) dataset [2] is used to evaluate Visual (VIS) light image versus NIR image heterogeneous recognition. This dataset contains images from 100 subjects, with four NIR and four VIS images per subject. The evaluation follows the standard Protocol II, i.e., the images from the first 70 subjects are used as training data and the images from the rest 30 subjects are used as testing data.

4.2 Experimental Settings

All images from Multi-PIE and CUFSF are cropped into 64 x 80 pixels and images from HFB dataset are cropped into 32×32 according to the standard protocol. In all experiments, each image is represented as a column vector by vectorizing its grey intensities. The proposed MvDA is compared to most related existing methods including Pairwise CCA (PW-CCA) [4], CDFE [19], CSR [21], PLS [6], Unified LDA [31] (U-LDA), Multiset CCA (MCCA) [26], MCCA [26]+LDA, MMDA [30] and GMA [28]. Among them, PW-CCA, CDFE, CSR and PLS are two-view methods; therefore we exploit the pairwise strategy for multi-view classification. The so called U-LDA is the traditional Fisherface [31] regardless of the view discrepancy. For CDFE, the parameter α and β are traversed in [0.01 1] and [0.0001 1] respectively to report the best result. For CSR, the parameter λ and η are traversed in [0.01 10] to obtain a best result. For GMA, following the suggestions in [28] we set $\mu = 1$, $\gamma =$ trace ratio, and tune the λ in [1 100]. For our MvDA without view consistency, there is no parameter needed to tune. For our MvDA-VC, the balance parameter λ is traversed in [0.001 0.3], and an illustration of the performance w.r.t different λ on Multi-PIE dataset is shown in Table 1. As seen, MvDA performs better when with a stronger constraint, but begins to degrade when with a very large constraint. This is because the structures of different views are similar but not exactly the same, which thus prefers a moderate constraint.

To reduce dimensionality, principal component analysis (PCA) [35] is first applied for all methods. For CCA, CDFE, CSR and PLS, the PCA dimensions are empirically set to achieve the best recognition accuracies via traversing possible dimensions. In contrast, for all other comparative methods, the reduced dimension is set to 100, 100 and 80 to preserve more than 95 percent energy on Multi-PIE, CUFSF and HFB datasets respectively.

4.3 Face Recognition Across Pose

Face recognition across pose is evaluated on Multi-PIE dataset by taking each pose as one view. The testing is conducted in pairwise manner, i.e., the images from one view are used as gallery while the ones from another view are used as probe. Here, note that the gallery and probe contain the enrolled faces to be recognized, which has no overlap with the training set for MvDA learning. The samples in Multi-PIE are from seven views, thus leading to $7 \times 6 = 42$ evaluations in terms of rank-1 recognition rate (as in the supplemental material, available online). Then, all 42 results are averaged as the mean accuracy (mACC) as shown in Table 2.

As seen, CCA and PLS perform poorly, which we argue can be ascribed to their ignorance of supervised information. Furthermore, CDFE that considers supervised information performs better. Although CSR is supervised method, it performs unexpectedly badly on this dataset, which may be due to the difficult regression between the appearance and the class labels.

Compared with the pairwise two-view methods, the multi-view ones such as MCCA, MCCA+LDA, GMA and MvDA perform much better. As seen, compared with PW-CCA, MCCA can significantly improve the recognition accuracy, up to 7.9 percent in terms of mACC. The unsupervised MCCA even outperforms the U-LDA which exploits the supervised information but disregards the view information. The method MCCA+LDA performs better, but still not good enough. We attribute this inferiority to the separately modeling of the cross-view gap and discriminancy, leading to some discriminancy loss in the first step (MCCA) that cannot be recalled by LDA anymore. Furthermore, GMA performs better than MCCA since it exploits the discriminant information within each view.

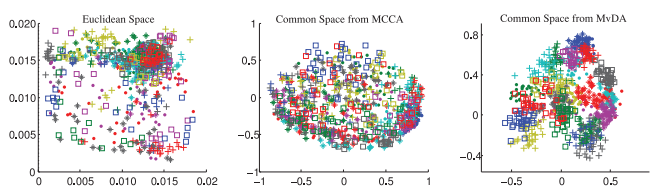

Fig. 2. The 2D embeddings of Euclidean space, common space from MCCA and MvDA for the samples from 7 views on Multi-PIE dataset. Different classes are denoted in different colors and shapes.

TABLE 3
Evaluation on CUFSF and HFB Datasets in Terms of rank-1 Recognition Rate

		CCA[4]**	CCA[4]+LDA	CDFE[19]	CSR[21]	PLS[6]	U-LDA[31]	GMA[28]	**MvDA**	**MvDA-VC**
CUFSF	Photo-Sketch	45.5%	45.0%	45.6%	50.2%	48.6%	46.8%	-	**53.4%**	**56.3%**
	Sketch-Photo	47.5%	50.6%	47.6%	49.0%	51.0%	53.4%	-	**55.5%**	**61.5%**
HFB	NIR-VIS	36.7%	40.0%	40.8%	26.7%	38.3%	39.1%	47.5%	**53.3%**	**59.2%**
	VIS-NIR	30.0%	40.0%	36.7%	32.5%	40.8%	40.0%	45.0%	**50.0%**	**59.2%**

**In [36], the constraints in Eq. (1) were not enforced to satisfy. Here, it is corrected and thus the results are slightly different.*

Furthermore, our MvDA even without view-consistency outperforms GMA with an absolute improvement up to 3.0 percent, due to the employment of more discriminant information embedded in both inter-view and intra-view variations. By adding the view-consistency, the MvDA-VC further improves the mean accuracy by 1.3 percent, which demonstrates the effectiveness of the view-consistency. It is worth noting that the above-mentioned improvements are very significant since they are the average of 42 cross-view accuracies (as shown in the supplemental material, available online). Especially, our MvDA and MvDA-VC outperform the competitive methods more significantly in case of larger pose deviations. For example, in case of large pose deviations, the improvements of MvDA-VC over MCCA and GMA are as high as 11.9 and 15.6 percent respectively. Fig. 2 displays the common spaces obtained by the MCCA and our MvDA for samples from seven views. As seen, the common space obtained by our method is more compact and discriminative.

4.4 Photo versus Sketch Face Recognition

Face recognition across photo and sketch is evaluated on CUFSF dataset. The results are shown in Table 3. Please note that on this dataset GMA fails to work, since there is only one sample per class per view. Besides, in this two-view case MCCA degenerates to PW-CCA.

As seen, CCA performs the worst and PLS performs much better benefited from the consideration of the intra-view variations besides the inter-view correlation considered in CCA. Moreover, CDFE and CSR also outperform CCA by employing the discriminant information. As expected, our MvDA and MvDA-VC performs the best. Especially, the MvDA-VC achieves improvements of 6.1 percent for Photo-Sketch and 8.1 percent for Sketch-Photo compared with the best results of competitive methods (i.e., CSR and U-LDA respectively). One can also find that MvDA-VC again achieves impressive gain over MvDA, which further validates the effectiveness of the view-consistency regularizer.

4.5 Visual Light versus Near Infrared Recognition

We also test MvDA for heterogeneous face recognition on HFB dataset. As in face recognition across photo and sketch, the samples in HFB dataset are only from two views, visual light image and near infrared image. The comparisons are shown in Table 3.

From Table 3, the same conclusion can be drawn even more safely. Besides, on this dataset MvDA achieve much larger improvements over all competitive methods than that on CUFSF, e.g., MvDA-VC has improved the recognition rates to 59.2 percent from the best known results of GMA. The larger gain can be attributed to the more intra-view discriminative information exploited by our MvDA, since there are more (i.e., 4) images per view per subject on HFB than that (i.e., 1) on CUFSF.

From the above evaluations, it can be seen that the common space obtained by the multi-view methods is more suitable for multi-view classification by jointly modeling multiple views. Furthermore, by taking advantages of both intra-view and inter-view variations, MvDA can obtain a more discriminant common space shared deeply by multiple views.

5 Conclusions and Future Works

To address the object recognition from multiple views problem, this work developed a multi-view discriminant analysis method that can obtain single discriminant common space shared by all views, in which the samples from different views can be readily matched. By exploiting both the intra-view and inter-view correlations, MvDA achieves better discriminability and generalizability. The problem is formulated as a generalized Rayleigh quotient leading to an analytical solution. Experiments on three heterogeneous face recognition tasks demonstrate the superiority of our method over the existing methods.

Obviously, our MvDA can be easily kernelized in future. We will also extend MvDA by modeling how each view originates from the commonality.

Acknowledgments

This work was partially supported by 973 Program under contract No. 2015CB351802, Natural Science Foundation of China under contracts Nos. 61173065, 61222211, 61402443 and 61390511. 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. S. Shan is the corresponding author.

References

[1] R. Gross, I. Matthews, J. Cohn, T. Kanada, and S. Baker, "The CMU multi-pose, illumination, and expression (multi-pie) face database," Robotics Inst., Carnegie Mellon Univ., Pittsburgh, PA, USA, Tech. Rep. TR-07-08, 2007.

[2] S. Z. Li, Z. Lei, and M. Ao, "The HFB face database for heterogeneous face biometrics research," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2009, pp. 1–8.

[3] A. Bellet, A. Habrard, and M. Sebban, "A survey on metric learning for feature vectors and structured data," arXiv preprint arXiv:1306.6709, Jun. 2013, http://arxiv.org/abs/1306.6709

[4] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, nos. 3/4, pp. 321–377, 1936.

[5] S. Akaho, "A kernel method for canonical correlation analysis," in *Proc. Int. Meeting Psychometric Soc.*, 2001.

[6] A. Sharma and D. W. Jacobs, "Bypassing synthesis: Pls for face recognition with pose, low-resolution and sketch," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 593–600.

[7] X. Cai, C. Wang, B. Xiao, X. Chen, and J. Zhou, "Regularized latent least square regression for cross pose face recognition," in *Proc. Int. Joint Conf. Artif. Intell.*, 2013, pp. 1247–1253.

[8] W. Zhang, X. Wang, and X. Tang, "Coupled information-theoretic encoding for face photo-sketch recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2011, pp. 513–520.

[9] S. Wang, L. Zhang, Y. Liang, and Q. Pan, "Semi-coupled dictionary learning with applications to image super-resolution and photo-sketch image synthesis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 2216–2223.

[10] D.-A. Huang and Y.-C. F. Wang, "Coupled dictionary and feature space learning with applications to cross-domain image synthesis and recognition," in *Proc. Int. Conf. Comput. Vis.*, 2013, pp. 2496–2503.

[11] X. Tang and X. Wang, "Face sketch recognition," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 14, no. 1, pp. 50–57, Jan. 2004.

[12] Q. Liu, X. Tang, H. Jin, H. Lu, and S. Ma, "A nonlinear approach for face sketch synthesis and recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2005, vol. 1, pp. 1005–1010.

[13] Y. Ma, S. Lao, E. Takikawa, and M. Kawade, "Discriminant analysis in correlation similarity measure space," in *Proc. Int. Conf. Mach. Learn.*, 2007, pp. 577–584.
[14] T. Sun, S. Chen, J. Yang, and P. Shi, "A novel method of combined feature extraction for recognition," in *Proc. IEEE Int. Conf. Data Mining*, 2008, pp. 1043–1048.
[15] T.-K. Kim, J. Kittler, and R. Cipolla, "Learning discriminative canonical correlations for object recognition with image sets," in *Proc. Eur. Conf. Comput. Vis.*, 2006, pp. 251–262.
[16] T. Diethe, D. R. Hardoon, and J. Shawe-Taylor, "Multiview Fisher discriminant analysis," in *Proc. NIPS Workshop Learn. Multiple Sources*, 2008.
[17] T. Diethe, D. R. Hardoon, and J. S. Taylor, "Constructing nonlinear discriminants from multiple data views," in *Proc. Eur. Conf. Mach. Learn. Knowl. Discovery Databases*, 2010, pp. 328–343.
[18] F. R. Bach and M. I. Jordan, "A probabilistic interpretation of canonical correlation analysis," Dept. Statist., Univ. California, Berkeley, CA, USA, Tech. Rep. 688, 2005.
[19] D. Lin and X. Tang, "Inter-modality face recognition," in *Proc. Eur. Conf. Comput. Vis.*, 2006, pp. 13–26.
[20] N. Chen, J. Zhu, and E. P. Xing, "Predictive subspace learning for multiview data: A large margin approach," in *Proc. Adv. Neural Inf. Process. Syst.*, 2010, vol. 23, pp. 361–369.
[21] Z. Lei and S. Z. Li, "Coupled spectral regression for matching heterogeneous faces," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2009, pp. 1123–1128.
[22] K. Wang, R. He, W. Wang, L. Wang, and T. Tan, "Learning coupled feature spaces for cross-modal matching," in *Proc. Int. Conf. Comput. Vis.*, 2013, pp. 2088–2095.
[23] B. F. Klare, Z. Li, and A. K. Jain, "Matching forensic sketches to mug shot photos," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 3, pp. 639–646, Mar. 2011.
[24] G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, "Subspace learning from image gradient orientations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 12, pp. 2454–2466, Dec. 2012.
[25] W. Li and X. Wang, "Locally aligned feature transforms across views," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2013, pp. 3594–3601.
[26] A. A. Nielsen, "Multiset canonical correlations analysis and multispectral, truly multitemporal remote sensing data," *IEEE Trans. Image Process.*, vol. 11, no. 3, pp. 293–305, Mar. 2002.
[27] J. Rupnik and J. Shawe-Taylor, "Multi-view canonical correlation analysis," in *Proc. Slovenian KDD Conf. Data Mining Data Warehouses*, 2010, pp. 1–4.
[28] A. Sharma, A. Kumar, H. Daume III, and D. W. Jacobs, "Generalized multiview analysis: A discriminative latent space," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 2160–2167.
[29] M. A. O. Vasilescu and D. Terzopoulos, "Multilinear analysis of image ensembles: Tensorfaces," in *Proc. Eur. Conf. Comput. Vis.*, 2002, pp. 447–460.
[30] T. Sim, S. Zhang, J. Li, and Y. Chen, "Simultaneous and orthogonal decomposition of data using multimodal discriminant analysis," in *Proc. Int. Conf. Comput. Vis.*, 2009, pp. 452–459.
[31] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.
[32] H. Wang, S. Yan, D. Xu, X. Tang, and T. Huang, "Trace ratio vs. ratio trace for dimensionality reduction," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2007, pp. 1–8.
[33] X. Wang and X. Tang, "Face photo-sketch synthesis and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 11, pp. 1955–1967, Nov. 2009.
[34] P. Phillips, H. Wechslerb, J. Huangb, and P. J. Rauss, "The FERET database and evaluation procedure for face-recognition algorithms," *Image Vis. Comput.*, vol. 16, no. 5, pp. 295–306, 1998.
[35] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 1991, vol. 591, pp. 586–591.
[36] M. Kan, S. Shan, H. Zhang, S. Lao, and X. Chen, "Multi-view discriminant analysis," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 808–821.