Robust Multiview Data Analysis Through Collective Low-Rank Subspace

Zhengming Ding, *Student Member, IEEE*, and Yun Fu, *Senior Member, IEEE*

***Abstract*— Multiview data are of great abundance in real-world applications, since various viewpoints and multiple sensors desire to represent the data in a better way. Conventional multiview learning methods aimed to learn multiple view-specific transformations meanwhile assumed the view knowledge of training, and test data were available in advance. However, they would fail when we do not have any prior knowledge for the probe data's view information, since the correct view-specific projections cannot be utilized to extract effective feature representations. In this paper, we develop a collective low-rank subspace (CLRS) algorithm to deal with this problem in multiview data analysis. CLRS attempts to reduce the semantic gap across multiple views through seeking a view-free low-rank projection shared by multiple view-specific transformations. Moreover, we exploit low-rank reconstruction to build a bridge between the view-specific features and those view-free ones transformed with the CLRS. Furthermore, a supervised cross-view regularizer is developed to couple the within-class data across different views to make the learned collective subspace more discriminative. Our CLRS makes our algorithm more flexible when addressing the challenging issue without any prior knowledge of the probe data's view information. To that end, two different settings of experiments on several multiview benchmarks are designed to evaluate the proposed approach. Experimental results have verified the effective performance of our proposed method by comparing with the state-of-the-art algorithms.**

***Index Terms*— Low rank, multiview data, transfer learning.**

I. Introduction

MULTIVIEW data analysis has attracted a great deal of attention recently [1]–[10], since multiview data are frequently seen in reality. Take face image as an example. Various viewpoints would generate cross-pose face images while different devices would generate different modalities, e.g., low-resolution face taken by a cellphone or even collected with near-infrared sensor. This results in the difficult issue that face images could be from various viewpoints, even heterogeneous [1]–[4], [11]. Such data with large view divergence would result in a challenging learning problem, in which data lying in different views show a large divergence, and therefore, they cannot be directly compared. In general, different views can be treated as different domains drawn from different distributions. Therefore, it is the key to adapt one view to another view to minimize the distribution divergences across them. In this paper, we mainly focus on the specific multiview learning problem, in which data have the same feature set but different probability distributions, e.g., the multipose image classification and multimodal image classification.

In general, three categories of techniques are proposed to deal with the multiview data problems, i.e., feature adaptation [12], [13], classifiers adaptation [14], and deep learning [15]. Specifically, feature adaptation methods are designed to find a common view-free space, in which the multiview data could be aligned well, while classifier adaptation approaches tend to generalize classifiers trained on some specific views to others. Deep learning algorithms focus on constructing deep structures to extract more discriminative features shared by different views to mitigate the view divergence. Our algorithm follows feature adaptation fashion, specifically the subspace learning scenario.

Conventional multiview subspace methods [2], [13] were developed to seek many view-specific projections, which transform different views into a common view-free space. Along this line, canonical correlation analysis (CCA) [16] was the most representative one, which learned two projections, each for one view, to align two-view data into the shared space, respectively. Furthermore, multiview CCA [17] was proposed and extended to multiple view cases based on CCA. Following this, Kan *et al.* [13] designed a multiview discriminant analysis (MvDA) algorithm, which sought an effective shared space by joining multiple view-specific linear projections learning and Fisher constraint in a unified framework. One common drawback is that those previous studies mainly dealt with the multiview learning tasks by applying one labeled view to predict another unlabeled view. Hence, we have to know the view knowledge of training and test data ahead of time. Only with view information at hand can the view-specific projections be adopted to the exact views; therefore, we need a lot of prior knowledge in real-world multiview learning scenarios.

Unfortunately, we cannot always obtain the test data's view information in advance at many real-world scenarios, since the test data are always accessible during evaluation. For example, a face image could be captured at running time with view-unknown camera so that we cannot get its exact view knowledge. In such cases, conventional multiview learning methods cannot work, since they only built multiple view-specific projections during training stage [2], [13], [16], which

Manuscript received May 7, 2016; revised October 30, 2016, January 3, 2017, and March 30, 2017; accepted March 31, 2017. Date of publication April 17, 2017; date of current version April 16, 2018. This work was supported in part by the NSF IIS Award under Grant 1651902, in part by the NIJ Graduate Research Fellowship under Grant 2016-R2-CX-0013, in part by the ONR Young Investigator Award under Grant N00014-14-1-0484, and in part by the U.S. Army Research Office Young Investigator Award under Grant W911NF-14-1-0218. *(Corresponding author: Zhengming Ding.)*

Z. Ding is with the Department of Electrical and Computer Engineering, Northeastern University, Boston, MA 02115 USA (e-mail: allanding@ece.neu.edu).

Y. Fu is with the Department of Electrical and Computer Engineering, College of Computer and Information Science, Northeastern University, Boston, MA 02115 USA (e-mail: yunfu@ece.neu.edu).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNNLS.2017.2690970

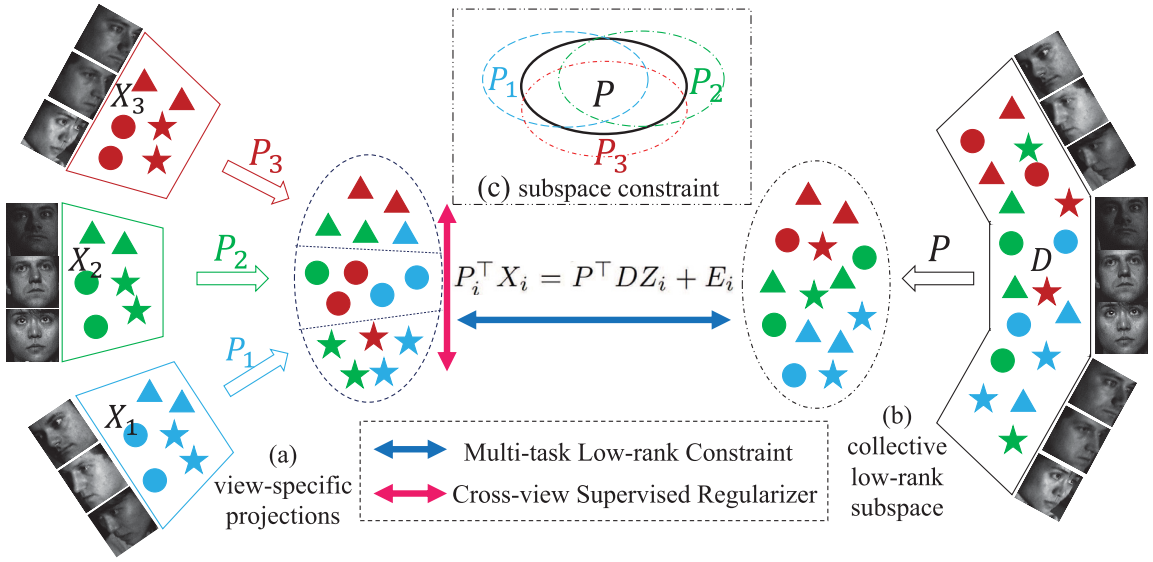


Fig. 1. Framework of our proposed CLRS algorithm. Here, we show three views (poses) and the same color represents the same view, while each view consists of three classes (the same shape denotes the data with same label). (a) We still adopt multiple view-specific projections $\{P_1, P_2, P_3\}$ for three views $\{X_1, X_2, X_3\}$ during the training stage. (b) Aiming to address the view-unknown testing data, we seek a surrogate by learning a low-rank shared transformation P for data D with three mixed views. (c) To further uncover more shared knowledge across multiple views, we adopt low-rank sparse decomposition to make the common P capture more shared information across those view-specific ones ($P_i = P + S_i$, where S_i is the sparse residue of the i-th view projection). Furthermore, multitask low-rank reconstruction $P_i^\top X_i = P^\top DZ_i + E_i$ is adopted to capture the classwise structures of the data so that the data projected with view-specific projections $P_i^\top X_i$ tend to be correlated to the mixed data drawn from the shared subspace $P^\top D$ in a classwise fashion. To this end, our proposed CLRS is able to uncover more view-invariant knowledge with the extra view information. With a supervised Fisher term $\Omega(P, Z)$ on reconstructed clean data $P^\top DZ = [P^\top DZ_1, \ldots, P^\top DZ_k] \approx [P_1^\top X_1, \ldots, P_k^\top X_k]$, our CLRS can well align cross-view data within the same class to make the learned CLRS P more discriminative.

are not helpful for each view-known test data. Another phenomenon is that the test images can be in the same distribution with the training data or totally different distributions from the training data. This leads to two scenarios: "traditional multiview learning" and "multiview transfer learning." When fighting off the multiview data with no prior knowledge either view information or label knowledge or both, we can ask help from an auxiliary multiview sources to facilitate the learning problem. In this scenario, transfer learning [18] has shown appealing performance in dealing with such a challenge. Along this line, feature adaption is a popular strategy in transfer learning, which aims to extract effective domain invariant features to reduce the domain shift so that the source knowledge could be transferred to the target [19]–[21].

Furthermore, low-rank modeling [22]–[24] has been well exploited in transfer learning [20] and robust subspace learning [25], [26] in the recent years. Low-rank constraint originally helps uncover the global structure of the data and detect noise or outliers. Robust subspace learning unifies low-rank modeling and dimensionality reduction to a framework by leveraging the merit of both [25], while low-rank transfer learning algorithms aim to uncover the intrinsic structure across source and target domains, which means each cluster in source domain is only reconstructed by one cluster in the target domain [20], [21]. To this end, marginal and conditional distribution discrepancy across source and target domains would be mitigated. Therefore, low-rank transfer learning can be an appealing data alignment tool for different distributions. In this way, low-rank reconstruction can build a bridge between view-known data and mixed view data, either in robust subspace learning or transfer learning scenario when addressing the multiview challenge.

In this paper, we develop a novel multiview learning algorithm, named collective low-rank subspace (CLRS), to deal with the challenge where the view knowledge of the test data is unavailable during the learning task (Fig. 1). Following conventional multiview subspace learning algorithms [2], [13], [17], we also learn the view-specific transformations for view-known training data to project the data into a latent view-free space in the training stage. Since we do not know the probe data's view information, we need to find a surrogate to preserve as much class information as possible, meanwhile reducing the impact of view divergence for mixed view-unknown test data, either in the same distribution or different distributions. On the account, the multiple view-specific projections all preserve the within-class knowledge for its specific view. In other words, those view-specific projections should have the similar discriminability for classification in different views. In other words, it is essential to find the consistent knowledge across multiple view-specific projections for view-unknown test data. This is also the core idea and uppermost contribution of this paper.

A. Our Contributions

To seek a more effective projection for view-unknown test data, we employ a collective low-rank projection to uncover most of the compatible structure across multiple view-specific projections, which are decomposed into the common part and sparse unique parts. To this end, our proposed algorithm is more flexible to solve real-world multiview problems when we

cannot have the view or even label information for the probe data at hand. Finally, we summarize our key contributions in three folds.

1) A CLRS is built through multiple view-specific projections by integrating unique parts into sparsity, so that our CLRS uncovers more shared information across different views. With low-rank constraints employed between the view-specific transformed data and commonly projected data under a multitask scheme, our method digs out more intrinsic information by gathering cross-view data within the same class together.
2) A cross-view discriminative regularizer is incorporated to align new representations of view-specific data from different tasks. This regularizer aims to align within-class data across multiple views by making full use of the supervised information at hand, therefore, our model can learn a more discriminative and robust collective subspace for multiview data analysis.
3) CLRS is a general method, which could be simply generalized to various learning problems (for example, conventional multiview learning and multiview transfer learning), by adapting different inputs. Two scenarios of experiments on several multiview benchmarks are conducted to evaluate the performance of our algorithm.

The remaining parts of this paper are presented as follows. In Section II, we provide a brief review of the related works and discuss the difference between theirs and ours. We present our novel CLRS algorithm in Section III, as well as the solution and complexity analysis of our method. Experimental analyses on two different scenarios are shown in Section IV, followed with the conclusion in Section V.

II. Related Work

In this section, we first briefly review several related works, and then, we highlight the difference of this paper to the related ones.

Conventional multiview data analysis assumes that the view knowledge of the test data is known ahead. Therefore, its setting is usually using one or more views with labels to predict another view. Those are the popular topics belonging to this scenario: cross-pose image recognition, heterogeneous image recognition, and domain adaptation/transfer learning. Generally, there are three strategies: feature adaptation [12], [13], classifiers adaptation [14], and deep learning [15].

The most popular way is adapting feature space, which is usually achieved by seeking for a shared space, through either subspace learning [12], [13] or dictionary learning [27]. For instance, Kan *et al.* [13] designed a discriminative multiview analysis model by seeking multiple view-specific projections under Fisher criteria. Zheng and Jiang [27] presented an approach to jointly build a set of view-specific dictionaries and a common dictionary, where view-specific sparse features and the view-shared sparse features are well aligned to transfer knowledge across multiple views and mitigate the view divergence.

Many methods attempt to adjust the classifier or boundary to fit the target data. A straightforward way is to model a shared classifier among different tasks, i.e., multitask learning. Recently, SVM has been widely discussed on transfer learning problems, such as remote sensing, images recognition, and video analysis, where either loss function or regularizer of SVM, or both of them are reformulated according to the specific problem. Hoffman *et al.* [28] presented a novel multi-domain model with mixed transformations, which then developed a constrained hierarchical clustering strategy to successfully uncover latent domains.

A deep structure for nonlinear representation learning is capable of disentangling different explanatory factors of variation behind data [29], which has been applied to formulate novel learning framework for multiview data analysis [30], [15]. Zhu *et al.* [30] aimed to recover the canonical view of face images by selecting a representative image for each identity, which is taken in the frontal view, under neutral lighting condition and with high resolution. Zhu *et al.* [15] designed a novel deep neural network to untangle the identity and view features.

However, we cannot always know the view information of the test data in reality. This paper is the extension to our previous conference work [12]. Specifically, our previous work [12] is a weakly supervised algorithm, which means we only need to know the view knowledge of the training data. It then builds multiple view-specific projections for training data, meanwhile a shared low-rank projection decomposed from them. Furthermore, the view-specific projections and the common one are integrated into a unified low-rank reconstruction framework. In the extension, we propose a supervised algorithm by making full advantage of the label information and the view information. The major difference is a novel cross-view supervised regularizer, which is developed to align multiple views to capture more discriminative information. As a result, another difference is the solution to those two algorithms. Specifically, in our previous work, we stacked multiple view-specific projections together as one to achieve the common projection, which is more stable, since there are less variables to be optimized. However, this solution may cost more computation, especially when involving more views' data. In the extension, we adopt multitask learning technique to solve the problem. Therefore, we can optimize multiple view-specific projections in a parallel manner to avoid a higher dimensional subspace optimization. However, it would introduce more variables to be optimized so that it becomes more flexible to achieve the local optimal solutions. Furthermore, we exploit a gradient descent strategy to optimize P instead of two-separate way. In the journal extension, we present more experimental results to verify our proposed algorithm on more data sets.

III. Robust Multiview Data Analysis

In this part, we briefly present our motivation, and then provide our CLRS for robust multiview learning. Finally, we design the optimization solution and complexity analysis.

A. Motivation

Research efforts on multiview learning are exploited to seek a shared view-free representation by seeking multiple

view-specific projections, so that the view divergence in the observed space could be well mitigated [2], [13]. However, conventional works only build multiple view-specific projections, since they all assume that the training and test data's view knowledge is already accessible. Unfortunately, we could confront such challenges in which the probe data's view knowledge is unknown ahead of time. For this reason, conventional multiview learning algorithms cannot work in such cases, because only multiple view-specific transformations are learned, which would be invalid for the view-unknown test data.

Fortunately, data collected from various views share the same class information, and therefore, each view-specific projection should have similar discriminability to separate different classes for individual view. Kan *et al.* [13] mentioned that the structure of each projection for multiview data is similar, that is, view consistency. In this way, it is reasonable to consider that the multiple view-specific projections should share a lot of consistent information [29]. Furthermore, Zhu *et al.* [15] provided an observation that "the identity representations of the same category tend to be similar, even though the data are collected in very different views, while the view representations of images in the same view are close, although they are across various identities."

To that end, we attempt to seek a collective projection to uncover more common intrinsic knowledge across multiple views. Therefore, in this paper, we assume that the multiple view-specific projections share a lot of consistent information, constrained to be low rank. Hence, the common low-rank projection can be well generalized to the view-unknown test data. Moreover, some recent works [20], [25] were designed to build a robust subspace through low-rank reconstruction to make merit from both techniques. Thus, it is helpful to incorporate low-rank reconstruction to mitigate the distribution shift between the view-specific features and the shared features. In other words, low-rank reconstruction would guide that the view-specific features from the same category tend to be correlated to the view-free features within the same class. Such strategy would capture the global structure of the data with the help of extra prior view knowledge.

B. Collective Low-Rank Subspace

Assume we have k-view training data as $X = [X_1, \ldots, X_k]$, and each view $X_i \in \mathbb{R}^{d\times m}$ contains the same c classes with m data samples. The view-specific transformation $\bar{\mathbf{P}}_i \in \mathbb{R}^{d\times d}$ would be learned for the ith view X_i following the conventional multiview learning [13]. Hence, each $\bar{\mathbf{P}}_i$ represents the basis to expand the space of each view X_i, i.e., $\bar{\mathbf{P}}_i = X_i A_i$, where A_i is the weight matrix [13]. As discussed before, multiple view-specific projections have the similar discriminability in their own view so that they should have a lot of shared knowledge. Then, we manage to seek as many common bases as possible across multiview data so that such common basis can be generalized to view-unseen test data. To this end, we adopt a collective low-rank transformation $\bar{\mathbf{P}} \in \mathbb{R}^{d\times d}$ to uncover such consistent knowledge that it can be extended to work for view-unknown test data. Specifically, we exploit low-rank sparse decomposition by assuming each $\bar{\mathbf{P}}_i$ is combined of $\bar{\mathbf{P}}$ and their unique sparse residue $\bar{\mathbf{S}}_i \in \mathbb{R}^{d\times d}$, so more common knowledge could be uncovered. Finally, the objective function for low-rank sparse decomposition is defined in the following:

$$\begin{aligned}&\min_{\bar{\mathbf{P}},\bar{\mathbf{S}}_i,\bar{\mathbf{P}}_i} \operatorname{rank}(\bar{\mathbf{P}}) + \lambda_0 \sum_{i=1}^{k} \|\bar{\mathbf{S}}_i\|_1 \\ &\text{s.t. } \bar{\mathbf{P}}_i = \bar{\mathbf{P}} + \bar{\mathbf{S}}_i, \quad i = 1, \ldots, k \end{aligned} \tag{1}$$

in which $\operatorname{rank}(\cdot)$ denotes the rank operator of a matrix, while $\|\cdot\|_1$ is l_1-norm that calculates the maximum absolute column sum of a matrix. $\lambda_0 > 0$ is the tradeoff to balance two parts. So far, we seek a low-rank common basis without dimensionality reduction. That is, we find all the d bases for feature learning.

Remark: There are already research activities on low-rank sparse decomposition, for example, Xia *et al.* [31] aimed to seek an optimal low-rank clustering matrix from multiple clustering results. Although the objective function is similar, the methodology, technical idea, and applications behind are very different. Specifically, Xia *et al.* [31] first achieved multiple clustering results for each view, and then they assumed that such multiple clustering results could generate a low-rank common clustering result, which is treated as the final optimal result. Therefore, they adopted low-rank sparse decomposition to multiple clustering results to obtain the low-rank optimal one. However, we assume that multiple view-specific transformations share a lot of information so that we adopt low-rank sparse decomposition to learn a CLRS for the challenge that the view knowledge of multiview test data is unavailable. Considering view consistency in multiview data, the collective low-rank projection can capture most common structure shared by multiple view-specific transformations. That is, the deviation error matrix tends to be sparse. Furthermore, we also evaluate different types of error, e.g., sparse norm and Frobenius norm, and we found that the results are almost the same. That is, even we do not know what types of the error across multiple views, we can still apply sparse norm to model it.

C. Multiview Low-Rank Subspace Learning

Since $\bar{\mathbf{P}}$ is low rank, there are many bases very similar, resulting in much redundant information within $\bar{\mathbf{P}}$. Assume the rank of $\bar{\mathbf{P}}$ is p ($p \ll d$), hence, we can adopt the p bases to extract effective features from multiview data, which could help well deal with the *curse of dimensionality*. Hence, we could transform the original problem into a fixed rank problem as

$$\begin{aligned}&\min_{P,S_i,P_i} \lambda_0 \sum_{i=1}^{k} \|S_i\|_1 \\ &\text{s.t.} P_i = P + S_i, \ i = 1, \ldots, k, \ P^\top P = I_p \end{aligned} \tag{2}$$

where $P \in \mathbb{R}^{d\times p}$, $P_i \in \mathbb{R}^{d\times p}$, and $S_i \in \mathbb{R}^{d\times p}$ are the p columns of $\bar{\mathbf{P}}$, $\bar{\mathbf{P}}_i$, and $\bar{\mathbf{S}}_i$, respectively, and we add an orthogonal constraint $P^\top P = I_p$ ($I_p \in \mathbb{R}^{p\times p}$ is an identity matrix) to make the P with the full rank of p.

Recently, low-rank modeling has been broadly studied in the global structure with the multiclass data [22]. Furthermore, low-rank representation has been integrated into dimensionality reduction framework to build a robust subspace for effective feature learning [20], [25]. Following the idea of low-rank subspace learning, we desire to exploit low-rank representation to build a bridge across the view-specific features and the shared features (Fig. 1). Hence, knowledge across multiple view-specific transformations could be transferred to the common subspace. Due to that the real-world data are always noisy, we design a sparse error term to figure out the noise or outliers. Finally, the objective function can be achieved by integrating (2) and low-rank reconstruction into a unified framework as

$$\begin{aligned}&\min_{P,Z_i,E_i,S_i,P_i}\sum_{i=1}^{k}(\text{rank}(Z_i)+\lambda_0\|S_i\|_1+\lambda_1\|E_i\|_{2,1})\\&\text{s.t. } {P_i}^\top X_i = P^\top DZ_i + E_i,\quad P_i = P + S_i\\&\qquad i=1,\ldots,k,\quad P^\top P = I_p\end{aligned}\tag{3}$$

where $Z_i \in \mathbb{R}^{\bar{m}\times m}$ is the ith low-rank reconstruction coefficient. $E_i \in \mathbb{R}^{p\times m}$ is the error term and $\|\cdot\|_{2,1}$ is the $L_{2,1}$-norm, i.e., $\|E_i\|_{2,1} = \sum_{k=1}^{p}(\sum_{j=1}^{m}([E_i]_{kj})^2)^{1/2}$, which aims to detect and remove outliers. In addition, $\lambda_1 > 0$ is the tradeoff to balance two parts. With the collective low-rank projection, we can alleviate the multiview learning by extracting effective features for the test data whatever view the data are. Since rank minimization problem is an NP-hard problem in (3), recent researches adopt *nuclear norm* as a good surrogate [22].

In the above objective function, $D \in \mathbb{R}^{d\times\bar{m}}$ denotes the data with mixed k views, which has different definitions in different scenarios. In feature learning setting, D means the dictionary ($\bar{m}$ is the atom size of dictionary) [22], which usually adopts the data itself X for simplicity. In this paper, we also directly use X as the basis. While in transfer learning setting, D denotes the unlabeled target domain and X represents the well-labeled source domain. We can easily understand that we are dealing with an unlabeled multiview data set by adapting the information from a well-learned source. Objective function (3) would help facilitate the target learning with the view/label knowledge of source domain.

D. Supervised Cross-View Alignment

To better utilize the label information in the training stage, we employ a supervised graph regularizer to align cross-view data within the same class. Model (3) only utilizes the view information of the training data so that it works in a same weakly supervised fashion to our previous work [12]. Moreover, model (3) exploits a multitask scheme, that is, data from each view are reconstructed by the commonly projected data in an individual manner. It is very important to align different views to make the learned collective subspace more discriminative. We first denote the projected low-dimensional data of each view $Y_i = {P_i}^\top X_i$ ($P^\top DZ_i \in \mathbb{R}^{p\times m}$ can be treated as its clean version), so the multiview projected data $Y = [Y_1,\ldots,Y_k] \approx P^\top DZ = P^\top D[Z_1,\ldots,Z_k] \in \mathbb{R}^{p\times km}$.

Since the fact that data from multiple views are from c different classes, these samples should be lying in c different subspaces. Hence, each view coefficient matrix Z_i tends to be low rank. Namely, the coefficient vectors within each view corresponding to samples within the same class should be highly correlated. For multiview learning, it is of great importance to couple within-class data across different views. We propose a supervised regularization $\Omega(P, Z)$ based on Fisher criterion as

$$\Omega(P,Z) = \frac{\text{tr}(\mathcal{S}_w)}{\text{tr}(\mathcal{S}_b)}\tag{4}$$

where $\text{tr}(\mathcal{M})$ is the trace of matrix $\mathcal{M}$. $\mathcal{S}_w$ and $\mathcal{S}_b$ are the within-class and between-class scatter matrices on $P^\top DZ$, respectively, defined as

$$\begin{aligned}\mathcal{S}_w &= \sum_{i=1}^{c}\sum_{j=1}^{n_i}(y_j^i-\mu_i)(y_j^i-\mu_i)^\top,\\ \mathcal{S}_b &= \sum_{i=1}^{c} n_i(\mu_i-\mu)(\mu_i-\mu)^\top\end{aligned}$$

in which μ_i is the mean of the ith class in Y, μ is the overall mean of Y, n_i is the size of the ith class, and y_j^i is the jth data in the ith class of Y. With Fisher criterion, the low-dimensional cross-view data from different classes should be far apart, while those from the same class should be close to each other. To better solve this problem, we convert trace-ratio problem into a trace difference problem [25]. Furthermore, we involve a regularization term to guarantee the convexity of $\Omega(P, Z)$ to Z and reformulate (4) as

$$\begin{aligned}\Omega(P,Z) &= \text{tr}(\mathcal{S}_w) - \text{tr}(\mathcal{S}_b) + \eta\|P^\top DZ\|_F^2\\ &= \text{tr}((P^\top DZ)(I_{km}-\mathcal{L}_w)(P^\top DZ)^\top)\\ &\quad - \text{tr}((P^\top DZ)\mathcal{L}_b(P^\top DZ)^\top) + \eta\|P^\top DZ\|_F^2\\ &= \text{tr}((P^\top DZ)((1+\eta)I_{km}-\mathcal{L}_w-\mathcal{L}_b)(P^\top DZ)^\top)\end{aligned}\tag{5}$$

where η is usually a small positive value (generally, we set $\eta = 10^{-3}$) and $\|\cdot\|_F^2$ is the matrix Frobenius norm. $I_{km} \in \mathbb{R}^{km\times km}$ is an identity matrix. The elements of $\mathcal{L}_w$ and $\mathcal{L}_b$ are defined as

$$\mathcal{L}_w[i,j] = \begin{cases}\dfrac{1}{n_c}, & \text{if } y_i \text{ and } y_j \text{ belong to class } c\\ 0, & \text{otherwise}\end{cases}$$

$$\mathcal{L}_b[i,j] = \begin{cases}\dfrac{1}{n_c}-\dfrac{1}{km}, & \text{if } y_i \text{ and } y_j \text{ belong to class } c\\ -\dfrac{1}{km}, & \text{otherwise.}\end{cases}$$

Finally, we come up with the final formulation for multiview data analysis as

$$\begin{aligned}&\min_{\substack{P,Z_i,E_i,\\ S_i,P_i}}\sum_{i=1}^{k}(\|Z_i\|_* + \lambda_0\|S_i\|_1 + \lambda_1\|E_i\|_{2,1}) + \lambda_2\Omega(P,Z)\\&\text{s.t. } {P_i}^\top X_i = P^\top DZ_i + E_i,\quad P_i = P + S_i,\\&\qquad i=1,\ldots,k,\quad P^\top P = I_p\end{aligned}\tag{6}$$

where λ_2 is the parameter to balance the weakly supervised multiview parts (3) and discriminative terms (5). $\|\cdot\|_*$ denotes

the nuclear norm of a matrix, which calculates the sum of singular values of a matrix. To sum up, we design a unified multiview learning framework by jointly building a discriminative collective subspace from multiple view-specific transformations and uncovering the data's global structure through low-rank reconstruction.

1) Discussion: Current multiview learning methods [2], [13] adopted multiple view-specific transformations to project the original data into a view-free space so that the view divergence across different views would be mitigated. It can be observed that such multiple view-specific projections are designed to preserve the similar discriminability for the same class across different views, and therefore, there exists a lot of common knowledge across multiple view-specific projections, which are independent of view variance [29]. Here, we exploit low-rank decomposition to seek a low-rank common projection, which aims to uncover most shared knowledge across different view-specific projections. Our collective low-rank projection could make our model more flexible in handling the challenges that we cannot achieve the view information of the probe data. Furthermore, a cross-view alignment term is incorporated into our framework to make our collective subspace more discriminative and robust.

We further discuss two most low-rank subspace learning algorithms, which are supervised regularization based robust subspace (SRRS) [25] and low-rank transfer subspace learning (LTSL) [20]. Specifically, SRRS focuses on seeking robust and effective features by integrating low-rank representation and linear discriminative analysis (LDA)-like discriminative regularizer into a unified framework. Differently, LTSL adopts low-rank subspace for transfer learning, aiming to transfer well-labeled knowledge from source data to the target one through locality aware reconstruction. LTSL aims to seek a domain-invariant subspace by adapting the knowledge of source to the target. Furthermore, LTSL incorporates general subspace learning algorithms, e.g., principle component analysis (PCA), LDA, and locality preserving projection (LPP) into the transfer learning framework. Compared with SRRS, LTSL and our CLRS both exploit the low-rank constraint using the low-dimensional features, and thus, we could save computational cost during the model training. Moreover, our algorithm is developed to deal with the challenging problem where the prior view information for the test data is unavailable.

E. Solving Objective Function

Since objective function (6) has many variables to be optimized, we adopt the popular alternating direction method of multipliers (ADMMs) algorithm [32] to address the problem (6). Recent studies show that ADMM converges well even with some variables nonsmooth. First of all, we introduce several relaxation variables, J_i and Q_i, and then reformulate (6) into its equivalent minimization problem as

$$\begin{aligned}&\min_{\substack{P,Z,E_i,S_i,\\ P_i,J_i,Z_i,Q_i}} \sum_{i=1}^{k}(\|J_i\|_* + \lambda_0\|S_i\|_1 + \lambda_1\|E_i\|_{2,1}) + \lambda_2\Omega(P,Z)\\ &\text{s.t. } P_i^\top X_i = P^\top D Q_i + E_i,\quad P_i = P + S_i,\ Z_i = J_i,\\ &\qquad Z_i = Q_i,\quad P^\top P = I_p,\ i = 1,\ldots,k \end{aligned} \tag{7}$$

whose augmented Lagrangian function is

$$\begin{aligned}&\sum_{i=1}^{k}\big(\|J_i\|_* + \lambda_0\|S_i\|_1 + \lambda_1\|E_i\|_{2,1} + \langle U_i, Z_i - J_i\rangle\\ &\quad + \langle \Upsilon_i, P_i{}^\top X_i - P^\top D Q_i - E_i\rangle + \langle V_i, P_i - P - S_i\rangle\\ &\quad + \langle R_i, Z_i - Q_i\rangle + \frac{\mu}{2}\big(\|P_i{}^\top X_i - P^\top D Q_i - E_i\|_F^2\\ &\quad + \|P_i - P - S_i\|_F^2 + \|Z_i - Q_i\|_F^2 + \|Z_i - J_i\|_F^2\big)\big)\\ &\quad + \lambda_2\mathrm{tr}((P^\top DZ)\mathcal{L}(P^\top DZ)^\top)\end{aligned}$$

where Υ_i, U_i, R_i, and V_i are Lagrange multipliers and μ is the positive penalty parameter. $\mathcal{L} = (1+\eta)I_{km} - \mathcal{L}_w - \mathcal{L}_b$. $\langle,\rangle$ denotes the inner product operator of two matrices.

As it can be seen, it is hard to jointly update the variables in objective function (7). Fortunately, we can achieve the optimization solution by iteratively updating each variable. Specifically, we alternately optimize the following variables J_i, Z_i, Q_i, S_i, E_i, P_i, S_i, Z, and P in a leave-one-out strategy. Moreover, assume that $J_{i,t}$, $Z_{i,t}$, $Q_{i,t}$, $E_{i,t}$, $P_{i,t}$, P_t, $S_{i,t}$, Z_t, $\Upsilon_{i,t}$, $R_{i,t}$, $V_{i,t}$, $U_{i,t}$, and μ_t are the solutions of the tth iteration, and hence, the solutions in the t+1 iteration are shown in the following.

Updating J_i:

$$J_{i,t+1} = \arg\min_{J_i}\frac{1}{\mu_t}\|J_i\|_* + \frac{1}{2}\left\|J_i - \left(Z_{i,t} + \frac{U_{i,t}}{\mu_t}\right)\right\|_F^2. \tag{8}$$

Updating Q_i:

$$Q_{i,t+1} = \left(D^\top P_t P_t^\top D + I_{\tilde{m}}\right)^{-1}\mathcal{Q}_{i,t} \tag{9}$$

where $\mathcal{Q}_{i,t} = D^\top P_t(P_{i,t}^\top X_i - E_{i,t}) + Z_{i,t} + (D^\top P_t\Upsilon_{i,t} + R_{i,t})/\mu_t$ and $I_{\tilde{m}} \in \mathbb{R}^{\tilde{m}\times\tilde{m}}$ is an identity matrix.

Updating E_i:

$$\begin{aligned}E_{i,t+1} = \arg\min_{E_i}\ &\frac{\lambda_1}{\mu_t}\|E_i\|_{2,1}\\ &+ \frac{1}{2}\left\|E_i - \left(P_{i,t}^\top X_i - P_t^\top D Q_{i,t+1} + \frac{\Upsilon_{i,t}}{\mu_t}\right)\right\|_F^2.\end{aligned} \tag{10}$$

Updating P_i:

$$\begin{aligned}P_{i,t+1} = (X_i X_i^\top + I_d)^{-1}\Big(&X_i(Q_{i,t+1}^\top D^\top P_t + E_{i,t+1}^\top)\\ &+ P\ P_t + S_{i,t+1} - \frac{X_i\Upsilon_{i,t}^\top + V_{i,t}}{\mu_t}\Big).\end{aligned} \tag{11}$$

Updating S_i:

$$S_{i,t+1} = \arg\min_{S_i}\frac{\lambda_0}{\mu_t}\|S_i\|_1 + \frac{1}{2}\left\|S_i - \left(P_{i,t+1} - P_t + \frac{V_{i,t}}{\mu_t}\right)\right\|_F^2. \tag{12}$$

Updating P:

$$P_{t+1} = \arg\min_{P}\mathcal{F}(P),\quad \text{s.t. } P^\top P = I_p \tag{13}$$

where $\mathcal{F}(P) = \sum_{i=1}^{K}\mu_t/2(\|P_{i,t+1} - P - S_{i,t+1} + V_{i,t+1}/\mu_t\|_F^2 + \|P_{i,t+1}{}^\top X_i - P^\top D Q_{i,t+1} - E_{i,t+1} + \Upsilon_{i,t}/\mu_t\|_F^2) + \lambda_2\mathrm{tr}((P^\top DZ_{t+1})\mathcal{L}(P^\top DZ_{t+1})^\top)$ and this optimization is a nonconvex problem. To fight off the

challenging nonconvex problem (13) caused by the orthogonal constraint, we adopt a gradient descent optimization procedure with curvilinear search for a local optimal solution [33]. Generally, we first calculate the gradient of $\mathcal{F}(P)$ with respect to P as $\partial\mathcal{F}(P)/\partial P = \mu_t(P-(P_{i,t+1}-S_{i,t+1}+V_{i,t+1}/\mu_t)+(DQ_{i,t+1})(Q_{i,t+1}^\top D^\top P - (P_{i,t+1}{}^\top X_i - E_{i,t+1}+\Upsilon_{i,t}/\mu_t)^\top)+2\lambda_2 DZ_{t+1}\mathcal{L}(DZ_{t+1})^\top P$. Then, we calculate skew-symmetric matrix and optimize P until Armijo–Wolfe conditions [34] meet. In this case, we need more iterations ($\tau > 1$) to update P compared with our conference version, i.e., two-step strategy [12].

Updating Z:

$$\begin{aligned} &Z_{t+1}+\lambda_2(D^\top P_{t+1}P_{t+1}^\top D)Z_{t+1}\mathcal{L}-G_t=0\\ &\quad\Rightarrow Z_{t+1}\mathcal{L}^{-1}+\lambda_2(D^\top P_{t+1}P_{t+1}^\top D)Z_{t+1}=G_t\mathcal{L}^{-1} \end{aligned} \tag{14}$$

where $G_t=[G_{1,t},\ldots,G_{k,t}]$ and $G_{i,t}=1/2(Q_{i,t+1}+J_{i,t+1}-U_{i,t}+R_{i,t}/\mu_t)$. When Z_{t+1} is learned, we could split it into $Z_{i,t+1}$.

Specifically, (14) is a conventional Sylvester equation [35]. Nuclear norm in (8) can be addressed with singular value thresholding [36]. Equations (10) and (12) are two sparse problems, which can be solved by the popular soft-thresholding operator [37]. To make it clear, we list the detailed steps of the optimization in Algorithm 1.

Algorithm 1 Solving Problem (7) by ADMM

Input: X_i $(i=1,\ldots,k)$, D, λ_0, λ_1, λ_2
Initialize: $\mathrm{J_{i,0}=Q_{i,0}=E_{i,0}=S_{i,0}=\Upsilon_{i,0}=U_{i,0}=V_{i,0}=0}$, $\mathrm{Z_0=0}$, $\mu_0=10^{-5}$, $\rho=1.3$, $\max_\mu=10^8$, $\epsilon=10^{-5}$, $t=0$.
while not converged **do**
1. Optimize $\mathrm{\{J/Q/E/P/S\}_{i,t+1}}$ in parallel.
   | 1-1 Optimize $\mathrm{J_{i,t+1}}$ via Eq. (8) by fixing others.
   | 1-2 Optimize $\mathrm{Q_{i,t+1}}$ via Eq. (9) by fixing others.
   | 1-3 Optimize $\mathrm{E_{i,t+1}}$ via Eq. (10) by fixing others.
   | 1-4 Optimize $\mathrm{P_{i,t+1}}$ via Eq. (11) by fixing others.
   | 1-5 Optimize $\mathrm{S_{i,t+1}}$ via Eq. (12) by fixing others.
2. Optimize $\mathrm{P_{t+1}}$ via Eq. (13) by fixing others.
3. Optimize $\mathrm{Z_{t+1}}$ via Eq. (14) by fixing others.
4. Optimize the multipliers $\mathrm{\Upsilon_{i,t+1}, U_{i,t+1}, V_{i,t+1}, R_{i,t+1}}$ via
   | $\mathrm{\Upsilon_{i,t+1}=\Upsilon_{i,t}+\mu_t(P_{i,t+1}^\top X_i - P_{t+1}^\top DQ_{i,t+1}-E_{i,t+1})}$;
   | $\mathrm{U_{i,t+1}=U_{i,t}+\mu_t(Z_{i,t+1}-J_{i,t+1})}$;
   | $\mathrm{R_{i,t+1}=R_{i,t}+\mu_t(Z_{i,t+1}-Q_{i,t+1})}$;
   | $\mathrm{V_{i,t+1}=V_{i,t}+\mu_t(P_{i,t+1}-P_{t+1}-S_{i,t+1})}$;
5. Optimize μ_{t+1} via $\mu_{t+1}=\min(\rho\mu_t,\max_\mu)$.
6. Check the convergence conditions
   | $\|P_{i,t+1}^\top X_i - P_{t+1}^\top DQ_{i,t+1}-E_{i,t+1}\|_\infty<\epsilon$;
   | $\|Z_{i,t+1}-J_{i,t+1}\|_\infty<\epsilon$, $\|Z_{i,t+1}-Q_{i,t+1}\|_\infty<\epsilon$;
   | $\|P_{i,t+1}-P_{t+1}-S_{i,t+1}\|_\infty<\epsilon$.
7. $t=t+1$.
end while
output: $\mathrm{Z_i, Q_i, J_i, E_i, S_i, P_i, P, Z}$.

So far, it is still difficult to guarantee the convergence of ADMM with three or more blocks [22]. Recent research efforts are exploited to prove the convergence for nonsmooth and nonconvex optimization problem [38], [39]. However, our problem is much more complex, especially, we have an orthogonal constraint on P, so we did not prove the convergence theoretically. We will show the convergence analysis in the experimental part. Furthermore, the parameters μ_0, ρ, ϵ, η, and $\max_\mu$ are set empirically, while the three tradeoff parameters λ_0, λ_1, λ_2, p, and τ can be selected through cross-validation strategy. Besides, P_i and P are randomly initialized.

1) Discussion: We propose a different objective function and optimization scheme, comparing with our conference version [12], which applies the low-rank constraint as $\widetilde{X}=P^\top DZ+E$ and $\widetilde{X}=[P_1^\top X_1,\ldots,P_k^\top X_k]$. In [12], we assume the new representations Z of multiview data together to be low rank, while here we constrain the new representation Z_i of each view data to be low rank. In our previous conference version [12], we stack multiple view-specific projections to a large new projection, so is the collective projection. When we employ low-rank constraint on the common subspace, we previously would need much more computational time and space than our current one.

Furthermore, we adopt a multitask low-rank framework to reconstruct view-specific projected data with mixed-view data in the collective subspace. Hence, variables J_i, Q_i, E_i, S_i, and P_i in each task can be solved in a parallel scheme, i.e., Steps 1-1 to 1-5. Therefore, we can save much time to solve those variables with new parallel techniques. However, we adopt a gradient descent strategy to update P, which would cost higher computational cost than our conference version [12]. Besides, we relax the low-rank constraint on big Z [12] to each small Z_i, while developing a novel cross-view alignment regularization to couple multiview data to take full advantage of label information. Another thing is we can only achieve local minimal solutions for both versions. In addition, we have more variables to be optimized in this journal version so that we may not achieve more optimal solutions than our conference version [12] when $\lambda_2=0$. We will show this phenomenon in the experiments.

F. Complexity Analysis

In this section, we provide a detail complexity analysis of our algorithm. Suppose $X_i\in\mathbb{R}^{d\times m}$ and $D\in\mathbb{R}^{d\times\bar{m}}$, where d is the original dimensionality of data, m is the size of the ith view data X_i, and $\bar{m}$ is the size of view-mixed data D. In addition, we assume that P_i and P are all $d\times p$ matrices, where p is the reduced dimensionality of subspace. The major time-consuming parts of Algorithm 1 are: 1) SVD operation in Step 1-1; 2) matrix inverse and multiplication in Steps 1-2 and 1-4; 3) subspace optimization in Step 2; and 4) Sylvester equation in Step 3.

We now discuss each part in detail. Since $k\ll m$, $\mathcal{O}(m)\approx\mathcal{O}(\bar{m})$. For simplicity, J_i, Q_i, and Z can be treated as $m\times m$ matrices, for P_i and P are $p\times d$ matrices. First of all, nuclear norm in Step 1-1 costs $\mathcal{O}(m^3)$ through SVD operation. Fortunately, according to [22, Th. 4.3], the SVD for Z_i could be speeded up to $\mathcal{O}(p^2m)$ where p is usually a small one. Second, we calculate the matrix multiplication and inverse. Step 1-2 would cost about $\mathcal{O}(m^3)$, while Step 1-4 will each approximately take $\mathcal{O}(d^3)$. Step 2 would cost $\mathcal{O}(\tau d^3)$, since there are τ iterations within Step 2. Finally, Step 3 generally

TABLE I

RECOGNITION PERFORMANCE (%) OF TEN ALGORITHMS ON THE ORIGINAL IMAGES FROM CMU-PIE FACE DATA SET, IN WHICH CASE 1: {C02, C14}, CASE 2: {C02, C27}, CASE 3: {C14, C27}, CASE 4: {C05, C07, C29}, CASE 5: {C05, C14, C29, C34}, AND CASE 6: {C02, C05, C14, C29, C31}

	PCA[40]	LDA[41]	LPP[42]	RSR[43]	TFRR[44]	SRRS[25]	LRCS [12]	MvDA[13]	RMSL[4]	Ours
Case 1	69.03±0.08	70.46±0.05	57.25±0.06	77.51±0.01	77.92±0.03	78.27±0.04	87.78±0.22	85.23±0.05	**88.15±0.06**	87.10±0.07
Case 2	69.21±0.08	71.32±0.02	58.83±0.07	74.74±0.17	76.24±0.12	78.74±0.23	86.67±0.09	85.81±0.09	**87.05±0.07**	86.70±0.12
Case 3	68.52±0.12	63.51±0.75	59.25±0.56	71.10±0.04	75.29±0.07	77.45±0.02	87.38±0.39	86.12±0.12	87.40±0.17	**87.48±0.10**
Case 4	52.65±0.04	56.53±0.02	43.56±0.08	67.57±0.01	69.74±0.05	71.44±0.03	74.84±0.04	**75.36±0.18**	75.16±0.12	72.38±0.09
Case 5	34.94±0.08	24.07±0.25	19.67±0.05	29.72±0.01	33.91±0.12	38.86±0.02	44.48±0.03	**54.13±0.16**	44.93±0.11	45.88±0.09
Case 6	29.09±0.01	07.06±0.01	13.11±0.01	09.44±0.02	28.36±0.04	30.16±0.02	36.17±0.11	**47.67±0.18**	37.14±0.08	38.28±0.11

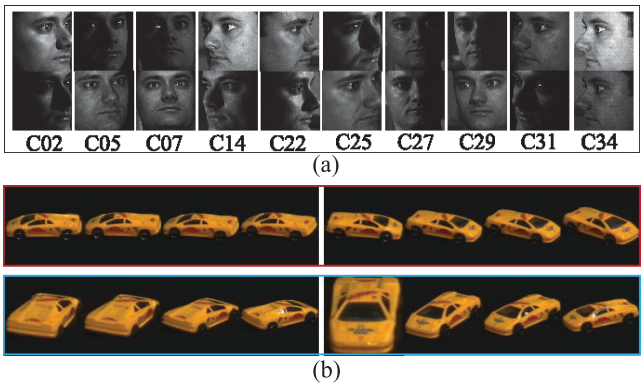


Fig. 2. (a) Face samples from different views of one individual in CMU-PIE Cross-Pose Face data set. It can be observed that the dissimilarity across different views of the same individual. (b) Object samples of COIL-100 object database, where the red rectangle shows two views of COIL1, while the blue one mean two views of COIL2.

takes $\mathcal{O}(m^3)$ to optimize $Z \in \mathbb{R}^{m\times\bar{m}}$ in Sylvester function. To sum up, we conclude that the time complexity of CLRS is $\mathcal{O}(d^3 + m^3)$.

IV. EXPERIMENT

In this section, we first introduce the real multiview data sets (e.g., cross-pose and cross-modality data) and experimental settings. Then, we compare with the state-of-the-art algorithms in two different scenarios. Finally, we evaluate some properties of our proposed CLRS.

A. Data Sets and Experimental Setting

CMU-PIE Face data set [45] totally consists of 68 subjects with different poses. There are 21 different illumination variations for the samples of each subject. Specifically, we adopt such different poses, which show large view variances within the same subject across different poses [Fig. 2(a)]. In the experiment, we select different numbers of views to build various evaluation scenarios. For each pose, we randomly choose ten samples for training while the left for testing. Furthermore, we crop faces into the size of 64×64 and adopt the gray-scale value as the input.

COIL-100 object database[1] includes 100 categories [Fig. 2(b)]. Images of each object were taken 5° apart as the object is rotated on a turntable and each object has 72 gray images. It is partitioned into two subsets: "COIL1" and "COIL2" [4]. COIL1 has all images taken in View 1 [0°, 85°] and View 2 [180°, 265°] while COIL2 contains those in View 3 [90°, 175°] and View 4 [270°, 355°]. The pixel values with 64×64 with 20% random corruption are applied.

ALOI object database[2] includes 1000 categories. In addition, we choose the first 100 categories with 7200 images as ALOI-100. Samples of each object were captured 5° apart so that each object contains 72 gray images in total. We apply the same setting with COIL-100 to four views of ALOI-100. The pixel values with 96×72 with 20% random corruption are used.

BUAA VIS-NIR Face data set [46] contains 150 different individuals, and every individual has two modalities, i.e., near-infrared (NIR) faces and visible (VIS) faces. Specifically, there are nine face images per modality per individuals. We further crop the faces and resize them to the size 200×200. The gray-scale value is used as the input feature.

B. Feature Representation Setting

In our experiment, we address the challenging problem, where the view knowledge of the probe data is unavailable. Thus, conventional multiview methods [2], [13] would fail. Therefore, we mainly compare with PCA [40], LDA [41], LPP [42], rotated sparse regression (RSR) [43], transformed fixed-rank representation (TFRR) [44], SRRS [25], robust multi-view subspace learning (RMSL) [4], and low-rank common subspace (LRCS) [12]. Specifically, LDA, RSR, SRRS, RMSL, and ours are five supervised algorithms; and PCA, LPP, TFRR, and LRCS are four unsupervised methods. Furthermore, we compare with one conventional multiview subspace learning algorithm, MvDA [13], by providing it extra view knowledge of the probe data to verify the effectiveness of our approach.

In this setting, we conduct experiment on the CMU-PIE Face data set and two object databases: COIL-100 and ALOI-100. The nearest neighbor classifier (NNC) is adopted to testify the final classification results. For CMU-PIE, we choose ten images per individual per pose to build the training set, and the remaining data are used for testing. We do five random selections and report the average performance. Tables I and II represent recognition performance on the original images and 10% corrupted images, respectively. Besides, we also evaluate their recognition performance under different dimensions in Figs. 3 and 4. For object databases, we choose one from COIL1 (ALOI1) and one from COIL2 (ALOI2) to construct two-view training set, while the left data are adopted for testing. In total, we have four cases for training. The results are presented in Fig. 5.

[1] http://www.cs.columbia.edu/CAVE/software/softlib/coil-100.php

[2] http://aloi.science.uva.nl/

TABLE II

RECOGNITION PERFORMANCE (%) OF TEN ALGORITHMS ON THE CORRUPTED IMAGES FROM CMU-PIE FACE DATA SET, IN WHICH CASE 1: {C02, C14}, CASE 2: {C02, C27}, CASE 3: {C14, C27}, CASE 4: {C05, C07, C29}, CASE 5: {C05, C14, C29, C34}, AND CASE 6: {C02, C05, C14, C29, C31}

	PCA[40]	LDA[41]	LPP[42]	RSR[43]	TFRR[44]	SRRS[25]	LRCS [12]	MvDA[13]	RMSL[4]	Ours
Case 1	64.87±0.32	26.71±0.20	31.26±0.26	37.02±0.03	68.10±0.07	72.27±0.05	78.98±0.03	75.34±0.09	**81.12±0.08**	80.16±0.09
Case 2	66.04±0.08	23.19±0.35	30.98±0.18	34.34±0.15	68.24±0.32	72.74±0.18	78.67±0.05	74.81±0.09	**81.67±0.09**	80.84±0.11
Case 3	65.21±0.04	20.34±0.75	32.21±0.36	31.69±0.09	67.85±0.12	71.45±0.08	78.38±0.26	76.24±0.15	81.08±0.17	**81.92±0.16**
Case 4	50.16±0.04	16.72±0.02	27.66±0.05	22.45±0.01	50.94±0.09	54.32±0.03	65.84±0.04	61.26±0.12	66.95±0.08	**67.26±0.10**
Case 5	31.74±0.08	06.67±0.25	14.34±0.04	10.02±0.01	29.26±0.12	32.34±0.02	39.48±0.03	**44.19±0.13**	43.87±0.11	41.54±0.12
Case 6	27.21±0.01	04.06±0.01	12.02±0.01	04.95±0.02	28.12±0.03	29.03±0.02	32.57±0.01	**34.17±0.21**	33.78±0.08	33.28±0.10

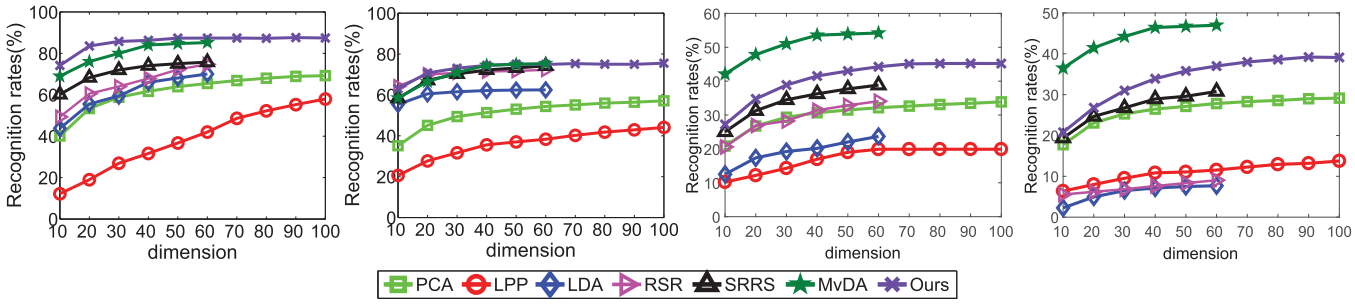


Fig. 3. Recognition performance of seven algorithms on the original images of CMU-PIE Face data set over different dimensions, which shows the performance of Case 2 and Cases 4–6, from left to right. We can only obtain at most 67 dimensions for the LDA-based algorithms (LDA, RSR, SRRS, and MvDA) (here we only present 60 dimensions for them).

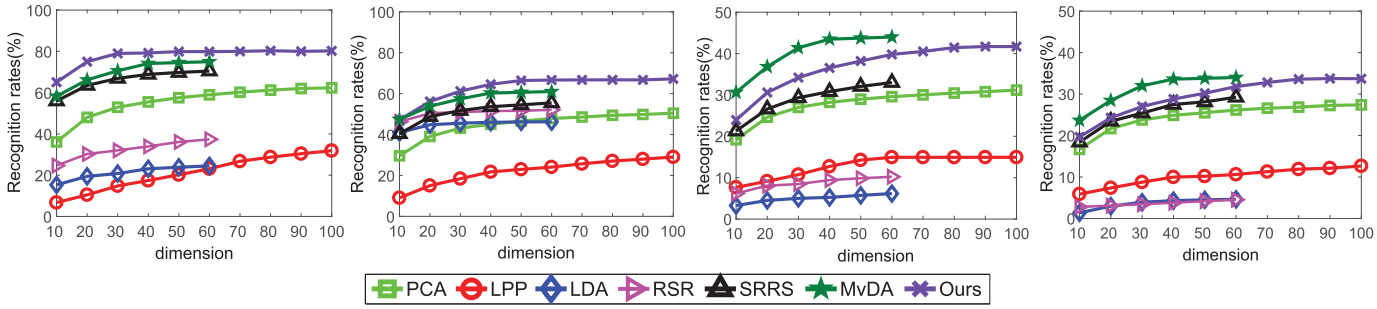


Fig. 4. Recognition performance of seven algorithms on the corrupted images of CMU-PIE Face data set over different dimensions, which shows the performance of Case 2 and Cases 4–6, from left to right. We can only obtain at most 67 dimensions for the LDA-based algorithms (LDA, RSR, SRRS, and MvDA) (here we only present 60 dimensions for them).

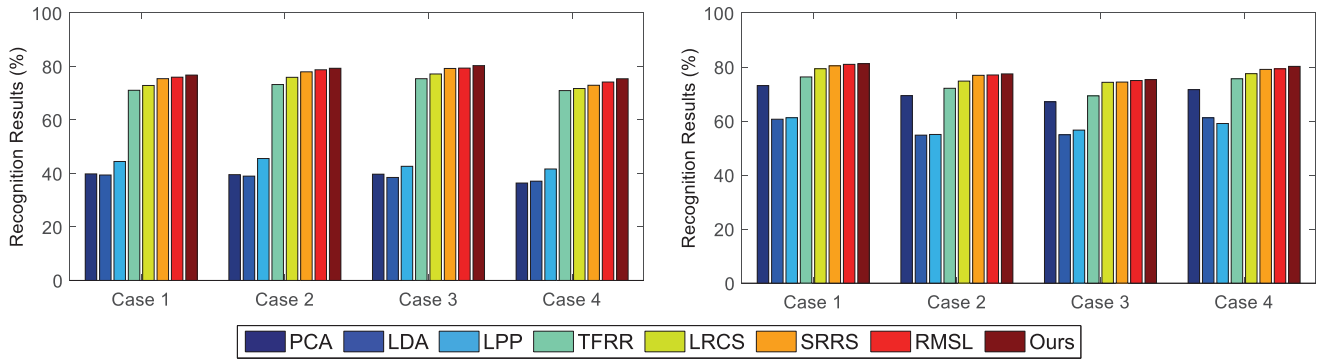


Fig. 5. Recognition results of seven methods on four cases of the 20% corrupted COIL-100 data set (left) and ALOI-100 one (right), where Case 1: View 1 and View 3; Case 2: View 1 and View 4; Case 3: View 2 and View 3; and Case 4: View 2 and View 4.

1) Discussion: Results demonstrate that our method outperforms others in most cases, except MvDA and RMSL. However, we can see that our algorithm could achieve competitive performance with MvDA, or even better in some cases. This demonstrates that our algorithm is an effective compromise when we are inaccessible to the view information for the evaluation data. However, when there are more views, MvDA has a superiority in performance, since multiple view-specific transformations could well fit each specific view data. Besides, MvDA is one kind of traditional subspace learning methods, which cannot work well in corrupted cases even though the view knowledge of the probe data is available.

Our algorithm unifies low-rank reconstruction and dimension reduction together, and therefore, it could well handle the corrupted data in reality.

With more poses involved, we could observe that all the algorithms suffer a decrease in terms of recognition performance. However, we notice that the performance of LDA and RSR decreases much faster, which shows that conventional subspace learning algorithms, e.g., LDA, would fail in multiview learning. For two-view scenarios (i.e., Cases 1–3), all the methods can obtain very similar results, which represent that the divergence across any two views is very similar. As we can see, these three cases effectively show the superiority of our model. This denotes that the collective low-rank common subspace, decomposed from two view-specific transformations, can uncover the most intrinsic information from the data. Moreover, for three-view case (Case 4), our model cannot improve with a large margin, and the reason we consider is that three views are relatively frontal faces. When involving more views (e.g., Cases 5 and 6), our proposed algorithm can still perform better than others.

Another observation is that low-rank-based algorithms (i.e., TFRR, SRRS, LRCS, RMLS, and ours) can work better than the other algorithms, especially when dealing with corrupted cases. The reason, we consider, is that low-rank-based algorithms are able to capture the data's intrinsic classwise structure. Another phenomenon is each pose show 21 different lighting variations in CMU-PIE Face data set and, therefore, PCA has the similar results in the original and randomly corrupted images.

Compared with our previous conference version, our CLRS works a little worse in two cases. The reason, we consider, is the solution to the objective function, since our current one is supervised. We adopt multitask learning technique to solve multiple view-specific transformations in parallel, which could make the optimal solutions more flexible, compared with our conference solution. This phenomenon is discussed more in Section IV-D2 parameter analysis.

Moreover, we notice that low-rank models work better than traditional subspace algorithms from the results of two corrupted object databases (Fig. 5). The reason is that random corruption would significantly hinder the conventional algorithms' recognition task. Besides, our CLRS performs best, since our method introduces the supervised cross-view alignment to make full use the label information, and therefore, our CLRS can achieve a more robust and discriminative subspace.

C. Transfer Learning Setting

In this part, experiments are conducted to evaluate the algorithms in transfer learning setting. We have six comparisons, i.e., joint domain adaptation (JDA) [50], LTSL [20], geodesic flow kernel (GFK) [48], domain adaptation via subspace alignment (DASA) [49], TSL [47], and LRCS [12]. In addition, we compare with some conventional dimensionality reduction techniques, e.g., PCA [40], LDA [41], and LPP [42]. For those methods, we apply them on source and target data together (except LDA, which is supervised, so only source data are used) to learn the projection in the training stage, and then predict the unlabeled target samples. Finally, the NNC is utilized to testify the effectiveness for target domain.

1) Group 1*:* To construct two domains, we first split CMU-PIE with 68 individuals into two subsets, each with 34 different individuals. To make source and target with different distributions, we utilize low-resolution process for the source domain. In target domain, we adopt the original 64×64 images. In source domain, the 64×64 images (high resolution) are resized to 16×16, and resized back to the original size (low resolution). The MATLAB function *imresize()* with default setting is used. We choose different poses in the same previous setting to "feature representation setting." Since there is no label overlap across two domains, therefore, we randomly choose two reference face images per view in the target domain in the evaluation stage, while other samples are used for evaluation. We randomly select ten times and report the average performance.

2) Group 2*:* we split the BUAA NIR-VIS into two subsets, one as source and the other as target. In this data set, we conduct two different cases. Case 1: choosing 50 individuals as source and the left 100 individuals as target; Case 2: choosing 75 individuals as source and the left 75 individuals as target. Every individual contains two modalities. Note that no identity overlap exits across source and target. To further differentiate source and target, we exploit down-sampling procession to source. Finally, we randomly select two target samples per individual as the reference, while the left target samples are used for evaluation. We randomly select for nine times, and then calculate the average results.

3) Discussion: It can be seen from Tables III and IV that our proposed model achieves better performance than others. Since we seek multiple view-specific transformations on the well-labeled source domain, our proposed algorithm could effectively transfer such multiview knowledge to the unlabeled target domain by coupling various views properly with the collective low-rank projection. Moreover, we could observe that low-rank-based algorithms, i.e., LTSL, LRCS, and ours, outperform the other algorithms. Furthermore, we could notice that the effectiveness of CLRS is significant in Group 1, but not obvious in Group 2. The reason may be that the domain divergence is larger in Group 1.

D. Property Evaluation

In this section, we evaluate several properties of our proposed algorithm, i.e., convergence analysis and parameter sensitivity.

1) Convergence Analysis: In this part, we mainly analyze the convergence of our model. Specifically, we evaluate on robust feature learning setting and use CMU-PIE database with Case 2 {C02, C14}. The convergence curve of our algorithm with different runs is shown in Fig. 6(a). From the results, we observe that our method converges well after 50 iterations.

2) Parameter Analysis: Moreover, we have three parameters λ_0, λ_1, and λ_2. Among them, λ_0 and λ_1 are two parameters to constrain the sparse errors parts for subspace decomposition and outliers noises. These two parameters are usually set as small values [22]. In our experiments, we set $\lambda_0 = \lambda_1 = 10^{-2}$.

TABLE III

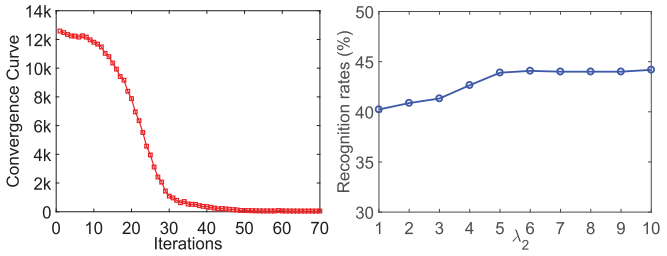
RECOGNITION PERFORMANCE (%) OF TEN ALGORITHMS ON GROUP 1 OF CMU-PIE DATA SET, IN WHICH CASE 1: {C02, C14}, CASE 2: {C02, C27}, CASE 3:{C14, C27}, CASE 4: {C05, C07, C29}, CASE 5: {C05, C14, C29, C34}, AND CASE 6: {C02, C05, C14, C29, C31}

Methods	PCA [40]	LDA[41]	LPP[42]	TSL [47]	GFK [48]	LTSL [20]	DASA [49]	JDA [50]	LRCS[12]	Ours
Case 1	34.87±0.32	23.71±0.20	29.26±0.26	49.43±0.09	56.21±0.03	58.29±0.01	57.34±0.03	60.91±0.08	61.23±0.12	**62.16±0.10**
Case 2	36.04±0.08	21.19±0.35	28.98±0.18	48.88±0.12	55.48±0.08	57.68±0.06	56.84±0.06	60.18±0.12	60.62±0.10	**61.68±0.09**
Case 3	34.21±0.04	19.84±0.75	30.21±0.36	49.72±0.08	56.12±0.08	58.36±0.08	57.42±0.10	61.23±0.11	61.37±0.09	**62.06±0.11**
Case 4	23.16±0.04	16.72±0.02	21.66±0.05	29.28±0.05	34.87±0.05	37.66±0.06	35.62±0.12	48.59±0.08	49.09±0.09	**49.68±0.11**
Case 5	13.74±0.08	06.67±0.25	12.34±0.04	19.34±0.08	23.29±0.08	28.11±0.05	26.23±0.09	32.86±0.08	33.05±0.04	**34.72±0.10**
Case 6	10.21±0.01	04.06±0.01	10.02±0.01	16.76±0.14	19.45±0.12	22.54±0.03	20.45±0.09	26.05±0.15	28.05±0.05	**30.14±0.11**

TABLE IV

RECOGNITION PERFORMANCE (%) OF TEN ALGORITHMS ON GROUP 2 OF BUAA NIR-VIS FACE DATA SET

Methods	PCA [40]	LDA[41]	LPP[42]	TSL [47]	GFK [48]	LTSL [20]	DASA [49]	JDA [50]	LRCS [12]	Ours
Case 1	46.23±0.21	45.83±0.27	48.07±0.18	56.02±0.42	65.82±0.26	68.07±0.15	62.12±0.32	68.01±0.21	70.56±0.20	**71.48±0.26**
Case 2	47.23±0.32	46.62±0.34	49.41±0.17	57.32±0.24	67.56±0.30	69.17±0.19	63.45±0.45	68.93±0.12	71.63±0.22	**72.68±0.32**



(a) (b)

Fig. 6. (a) Convergence curve of CLRS for Case 2 {C02, C14} of CMU-PIE Face data set, where $p = 200$, $\lambda_0 = \lambda_1 = 10^{-2}$, and $\lambda_2 = 10^1$. (b) Influence of parameter λ_2 on the case of {C05, C14, C29, C34}. The value of x-axis from 1 to 10 represents [0, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^2, 10^3, 10^4].

On the other hand, λ_2 is used to constrain the discriminative term. In this section, we mainly evaluate it on one four-view case [Fig. 6(b)]. From the results, we notice that λ_2 slightly influences the final recognition performance. Without the loss of generality, we set $\lambda_2 = 10$ in our experiments.

Remark: When λ_2 becomes zero, the proposed method degenerates to our conference version, i.e., LRCS [12]. However, we apply multitask learning technique to solve the problem in the journal extension to speed up the optimization, which is different from our conference version. When we adopt multitask learning technique, multiple view-specific projections are optimized in an individual way, which results in more variables needed to be updated. Since we can only achieve local minimum solutions through ADMM, more variables may produce more flexible optimization, and therefore, the proposed approach may not find better optimal solutions than our conference version, where multiple view-specific projections are stacked together and optimized as a whole. These are the cases when the proposed method fails to outperform our conference version, i.e., $\lambda_2 = 0$.

V. CONCLUSION

In this paper, we developed an effective CLRS framework to solve multiview learning problem. Our proposed model was designed to address the challenging case where the view knowledge of testing data is unavailable. Specifically, we built a CLRS shared by multiple view-specific projections. Moreover, our designed approach obtained a discriminative subspace through a supervised cross-view alignment and low-rank reconstruction. Furthermore, our method was easily generalized to different cases by adapting the input of our algorithm. Experiments were evaluated on several multiview databases in two different settings. The experimental results on several multiview benchmarks demonstrated that our algorithm achieved better performance than other comparisons.

REFERENCES

[1] M. Shao and Y. Fu, "Cross-modality feature learning through generic hierarchical hyperlingual-words," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 2, pp. 451–463, Feb. 2016.

[2] X. Cai, C. Wang, B. Xiao, X. Chen, and J. Zhou, "Regularized latent least square regression for cross pose face recognition," in *Proc. 23rd Int. Joint Conf. Artif. Intell.*, 2013, pp. 1247–1253.

[3] M. Du, A. C. Sankaranarayanan, and R. Chellappa, "Robust face recognition from multi-view videos," *IEEE Trans. Image Process.*, vol. 23, no. 3, pp. 1105–1117, Mar. 2014.

[4] Z. Ding and Y. Fu, "Robust multi-view subspace learning through dual low-rank decompositions," in *Proc. 30th AAAI Conf. Artif. Intell.*, 2016, pp. 1181–1187.

[5] Y. Wang, W. Zhang, L. Wu, X. Lin, and X. Zhao, "Unsupervised metric fusion over multiview data by graph random walk-based cross-view diffusion," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 1, pp. 57–70, Jan. 2017.

[6] Y. Wang, X. Lin, L. Wu, W. Zhang, Q. Zhang, and X. Huang, "Robust subspace clustering for multi-view data by exploiting correlation consensus," *IEEE Trans. Image Process.*, vol. 24, no. 11, pp. 3939–3949, Nov. 2015.

[7] L. Niu, W. Li, D. Xu, and J. Cai, "An exemplar-based multi-view domain generalization framework for visual recognition," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published. [Online]. Available: http://ieeexplore.ieee.org/abstract/document/7733141/

[8] W. Yang, Y. Gao, Y. Shi, and L. Cao, "MRM-lasso: A sparse multiview feature selection method via low-rank analysis," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 11, pp. 2801–2815, Nov. 2015.

[9] H. Zhao and Y. Fu, "Dual-regularized multi-view outlier detection," in *Proc. 24th Int. Joint Conf. Artif. Intell.*, 2015, pp. 4077–4083.

[10] S. Wang, Z. Ding, and Y. Fu, "Coupled marginalized auto-encoders for cross-domain multi-view learning," in *Proc. 25th Int. Joint Conf. Artif. Intell.*, 2016, pp. 2125–2131.

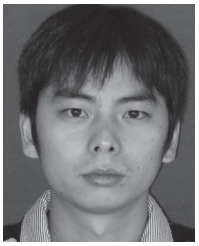
[11] X. Huang, Z. Lei, M. Fan, X. Wang, and S. Z. Li, "Regularized discriminative spectral regression method for heterogeneous face matching," *IEEE Trans. Image Process.*, vol. 22, no. 1, pp. 353–362, Jan. 2013.

[12] Z. Ding and Y. Fu, "Low-rank common subspace for multi-view learning," in *Proc. IEEE Int. Conf. Data Mining*, Dec. 2014, pp. 110–119.

[13] M. Kan, S. Shan, H. Zhang, S. Lao, and X. Chen, "Multi-view discriminant analysis," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 1, pp. 188–194, Jan. 2016.

[14] M. R. Hestenes, "Multiplier and gradient methods," *J. Optim. Theory Appl.*, vol. 4, no. 5, pp. 303–320, 1969.

[15] Z. Zhu, P. Luo, X. Wang, and X. Tang, "Multi-view perceptron: A deep model for learning face identity and view representations," in *Proc. Adv. Neural Inf. Process. Syst.*, 2014, pp. 217–225.
[16] H. Hotelling, "Relations between two sets of variates," *Biometrika*, vol. 28, nos. 3–4, pp. 321–377, 1936.
[17] J. Rupnik and J. Shawe-Taylor, "Multi-view canonical correlation analysis," in *Proc. Conf. Data Mining Data Warehouses*, 2010, pp. 1–4.
[18] L. Shao, F. Zhu, and X. Li, "Transfer learning for visual categorization: A survey," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 5, pp. 1019–1034, May 2014.
[19] S. Shekhar, V. M. Patel, H. V. Nguyen, and R. Chellappa, "Generalized domain-adaptive dictionaries," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 361–368.
[20] M. Shao, D. Kit, and Y. Fu, "Generalized transfer subspace learning through low-rank constraint," *Int. J. Comput. Vis.*, vol. 109, no. 1, pp. 74–93, 2014.
[21] Z. Ding, M. Shao, and Y. Fu, "Incomplete multisource transfer learning," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published. [Online]. Available: http://ieeexplore.ieee.org/document/7742923/
[22] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 171–184, Jan. 2013.
[23] J. Li, Y. Kong, H. Zhao, J. Yang, and Y. Fu, "Learning fast low-rank projection for image classification," *IEEE Trans. Image Process.*, vol. 25, no. 10, pp. 4803–4814, Oct. 2016.
[24] Y. Kong, M. Shao, K. Li, and Y. Fu, "Probabilistic low-rank multitask learning," *IEEE Trans. Neural Netw. Learn. Syst.*, to be published. [Online]. Available: http://ieeexplore.ieee.org/document/7805309/
[25] S. Li and Y. Fu, "Learning robust and discriminative subspace with low-rank constraints," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 11, pp. 2160–2173, Nov. 2016.
[26] Z. Ding, S. Suh, J.-J. Han, C. Choi, and Y. Fu, "Discriminative low-rank metric learning for face recognition," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recognit.*, vol. 1. May 2015, pp. 1–6.
[27] J. Zheng and Z. Jiang, "Learning view-invariant sparse representations for cross-view action recognition," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 3176–3183.
[28] J. Hoffman, B. Kulis, T. Darrell, and K. Saenko, "Discovering latent domains for multisource domain adaptation," in *Proc. Eur. Conf. Comput. Vis.*, 2012, pp. 702–715.
[29] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013.
[30] Z. Zhu, P. Luo, X. Wang, and X. Tang. (Apr. 2014). "Recover canonical-view faces in the wild with deep neural networks." [Online]. Available: https://arxiv.org/abs/1404.3543
[31] R. Xia, Y. Pan, L. Du, and J. Yin, "Robust multi-view spectral clustering via low-rank and sparse decomposition," in *Proc. 28th AAAI Conf. Artif. Intell.*, 2014, pp. 2149–2155.
[32] D. Gabay and B. Mercier, "A dual algorithm for the solution of nonlinear variational problems via finite element approximation," *Comput. Optim. Appl.*, vol. 2, no. 1, pp. 17–40, 1976.
[33] Z. Wen and W. Yin, "A feasible method for optimization with orthogonality constraints," *Math. Program.*, vol. 142, no. 1, pp. 397–434, 2013.
[34] W. Sun and Y.-X. Yuan, *Optimization Theory and Methods: Nonlinear Programming*, vol. 1. New York, NY, USA: Springer, 2006. [Online]. Available: https://link.springer.com/book/10.1007%2Fb106451
[35] R. H. Bartels and G. Stewart, "Solution of the matrix equation AX+ XB = C [F4]," *Commun. ACM*, vol. 15, no. 9, pp. 820–826, 1972.
[36] J.-F. Cai, E. J. Candes, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. Optim.*, vol. 20, no. 4, pp. 1956–1982, 2010.
[37] D. L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inf. Theory*, vol. 41, no. 3, pp. 613–627, May 1995.
[38] Y. Wang, W. Yin, and J. Zeng. (Nov. 2015). "Global convergence of ADMM in nonconvex nonsmooth optimization." [Online]. Available: https://arxiv.org/abs/1511.06324
[39] M. Hong, Z.-Q. Luo, and M. Razaviyayn, "Convergence analysis of alternating direction method of multipliers for a family of nonconvex problems," *SIAM J. Optim.*, vol. 26, no. 1, pp. 337–364, 2016.
[40] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cognit. Neurosci.*, vol. 3, no. 1, pp. 71–86, 1991.
[41] P. N. Belhumeur, J. P. Hespanha, and D. 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.
[42] X. He and P. Niyogi, "Locality preserving projections," in *Proc. Neural Inf. Process. Syst.*, vol. 16. 2004, p. 153.
[43] D. Chen, X. Cao, F. Wen, and J. Sun, "Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 3025–3032.
[44] R. Liu, Z. Lin, F. De la Torre, and Z. Su, "Fixed-rank representation for unsupervised visual learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 598–605.
[45] T. Sim, S. Baker, and M. Bsat, "The CMU pose, illumination, and expression (PIE) database of human faces," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recognit.*, 2002, pp. 46–51.
[46] H. Di, S. Jia, and W. Yunhong, "The buaa-visnir face database instructions," School Comput. Sci. Eng., Beihang Univ., Beijing, China, Tech. Rep. IRIP-TR-12-FR-001, 2012.
[47] S. Si, D. Tao, and B. Geng, "Bregman divergence-based regularization for transfer subspace learning," *IEEE Trans. Knowl. Data Eng.*, vol. 22, no. 7, pp. 929–942, Jul. 2010.
[48] B. Gong, Y. Shi, F. Sha, and K. Grauman, "Geodesic flow kernel for unsupervised domain adaptation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2066–2073.
[49] B. Fernando, A. Habrard, M. Sebban, and T. Tuytelaars, "Unsupervised visual domain adaptation using subspace alignment," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 2960–02967.
[50] M. Long, J. Wang, G. Ding, J. Sun, and P. S. Yu, "Transfer feature learning with joint distribution adaptation," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 2200–2207.



Zhengming Ding (S'14) received the B.Eng. degree in information security and the M.Eng. degree in computer software and theory from the University of Electronic Science and Technology of China, Chengdu, China, in 2010 and 2013, respectively. He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, Northeastern University, Boston, MA, USA.

His current research interests include machine learning and computer vision. Specifically, he devotes himself to develop scalable algorithms for challenging problems in transfer learning and deep learning scenarios.

Mr. Ding is an AAAI Student Member. He was a recipient of the Student Travel Grant of ACM MM 14, ICDM 14, AAAI 16, IJCAI 16, and FG 17. He received the National Institute of Justice Fellowship. He was a recipient of the best paper award. He has served as the Reviewer for the IEEE journals: the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEARNING SYSTEMS and the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE.

Yun Fu (S'07–M'08–SM'11) received the B.Eng. degree in information engineering and the M.Eng. degree in pattern recognition and intelligence systems from Xi'an Jiaotong University, Xi'an, China, respectively, and the M.S. degree in statistics and the Ph.D. degree in electrical and computer engineering from the University of Illinois at Urbana–Champaign, Champaign, IL, USA, respectively.

He has been an Interdisciplinary Faculty Member with the College of Engineering and the College of Computer and Information Science, Northeastern University, Boston, MA, USA, since 2012. He has extensive publications in leading journals, books/book chapters, and international conferences/workshops. His current research interests include machine learning, computational intelligence, big data mining, computer vision, pattern recognition, and cyber-physical systems.

Dr. Fu is a fellow of the IAPR, a Lifetime Senior Member of the ACM and the SPIE, a Lifetime Member of the AAAI, the OSA, and the Institute of Mathematical Statistics, a member of the Global Young Academy and INNS and a Beckman Graduate Fellow from 2007 to 2008. He serves as an Associate Editor, the Chair, a PC Member, and a Reviewer of many top journals and international conferences/workshops. He received seven Prestigious Young Investigator Awards from the NAE, the ONR, the ARO, the IEEE, the INNS, the UIUC, and the Grainger Foundation; seven Best Paper Awards from the IEEE, the IAPR, the SPIE, and the SIAM; and three major Industrial Research Awards from Google, Samsung, and Adobe. He is currently an Associate Editor of the IEEE TRANSACTIONS ON NEURAL NETWORKS AND LEANING SYSTEMS.