Robust Multi-View Subspace Learning through Dual Low-Rank Decompositions

Zhengming Ding[†] and **Yun Fu**[†‡]
[†]Department of Electrical & Computer Engineering, Northeastern University, Boston, USA
[‡]College of Computer & Information Science, Northeastern University, Boston, USA
allanding@ece.neu.edu, yunfu@ece.neu.edu

Abstract

Multi-view data is highly common nowadays, since various view-points and different sensors tend to facilitate better data representation. However, data from different views show a large divergence. Specifically, one sample lies in two kinds of structures, one is class structure and the other is view structure, which are intertwined with one another in the original feature space. To address this, we develop a Robust Multi-view Subspace Learning algorithm (RMSL) through dual low-rank decompositions, which desires to seek a low-dimensional view-invariant subspace for multi-view data. Through dual low-rank decompositions, RMSL aims to disassemble two intertwined structures from each other in the low-dimensional subspace. Furthermore, we develop two novel graph regularizers to guide dual low-rank decompositions in a supervised fashion. In this way, the semantic gap across different views would be mitigated so that RMSL can preserve more within-class information and reduce the influence of view variance to seek a more robust low-dimensional subspace. Extensive experiments on two multi-view benchmarks, e.g., face and object images, have witnessed the superiority of our proposed algorithm, by comparing it with the state-of-the-art algorithms.

Introduction

Multi-view data analysis has caught an increasing attention in the recent years (Kan et al. 2012; Liu et al. 2012; Wu and Jia 2012; Cai et al. 2013; Shekhar et al. 2014; Zhao and Fu 2015), since multi-view data (e.g., multi-pose images and multi-modal data) is frequently seen in reality (Ding and Tao 2015) when data is taken from various view-points (Cai et al. 2013) or captured with different types of sensors (Liu et al. 2012; Shekhar et al. 2014). Multi-view data brings in the challenge that the data with the same label tend to be multiple different samples, even heterogeneous. This results in a difficult learning problem, where within-class data across multiple views show a lower similarity than that within the same view but from different classes. In other words, one sample lies in two kinds of structures, one is class structure and the other is view structure, which are intertwined in the original high-dimensional space.

Generally, there are three categories of techniques to handle multi-view data analysis problems, including feature adaptation (Kan et al. 2012; Zheng and Jiang 2013; Zhang et al. 2013), classifiers adaptation (Hestenes 1969; Wu and Jia 2012) and deep learning (Zhu et al. 2014). Specifically, feature adaptation algorithms aim to seek a common space where features across different views could be aligned well. Classifier adaptation algorithms are designed to adapt classifier trained on one view to another view. Whilst deep learning algorithms focus on building a hierarchical structure to capture more discriminative features to mitigate the view divergence.

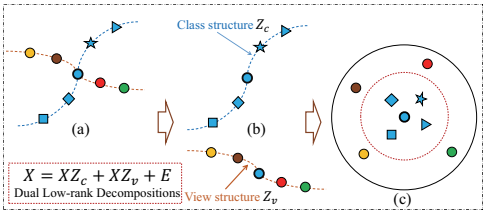


Figure 1: Framework Illustration. Note that the same color denotes same class, while same shape means same view. (a) represents the two intertwined structures of multi-view data X at one data point (blue ○). (b) Dual low-rank decompositions tend to separate view structure Z_v and class structure Z_c from each other through $X = X(Z_c + Z_v) + E$. (c) denotes the desired results with the guidance of two supervised graph regularizers, and therefore, the same class data points are pulled close, while different class data points lying in the same view are pushed far away.

Recently, low-rank constraint is widely adopted in data representation, which originally helps uncover the multiple structures of the data by finding the lowest rank representation and detecting noise or outliers. Along this line, recent works (Bao et al. 2013; Zhang et al. 2015) adopt dual low-rank decompositions to handle largely corrupted data situations. However, they usually apply low-rank decomposition in the original high-dimensional space and none of these works consider two intertwined structures for multi-view data. Most recently, robust subspace learning smoothly integrates subspace learning and low-rank decomposition in a unified framework to deal with high-dimensional data (Li and Fu 2014; Ding et al. 2015; Ding, Shao, and Fu 2014).

In this paper, we develop a novel multi-view learning al-

gorithm, named as Robust Multi-view Subspace Learning (RMSL), to seek a view-invariant subspace(Figure 1). Since there are two kinds of intertwined structures within multi-view data, it is essential to preserve more within-class information while removing the influence of view-variance from the same class. As far as we know, we are the first to consider dual low-rank decompositions in multi-view data analysis. To sum up, the major contributions are highlighted as:

- Dual low-rank decompositions are proposed to disassemble two intertwined structures within multi-view data. Therefore, a more robust view-invariant subspace is learned to capture more label information while reducing the influence from view-variance. It is of great importance for such case in reality, where we cannot have the view knowledge of the testing data at hand, since we only seek a common view-invariant subspace, which is applicable for different views.
- Two novel graph regularizers are designed to guide the dual low-rank decompositions in a supervised manner. Specifically, within-class feature tends to be compact while within-view feature among different classes will be discriminative. This practice would further boost the classification performance by better handling view-variance influence within each class.

Related Work

In this section, we first present two lines of the most related work, then highlight the differences of our work.

Low-rank representation (Candès et al. 2011; Liu, Lin, and Yu 2010; Liu et al. 2013) becomes very popular and widely applied in many fields. Among them, Robust PCA (RPCA) (Candès et al. 2011) is a well-known algorithm, which assumes the data is drawn from one single subspace. However, real-world data is generally lying in multiple subspaces. Therefore, Low-Rank Representation (LRR) (Liu, Lin, and Yu 2010; Liu et al. 2013) is developed to uncover the global class structures within the data meanwhile detect sparse noise or outliers. LRR employs low-rank constraint on the original space of data, therefore, it is very time-consuming when the data is very high-dimensional. Besides, for multi-view data, there exist two kinds of global structures (i.e., class structure and view structure) mixed together so that LRR cannot well handle such cases. To this end, we propose dual low-rank decompositions to learn a view-invariant subspace for multi-view data.

Low-rank subspace learning attempts to seek a robust low-dimensional subspace by integrating low-rank constraint with subspace learning (Li and Fu 2014; Ding et al. 2015; Ding, Shao, and Fu 2014; Shao, Kit, and Fu 2014). Specifically, subspace learning aims to seek a relatively low-dimensional subspace, which can preserve the intrinsic structure within the data. Conventional subspace learning algorithms are generally separated into unsupervised and supervised fashion. Those conventional subspace learning algorithms, however, are heavily sensitive to noisy or corrupted data, which leads to a very poor classification performance in dealing with real-world data. Two most representative low-rank subspace learning methods are DLML (Ding et al. 2015) and SRRS (Li and Fu 2014), which tend to take the advantages of both subspace learning and low-rank representation to generate a robust subspace. Our proposed one also adopts this idea to analyze multi-view data, however, RMSL aims to learn a view-invariant subspace through dual low-rank decompositions with the guidance of supervised graph regularizers, which aims to preserve more within-class information while reducing the influence of view-variance.

Robust Multi-view Subspace Learning

In this section, we propose our novel Robust Multi-view Subspace Learning. Then, we develop an efficient solution with its complexity analysis.

Dual Low-Rank Decompositions

Assume there is multi-view data $X = [X_1, \cdots, X_k]$ with k views, and each view $X_i \in \mathbb{R}^{d \times m_i}$ contains the same c classes, where d is the dimensionality of the original feature and m_i is the sample-size of each view data ($m = \sum_i m_i$). Conventional low-rank representation methods (Liu, Lin, and Yu 2010; Liu et al. 2013) aim to seek a robust representation Z to seek the global multiple subspaces structure within the data as follows:

$$\min_{Z,E} \operatorname{rank}(Z) + \lambda \|E\|_1, \ \text{s.t.}\ X = XZ + E, \tag{1}$$

where $\operatorname{rank}(\cdot)$ is the rank operator of a matrix. $Z \in \mathbb{R}^{m \times m}$ is the low-rank coefficient matrix and $E \in \mathbb{R}^{d \times m}$ is the sparse error part constrained with l_1-norm in order to handle noisy data. λ is the trade-off parameter. Generally, Z uncovers class structure of the data X. However, for multi-class multi-view images, it is difficult for Z to uncover the class structure due to the large divergence across different views within one class. Therefore, the assumption that Z is low-rank cannot hold. Besides, the data within the same view but from different classes would gather very close.

In fact, there are two kinds of structures behind multi-view data (Su et al. 2014), one is class structure while the other is view-variance structure. These two independent structures are mixed together, that is, each data would lie in two structures. Both structures should be low-rank, since class structure aims to uncover global structure for class information, while view-variance structure preserves the view information across different classes. Therefore, Z can be decomposed into two low-rank parts as follows:

$$\begin{aligned} \min_{Z_c, Z_v, E} & \operatorname{rank}(Z_c) + \operatorname{rank}(Z_v) + \lambda \|E\|_1, \\ \text{s.t.}\ & X = X(Z_c + Z_v) + E, \end{aligned} \tag{2}$$

where $Z_c \in \mathbb{R}^{m \times m}$ and $Z_v \in \mathbb{R}^{m \times m}$ are the low-rank representations for class structure and view-variance structure, respectively. In this way, two structures can be dug out from each other. Therefore, the redundant part from view structure can be removed from class structure so that we can achieve a better global class structure. However, such an unsupervised manner cannot separate the two structures in the way we expect. Therefore, it is essential to guide the dual low-rank decompositions with supervised information.

Robust Multi-view Subspace Learning

To better guide the dual low-rank decompositions in our previous model (2), we develop two supervised graph regularizers in order to disassemble those two intertwined structures in an expected way. Moreover, our goal is to seek a robust view-invariant subspace for multi-view data. Following recent low-rank subspace learning methods (Li and Fu 2014; Ding et al. 2015), we develop our view-invariant robust subspace learning framework as:

$$\begin{aligned}\min_{P,Z_c,Z_v,E} &\|Z_c\|_* + \|Z_v\|_* + \lambda\|E\|_1 + \alpha\mathcal{G}(P,Z_c,Z_v)\\ \text{s.t. } &P^{\mathrm{T}}X = P^{\mathrm{T}}X(Z_c+Z_v)+E,\ P^{\mathrm{T}}P = \mathrm{I}_{\mathrm{p}},\end{aligned} \tag{3}$$

where α is the balanced parameter for the supervised graph regularizer $\mathcal{G}(P,Z_c,Z_v)$. $P \in \mathbb{R}^{d\times p}$ is the learned subspace (p is reduced dimensionality). The previous rank minimization problem is solved by the nuclear norm $\|\cdot\|_*$ as a good surrogate (Liu, Lin, and Yu 2010; Liu et al. 2013). Note that the orthogonal constraint $P^{\mathrm{T}}P = \mathrm{I}_{\mathrm{p}}$ ($\mathrm{I}_{\mathrm{p}} \in \mathbb{R}^{p\times p}$) is imposed to avoid some trivial solutions to P.

Next, we present how to define the graph regularizer. Specifically, we incorporate supervised information, e.g., class information, view information, to guide the dual low-rank decompositions. To this end, we design two graphs for class manifold structure and view manifold structure. Since we aim to preserve more within-class information while moving out the influence of view-variance, we propose to minimize within-class similarity on the new low-dimensional within-class feature $Y_c = P^{\mathrm{T}}XZ_c$ ($Y_c \in \mathbb{R}^{p\times m}$) while maximizing between-view dissimilarity on the new low-dimensional within-view feature $Y_v = P^{\mathrm{T}}XZ_v$ ($Y_v \in \mathbb{R}^{p\times m}$). Therefore, we develop the following two graph terms:

$$\begin{aligned}\mathcal{G}_c &= \textstyle\sum_{i,j}(Y_{c,i}-Y_{c,j})^2 W^c_{i,j}\\ \mathcal{G}_v &= \textstyle\sum_{i,j}(Y_{v,i}-Y_{v,j})^2 W^v_{i,j}\end{aligned} \tag{4}$$

where $Y_{c,i}, Y_{c,j}$ are the i-th and j-th column of Y_c, while $Y_{v,i}, Y_{v,j}$ are the i-th and j-th column of Y_v. W^c and W^v are two weight matrices of two graphs, whose elements are defined as follows:

$$W^c_{i,j} = \begin{cases} 1, & \text{if } x_i \in \mathcal{N}_{k_1}(x_j), \text{ and } l_i = l_j,\\ 0, & \text{otherwise}\end{cases} \tag{5}$$

$$W^v_{i,j} = \begin{cases} 1, & \text{if } x_i \in \mathbf{N}_{k_2}(x_j), \text{but } l_i \neq l_j,\\ 0, & \text{otherwise}\end{cases} \tag{6}$$

where l_i, l_j are the labels of sample x_i, x_j, respectively. $x_i \in \mathcal{N}_{k_1}(x_j)$ means x_i is the k_1 nearest neighbor of the same class data x_j, while $x_i \in \mathbf{N}_{k_2}(x_j)$ denotes x_i belongs to the k_2 nearest neighbor of the same view data x_j. In this way, we can preserve the local manifold structure within the same class and dig out the influence of view manifold.

To this end, we design the graph regularizer $\mathcal{G}(P,Z_c,Z_v)$ to minimize the within-class variance while maximizing the margin for different classes but within the same view. We formulate $\mathcal{G}(P,Z_c,Z_v)$ in LDA fashion as:

$$\mathcal{G}(P,Z_c,Z_v) = \frac{\mathcal{G}_c}{\mathcal{G}_v} = \frac{\mathrm{tr}(P^{\mathrm{T}}XZ_cL_c(P^{\mathrm{T}}XZ_c)^{\mathrm{T}})}{\mathrm{tr}(P^{\mathrm{T}}XZ_vL_v(P^{\mathrm{T}}XZ_v)^{\mathrm{T}})}, \tag{7}$$

where L_c and L_v are the graph Laplacian of W^c and W^v, respectively (He and Niyogi 2003). To make our solution to Eq. (3) simple, we convert the trace ratio to trace difference (Li and Fu 2014) and achieve:

$$\begin{aligned}\mathcal{G}(P,Z_c,Z_v) \quad &= \mathrm{tr}(P^{\mathrm{T}}XZ_cL_c(P^{\mathrm{T}}XZ_c)^{\mathrm{T}})\\ &\quad -\beta\mathrm{tr}(P^{\mathrm{T}}XZ_vL_v(P^{\mathrm{T}}XZ_v)^{\mathrm{T}}),\end{aligned} \tag{8}$$

where we directly set the trade-off β between $\mathcal{G}_c$ and $\mathcal{G}_v$ as 1 for simplicity throughout this work.

Discussion: Our dual low-rank decompositions are designed to separate the intertwined two structures in multi-view data, one is class structure, and the other is view structure. With the supervision of two novel graph regularizers, our new learned subspace would keep the within-class data more compact while maximizing the margin between two different data within the same view. In this way, two intertwined structures could be disassembled so that the view-variance influence can be minimized. Furthermore, a more robust view-invariant subspace is learned to facilitate multi-view data learning task with the merits of subspace learning and low-rank representation.

Solving Objective Function

Problem (3) could be addressed by Augmented Lagrange Methods (ALM) (Liu, Lin, and Yu 2010; Liu and Yan 2011). However, ALM would introduce extra relax variables, which results in complex matrix operations during optimization, e.g., matrix inverse and multiplications. We adopt the first order Taylor expansion like approximation in order to save computational cost of the original quadratic term. In this way, we can achieve a simpler solution, similar to ALM for the original problem (3). To clarify it, we first formulate problem (3) in the augmented Lagrangian function as:

$$\begin{aligned}\min_{P,E,Z_c,Z_v,Q} &\|Z_c\|_* + \|Z_v\|_* + \lambda\|E\|_1 + \alpha\mathcal{G}(P,Z_c,Z_v)\\ &+\langle Q, P^{\mathrm{T}}X - P^{\mathrm{T}}X(Z_c+Z_v) - E\rangle\\ &+\tfrac{\mu}{2}\|P^{\mathrm{T}}X - P^{\mathrm{T}}X(Z_c+Z_v) - E\|^2_{\mathrm{F}},\end{aligned} \tag{9}$$

where Q is lagrange multiplier while $\mu > 0$ is the positive penalty. $\langle,\rangle$ denotes the matrix inner product operator, i.e., $\langle U,V\rangle = \mathrm{tr}(U^{\mathrm{T}}V)$. $\|\cdot\|^2_{\mathrm{F}}$ means Frobenius norm of a matrix.

Then we reformulate Eq. (9) by merging the last three terms into a quadratic term as follows:

$$\begin{aligned}\min_{P,E,Z_c,Z_v,Q} &\|Z_c\|_* + \|Z_v\|_* + \lambda\|E\|_1\\ &+h(P,Z_c,Z_v,E,Q,\mu) - \tfrac{1}{\mu}\|Q\|^2_{\mathrm{F}},\end{aligned} \tag{10}$$

where $h(P,Z_c,Z_v,E,Q,\mu) = \alpha\mathcal{G}(P,Z_c,Z_v) + \frac{\mu}{2}\|P^{\mathrm{T}}X - P^{\mathrm{T}}X(Z_c+Z_v) - E + \frac{Q}{\mu}\|^2_{\mathrm{F}}$. Similar to the conventional ALM, variables Z_c, Z_v, P and E in Eq. (10) cannot be solved jointly, but they are solvable one by one when fixing others. To this end, we address each subproblem separately by approximating the quadratic term h to first order Taylor expansion, while treating others variables as constant. We define the t-th iteration optimized variables as $Z_{c,t}, Z_{v,t}, E_t, P_t$ and Q_t. Superficially, we can achieve each sub-solution at the $t+1$ ($t \geq 0$)-th iteration as:

Updating Z_c:

$$\begin{aligned} Z_{c,t+1} &= \arg\min_{Z_c} \|Z_c\|_* + h(Z_c, Z_{v,t}, E_t, P_t, Q_t, \mu) \\ &= \arg\min_{Z_c} \|Z_c\|_* + \tfrac{\eta\mu}{2}\|Z_c - Z_c^{(t)}\|_{\mathrm{F}}^2 \\ &\quad + \langle \nabla_{Z_c} h, Z_c - Z_c^{(t)} \rangle \\ &= \arg\min_{Z_c} \tfrac{1}{\eta\mu}\|Z_c\|_* + \tfrac{1}{2}\|Z_c - Z_{c,t} + \nabla_{Z_c} h\|_{\mathrm{F}}^2, \end{aligned} \tag{11}$$

where $\nabla_{Z_c} h = \nabla_{Z_c} h(Z_{c,t}, Z_{v,t}, E_t, P_t, Q_t, \mu) = 2\alpha X^{\mathrm{T}} P_t P_t^{\mathrm{T}} X Z_{c,t} L_c - Q_t^{\mathrm{T}} P_t^{\mathrm{T}} X - \mu X^{\mathrm{T}} P_t (P_t^{\mathrm{T}} X - P_t^{\mathrm{T}} X (Z_{c,t} + Z_{v,t}) - E_t)$ and $\eta = \|{P_t}^{\mathrm{T}} X\|_2^2$. Problem (11) can be solved with singular value thresholding effectively (Cai, Candès, and Shen 2010).

Updating Z_v:

$$\begin{aligned} Z_{v,t+1} &= \arg\min_{Z_v} \|Z_v\|_* + h(Z_{c,t+1}, Z_v, E_t, P_t, Q_t, \mu) \\ &= \arg\min_{Z_v} \|Z_v\|_* + \tfrac{\eta\mu}{2}\|Z_v - Z_v^{(t)}\|_{\mathrm{F}}^2 \\ &\quad + \langle \nabla_{Z_v} h, Z_v - Z_v^{(t)} \rangle \\ &= \arg\min_{Z_v} \tfrac{1}{\eta\mu}\|Z_v\|_* + \tfrac{1}{2}\|Z_v - Z_{v,t} + \nabla_{Z_v} h\|_{\mathrm{F}}^2, \end{aligned} \tag{12}$$

where $\nabla_{Z_v} h = \nabla_{Z_v} h(Z_{c,t+1}, Z_{v,t}, E_t, P_t, Q_t, \mu) = -2\alpha X^{\mathrm{T}} P_t P_t^{\mathrm{T}} X Z_{v,t} L_v - Q_t^{\mathrm{T}} P_t^{\mathrm{T}} X - \mu X^{\mathrm{T}} P_t (P_t^{\mathrm{T}} X - P_t^{\mathrm{T}} X (Z_{c,t+1} + Z_{v,t}) - E_t)$. Problem (12) can be solved in the same way to problem (11).

Updating E:

$$\begin{aligned} E_{t+1} = \quad & \arg\min_{E} \tfrac{\lambda}{\mu}\|E\|_1 \\ & + \tfrac{1}{2}\|E - (P_t^{\mathrm{T}}(\widetilde{X}_{t+1} + \tfrac{Q_t}{\mu})\|_{\mathrm{F}}^2, \end{aligned} \tag{13}$$

where we define $\widetilde{X}_{t+1} = X - X(Z_{c,t+1} + Z_{v,t+1})$ for simplicity. Problem (13) can be solved by using the shrinkage operator (Lin, Chen, and Ma 2010).

Updating P:

$$\begin{aligned} P_{t+1} = \quad & \big(2\alpha X \widetilde{Z}_{t+1} X^{\mathrm{T}} + \mu \widetilde{X}_{t+1} \widetilde{X}_{t+1}^{T}\big)^{-1} \\ & \big(\widetilde{X}_{t+1}(E_{t+1} - \tfrac{Q_t}{\mu})^{\mathrm{T}}\big), \end{aligned} \tag{14}$$

where we define $\widetilde{Z}_{t+1} = Z_{c,t+1} L_c Z_{c,t+1}^{\mathrm{T}} - Z_{v,t+1} L_v Z_{v,t+1}^{\mathrm{T}}$ for simplicity.

Algorithm 1 lists the detailed solutions to problem (10), where we set those parameters μ_0, ρ, ϵ, $t_{\max}$ and $\mu_{\max}$ empirically, while tuning the two trade-offs, i.e., λ and α throughout the experiment, which is further discussed in experimental part. Moreover, P gets initialized with random matrixes. We have adopted traditional subspace learning methods to initialize P in different types, and found that the final evaluation performance tends to be almost the same.

Complexity Analysis

For simplicity, we mainly analyze the complexity of optimization parts listed in **Algorithm 1**. Note that $X \in \mathbb{R}^{d\times m}$, $P \in \mathbb{R}^{d\times p}$ and $Z_c \in \mathbb{R}^{m\times m}$, $Z_v \in \mathbb{R}^{m\times m}$. In **Algorithm 1**, we can observe that the most time-consuming components are the trace norm computation in Step 1& 2, and matrix multiplications & inverse in Step 4.

Algorithm 1 Solution to Problem (3)

Input: data X, variables $\lambda, \alpha, L_c, L_v$

Initialize: $E_0 = Y_0 = 0, \epsilon = 10^{-6}, \rho = 1.3, \mu = 10^{-6}$, $\mu_{\max} = 10^6, t_{\max} = 10^3, t = 0$.

while not converged **or** $t \le t_{\max}$ **do**

1. Optimize $Z_{c,t+1}$ according to (11) by fixing others;
2. Optimize $Z_{v,t+1}$ according to (12) by fixing others;
3. Optimize E_{t+1} according to (13) by fixing others;
4. Optimize P_{t+1} according to (14) by fixing others, then $P_{t+1} \leftarrow$ **orthogonal**(P_{t+1})
5. Optimize the multiplier Q_{t+1}
   $Q_{t+1} = Q_t + \mu\big(P_{t+1}^{\mathrm{T}}(X - X(Z_{c,t+1} + Z_{v,t+1})) + E_{t+1}\big)$;
6. Update the parameter μ by $\mu = \min(\rho\mu, \mu_{\max})$;
7. Check the convergence conditions
   $\|P_{t+1}^{\mathrm{T}}(X - X(Z_{c,t+1} + Z_{v,t+1})) - E_{t+1}\|_\infty < \epsilon$.
8. $t = t + 1$.

end while

output: Z_c, Z_v, E, P

Next we present the computational cost of each part in detail.Since conventional SVD operator in Step 1&2 would cost $O(m^3)$ for low-rank matrices Z_c, Z_v, repetitively. Fortunately, step 1& 2 can be accelerated to $O(rm^2)$, where r is the rank of the low-rank matrix, by the recent fast low-rank method(Liu et al. 2013). Generally, the rank of Z_v is less than that of Z_c for multi-view multi-class data analysis tasks. On the other hand, each matrix multiplication costs close to $O(d^3)$ and the matrix inverse takes $O(d^3)$ for $d \times d$ matrixes. Therefore, step 4 costs nearly $(k+1)O(d^3)$ in total, when there are k multiplication operations.

Experiments

In this part, we will describe the used datasets and experimental setting first. Then we will show the comparison results of our algorithm and others algorithms, followed by the evaluation on some properties of our algorithm.

Datasets & Experimental Setting

CMU-PIE Face database is consisted of 68 subjects in total, which is a multi-view face dataset [1] and show large variances within the same subject but in different poses. Each pose per subject has 21 different illumination variations. We adopt face images from 9 different poses, e.g., C02, C05, C22, C27, C07, C14, C29, C31, C34. We select different numbers of poses to construct various evaluation subsets. The face images are cropped to 64×64 size and only the pixel value features are adopted as the input.

COIL-100 object database[2] contains 100 categories with 7200 images. Each object has 72 images and each was captured with 5-degree rotation. We further partition the

[1] http://vasc.ri.cmu.edu/idb/html/face/

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

Table 1: Comparison Results (%) of 8 algorithms on the **original** CMU-PIE multi-pose face database. **Bold** denotes the best performance.

Algorithms	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7
PCA	69.03±0.08	69.21±0.08	68.52±0.12	71.58±0.14	52.65±0.04	34.94±0.08	29.09±0.01
LDA	70.46±0.05	71.32±0.02	63.51±0.75	72.12±0.09	56.53±0.02	24.07±0.25	7.06±0.01
LPP	57.25±0.06	58.83±0.07	59.25±0.56	65.58±0.13	43.56±0.08	19.67±0.05	13.11±0.01
RPCA+PCA	74.39±0.08	75.55±0.12	75.29±0.09	78.27 ±0.09	61.17±0.12	38.66±0.08	31.94±0.12
LatLRR	77.92±0.03	76.24±0.12	75.29±0.07	83.68±0.07	69.74±0.05	42.54±0.12	35.33±0.04
SRRS	78.27±0.04	78.74±0.23	77.45±0.02	86.28±0.09	71.44±0.03	38.86±0.02	30.16±0.02
LRCS	87.78±0.02	86.67±0.01	87.38±0.19	89.12±0.12	74.84±0.04	44.48±0.03	36.17±0.01
Ours	**89.15±0.06**	**88.05±0.07**	**88.40±0.17**	**93.95±0.11**	**75.16±0.12**	**44.93±0.11**	**37.14±0.08**

Table 2: Comparison Results (%) of 8 algorithms on **corrupted** CMU-PIE multi-pose face database. **Bold** denotes the best performance.

Algorithms	Case 1	Case 2	Case 3	Case 4	Case 5	Case 6	Case 7
PCA	64.87±0.32	66.04±0.08	65.21±0.04	69.32±0.09	50.16±0.04	31.74±0.08	27.21±0.01
LDA	26.71±0.20	23.19±0.35	20.34±0.75	35.12±0.08	46.72±0.02	6.67±0.25	4.06±0.01
LPP	31.26±0.26	30.98±0.18	32.21±0.36	40.34±0.14	27.66±0.05	14.34±0.04	12.02±0.01
RPCA+PCA	73.07±0.11	74.28±0.12	73.92±0.12	73.98±0.10	60.18±0.14	37.65±0.09	31.34±0.06
LatLRR	73.10±0.07	73.24±0.32	73.85±0.12	75.21±0.08	58.94±0.09	39.26±0.12	32.07±0.03
SRRS	72.27±0.05	72.74±0.18	71.45±0.08	74.19±0.13	54.32±0.03	32.34±0.02	29.03±0.02
LRCS	78.98±0.03	78.67±0.05	78.38±0.26	80.54±0.12	65.84±0.04	39.48±0.03	32.57±0.01
Ours	**82.12±0.08**	**82.67±0.09**	**82.38±0.17**	**84.18±0.12**	**69.84±0.09**	**43.87±0.11**	**35.78±0.08**

dataset into two subsets as "COIL1" and "COIL2". Specifically, COIL1 includes the images in View 1 $[0^o, 85^o]$ and View 2 $[180^o, 265^o]$ while COIL2 contains those in View 3 $[90^o, 175^o]$ and View 4 $[270^o, 355^o]$. The raw feature with image size 64×64 with 20% corruption are adopted in the experiments.

Comparison Experiments

We mainly compare with feature extraction algorithms, which are: LDA (Belhumeur, Hespanha, and Kriegman 1997), LatLRR (Liu and Yan 2011), SRRS (Li and Fu 2014), LPP (He and Niyogi 2003), PCA (Turk and Pentland 1991), RPCA (Wright et al. 2009)+PCA and LRCS (Ding and Fu 2014). Among them, LDA, SRRS and our proposed method belong to the supervised fashion; PCA, LPP, RPCA and LatLRR are totally unsupervised; while LRCS can be treated as weakly supervised, which needs to know the view information of the data in the training stage. To all compared algorithms, we evaluate the final performance in terms of recognition accuracy by using the Nearest Neighbor Classifier (NNC). For CMU-PIE faces, we randomly select 10 face images from each subject per view to build the dataset for training, while the rest face images are used for final evaluation. In total, we conduct 5 random selections and achieve the average performance. Table 1 &2 present the comparison results of 8 algorithms on both clean and noisy face samples with 10% corruptions, where Case 1: {C02, C14}, Case 2: {C02, C27}, Case 3: {C14, C27}, Case 4: {C05, C29}, Case 5: {C05, C07, C29}, Case 6: {C05, C14, C29, C34}, Case 7: {C02, C05, C14, C29, C31}. For COIL-100 objects, we select one from COIL1 and one from COIL2 as the training

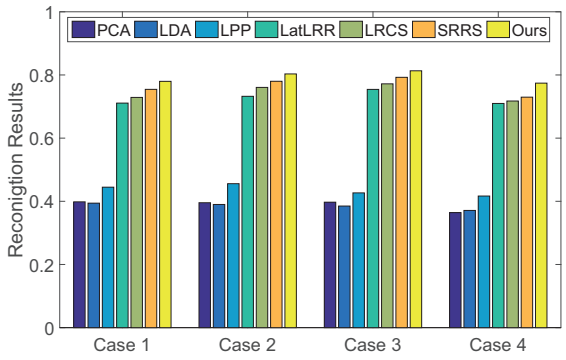
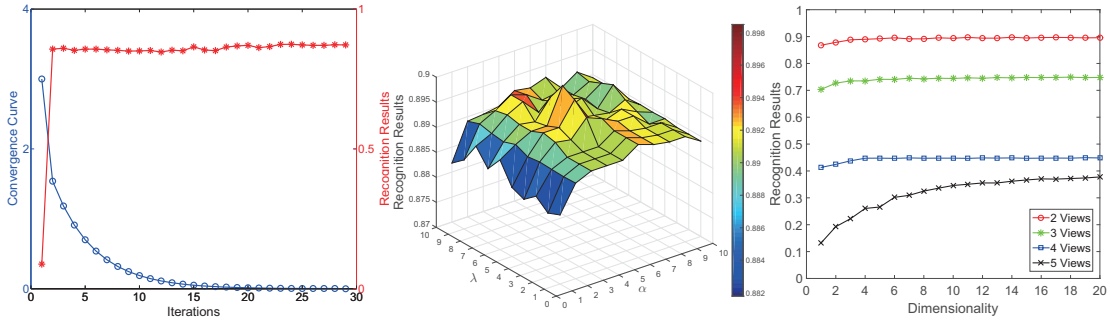


Figure 2: Recognition results of seven algorithms on 4 Cases of COIL-100 object dataset, where Case 1: View 1 & 3; Case 2: View 1 & 4; Case 3: View 2 & 3; Case 4: View 2 & 4.

with 2 views and the remaining samples are used for final evaluation. In total, there are 4 cases to evaluate the performance of the algorithms, which are shown in Figure 2.

From the recognition results (Table 1,2 and Figure 2), we can observe that our proposed algorithm achieves better recognition performance than other comparisons both in face and object images. For CMU-PIE face dataset, the performance of all the algorithms would decrease when there are more views involving, since more within-class variance are involved. Besides, we notice that LDA degrades much quicker than others. This represents that the class information may not do a favor sometimes, due to a large variance



(a) (b) (c)

Figure 3: (a) Recognition curve (red '*') and convergence curve (Blue 'o') of our proposed method in 2-view case (C02&C27), where we set the dimensionality to 400 and the parameter values as $\lambda = 10^{-2}, \alpha = 10^{2}$, respectively. (b) The performance of our algorithm is evaluated on the two parameters influence $\{\alpha, \lambda\}$ using 2-view case (C02&C27), where the value from 0 to 10 denotes $[10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^{2}, 10^{3}, 10^{4}, 10^{5}]$, respectively. (c) Dimensionality influence of our algorithm on four cases, where the value from 1 to 20 represents 50 to 1000, respectively.

within each class. For the four two-view combinations, all the algorithms obtain almost the same results in each combination, i.e., the divergence between two views is equivalent to some extent. These four combinations can well demonstrate the superiority of our proposed method, which represents our view-invariant subspace captures the most intrinsic information from two-view face images. In three-view combination, our proposed method could not achieve a large improvement, since there is a high similarity across the three poses. With more view involved in the evaluation, our algorithm can still outperform other algorithms, even with more views. Besides, low-rank based methods perform better than the non-low-rank others, especially in corrupted cases, as they can detect noise with the sparse error terms.

Moreover, we observe that the images within each view of CMU PIE have 21 different illuminations, some even invisible. This phenomenon leads to the similar performance of PCA in clean and noisy situations. This also results in traditional supervised methods cannot outperform unsupervised ones. Furthermore, from the corrupted object database, we can observe low-rank based methods outperform traditional subspace learning methods, as we manually introduce random noise, which would definitely decrease the performance of traditional subspace learning methods. However, our algorithm incorporates dual low-rank decompositions to dig out the view-variance structure within each class so that our algorithm could better handle multi-view data. Moreover, with two supervised graph regularizers to guide the decompositions, our algorithm can achieve even better performance in a supervised fashion.

Property Analysis

In this section, we will evaluate some properties of our algorithm, e.g., convergence, parameter analysis, dimensionality influence, and training time cost.

First, we conduct some experiments on convergence curve and recognition results of different iterations. We evaluate on two-view case {C02,C27} and the results are shown as Figure 3(a). From the results, we can observe our algorithm converges very fast. Also we notice that the recognition results go up quickly and stay at a stable value.

Table 3: Training time (*second*) of three algorithms on CMU-PIE face dataset.

Config	2 Views	3 Views	4 Views	5 Views
LatLRR	291.5	817.7	1635.4	2736.9
LRCS	184.0	547.3	1305.3	2311.1
Ours	72.3	162.6	311.3	510.4

Second, since there are two parameters λ, α in our algorithm, we evaluate them simultaneously on case and the results are shown in Figure 3(b). From the results, we can observe the performance is not good when α is close to 0, that is, two novel graph regularizers definitely help a lot. Generally, the performance would be best when α is around 10^{2}. However, the error term influences very small. Therefore, we set $\lambda = 10^{-2}$ and $\alpha = 10^{2}$ throughout the experiments.

Third, we testify the dimensionality influence of our algorithm on several cases, whose recognition results are presented in Figure 3(c). From the results, we can see that recognition rates increase when dimensionality goes up. It would reach its highest performance around 400.

Finally, we also calculate the training time cost of our algorithm by comparing several others. We apply on different views of CMU-PIE dataset and run 10 iterations to calculate the training time. Experiments are conducted with Matlab 2014b, CPU i7-3770 and 32 GB memory size. The computational costs for training are shown in Table 3 (unit is *second*). as we can see from the results, our proposed algorithm is more efficiently than LRCS and LatLRR. The reason mainly attribute to the efficient solution for the optimization, which avoids introducing relaxing variables with their extra matrix multiplications. Also we deploy the dual low-rank decompositions in the low-dimensional subspace.

Conclusion

In this paper, we developed a Robust Multi-view Subspace Learning to seek a view-invariant for multi-view data analysis. Specifically, we proposed a dual low-rank decompositions to separate two intertwined structures and therefore preserved more within-class compactness by degrading the influence of large view variance from the same class. Furthermore, two supervised graph regularizers were incorporated into the low-rank decompositions so that it could guide the decompositions to further preserve more within-class local manifold information and within-view marginal structure. Experimental results on multi-view datasets, witnessed the effectiveness and efficiency of our proposed algorithm, compared to the existing feature extraction algorithms.

Acknowledgement

This research is supported in part by the NSF CNS award 1314484, ONR award N00014-12-1-1028, ONR Young Investigator Award N00014-14-1-0484, NPS award N00244-15-1-0041, and U.S. Army Research Office Young Investigator Award W911NF-14-1-0218.

References

Bao, B.-K.; Liu, G.; Hong, R.; Yan, S.; and Xu, C. 2013. General subspace learning with corrupted training data via graph embedding. *IEEE Transactions on Image Processing* 22(11):4380–4393.

Belhumeur, P. N.; Hespanha, J. P.; and Kriegman, D. J. 1997. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 19(7):711–720.

Cai, X.; Wang, C.; Xiao, B.; Chen, X.; and Zhou, J. 2013. Regularized latent least square regression for cross pose face recognition. In *Proceedings of the Twenty-Third international joint conference on Artificial Intelligence*, 1247–1253.

Cai, J.-F.; Candès, E. J.; and Shen, Z. 2010. A singular value thresholding algorithm for matrix completion. *SIAM Journal on Optimization* 20(4):1956–1982.

Candès, E. J.; Li, X.; Ma, Y.; and Wright, J. 2011. Robust principal component analysis? *Journal of the ACM* 58(3):11.

Ding, Z., and Fu, Y. 2014. Low-rank common subspace for multi-view learning. In *IEEE International Conference on Data Mining*, 110–119. IEEE.

Ding, C., and Tao, D. 2015. A comprehensive survey on pose-invariant face recognition. *arXiv preprint arXiv:1502.04383*.

Ding, Z.; Suh, S.; Han, J.-J.; Choi, C.; and Fu, Y. 2015. Discriminative low-rank metric learning for face recognition. In *12th IEEE International Conference on Automatic Face and Gesture Recognition*.

Ding, Z.; Shao, M.; and Fu, Y. 2014. Latent low-rank transfer subspace learning for missing modality recognition. In *Twenty-Eighth AAAI Conference on Artificial Intelligence*, 1192–1198.

He, X., and Niyogi, P. 2003. Locality preserving projections. In *Neural information processing systems*, volume 16, 153.

Hestenes, M. R. 1969. Multiplier and gradient methods. *Journal of optimization theory and applications* 4(5):303–320.

Kan, M.; Shan, S.; Zhang, H.; Lao, S.; and Chen, X. 2012. Multi-view discriminant analysis. In *Proceedings of European Conference on Computer Vision*. Springer. 808–821.

Li, S., and Fu, Y. 2014. Robust subspace discovery through supervised low-rank constraints. In *Proceedings of SIAM International Conference on Data Mining*, 163–171.

Lin, Z.; Chen, M.; and Ma, Y. 2010. The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices. *arXiv preprint arXiv:1009.5055*.

Liu, G., and Yan, S. 2011. Latent low-rank representation for subspace segmentation and feature extraction. In *IEEE International Conference on Computer Vision*, 1615–1622.

Liu, S.; Yi, D.; Lei, Z.; and Li, S. Z. 2012. Heterogeneous face image matching using multi-scale features. In *Fifth IAPR International Conference on Biometrics*, 79–84. IEEE.

Liu, G.; Lin, Z.; Yan, S.; Sun, J.; Yu, Y.; and Ma, Y. 2013. Robust recovery of subspace structures by low-rank representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 171–184.

Liu, G.; Lin, Z.; and Yu, Y. 2010. Robust subspace segmentation by low-rank representation. In *Proceedings of the Twenty-Seventh International Conference on Machine Learning*, 663–670.

Shao, M.; Kit, D.; and Fu, Y. 2014. Generalized transfer subspace learning through low-rank constraint. *International Journal of Computer Vision* 109(1-2):74–93.

Shekhar, S.; Patel, V.; Nasrabadi, N.; and Chellappa, R. 2014. Joint sparse representation for robust multimodal biometrics recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 36(1):113–126.

Su, Y.; Li, S.; Wang, S.; and Fu, Y. 2014. Submanifold decomposition. *IEEE Transactions on Circuits and Systems for Video Technology* 24(11):1885–1897.

Turk, M., and Pentland, A. 1991. Eigenfaces for recognition. *Journal of cognitive neuroscience* 3(1):71–86.

Wright, J.; Ganesh, A.; Rao, S.; Peng, Y.; and Ma, Y. 2009. Robust principal component analysis: Exact recovery of corrupted low-rank matrices via convex optimization. In *Advances in neural information processing systems*, 2080–2088.

Wu, X., and Jia, Y. 2012. View-invariant action recognition using latent kernelized structural svm. In *European Conference on Computer Vision*. Springer. 411–424.

Zhang, Y.; Shao, M.; Wong, E. K.; and Fu, Y. 2013. Random faces guided sparse many-to-one encoder for pose-invariant face recognition. In *IEEE International Conference on Computer Vision*, 2416–2423. IEEE.

Zhang, F.; Yang, J.; Tai, Y.; and Tang, J. 2015. Double nuclear norm-based matrix decomposition for occluded image recovery and background modeling. *IEEE Transactions on Image Processing* 24(6):1956–1966.

Zhao, H., and Fu, Y. 2015. Dual-regularized multi-view outlier detection. In *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence*, 4077–4083.

Zheng, J., and Jiang, Z. 2013. Learning view-invariant sparse representations for cross-view action recognition. In *IEEE International Conference on Computer Vision*, 3176–3183. IEEE.

Zhu, Z.; Luo, P.; Wang, X.; and Tang, X. 2014. Multi-view perceptron: a deep model for learning face identity and view representations. In *Advances in Neural Information Processing Systems*, 217–225.