Robust Multimodal Recognition via Multitask Multivariate Low-Rank Representations

Heng Zhang, Vishal M. Patel and Rama Chellappa

Center for Automation Research, UMIACS, University of Maryland, College Park, MD 20742 USA

{hzhang98, pvishalm, rama}@umiacs.umd.edu

Abstract— We propose multi-task, multivariate low-rank representation-based methods for multimodal biometrics recognition. Our methods can be viewed as a generalized version of multivariate low-rank regression, where low-rank representation across all the modalities is imposed. One of our methods takes into account coupling information among different biometric modalities simultaneously by enforcing the common low-rank representation within each biometric's observations. We further modify our methods by including a background occlusion term that is assumed to be sparse. Alternating direction method of multipliers is proposed to solve the proposed optimization problems. Extensive experiments using face and touch gesture dataset show that our method compares favorably with other feature level fusion-based methods.

I. INTRODUCTION

Developments in sensing and communication technology have let to an explosion in the availability of visual data from multiple sources and modalities. Millions of cameras have been installed in buildings, streets, and airports around the world that are capable of capturing multimodal information such as light, depth and heat. This has resulted in the development of various multi sensor fusion algorithms. In particular, multimodal classification has received a lot of attention in recent years where one uses information from various modalities recording the same physical event to achieve an improved classification performance [1], [2].

Multimodal biometrics recognition [3], [4] can be regarded as a special case of multi sensor classification. In a multimodal biometrics system, the evidence presented by multiple sources of information such as face, fingerprints and iris are integrated for recognition. One of the advantages of multibiometric systems is that they are less vulnerable to spoof attacks. Fusion can be achieved at multiple levels, which can be broadly divided into sensor level, feature level, score level and decision or rank level fusion. Since feature level fusion preserves the raw information, it can be more discriminative than score or rank level fusion. However, the differences in features extracted from different sensors in terms of types and dimensions make the feature level fusion very difficult. One of the simplest methods for feature level fusion is feature concatenation. While feature concatenation has been used for multibiometric fusion [5], [6], they often tend to be non-robust and inefficient.

Multiple Kernel Learning (MKL) can also be used to integrate information from multiple features by learning a weighted combination of appropriate kernels. See [7] for details on various MKL algorithms. Recently, various sparse representation-based multimodal fusion algorithms have been proposed in the literature [8], [9], [10]. In [8], a multi-task sparse linear regression model is proposed for image classification. In order to recognize the same object viewed from multiple observations (e.g. poses), a joint dynamic sparse representation method was proposed in [9]. In [10] a joint sparse representation-based method was proposed for fusing multiple biometrics features. This method is based on multi-task multivariate Lasso [11]. It imposes common sparsities both within each biometric modality and across different modalities. This method produced state-of-the-art results on various multimodal biometrics recognition problems. One of the advantages of using sparse representation-based fusion for classification is that they are robust to noise and occlusion [12], [13].

Different from sparse representation-based fusion methods, we propose low-rank representation-based methods for multimodal biometrics recognition. Our method seeks the lowest-rank representation when the data is represented as a linear combination of the training data. Low-rank representations come with good theory, and powerful computational tools [14], [15] and hence have had impact on a wide variety of problems such as subspace clustering [16], [17], matrix competition [14], image alignment [18] and object recognition [19]. We propose multimodal, multivariate, low-rank representation for fusing multiple biometric traits. One of the proposed methods is based on multivariate low-rank regression [20], [21]. It can deal with multimodal as well as multivariate low-rank representations. The other method imposes common low-rank representation across different modalities. The resulting optimization problems are solved using the classical Alternating Direction Method of Multipliers (ADMM) [22].

This paper makes the following contributions:

- Robust multi-modal recognition methods using Multitask Multivariate Low-Rank Representations are proposed for multi-modal biometric fusion.
- Efficient iterative methods based on the ADMM are proposed for solving the proposed optimization problems.
- The algorithms are evaluated on a new dataset consisting of face and touch gestures collected from 50 mobile phone users for active authentication.

The paper is organized as follows. In Section II,

we define the proposed multitask, multivariate, low-rank representation-based multimodal recognition problems. Optimization procedures are described in Section III. Experimental evaluations are described in Section IV. Finally, concluding remarks are presented in Section V with a brief summary and discussion.

II. BACKGROUND AND PROBLEM FORMULATION

Suppose we are given a C-class classification problem with D different modalities. Assume that there are m training samples in each modality. For each modality, $i = 1, \cdots, D$, we denote $\mathbf{X}^i = [\mathbf{X}_1^i, \mathbf{X}_2^i, \cdots, \mathbf{X}_C^i]$ as an $n_i \times m$ matrix of training samples containing C sub-matrices $\mathbf{X}_j^i$'s corresponding to C different classes. Each sub-matrix $\mathbf{X}_j^i = [\mathbf{x}_{j,1}^i, \mathbf{x}_{j,2}^i, \cdots, \mathbf{x}_{j,m_j}^i] \in \mathbb{R}^{n_i \times m_j}$ contains a set of training samples from the ith modality corresponding to the jth class. Here, m_j is the number of training samples in class j and n_i is the feature dimension of each sample. As a result, there are in total $m = \sum_{j=1}^C m_j$ many samples in $\mathbf{X}_C$. Given a test matrix $\mathbf{Y}$, which consists of D different modalities, $\{\mathbf{Y}^1, \cdots, \mathbf{Y}^D\}$, where each sample $\mathbf{Y}^i$ consists of d_i observations $\mathbf{Y}^i = [\mathbf{y}_1^i, \mathbf{y}_2^i, \cdots, \mathbf{y}_{d_i}^i] \in \mathbb{R}^{n_i \times d_i}$, the objective is to identify the class to which a test sample $\mathbf{Y}$ belongs to.

A. Multitask Multivariate Low-Rank Representation (MLRR)

Let $\{\mathbf{Y}^i\}_{i=1}^D$ be a set of D observations each consisting of d_i samples from each modality. Let $\mathbf{\Gamma} = [\mathbf{\Gamma}^1, \mathbf{\Gamma}^2, \cdots, \mathbf{\Gamma}^D] \in \mathbb{R}^{m \times d}$ be the coefficient matrix formed by concatenating D representation matrices with $d = \sum_{i=1}^D d_i$. We wish to solve for the low-rank matrix $\mathbf{\Gamma}$ by solving the following multivariate, low-rank representation problem

$$\hat{\mathbf{\Gamma}} = \arg\min_{\mathbf{\Gamma}} \frac{1}{2} \sum_{i=1}^D \|\mathbf{Y}^i - \mathbf{X}^i \mathbf{\Gamma}^i\|_F^2 + \lambda \|\mathbf{\Gamma}\|_*, \quad (1)$$

where $\|\mathbf{A}\|_F = \sqrt{\sum_{i,j} A_{i,j}^2}$ is the Frobenius norm of $\mathbf{A}$, λ is a positive regularization parameter and $\|\mathbf{A}\|_* = \sum_i \sigma_i(A)$ is the sum of the singular values of $\mathbf{A}$ (i.e. the nuclear norm of $\mathbf{A}$). Once the low-rank matrix $\hat{\mathbf{\Gamma}}$ is obtained, the class label associated with an observation vector is declared as the one that produces the smallest approximation error

$$\hat{\ell} = \arg\min_{\ell} \sum_{i=1}^D \|\mathbf{Y}^i - \mathbf{X}^i \delta_\ell(\hat{\mathbf{\Gamma}}^i)\|_F^2, \quad (2)$$

where $\delta_\ell(\cdot)$ is the matrix indicator function that keeps rows corresponding to the ℓth class and sets all other rows equal to zero. In the case when $D = 1$, (1) reduces to multivariate, low-rank regression problem [20], [21]. See [23], [24], [25] and the references therein for more details on low-rank regression[1].

Ideally, the learned coefficients corresponding to the correct class should exhibit relative larger values compared to the coefficients corresponding to the incorrect classes. In

order take this assumption into the classification mechanism, for a given coefficient vector obtained from the ith modality, we define $\mathbf{w}_\ell^i$ as:

$$\mathbf{w}_\ell^i = \frac{C^{\frac{\|\delta_\ell(\hat{\mathbf{\Gamma}}^i)\|_*}{\|\hat{\mathbf{\Gamma}}^i\|_*}} - 1}{C - 1}. \quad (3)$$

Therefore, the classification rule (2) can be modified as:

$$\hat{\ell} = \arg\min_{\ell} \sum_{i=1}^D \mathbf{w}_\ell^i \|\mathbf{Y}^i - \mathbf{X}^i \delta_\ell(\hat{\mathbf{\Gamma}}^i)\|_F^2. \quad (4)$$

Note that similar idea has been explored in [12] and [10].

B. Robust Multitask Multivariate Low-Rank Representation (RMLRR)

In the case when the data is contaminated by noise and occlusion, the observation can be modeled as follows

$$\mathbf{Y}^i = \mathbf{X}^i \mathbf{\Gamma}^i + \mathbf{N}^i + \mathbf{E}^i,$$

where $\mathbf{N}^i$ is a small dense additive noise and $\mathbf{E}^i$ is a matrix of sparse occlusion (background noise) with arbitrary large magnitude. By taking advantage of the fact that $\mathbf{E}^i$ is sparse, one can simultaneously estimate $\mathbf{\Gamma}^i$ and $\mathbf{E}^i$ by solving the following optimization problem

$$\hat{\mathbf{\Gamma}}, \hat{\mathbf{E}} = \arg\min_{\mathbf{\Gamma}, \mathbf{E}} \frac{1}{2} \sum_{i=1}^D \|\mathbf{Y}^i - \mathbf{X}^i \mathbf{\Gamma}^i - \mathbf{E}^i\|_F^2$$
$$+ \lambda_1 \|\mathbf{\Gamma}\|_* + \lambda_2 \|\mathbf{E}\|_1, \quad (5)$$

where $\mathbf{E} = [\mathbf{E}^1, \mathbf{E}^2, \cdots, \mathbf{E}^D]$ is the spare occlusion matrix and $\|\mathbf{A}\|_1 = \sum_{i,j} |A_{i,j}|$ is the ℓ_1-norm of $\mathbf{A}$. Note that $\mathbf{E}$ is just a compact representation and we solve each $\mathbf{E}^i$ separately since their dimensions are different because of different modalities they are related to. Here, λ_1 and λ_2 are positive parameters that control the rank and sparsity, respectively. Once $\mathbf{\Gamma}$ and $\mathbf{E}$ are estimated, the effect of occlusion can be removed by setting $\hat{\mathbf{Y}}^i = \mathbf{Y}^i - \hat{\mathbf{E}}^i$. Finally, one can declare the class label associated to an observation vector as

$$\hat{\ell} = \arg\min_{\ell} \sum_{i=1}^D \mathbf{w}_\ell^i \|\mathbf{Y}^i - \mathbf{X}^i \delta_\ell(\hat{\mathbf{\Gamma}}^i) - \mathbf{E}^i\|_F^2, \quad (6)$$

where $\mathbf{w}_\ell^i$ is defined in (3).

C. Multitask Multivariate Common Low-Rank Representation (MCLRR)

In this section, we propose a different formulation in which a common low-rank representation is enforced for different modalities. By sharing the same low-rank representation across the modalities, we can enforce the similarity of the representations among the different modalities. As a results, we can obtain a more robust low-rank representation. Similar ideas have been explored in [26] for jointly learning sparse representation for image super-resolution.

In this case, we assume that the observations are of the following form:

$$\mathbf{Y}^i = \mathbf{X}^i \mathbf{\Gamma} + \mathbf{N}^i.$$

---

[1]also known as reduced rank regression

Note that, the same representation is used for all the modalities in the above model and we assume that the number of samples from each modality is the same, i.e $\mathbf{Y}^i \in \mathbb{R}^{n_i \times d}$. We propose the following optimization problem

$$\hat{\mathbf{\Gamma}} = \arg\min_{\mathbf{\Gamma}} \frac{1}{2}\sum_{i=1}^{D} \|\mathbf{Y}^i - \mathbf{X}^i\mathbf{\Gamma}\|_F^2 + \lambda\|\mathbf{\Gamma}\|_* \qquad (7)$$

for jointly learning the lowest rank representation. Once $\hat{\mathbf{\Gamma}}$ is estimated, it can be used to declare the class label of the observation by the minimum reconstruction error criteria as follows

$$\hat{\ell} = \arg\min_{\ell} \sum_{i=1}^{D} \mathbf{w}_\ell \|\mathbf{Y}^i - \mathbf{X}^i\delta_\ell(\hat{\mathbf{\Gamma}})\|_F^2, \qquad (8)$$

where $\mathbf{w}_\ell$ is defined as

$$\mathbf{w}_\ell = \frac{C^{\frac{\|\delta_\ell(\hat{\mathbf{\Gamma}})\|_*}{\|\hat{\mathbf{\Gamma}}\|_*}} - 1}{C - 1}. \qquad (9)$$

.

D. Robust Multitask Multivariate Common Low-Rank Representation (RMCLRR)

Similar to the MLRR method, the joint low-rank representation model can be extended for the case when the data is contaminated by sparse errors. Assuming the following observation model

$$\mathbf{Y}^i = \mathbf{X}^i\mathbf{\Gamma} + \mathbf{N}^i + \mathbf{E}^i, \qquad (10)$$

one can recover the common low-rank representation by solving the following optimization problem

$$\hat{\mathbf{\Gamma}}, \hat{\mathbf{E}} = \arg\min_{\mathbf{\Gamma},\mathbf{E}} \frac{1}{2}\sum_{i=1}^{D} \|\mathbf{Y}^i - \mathbf{X}^i\mathbf{\Gamma} - \mathbf{E}^i\|_F^2$$
$$+ \lambda_1\|\mathbf{\Gamma}\|_* + \lambda_2\|\mathbf{E}\|_1. \qquad (11)$$

After estimating $\mathbf{\Gamma}$, the class label associated with an observation can be declared as follows

$$\hat{\ell} = \arg\min_{\ell} \sum_{i=1}^{D} \mathbf{w}_\ell \|\mathbf{Y}^i - \mathbf{X}^i\delta_\ell(\hat{\mathbf{\Gamma}}) - \mathbf{E}^i\|_F^2, \qquad (12)$$

where $\mathbf{w}_\ell$ is defined in Eq. (9).

III. OPTIMIZATION

In this section we propose an approach based on the ADMM method [22] to solve the optimization problems denoted by Eqs. (1), (5), (7) and (11). Due to the similarity of these problems and page limitations, we only provide details on the optimization of the RMLRR problem. In ADMM, appropriate auxiliary variables are introduced into the optimization program, the constraints are augmented into the objective function and the Lagrangian is iteratively minimized with respect to the primal variables and maximized with respect to the Lagrange multipliers. Note that we let all the primal variables and lagrange multipliers be zero to initialize the algorithm.

A. Optimization of RMLRR

The problem (5) can be reformulated by introducing the auxiliary variables as follows

$$\arg\min_{\mathbf{\Gamma},\mathbf{E},\mathbf{V},\mathbf{U}} \frac{1}{2}\sum_{i=1}^{D} \|\mathbf{Y}^i - \mathbf{X}^i\mathbf{\Gamma}^i - \mathbf{E}^i\|_F^2 + \lambda_1\|\mathbf{V}\|_*$$
$$+ \lambda_2\|\mathbf{U}\|_1 \text{ s.t. } \mathbf{\Gamma} = \mathbf{V}, \mathbf{E} = \mathbf{U}. \qquad (13)$$

Note that Eq.(13) is equally constrained problem which can be solved using the Augmented Lagrangian Method (ALM) [22]. The augmented Lagrangian function $f_{\alpha_\Gamma,\alpha_E}(\mathbf{\Gamma}, \mathbf{E}, \mathbf{V}, \mathbf{U}; \mathbf{A}_E, \mathbf{A}_\Gamma)$ is defined as

$$\arg\min_{\mathbf{\Gamma},\mathbf{E},\mathbf{V},\mathbf{U}} \frac{1}{2}\sum_{i=1}^{D} \|\mathbf{Y}^i - \mathbf{X}^i\mathbf{\Gamma}^i - \mathbf{E}^i\|_F^2 + \lambda_1\|\mathbf{V}\|_*$$
$$\langle\mathbf{A}_\Gamma, \mathbf{\Gamma} - \mathbf{V}\rangle + \frac{\alpha_\Gamma}{2}\|\mathbf{\Gamma} - \mathbf{V}\|_F^2 + \lambda_2\|\mathbf{U}\|_1 + \langle\mathbf{A}_E, \mathbf{E} - \mathbf{U}\rangle$$
$$+\frac{\alpha_E}{2}\|\mathbf{E} - \mathbf{U}\|_F^2, \qquad (14)$$

where $\mathbf{A}_E$ and $\mathbf{A}_\Gamma$ are the multipliers of the two linear constrains, α_E and α_Γ are the positive parameters, and $\langle\mathbf{A}, \mathbf{B}\rangle$ denotes $tr(\mathbf{A}^T\mathbf{B})$. In the ALM algorithm, $f_{\alpha_\Gamma,\alpha_E}$ is solved with respect to $\mathbf{\Gamma}, \mathbf{E}, \mathbf{U}$ and $\mathbf{V}$ jointly, while keeping $\mathbf{A}_\Gamma$ and $\mathbf{A}_E$ fixed and then updating $\mathbf{A}_\Gamma$ and $\mathbf{A}_E$ keeping the remaining variables fixed.

1) Update step for $\mathbf{\Gamma}$: Obtain $\mathbf{\Gamma}_{t+1}$ by minimizing $f_{\alpha_\Gamma,\alpha_E}$ with respect to $\mathbf{\Gamma}$. This can be done by taking the first-order derivative of $f_{\alpha_\Gamma,\alpha_E}$ and setting it equal to zero. Furthermore, the first term of $f_{\alpha_\Gamma,\alpha_E}$ is a sum of convex functions associated with sub-matrices $\mathbf{\Gamma}^i$, one can find $\mathbf{\Gamma}_{t+1}^i, i = 1, \cdots, D$, by solving the following linear system

$$(\mathbf{X}^{i^T}\mathbf{X}^i + \alpha_\Gamma\mathbf{I})\mathbf{\Gamma}_{t+1}^i = \mathbf{X}^{i^T}(\mathbf{Y}^i - \mathbf{E}_t^i) + \alpha_\Gamma\mathbf{V}_t^i - \mathbf{A}_{\Gamma,t}^i \quad (15)$$

where $\mathbf{I}$ is $m \times m$ identity matrix and $\mathbf{E}_t^i, \mathbf{V}_t^i$ and $\mathbf{A}_{\Gamma,t}^i$ are submatrices of $\mathbf{\Gamma}_t, \mathbf{V}_t$ and $\mathbf{A}_{\Gamma,t}$, respectively. When m is is not very large, one can simply apply matrix inversion to obtain $\mathbf{\Gamma}_{t+1}^i$ from Eq.(15). For large values of m, gradient-based methods should be employed to obtain $\mathbf{\Gamma}_{t+1}^i$.

2) Update step for $\mathbf{E}$: The second optimization is similar in nature whose solution is give as follows

$$\mathbf{E}_{t+1}^i = (1 + \alpha_E)^{-1}(\mathbf{Y}^i - \mathbf{X}^i\mathbf{\Gamma}_{t+1}^i + \alpha_E\mathbf{U}_t^i - \mathbf{A}_{E,t}^i),$$

where $\mathbf{U}_t^i$ and $\mathbf{A}_{E,t}^i$ are sub-matrices of $\mathbf{U}_t$ and $\mathbf{A}_{E,t}$, respectively.

3) Update step for $\mathbf{U}$: In order to update $\mathbf{U}$, one needs to solve the following ℓ_1 minimization problem

$$\min \frac{1}{2}\|\mathbf{E}_{t+1} + \alpha_E^{-1}\mathbf{A}_{E,t} - \mathbf{U}\|_F^2 + \frac{\lambda_2}{\alpha_E}\|\mathbf{U}\|_1 \qquad (16)$$

whose solution is given by [27]

$$\mathbf{U}_{t+1} = \mathcal{S}\left(\mathbf{E}_{t+1} + \alpha_E^{-1}\mathbf{A}_{E,t}, \frac{\lambda_2}{\alpha_E}\right),$$

where $\mathcal{S}(a, b) = sgn(a)(|a| - b)$ for $|a| \geq b$ and zero otherwise.

4) Update step for $\mathbf{V}$*:* The final suboptimization for updating $\mathbf{V}$ has the following form

$$\min \frac{1}{2}\|\mathbf{\Gamma}_{t+1} + \alpha_\Gamma^{-1}\mathbf{A}_{\Gamma,t} - \mathbf{V}\|_F^2 + \frac{\lambda_1}{\alpha_\Gamma}\|\mathbf{V}\|_*. \quad (17)$$

Solution to this optimization problem is obtained by shrinking the singular values of $\mathbf{\Gamma}_{t+1} + \alpha_\Gamma^{-1}\mathbf{A}_{\Gamma,t}$ [14], [15]. As a result, we obtain the following update for $\mathbf{V}$

$$\mathbf{V}_{t+1} = \mathbf{F}\mathcal{L}_{\frac{\lambda_1}{\alpha_\Gamma}}(\mathbf{\Sigma})\mathbf{B}^T,$$

where $\mathbf{F}\mathbf{\Sigma}\mathbf{B}^T$ is the Singular Value Decomposition (SVD) of $\mathbf{\Gamma}_{t+1} + \alpha_\Gamma^{-1}\mathbf{A}_{\Gamma,t}$ and

$$\mathcal{L}_{\frac{\lambda_1}{\alpha_\Gamma}}(x) = \begin{cases} x - \frac{\lambda_1}{\alpha_\Gamma}, & x > \frac{\lambda_1}{\alpha_\Gamma} \\ x + \frac{\lambda_1}{\alpha_\Gamma}, & x < -\frac{\lambda_1}{\alpha_\Gamma} \\ 0, & \text{otherwise.} \end{cases}$$

5) Update steps for $\mathbf{A}_\Gamma$ *and* $\mathbf{A}_E$*:* Finally, the Lagrange multipliers are updated as

$$\mathbf{A}_{\Gamma,t+1} = \mathbf{A}_{\Gamma,t} + \alpha_\Gamma(\mathbf{\Gamma}_{t+1} - \mathbf{V}_{t+1}) \quad (18)$$

$$\mathbf{A}_{E,t+1} = \mathbf{A}_{E,t} + \alpha_E(\mathbf{E}_{t+1} - \mathbf{U}_{t+1}). \quad (19)$$

The proposed ADMM algorithm for solving the RMLRR problem is summarized in Algorithm 1.

Algorithm 1: Robust Multitask Multivariate Low-Rank Representation (RMLRR) using ADMM.

Input: Training samples $\{\mathbf{X}_i\}_{i=1}^D$, test sample $\{\mathbf{Y}_i\}_{i=1}^D, \lambda_1, \lambda_2$
Initialization:
$\mathbf{\Gamma}_0, \mathbf{V}_0, \mathbf{U}_0, \mathbf{A}_{E,0}, \mathbf{A}_{\Gamma,0}, \alpha_\Gamma, \alpha_E$
While not converged do
1. Update $\mathbf{\Gamma}$: $\mathbf{\Gamma}_{t+1} = [\mathbf{\Gamma}_{t+1}^1, \cdots, \mathbf{\Gamma}_{t+1}^D]$, where

$$\mathbf{\Gamma}_{t+1}^i = (\mathbf{X}^{i^T}\mathbf{X}^i + \alpha_\Gamma\mathbf{I})^{-1}(\mathbf{X}^{i^T}(\mathbf{Y}^i - \mathbf{E}_t^i) + \alpha_\Gamma\mathbf{V}_t^i - \mathbf{A}_{\Gamma,t}^i)$$

2. Update $\mathbf{E}$: $\mathbf{E}_{t+1} = [\mathbf{E}_{t+1}^i, \cdots, \mathbf{E}_{t+1}^D]$, where

$$\mathbf{E}_{t+1}^i = (1 + \alpha_E)^{-1}(\mathbf{Y}^i - \mathbf{X}^i\mathbf{\Gamma}_{t+1}^i + \alpha_E\mathbf{U}_t^i - \mathbf{A}_{E,t}^i)$$

3. Update $\mathbf{U}$:

$$\mathbf{U}_{t+1} = \mathcal{S}\left(\mathbf{E}_{t+1} + \alpha_E^{-1}\mathbf{A}_{E,t}, \frac{\lambda_2}{\alpha_E}\right)$$

4. Update $\mathbf{V}$:

$$\mathbf{V}_{t+1} = \mathbf{F}\mathcal{L}_{\frac{\lambda_1}{\alpha_\Gamma}}(\mathbf{\Sigma})\mathbf{B}^T$$

5. Update $\mathbf{A}_\Gamma$: $\mathbf{A}_{\Gamma,t+1} = \mathbf{A}_{\Gamma,t} + \alpha_\Gamma(\mathbf{\Gamma}_{t+1} - \mathbf{V}_{t+1})$
6. Update $\mathbf{A}_E$: $\mathbf{A}_{E,t+1} = \mathbf{A}_{E,t} + \alpha_E(\mathbf{E}_{t+1} - \mathbf{U}_{t+1})$
Output: $\hat{\mathbf{E}} = \mathbf{E}_{t+1}$ and $\hat{\mathbf{\Gamma}} = \mathbf{\Gamma}_{t+1}$.

B. Optimization of RMCLRR

The RMCLRR problem (11) can be optimized in a similar way using the ADMM method. However, there are a few key differences in the implementation details. In particular, the update steps of $\mathbf{\Gamma}$ cannot be separated into D different sub optimization problems. Different steps of the optimization algorithm are summarized in Algorithm 2.

Algorithm 2: Robust Multitask Multivariate Common Low-Rank Representation (RMCLRR) using ADMM.

Input: Training samples $\{\mathbf{X}_i\}_{i=1}^D$, test sample $\{\mathbf{Y}_i\}_{i=1}^D, \lambda_1, \lambda_2$
Initialization:
$\mathbf{\Gamma}_0, \mathbf{V}_0, \mathbf{U}_0, \mathbf{A}_{E,0}, \mathbf{A}_{\Gamma,0}, \alpha_\Gamma, \alpha_E$
While not converged do
1. Update $\mathbf{\Gamma}$: $\mathbf{\Gamma}_{t+1} =$

$$\left(\sum_{i=1}^D \mathbf{X}^{i^T}\mathbf{X}^i + \alpha_\Gamma\mathbf{I}\right)^{-1}\left(\sum_{i=1}^D \mathbf{X}^{i^T}(\mathbf{Y}^i - \mathbf{E}^i) + \alpha_\Gamma\mathbf{V}_t - \mathbf{A}_{\Gamma,t}\right)$$

2. Update $\mathbf{E}$: $\mathbf{E}_{t+1} = [\mathbf{E}_{t+1}^i, \cdots, \mathbf{E}_{t+1}^D]$, where

$$\mathbf{E}_{t+1}^i = (1 + \alpha_E)^{-1}(\mathbf{Y}^i - \mathbf{X}^i\mathbf{\Gamma}_{t+1} + \alpha_E\mathbf{U}_t^i - \mathbf{A}_{E,t}^i)$$

3. Update $\mathbf{U}$:

$$\mathbf{U}_{t+1} = \mathcal{S}\left(\mathbf{E}_{t+1} + \alpha_E^{-1}\mathbf{A}_{E,t}, \frac{\lambda_2}{\alpha_E}\right)$$

4. Update $\mathbf{V}$:

$$\mathbf{V}_{t+1} = \mathbf{F}\mathcal{L}_{\frac{\lambda_1}{\alpha_\Gamma}}(\mathbf{\Sigma})\mathbf{B}^T$$

5. Update $\mathbf{A}_\Gamma$: $\mathbf{A}_{\Gamma,t+1} = \mathbf{A}_{\Gamma,t} + \alpha_\Gamma(\mathbf{\Gamma}_{t+1} - \mathbf{V}_{t+1})$
6. Update $\mathbf{A}_E$: $\mathbf{A}_{E,t+1} = \mathbf{A}_{E,t} + \alpha_E(\mathbf{E}_{t+1} - \mathbf{U}_{t+1})$
Output: $\hat{\mathbf{E}} = \mathbf{E}_{t+1}$ and $\hat{\mathbf{\Gamma}} = \mathbf{\Gamma}_{t+1}$.

IV. EXPERIMENTAL RESULTS

In this section, we evaluate the proposed algorithm on a face and touch gesture dataset collected by the authors' group using a mobile device for active authentication in different ambient conditions. We compare our proposed methods with several sate-of-the-art feature level multimodal fusion methods including MKL [28] and Sparsity-based Multimodal Biometrics Recognition (SMBR-WE and SMBR-E) [10] and score-level fusion method based on Sparse logistic Regression (SLR) [29] which is named as SLR-sum in [10].

A. Face and Screen Touch Gesture Dataset

Most mobile devices use passwords, pin numbers, or secret patterns for authenticating users. As long as the device remains active, there is no mechanism to verify that the user originally authenticated is still the user in control of the device. As a result, unauthorized individuals may improperly gain access to personal information of the user if the password is compromised. Active authentication systems deal with this issue by continuously monitoring the user identity after the initial access has been granted. Examples of such systems include screen touch gesture-based recognition [30] and gait-based recognition [31].

Faces have shown to be a promising biometric for verifying a user identity. In order to study the effectiveness of both faces and touch gestures for active authentication on mobile devices, we collected data from 50 users in an application environment on iPhone 5s. The users were asked to perform different tasks such as scrolling a document, viewing pictures, reading a long article etc. While users performed these tasks, their touch data sensed by the screen and face images acquired by the front-facing camera were simultaneously captured. The users were asked to perform these tasks in different sessions with different ambient conditions, namely
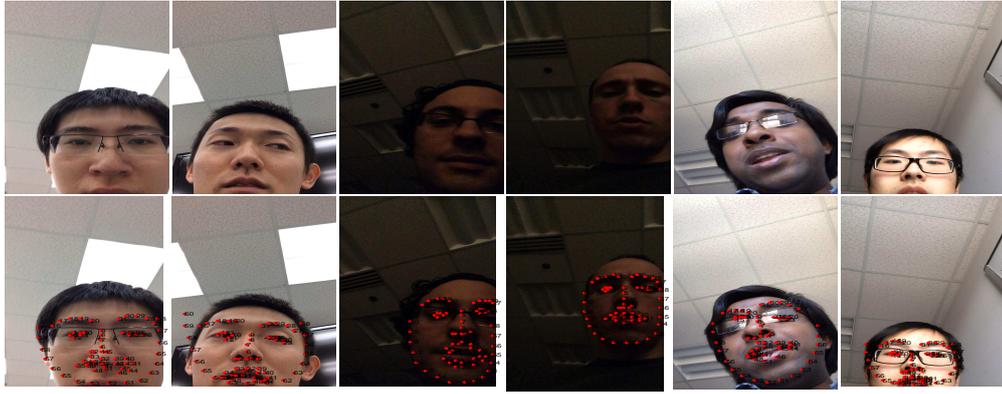
Fig. 1: First row: Example faces in this dataset. Second row: Detected landmarks on the images shown on the first row.

in a well-lit room, in a dim-lit room, and in a room with natural daytime illumination. During data collection, users were free to use the phone in either orientation mode and hold the phone in any position of their choice. The goal was to simulate real-world scenarios to study how ambient changes can influence users' face data captured by the frontal camera and can possibly influence touch gestures as users might swipe differently under different lighting conditions. Data collection from 50 users over 3 sessions resulted in 750 videos consisting of facial data with each video lasting between 0.5 minute to 2 minutes and 15490 touch swipes. It is a very challenging dataset. Since facial video data were collected in an unconstrained manner, many faces exhibit different poses, rotations and illuminations. In particular, partial faces are common in this dataset. First row of Figures 1 shows sample face images from this dataset. Each row shows images from a particular ambient condition. Raw touch swipes on the screen from this dataset are shown in Figure 2.
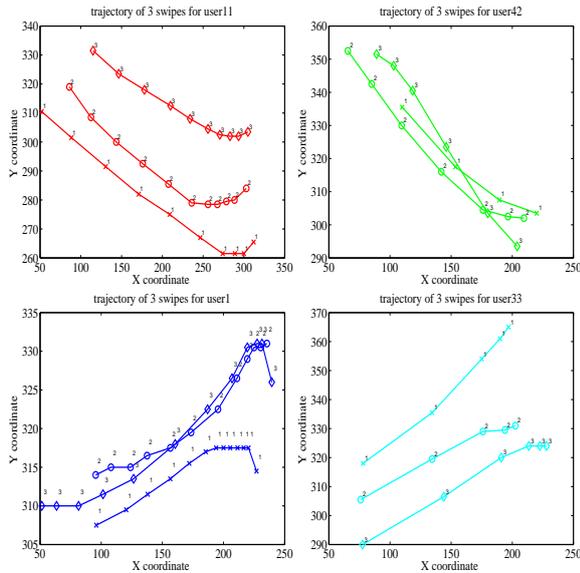

Fig. 2: Trajectories of sample touch swipes from our dataset.

B. Preprocessing and Feature Extraction

Since this dataset consists of two modalities, we perform preprocessing and feature extraction for face and screen touch data separately.

1) Faces: For the face data, we first detect the landmarks of the face images frame by frame from the videos using the tree-based landmarks detector [32]. These detected landmarks are shown in the second row of Figure 1. We then crop and align the faces using the method described in [33] based on the landmarks' locations. We then applied the illumination normalization method described in [34] to the cropped face images. Finally the face images were rescaled to dimension $192 \times 168 \times 3$ and converted to grayscale images. After preprocessing, we down sampled the preprocessed face images to 24 by 21 and used the whole image as a feature vector of dimension 504.

2) Touch Gestures: Every touch swipe $\mathbf{S}$ is encoded as a sequence of vectors

$$\mathbf{s}_i = (x_i, y_i, t_i, A_i, o_i^{ph}),$$

$i \in \{1, \cdots, N_c\}$ where x_i, y_i are the location points, t_i is the time stamp, A_i is the area occluded by the finger and o_i^{ph} is the orientation of the phone (e.g. landscape or portrait). Given these touch data, we extracted a 27 dimensional feature vector for every single swipe in the dataset using the method described in [30]. These features are summarized in Table I.

The new dataset is a relative large dataset. In order to evaluate the proposed multi-modal fusion methods, we sampled a small subset from this dataset. For the face component, for each user, we selected 30 faces from each session. As a result, in total we selected 4500 face images for 50 users across 3 different sessions. Similarly, for the touch component, we also selected the corresponding 4500 touch swipes. All the experiment done will be based on these selected 4500 face images and 4500 touch swipes. This part of the data and the Matlab implementation of our low-rank fusion methods will be made available for research purposes.

FeatureID	Description
feature 1	inter-stroke time
feature 2	stroke duration
feature 3	start x
feature 4	start y
feature 5	stop x
feature 6	stop y
feature 7	direct end-to-end distance
feature 8	mean resultant length
feature 9	up/down/left/right flag
feature 10	direction of end-to-end line
feature 11	20%-perc. pairwise velocity
feature 12	50%-perc. pairwise velocity
feature 13	80%-perc. pairwise velocity
feature 14	20%-perc. pairwise acceleration
feature 15	50%-perc. pairwise acceleration
feature 16	80%-perc. pairwise acceleration
feature 17	median velocity at last 3 points
feature 18	largest deviation from end-to-end line
feature 19	20%-perc. dev. from end-to-end line
feature 20	50%-perc. dev. from end-to-end line
feature 21	80%-perc. dev. from end-to-end line
feature 22	average direction
feature 23	length of trajectory
feature 24	ratio end-to-end dist and length of trajectory
feature 25	average velocity
feature 26	median acceleration at first 5 points
feature 27	mid-stroke area covered

TABLE I: Description of the 27 dimensional feature vector.

C. Fusion of Face and Screen Touch Gestures

In this experiment, we select 10 (15) samples for each user to form training data, and use rest of the data for testing. In total, there are 500 (750) samples for training and 4000 (3750) samples for testing. Each sample contains a 504 dimensional feature vector for one face image and a 27 dimensional feature vector for one screen touch gesture. By randomly splitting data for training and testing, we repeated each experiment 10 times and report mean and standard deviation of the rank 1 recognition accuracy. Rank K recognition accuracy is often used to evaluate biometric recognition algorithms. The experimental results comparing our proposed methods with the other multi-modal fusion methods are shown in Table II and Table III, respectively, when we use 10 and 15 training samples for each user.

Methods	Face	Touch	Face & Touch
SLR-sum	57.74 $\pm$ 0.96	22.05 $\pm$ 1.03	58.26 $\pm$0.92
MKL	72.58 $\pm$ 1.08	36.02 $\pm$ 0.49	75.13 $\pm$2.22
SMBR-WE	75.37 $\pm$ 1.13	30.40 $\pm$ 1.59	66.69 $\pm$0.78
SMBR-E	73.05 $\pm$ 1.29	27.72 $\pm$ 1.50	64.49 $\pm$1.61
MLRR	76.04 $\pm$ 0.92	21.95 $\pm$ 1.41	69.24 $\pm$0.85
RMLRR	75.94 $\pm$ 1.16	21.88 $\pm$ 1.35	69.21 $\pm$1.17
MCLRR	75.49 $\pm$ 1.03	22.02 $\pm$ 1.37	**78.58 $\pm$ 1.21**
RMCLRR	72.72 $\pm$ 1.49	21.88 $\pm$ 1.34	**77.93 $\pm$ 1.35**

TABLE II: Rank 1 recognition accuracy (in %) for different fusion methods using 10 samples from each user for training.

From the results shown in Table II and Table III, we make the following observations: (1) All the algorithm compared can achieve better recognition accuracy on face modality than on the touch modality. Faces as physical biometrics are more robust and reliable while screen touch gestures, as a kind

Methods	Face	Touch	Face & Touch
SLR-sum	74.77 $\pm$ 0.78	23.64 $\pm$ 1.23	75.26 $\pm$0.92
MKL	77.23 $\pm$ 0.57	39.19 $\pm$ 1.25	80.80 $\pm$1.22
SMBR-WE	81.44 $\pm$ 0.49	32.42 $\pm$ 1.13	74.31 $\pm$1.10
SMBR-E	79.12 $\pm$ 0.61	30.18 $\pm$ 1.22	71.90 $\pm$1.36
MLRR	81.04 $\pm$ 0.60	23.26 $\pm$ 1.57	75.82 $\pm$1.06
RMLRR	81.19 $\pm$ 0.63	23.27 $\pm$ 1.69	76.28 $\pm$1.06
MCLRR	80.60 $\pm$ 0.52	23.26 $\pm$ 1.58	**83.68 $\pm$ 0.53**
RMCLRR	79.19 $\pm$ 0.72	23.27 $\pm$ 1.65	**83.75 $\pm$ 0.66**

TABLE III: Rank 1 recognition accuracy (in %) for different fusion methods using 15 samples from each user for training.

of behavioral biometric, exhibits more variations and can changes more easily. (2) When fusing two modalities, not all the methods perform better than using any single modality alone for recognition. Compared to the other methods, our proposed fusion methods, in particular RMCLRR performs the best. SMBR-WE, SMBR-E, MLRR and RMLRR do not achieve the best performance in this experiment. When we try to learn a low-rank representation from these two modalities (one strong, one weak), MLRR and RMLRR may not be able to correctly obtain the correct internal structure of the data. The reason may be the fact that one part of the data (touch gestures) are too noisy. From the formulations of MCLRR and RMCLRR, these methods try to learn a common low-rank representation from different modalities. The learned common representation can correctly capture the low-rank structure of the input data with respect to the training data. (3) When we use more training samples, all methods can perform much better in terms of both single modality and the fusion of two modalities.

D. Fusion of Facical Components

As we can see from Figure 1, there are many variations in the dataset especially out of plane rotations and partial faces which make the face recognition difficult. As a result, instead of extracting features from holistic face images, we focus on different parts of the face such as right eye, left eye, nose and mouth as shown in Figure 3. We treat each face part as one modality and apply our proposed fusion methods to perform face recognition. Similar ideas have been proposed in [10] and [35]. Note that in this experiment, we do not align the face image before extracting each face part. Furthermore, we do not align the face components. For the right and left eye components, we extract them using the landmarks we obtained and resize the extracted components to $20 \times 30 \times 3$, and transformed it to grayscale image of size 20×30. We then apply the illumination normalization [34] to alleviate extreme illumination effects and finally get 600 dimensional feature vectors by concatenating each columns of the processed image part from. Similarly, for the nose and the mouth parts, we get 450 (15×30) dimensional feature vectors.

We select 10 (15) samples for each user to form the training data, and use rest of the data for testing. In total, there are 500 (750) samples for training and 4000 (3750) samples for testing. Each sample contains 4 feature vectors representing 4 different facial components with dimension

Fig. 3: Different facial components extracted using the detected landmarks.

600, 600, 450 and 450 for left eye, right eye, nose and mouth, respectively. By randomly splitting data into training and testing, we repeated each experiment 10 times and report the mean and standard deviation of the rank 1 recognition accuracy. The experimental results comparing the proposed methods with other multi-modal fusion methods are shown in Table IV and Table V, respectively, when we use 10 and 15 training samples for each user.

From the results shown in Table IV and Table V, we make several observations. (1) All the algorithms compared can achieve better recognition accuracy when fusing 4 facial components than performing recognition using single face part alone. Our proposed RMLRR method performs the best since the learned low-rank representation from 4 modalities can efficiently and correctly capture the structure of input data samples. The SLR-sum method performs the worst on this experiment since this method is performing score-level fusion rather than feature level fusion. This method does not fully utilize the information in the feature vectors for each facial component. (2) The recognition performance based on the right eye or left eye alone show relative high accuracy. This result is consistent with the findings in psychological studies of face recognition by humans as discussed in [36]. (3) Comparing the face recognition results based on the holistic face image, fusing different face components can perform much better. This is because the face images obtained from mobile devices shows many variations that typical face preprocessing procedures struggle to tackle especially when presented with pose variations and partial faces. This suggests fusing face parts may be a good solution for mobile face recognition.

E. Runtime Analysis and Convergence

To empirically show the convergence of our method, in Fig 4 (a) and (b), we show the objective function vs iteration plots of the ADMM method for solving Algorithm 1 (RMLRR) and Algorithm 2 (RMCLRR), respectively. As can be seen from this figure, the proposed algorithms do converge in a few iterations. Furthermore, our method is very efficient compared to other feature level fusion methods. For the fusion of 4 facial parts experiment using 10 samples from each subjects for training, on average MLRR, RMLRR, MCLRR and RMCLRR take about $0.22s$, $0.54s$, 0.06 and $0.43s$ respectively to classify a test sample, compared to $0.53s$ for SMBR-WE [10] and $0.87s$ for SMBR-E [10]. Note that, even tough MKL Algorithm [28] reqires $0.04s$ to classify a test sample, it requires a lot of time to train

the model while our methods do not require any training. Experiments were done in 64bit Matlab R2013a environment on laptop with 2.9GHz Intel Core i7-3520M CPU and 8GB Memory.
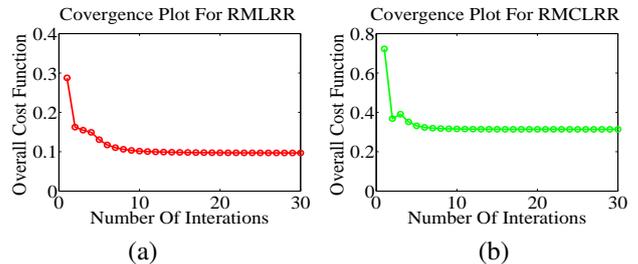

Fig. 4: Objective function versus number of iterations. (a)Algorithm 1 (RMLRR). (b)Algorithm 2 (RMCLRR).

V. CONCLUSION

In this paper, we proposed four new multi-modal fusion methods based on low-rank representations, namely MLRR, RMLRR, MCLRR and RMCLRR. Efficient optimization procedures are proposed for solving the proposed problems using the ADMM method. Extensive experiments on a dataset consisting of face and touch gestures collected using a mobile device show that the proposed methods can perform better than recently proposed sparsity-based fusion methods.

Our future research will focus on applying our proposed fusion methods for more general computer vision and patter recognition problems which are associated with multiple feature descriptors, like object recognition. Also, we will extend the proposed methods to the nonlinear case for application where the nonlinearity lies within the data.

REFERENCES

[1] D. L. Hall and J. Llinas, "An introduction to multisensor data fusion," *Proceedings of the IEEE*, vol. 85, no. 1, pp. 6–23, 1997.

[2] P. K. Varshney, "Multisensor data fusion," *Electronics and Communication Engineering Journal*, vol. 9, no. 6, pp. 245–253, 1997.

[3] A. Ross and A. K. Jain, "Multimodal biometrics: an overview," in *European Signal Processing Conference*, 2004, pp. 1221–1224.

[4] A. Ross, K. Nandakumar, and A. K. Jain, *Handbook of Multibiometrics*. Springer, 2006.

[5] A. Rattani, D. Kisku, M. Bicego, and M. Tistarelli, "Feature level fusion of face and fingerprint biometrics," in *IEEE International Conference on Biometrics: Theory, Applications, and Systems*, 2007, pp. 1–6.

[6] X. Zhou and B. Bhanu, "Feature fusion of face and gait for human recognition at a distance in video," in *International Conference on Pattern Recognition*, 2006, pp. 529–532.

[7] M. Gönen and E. Alpaydın, "Multiple kernel learning algorithms," *Journal of Machine Learning Research*, vol. 12, pp. 2211–2268, July 2011.

[8] X.-T. Yuan and S. Yan, "Visual classification with multi-task joint sparse representation," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2010, pp. 3493–3500.

[9] H. Zhang, N. M. Nasrabadi, Y. Zhang, and T. S. Huang, "Multiobservation visual recognition via joint dynamic sparse representation," in *International Conference on Computer Vision*, 2011, pp. 595–602.

[10] S. Shekhar, V. M. Patel, N. M. Nasrabadi, and R. Chellappa, "Joint sparse representation for robust multimodal biometrics recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 1, pp. 113–126, Jan 2014.

Methods	modality 1 (right eye)	modality 2 (left eye)	modality 3 (nose)	modality 4 (mouth)	Fusion of 4 modalities
SLR-sum	30.92 ± 1.14	32.97 ± 1.17	16.10 ±0.69	30.40 ± 1.13	48.94 ± 1.30
MKL	73.97 ± 0.74	74.70 ± 0.74	41.60 ±0.76	50.23 ± 1.02	88.23 ± 1.07
SMBR-WE	71.44 ± 0.77	72.12 ± 1.00	41.32 ± 0.78	50.08 ± 0.82	89.42 ± 0.80
SMBR-E	71.19 ± 0.69	71.99 ± 0.93	40.92 ± 0.69	49.94 ± 0.84	89.43 ± 0.88
MLRR	74.43 ± 0.57	74.96 ± 1.26	42.38 ± 0.64	55.92 ± 0.91	**91.62 ± 0.85**
RMLRR	77.23 ± 0.66	77.15 ± 0.98	44.58 ± 0.65	59.31 ± 0.84	**92.57 ± 0.72**
MCLRR	72.76 ± 0.66	73.46 ± 1.21	41.49 ± 0.72	51.26 ± 0.73	86.51 ± 1.02
RMCLRR	75.44 ± 0.87	75.54 ± 1.09	43.98 ± 0.89	54.54 ± 0.81	87.69 ± 0.86

TABLE IV: rank 1 recognition accuracy (in %) for different fusion methods on 4 facial parts using 10 samples from each user for training.

Methods	modality 1 (right eye)	modality 2 (left eye)	modality 3 (nose)	modality 4 (mouth)	Fusion of 4 modalities
SLR-sum	44.72 ± 1.00	47.94.4±1.13	22.52 ± 0.87	40.87 ± 0.66	67.06 ±0.87
MKL	79.96 ± 1.19	80.40 ±0.61	48.55 ± 1.34	57.44 ± 0.92	91.46 ±0.42
SMBR-WE	76.79 ± 1.26	77.11 ± 0.72	47.21 ± 1.06	55.50 ± 0.89	92.62 ±0.52
SMBR-E	76.52 ± 1.08	76.81 ± 0.71	46.39 ± 1.03	55.45 ± 1.05	92.71 ±0.45
MLRR	79.35 ± 0.88	79.48 ± 0.47	47.74 ± 0.83	60.94 ± 0.85	**94.37 ± 0.42**
RMLRR	81.97 ± 0.83	81.61 ± 0.41	50.22 ± 0.76	65.23 ± 0.71	**95.07 ± 0.37**
MCLRR	77.16 ± 1.20	77.35 ± 0.74	46.55 ± 0.94	55.11 ± 0.93	90.95 ±0.49
RMCLRR	80.21 ± 1.18	79.94 ± 0.54	49.90 ± 0.76	60.14 ± 0.88	91.91 ±0.39

TABLE V: rank 1 recognition accuracy (in %) for different fusion methods on 4 facial parts using 15 samples from each user for training.

[11] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society: Series B*, vol. 68, pp. 49–67, Feb 2006.

[12] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, Feb 2009.

[13] J. Pillai, V. M. Patel, R. Chellappa, and N. Ratha, "Secure and robust iris recognition using random projections and sparse representations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 9, pp. 1877–1893, Sept 2011.

[14] J.-F. Cai, E. J. Candès, and Z. Shen, "A singular value thresholding algorithm for matrix completion," *SIAM J. on Optimization*, vol. 20, no. 4, pp. 1956–1982, Mar 2010.

[15] E. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *Journal of the ACM*, vol. 58, no. 3, May 2011.

[16] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 171–184, 2013.

[17] R. Vidal and P. Favaro, "Low rank subspace clustering (LRSC)," *Pattern Recognition Letters*, 2013.

[18] Y. Peng, A. Ganesh, J. Wright, W. Xu, and Y. Ma, "Rasl: Robust alignment by sparse and low-rank decomposition for linearly correlated images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 11, pp. 2233–2246, 2011.

[19] I.-H. Jhuo, D. Liu, D. Lee, and S.-F. Chang, "Robust visual domain adaptation with low-rank reconstruction," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2012, pp. 2168–2175.

[20] A. Izenman, "Reduced-rank regression for the multivariate linear model," *Journal of Multivariate analysis*, vol. 5, pp. 248–262, 1975.

[21] T. Anderson, "Estimating linear restrictions on regression coefficients for multivariate normal distribution," *Annals of Mathematical Statistics*, vol. 22, pp. 327–351, 1951.

[22] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, Jan 2011.

[23] F. Bach, "Consistency of trace norm minimization," *Journal of Machine Learning Research*, vol. 9, pp. 1019–1048, 2008.

[24] M. Yuan, A. Ekici, Z. Lu, and R. Monteiro, "Dimension reduction and coefficient estimation in multivariate linear regression," *Journal of the Royal Statistical Society Series B*, vol. 69, pp. 329–346, 2007.

[25] A. Rohde and A. Tsybakov, "Estimation of high-dimensional low-rank matrices," *The Annals of Statistics*, vol. 39, no. 2, pp. 887–930, 2011.

[26] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Transactions on Image Processing*, vol. 19, no. 11, pp. 2861–2873, November 2010.

[27] M. Elad, *Sparse and Redundant Representations: From theory to applications in Signal and Image processing*. Springer, 2010.

[28] A. Rakotomamonjy, F. Bach, S. Canu, and Y. Grandvalet, "SimpleMKL," *Journal of Machine Learning Research*, vol. 9, pp. 2491–2521, 2008.

[29] B. Krishnapuram, L. Carin, M. A. T. Figueiredo, and A. J. Hartemink, "Sparse multinomial logistic regression: Fast algorithms and generalization bounds," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 6, pp. 957–968, June 2005.

[30] M. Frank, R. Biedert, E. Ma, I. Martinovic, and D. Song, "Touchalytics: On the applicability of touchscreen input as a behavioral biometric for continuous authentication," *IEEE Transactions on Information Forensics and Security*, vol. 8, no. 1, pp. 136–148, Jan 2013.

[31] M. Derawi, C. Nickel, P. Bours, and C. Busch, "Unobtrusive user-authentication on mobile phones using biometric gait recognition," in *International Conference on Intelligent Information Hiding and Multimedia Signal Processing*, Oct 2010, pp. 306–311.

[32] D. Ramanan, "Face detection, pose estimation, and landmark localization in the wild," in *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, ser. CVPR '12, 2012, pp. 2879–2886.

[33] V. Štruc and N. Pavešić, "The complete gabor-fisher classifier for robust face recognition," *EURASIP Advances in Signal Processing*, vol. 2010, p. 26, 2010.

[34] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," in *International Conference on Analysis and Modeling of Faces and Gestures*, 2007, pp. 168–182.

[35] S. Liao, A. K. Jain, and S. Z. Li, "Partial face recognition: Alignment-free approach," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 5, pp. 1193–1205, 2013.

[36] P. Sinha, B. Balas, Y. Ostrovsky, and R. Russell, "Face recognition by humans: Nineteen results all computer vision researchers should know about," *Proceedings of the IEEE*, vol. 94, no. 11, pp. 1948–1962, Nov 2006.