MULTITASK MULTIVARIATE COMMON SPARSE REPRESENTATIONS FOR ROBUST MULTIMODAL BIOMETRICS RECOGNITION

Heng Zhang, Vishal M. Patel and Rama Chellappa

Center for Automation Research
University of Maryland, Collage Park 20742
{hzhang98, pvishalm, rama}@umiacs.umd.edu

ABSTRACT

In this paper, we propose multitask multivairate common sparse representations for robust multimodal biometrics recognition. The proposed approach can be viewed as an extension of previous work on joint sparse representations for robust multimodal biometrics recognition. The proposed algorithm utilizes the discriminative information among different modalities simultaneously by enforcing the common sparse representation across all the modalities and achieves more robust multimodal recognition especially when all modalities are noisy and "weak". Alternating direction method of multipliers is proposed to solve the resulting optimization problem. Experiments on two biometric datasets show that our method performs better than the state-of-the-art fusion methods.

***Index Terms*—** multimodal biometrics recognition, common sparse representation

1. INTRODUCTION

Multimodal biometrics recognition [1], [2] as a special case of multi sensor classification problem has drawn a lot of attention from both academia and industry. Compared with unimodal biometric recognition system, like face recognition or fingerprint recognition, a multimodal biometrics system tries to integrate signatures such as face, fingerprints and iris and so on for recognition. One advantage of multimodal biometric system over unimodal bioemtric system is that multimodal systems are more robust to spoof attacks. Fusion of data from multi modalities for can be achieved at multiple levels including sensor level, feature space level,score level and decision or rank level. Since feature level fusion preserves the raw information, it can be more discriminative than score or rank level fusion. Our proposed methods belong to feature level fusion.

The state of the art results for several multimodal fusion problems are achieved by feature level fusion [3]. In [3], joint sparse representation-based multimodal fusion algorithms have been proposed. This method is based on multi-task multivariate Lasso [4]. It imposes the same sparse patterns for different representations corresponding to different biometric modalities. One of the advantages of using sparse representation-based fusion for classification is that it is robust to noise and occlusion [5], [6].

Different from joint sparse representation-based fusion methods, in this paper, we propose multitask and multivariate common sparse representation across different modalities in order to fuse multiple biometric traits. The key difference between the proposed methods in this paper and those in [3] is that for our methods, the sparse presentations of different modalities are enforced to be the same while for joint sparse representation-based fusion methods, the sparse representations of different modalities are different but share the same structure across different modalities. The resulting optimization problems of the proposed methods are solved efficiently using the classical Alternating Direction Method of Multipliers (ADMM) [7].

The rest of the paper is organized as follows: in Section 2, we define the proposed multitask and multivariate common sparse representations-based multimodal biometric recognition problem. Optimization procedures are described in Section 3. Experimental evaluations are described in Section 4. Finally, concluding remarks are presented in Section 5 with a brief summary and discussion.

This work was supported by cooperative agreement FA8750-13-2-0279 from DARPA.

2. PROBLEM FORMULATION

Suppose we are given a C-class classification problem with D different modalities. Assume that there are m_i 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 $n_i \times m_i$ matrix of training samples containing C sub-matrices $\mathbf{X}_j^i$'s corresponding to C different classes. Each sub-matrix $\mathbf{X}_j^i = [\mathbf{x}_{j,1}^i, \mathbf{x}_{j,2}^i, \cdots, \mathbf{x}_{j,m_j}^i] \in \mathbb{R}^{n \times m_j}$ 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_j is the feature dimension of each sample. As a result, there are in total $m = \sum_{i=1}^{C} m_i$ many samples in $\mathbf{X}_C^i$. 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 observations $\mathbf{Y}^i = [\mathbf{y}_1^i, \mathbf{y}_2^i, \cdots, \mathbf{y}_d^i] \in \mathbb{R}^{n \times d}$, the objective is to identify the class to which a test sample $\mathbf{Y}$ belongs to.

2.1. Multitask Multivariate Common Sparse Representation (MCSR)

In this section, we propose a formulation in which a common sparse representation is enforced for different modalities. By sharing the same sparse representation across all the modalities, we can leverage discriminative information from all the modalities and obtain more robust sparse representations for recognition. Similar ideas have been explored in [8] for jointly learning a sparse representation for image super-resolution.

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

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

Note that, the same representation is used for all the modalities in the above model. We propose the following optimization problem

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

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 $\|\mathbf{A}\|_{1,2} = \sum_k \|\mathbf{a}^k\|_2$ and $\mathbf{a}^k$ is the kth row vector of the matrix $\mathbf{A}$ (i.e the row sparsity of $\mathbf{A}$). Once the common sparse 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

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

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 lasso problem [9].

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:

$$\mathbf{w}_\ell = \frac{C^{\frac{\|\delta_\ell(\hat{\mathbf{\Gamma}})\|_{1,2}}{\|\hat{\mathbf{\Gamma}}\|_{1,2}}} - 1}{C - 1}. \tag{3}$$

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

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

Note that this idea has been explored in [5] and [3].

2.2. Robust Multitask Multivariate Common Sparse Representation (RMCSR)

When data is contaminated by noise and occlusion, the observation can be modeled as follows

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

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 $\mathbf{\Gamma}$ and $\mathbf{E}^i$ by solving the following optimization problem

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

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 need to 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 row sparsity of the coefficient and sparsity of occlusion term, 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 with an observation vector as

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

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

3. OPTIMIZATION

In this section, we propose an approach based on the ADMM method [7] to solve the optimization problems (1), (5). Due to the similarity of these problems, we only provide details on the optimization of the RMCSR 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.

3.1. Optimization of RMCSR

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

$$\arg\min_{\mathbf{\Gamma},\mathbf{E},\mathbf{V},\mathbf{U}} \frac{1}{2}\sum_{i=1}^{D} \|\mathbf{Y}^i - \mathbf{X}^i\mathbf{\Gamma} - \mathbf{E}^i\|_F^2 + \lambda_1\|\mathbf{V}\|_{1,2} + \lambda_2\|\mathbf{E}\|_1 \;\text{ s.t. }\; \mathbf{\Gamma} = \mathbf{V}, \mathbf{E} = \mathbf{U}. \tag{7}$$

Note that (7) is an equally constrained problem which can be solved using the Augmented Lagrangian Method (ALM) [7]. 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

$$\arg\min_{\mathbf{\Gamma},\mathbf{E},\mathbf{V},\mathbf{U}} \frac{1}{2}\sum_{i=1}^{D} \|\mathbf{Y}^i - \mathbf{X}^i\mathbf{\Gamma} - \mathbf{E}^i\|_F^2 + \lambda_1\|\mathbf{V}\|_{1,2} + \langle \mathbf{A}_\Gamma, \mathbf{\Gamma} - \mathbf{V}\rangle + \frac{\alpha_\Gamma}{2}\|\mathbf{\Gamma} - \mathbf{V}\|_F^2 + \lambda_2\|\mathbf{U}\|_1 + \langle \mathbf{A}_E, \mathbf{E} - \mathbf{U}\rangle + \frac{\alpha_E}{2}\|\mathbf{E} - \mathbf{U}\|_F^2, \tag{8}$$

where $\mathbf{A}_E$ and $\mathbf{A}_\Gamma$ are the multipliers of the two linear constrains and α_E and α_Γ are the positive parameters. 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.

In the above formula, we let $\mathbf{U} = [\mathbf{U}^1, \mathbf{U}^2, \cdots, \mathbf{U}^D]$ and $\mathbf{A}_E = [\mathbf{A}_E^1, \mathbf{A}_E^2, \cdots, \mathbf{A}_E^D]$; similar to $\mathbf{E}$, $\mathbf{U}$ and $\mathbf{A}_E$ are just compact representations and we need to solve each $\mathbf{U}^i$ and $\mathbf{A}_E^i$.

3.1.1. *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 setting 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 $\mathbf{\Gamma}_{t+1}$, by solving the following linear system

$$\left(\sum_{i=1}^{D} \mathbf{X}^{i^T}\mathbf{X}^i + \alpha_\Gamma\mathbf{I}\right)\mathbf{\Gamma}_{t+1} = \sum_{i=1}^{D} \mathbf{X}^{i^T}(\mathbf{Y}^i - \mathbf{E}^i) + \alpha_\Gamma\mathbf{V}_t - \mathbf{A}_{\Gamma,t} \tag{9}$$

where $\mathbf{I}$ is $m \times m$ identity matrix and $\mathbf{E}_t^i$ are submatrices of $\mathbf{E}_t$. When m is not very large, one can simply apply matrix inversion to obtain $\mathbf{\Gamma}_{t+1}$ from (9). For large values of m, gradient-based methods should be employed to obtain $\mathbf{\Gamma}_{t+1}$.

3.1.2. *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.

3.1.3. Update step for $\mathbf{U}$

In order to update $\mathbf{U}$, one needs to solve the following ℓ_1 minimization problem

$$\min \frac{1}{2}\|\mathbf{E}^i_{t+1} + \alpha_E^{-1}\mathbf{A}^i_{E,t} - \mathbf{U}^i\|_F^2 + \frac{\lambda_2}{\alpha_E}\|\mathbf{U}^i\|_1 \tag{10}$$

whose solution is given by [10]

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

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

3.1.4. Update step for $\mathbf{V}$

The final suboptimization for updating $\mathbf{V}$ has the following form

$$\min \frac{1}{2}\|\mathbf{\Gamma}_{t+1} + \alpha_\Gamma^{-1}\mathbf{A}_{\Gamma,t} - \mathbf{V}\|_F^2 + \frac{\lambda_1}{\alpha_\Gamma}\|\mathbf{V}\|_{1,2}. \tag{11}$$

Due to the separable structure of (11), it can be solved by minimizing with respect to each row of V separately. Follow the method used in [3] to solve V, we let $\gamma_{i,t+1}$, $\mathbf{a_\Gamma}_{i,t}$ and $\mathbf{v_{i,t+1}}$ be the ith row of matrices $\mathbf{\Gamma}_{t+1}$, $\mathbf{A}_{\Gamma,t}$ and $\mathbf{V}_{t+1}$ respectively. Then for each row, we solve the following subproblem:

$$\mathbf{v}_{i,t+1} = \min \frac{1}{2}\|\mathbf{z} - \mathbf{v}\|_2^2 + \frac{\lambda_1}{\alpha_\Gamma}\|\mathbf{v}\|_2 \tag{12}$$

where $\mathbf{z} = \gamma_{i,t+1} + \mathbf{a_\Gamma}_{i,t}\alpha_\Gamma^{-1}$. The closed form solution of (12) is:

$$\mathbf{v}_{i,t+1} = \left(\mathbf{1} - \frac{\boldsymbol{\lambda_1}}{\alpha_\mathbf{\Gamma}\|\mathbf{z}\|_\mathbf{2}}\right)_+ \mathbf{z}$$

where $(\mathbf{v})_+$ is the vector with entries receiving values $\max(v_i, 0)$

3.1.5. Update steps for $\mathbf{A}_\Gamma$ *and* $\mathbf{A}_E$

Finally, the Lagrange multipliers are updated as

$$\mathbf{A}_{\Gamma,t+1} = \mathbf{A}_{\Gamma,t} + \alpha_\Gamma(\mathbf{\Gamma}_{t+1} - \mathbf{V}_{t+1}) \tag{13}$$
$$\mathbf{A}^i_{E,t+1} = \mathbf{A}^i_{E,t} + \alpha_E(\mathbf{E}^i_{t+1} - \mathbf{U}^i_{t+1}). \tag{14}$$

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

Algorithm 1: Robust Multitask Multivariate Common Sparse Representation (RMCSR) using ADMM.

Input: Training samples $\{\mathbf{X}_i\}_{i=1}^D$, test sample $\{\mathbf{Y}_i\}_{i=1}^D$, λ_1, λ_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 $\mathbf{\Gamma}$: $\mathbf{\Gamma}_{t+1} =$

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

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

$$\mathbf{E}^i_{t+1} = (1 + \alpha_E)^{-1}(\mathbf{Y}^i - \mathbf{X}^i\mathbf{\Gamma_{t+1}} + \alpha_E\mathbf{U}^i_t - \mathbf{A}^i_{E,t})$$

3. Update $\mathbf{U}$: $\mathbf{U}_{t+1} = [\mathbf{U}^i_{t+1}, \cdots, \mathbf{U}^D_{t+1}]$, where

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

4. Update $\mathbf{V}$:

$$\mathbf{v}_{i,t+1} = \left(\mathbf{1} - \frac{\boldsymbol{\lambda_1}}{\alpha_\mathbf{\Gamma}\|\mathbf{z}\|_\mathbf{2}}\right)_+ \mathbf{z}$$

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 $\mathbf{A}_E$: $\mathbf{A}_{E,t+1} = [\mathbf{A}^1_{E,t+1}, \mathbf{A}^2_{E,t+1}, \cdots, \mathbf{A}^D_{E,t+1}]$,

$$\mathbf{A}^i_{E,t+1} = \mathbf{A}^i_{E,t} + \alpha_E(\mathbf{E}^i_{t+1} - \mathbf{U}^i_{t+1})$$

Output: $\hat{\mathbf{E}} = \mathbf{E}_{t+1}$ and $\hat{\mathbf{\Gamma}} = \mathbf{\Gamma}_{t+1}$.

4. EXPERIMENT

In this section, we evaluate the proposed algorithms and reported results on two publicly available biometric datasets, namely the WVU multimodal dataset [11] and the AR face dataset [12].

4.1. WVU dataset

The WVU multimodal dataset is a comprehensive collection of different biometric modalities such as fingerprint, iris, palmprint, hand geometry, and voice from subjects of different age, gender, and ethnicity. It is a challenging dataset, as many of these samples are corrupted with blur, occlusion, and sensor noise. Figure1 shows some examples of fingerprint and iris from WVU dataset.

4.1.1. Feature Extraction and Experiment Setup

Following the exact same signatures extracted and experiment setups in [3], we use two iris(right and left iris) and 4 fingerprint modalities on a subset of 219 subjects having samples in both modalities. After feature extraction, the dimension of the feature vector for iris samples is 6000 and the dimension of the feature vector for fingerprint samples is 3600.

The data samples(one sample includes 6 observations corresponding to 6 modalities) were randomly divided into four training samples per class and the remaining samples were used for testing. As a result, 876 samples were used for training and 519 samples were used for testing. The recognition result was averaged over five runs and we reported the mean and standard deviation of rank one recognition accuracy. We compare our proposed methods with several sate-of-the-art feature level multimodal fusion methods including MKL [13], Sparsity-based Multimodal Biometrics Recognition (SMBR-WE and SMBR-E) [3] and score-level fusion methods based on Sparse ogistic Regression (SLR) [14] and SVM [15]. Short names for these different methods are used as suggested in [3].

4.1.2. Experiments result

The rank one recognition results comparing the proposed methods with other multi-modal fusion methods are shown in Table 1 for each single modality and in Table 2 for fusion of modalities.

From the results shown in Table 1 and Table 2, we make the following observations: (1) All the considered algorithms achieve better recognition accuracy when fusing multiple modalities than a single modality. (2) Compared to other methods, the proposed fusion methods, in particular, the RMCSR performs the best. The reason why the proposed methods are better than joint sparsity-based methods is when some(all) modalities are noisy or of low quality,

Methods	Finger 1	Finger 2	Finger 3	Finger 4	Iris 1	Iris 2
MCSR	$\mathbf{70.3 \pm 1.0}$	$\mathbf{90.1 \pm 0.8}$	$\mathbf{69.2 \pm 2.3}$	$\mathbf{89.5 \pm 1.4}$	$\mathbf{62.6 \pm 1.8}$	$\mathbf{64.6 \pm 1.0}$
RMCSR	$\mathbf{69.8 \pm 1.4}$	$\mathbf{89.4 \pm 1.0}$	$\mathbf{69.2 \pm 2.3}$	$\mathbf{89.2 \pm 1.1}$	$\mathbf{70.5 \pm 1.1}$	$\mathbf{71.7 \pm 0.5}$
SMBR-WE	68.1 ± 1.1	88.4 ± 1.2	69.2 ± 1.5	87.5 ± 1.5	60.0 ± 1.5	62.1 ± 0.4
SMBR-E	67.1 ± 1.0	87.9 ± 0.8	67.4 ± 1.9	86.9 ± 1.5	62.5 ± 1.2	64.3 ± 1.0
SLR	67.4 ± 1.9	87.9 ± 1.3	66.0 ± 2.2	87.5 ± 1.3	57.1 ± 3.0	57.9 ± 2.7
SVM	41.1 ± 5.0	75.5 ± 2.2	49.2 ± 1.6	67.0 ± 8.3	44.3 ± 1.2	45.0 ± 2.9

Table 1: rank one recognition accuracy (in %) on WVU dataset for individual modalities.

Methods	**MCSR**	**RMCSR**	SMBR-WE	SMBR-E	SLR-Sum	SVM-Sum	MKLFusion
4 Fingerprints	95.6 ± 0.4	96.1 ± 0.6	97.9 ± 0.4	$\mathbf{97.6 \pm 0.6}$	96.3 ± 0.8	90.0 ± 2.2	86.2 ± 1.2
2 Irises	$\mathbf{78.3 \pm 0.2}$	$\mathbf{85.3 \pm 1.9}$	76.5 ± 1.6	78.2 ± 1.2	72.7 ± 4.0	62.8 ± 2.6	76.8 ± 2.5
All modalities	98.2 ± 0.4	$\mathbf{99.4 \pm 0.5}$	98.7 ± 0.2	98.6 ± 0.5	97.6 ± 0.4	94.9 ± 1.5	89.8 ± 0.9

Table 2: rank one recognition accuracy (in %) on WVU dataset for fusion of modalities.

Fig. 1: examples of fingerprint and iris from WVU dataset.

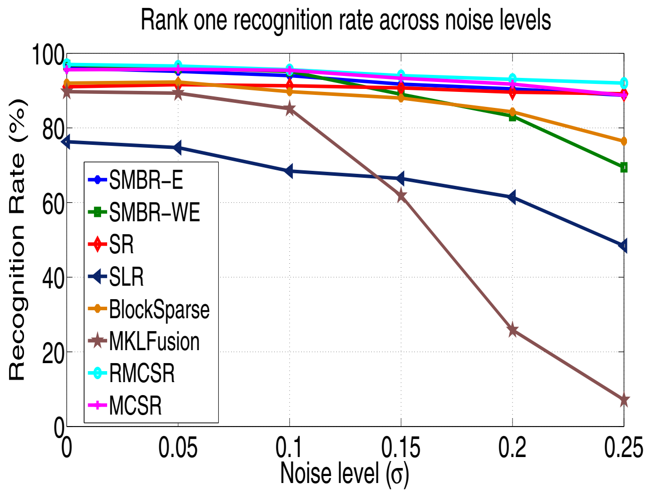

Fig. 2: Rank 1 recognition performance on AR dataset under different noise level.

the learned common sparse representation from different modalities can be more robust and discriminative by trying to leverage all the information embedded in all the modalities.

4.2. AR dataset

The AR face dataset [12] was collected over 2 sessions and in each session, face images exhibits different illumination, expression and occlusions(sun-glasses and scarf). Since our goal is to mainly compare the proposed algorithms with joint sparsity-based methods(SMBR-WE and SMBR-E), we followed the exact same protocol used in [3] in terms of selecting training and testing images and subsequent feature extraction from the face images. Specifically, we used a set of 100 subjects. For each subject, 7 images with different illumination and expression in session 1 were used for training and 7 images with same kinds of variations from the session 2 were used for testing. The rectangular masks were applied to extract left eye, right eye, nose and mouth from every face image. Face recognition was achieved by fusing 4 extracted face parts and the holistic face image with their normalized pixel values as feature vectors corresponding to these 5 modalities.

In addition, to study the noise effect on the fusion algorithms, test images were corrupted with Gaussian noise of different variance σ^2. The rank one recognition results comparing our proposed methods with the other multi-modal fusion methods under different noise level are shown in Figure 2. The proposed MCSR algorithm achieves very robust results and the recognition rates are 95.57%, 95.71%, 95.43%, 93.29%, 91.71%, and 88.71% respectively, corresponding to each noise level. The proposed RMCSR algorithm perform the best and the recognition rates are 97%, 96.57%, 95.57%, 94%, 93%, and 92% respectively, corresponding to each noise level.

5. CONCLUSION

In this paper, we proposed 2 multi-modal feature fusion methods based on common sparse representations, namely MCSR and RMCSR. Efficient optimization procedures are proposed for solving the proposed problems using the ADMM method. Experiments on WVU dataset using fingerprint and iris modalities and fusion of face parts along with holistic face image for recognition on AR dataset show that the proposed methods can perform better than currently state of the art fusion methods based on joint sparse representations.

6. REFERENCES

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

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

[3] 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.

[4] 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.

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

[6] J.K. Pillai, V. M. Patel, R. Chellappa, and N.K. 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.

[7] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan 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.

[8] 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.

[9] 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.

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

[11] S. Crihalmeanu, A. Ross, S. Schuckers, and L. Hornak, ," Tech. Rep., WVU, Lane Department of Computer Science and Electrical Engineering.

[12] Aleix Martínez and Robert Benavente, "The ar face database," Jun 1998, Cites in Scholar Google: http://scholar.google.com/scholar?hl=en&lr=&client=firefox-a&cites=1504264687621469812.

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

[14] Balaji Krishnapuram, Lawrence Carin, Mario A. T. Figueiredo, and Alexander 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.

[15] Christopher J. C. Burges, "A tutorial on support vector machines for pattern recognition," *Data Min. Knowl. Discov.*, vol. 2, no. 2, June 1998.