HIERARCHICAL SPARSE AND COLLABORATIVE LOW-RANK REPRESENTATION FOR EMOTION RECOGNITION

Xiang Xiang, Minh Dao, Gregory D. Hager, Trac D. Tran

Johns Hopkins University, 3400 N. Charles Street, Baltimore, MD 21218, USA
{xxiang, minh.dao, ghager1, trac}@jhu.edu

ABSTRACT

In this paper, we design a Collaborative-Hierarchical Sparse and Low-Rank (C-HiSLR) model that is natural for recognizing human emotion in visual data. Previous attempts require explicit expression components, which are often unavailable and difficult to recover. Instead, our model exploits the low-rank property to subtract neutral faces from expressive facial frames as well as performs sparse representation on the expression components with group sparsity enforced. For the CK+ dataset, C-HiSLR on raw expressive faces performs as competitive as the Sparse Representation based Classification (SRC) applied on manually prepared emotions. Our C-HiSLR performs even better than SRC in terms of true positive rate.

Index Terms— Low-rank, group sparsity, multichannel

1. INTRODUCTION

In this paper, the problem of interest is to recognize the emotion given a video of a human face and emotion category [1]. As shown in Fig.1, an expressive face can be separated into a dominant neutral face and a sparse **expression component**, which we term **emotion** and is usually encoded in a sparse noise term $\mathbf{e}$. We investigate *if we can sparsely represent the emotion over a dictionary of emotions* [2] rather than expressive faces [3], which may confuse a similar expression with a similar identity [2]. Firstly, *how to get rid of the neutral face?* Surely we can prepare an expression with a neutral face explicitly provided as suggested in [2]. Differently, we treat an emotion as an action and assume neutral faces stay the same. If we stack vectors of neutral faces as a matrix, it should be low-rank (ideally with rank 1). Similarly, over time sparse vectors of emotions form a sparse matrix. Secondly, *how to recover the low-rank and sparse components?* In [4], the (low-rank) Principal Component Pursuit (PCP) [5] is performed explicitly. While theoretically the recovery is exact under conditions [5], it is of approximate nature in practice. Finally, since we only care about the sparse component, *can we avoid such an approximate explicit PCP step?* This drives us to exploit Sparse representation and Low-Rank property jointly in one model named SLR (Sec. 3.1).

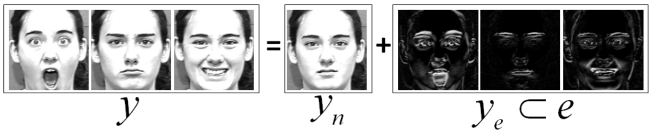

Fig. 1. The separability of the neutral face $\mathbf{y}_n$ and emotion $\mathbf{y}_e$. Given a different expressive face $\mathbf{y}$ (*e.g., surprise, sadness, happiness*), the difference is $\mathbf{y}_e$, which is encoded in error $\mathbf{e}$.

Different from image-based methods [2, 4], we treat an emotion video as a multichannel signal. If we just use a single channel such as one frame to represent an emotion, much information is lost since all frames collaboratively represent an emotion. Therefore, we prefer using all or most of them. *Should we treat them separately or simultaneously?* The former just needs to recover the sparse coefficient vector for each frame. The latter gives a spatial-temporal representation, while it requires the recovery of a sparse coefficient matrix, which should often exhibit a specific structure. *Should we enforce a class-wise sparsity separately or enforce a group sparsity collaboratively?* [4] models the class-wise sparsity separately for the recognition of a neutral face's identity and an expression image's emotion once they have been separated. Alternatively, we can exploit the low-rankness as well as structured sparsity by inter-channel observation. Since class decisions may be inconsistent, we prefer a collaborative model [6] with group sparsity enforced [7]. This motivates us to introduce the group sparsity as a root-level sparsity to the SLR model embedded with a leaf-level atom-wise sparsity. The reason of keeping both levels is that signals over frames share class-wise yet not necessarily attom-wise sparsity patterns [8]. Therefore, we term this model Collaborative-Hierarchical Sparse and Low-Rank (C-HiSLR) model.

In the remainder of this paper, we review sparse and low-rank representation literature in Sec. 2, elaborate our model in Sec. 3, discuss the optimization in Sec. 4, empirically validate the model in Sec. 5, and draw a conclusion in Sec. 6.

2. RELATED WORKS

When observing a random signal $\mathbf{y}$ for recognition, we hope to send the classifier a *discriminative compact* representation $\mathbf{x}$, which satisfies $\mathbf{A}\mathbf{x} = \mathbf{y}$ and is yet computed by pursuing the best *reconstruction*. When $\mathbf{A}$ is under-complete, a closed-form approximate solution can be obtained by Least-Squares:

978-1-4673-6997-8/15/$31.00 ©2015 IEEE

$$\mathbf{x}^* = \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 \approx (\mathbf{A}^T\mathbf{A})^{-1}\mathbf{A}^T\mathbf{y}.$$

When $\mathbf{A}$ is over-complete, we add a Tikhonov regularizer [9]: $\mathbf{x}^* = \arg\min_{\mathbf{x}} \|\mathbf{y} - \mathbf{A}\mathbf{x}\|_2^2 + \lambda_r\|\mathbf{x}\|_2^2 = \arg\min_{\mathbf{x}} \|\widetilde{\mathbf{y}} - \widetilde{\mathbf{A}}\mathbf{x}\|_2^2$ $\approx (\mathbf{A}^T\mathbf{A}+\lambda_r\mathbf{I})^{-1}\mathbf{A}^T\mathbf{y}$ where $\widetilde{\mathbf{y}} = [\mathbf{y}, \mathbf{0}]^T$; $\widetilde{\mathbf{A}} = [\mathbf{A}, \sqrt{\lambda_r}\mathbf{I}]^T$ is always under-complete. But $\mathbf{x}^*$ is not necessarily compact yet generally dense. Alternatively, we can seek a sparse usage of $\mathbf{A}$. Sparse Representation based Classification [10] (SRC) expresses a test sample $\mathbf{y}$ as a linear combination $\mathbf{y} = \mathbf{D}\mathbf{x}$ of *training samples* stacked columnwise in a dictionary $\mathbf{D}$. *Since non-zero coefficients should all drop to the ground-truth class, ideally not only* $\mathbf{x}$ *is sparse but also the class-level sparsity is* 1. In fact, non-zero coefficients also drop to other classes due to noises and correlations among classes. By adding a sparse error term $\mathbf{e}$, SRC simply employs an atom-wise sparsity:

$$[\mathbf{x}^*, \mathbf{e}^*]^T = \arg\min_{\widetilde{\mathbf{x}}} sparsity(\widetilde{\mathbf{x}})$$

$$s.t. \quad \mathbf{y} = \mathbf{D}\mathbf{x} + \mathbf{e} = [\,\mathbf{D} \mid \mathbf{I}\,] \times \begin{bmatrix} \mathbf{x} \\ \mathbf{e} \end{bmatrix} = \widetilde{\mathbf{D}}\widetilde{\mathbf{x}},$$

where $\widetilde{\mathbf{D}}$ is over-complete and needs to be sparsely used. SRC evaluates which class leads to the minimum reconstruction error, which can be seen as a max-margin classifier [11]. Using a fixed $\mathbf{D}$ without dictionary learning [12] or sparse coding, SRC still performs robustly well for denoising and coding tasks such as well-aligned noisy face identifications. But there is a lack of theoretical justification why a sparser representation is more discriminative. [13] incorporates the Fisher's discrimination power into the objective. [14] follows the regularized Least-Squares [9] and argues SRC's success is due to the linear combination as long as the ground-truth class dominates coefficient magnitudes. SRC's authors clarify this confusion using more tests on robustness to noises [15].

In practice, we care more about how to recover $\mathbf{x}$ [16]. Enforcing sparsity is feasible since $\mathbf{x}$ can be exactly recovered from $\mathbf{y} = \mathbf{D}\mathbf{x} + \mathbf{e}$ under conditions for $\mathbf{D}$ [17]. However, finding the sparsest solution is NP-hard and difficult to solve exactly [18]. But now, it is well-known that the ℓ_1 norm is a good convex relaxation of sparsity -- minimizing the ℓ_1 norm induces the sparsest solution under mild conditions [19]. Exact recovery is also guaranteed by ℓ_1-minimization under suitable conditions [20]. Typically, an iterative greedy algorithm is the Orthogonal Matching Pursuit (OMP) [16].

For multichannel $\mathbf{Y}$ with dependant coefficients across channels [21], $\mathbf{Y} = \mathbf{D}\mathbf{X}$ where $\mathbf{X}$ is low-rank. In a *unsupervised* manner, Sparse Subspace Clustering [22] of $\mathbf{Y}$ solves $\mathbf{Y} = \mathbf{Y}\mathbf{X}$ where $\mathbf{X}$ is sparse and Principal Component Analysis is $\min_{\mathbf{A}} \|\mathbf{Y} - \mathbf{A}\mathbf{Y}\|^2$ where $\mathbf{A}$ is a projection matrix.

3. REPRESENTATION MODELS

In this section, we explain how to model $\mathbf{X}$ using $\mathbf{Y}$ and training data $\mathbf{D}$, which contains $K \in \mathbb{Z}^+$ types of **emotions**. We would like to classify a test video as one of the K classes.

3.1. SLR: joint Sparse representation and Low-Rankness

First of all, we need an explicit representation $\mathbf{Y}$ of an expressive face. The matrix $\mathbf{Y} \in \mathbb{R}^{d\times\tau}$ can be an arrangement of d-dimensional feature vectors $\mathbf{y} \in \mathbb{R}^d$ $(i = 1, 2, ..., \tau)$ such as Gabor features [23] or concatenated image raw intensities [10] of the τ frames: $\mathbf{Y} = \left[\mathbf{Y}_1|...|\mathbf{Y}_\tau\right]_{d\times\tau}$. We emphasize our model's power by simply using the raw pixel intensities.

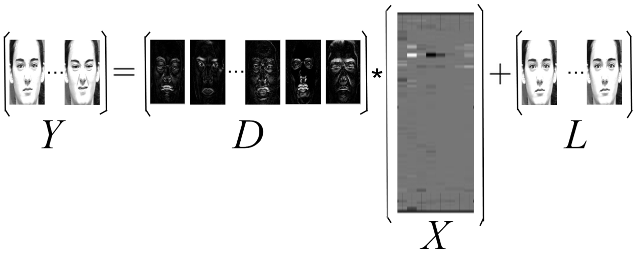

Fig. 2. Pictorial illustration of the constraint in SLR and C-HiSLR for recognizing *disgust*. $\mathbf{D}$ is prepared and fixed.

Now, we seek an implicit latent representation $\mathbf{X} \in \mathbb{R}^{n\times\tau}$ of an input test face's emotion $\mathbf{Y}_e \in \mathbb{R}^{d\times\tau}$ as a sparse linear combination of prepared fixed training emotions $\mathbf{D} \in \mathbb{R}^{d\times n}$:

$$\mathbf{Y}_e = \mathbf{D}\mathbf{X}.$$

Since an expressive face $\mathbf{y} = \mathbf{y}_e + \mathbf{y}_n$ is a superposition of an emotion $\mathbf{y}_e \in \mathbb{R}^d$ and a neutral face $\mathbf{y}_n \in \mathbb{R}^d$, we have

$$\mathbf{Y} = \mathbf{Y}_e + \mathbf{L},$$

where $\mathbf{L} \in \mathbb{R}^{d\times\tau}$ is ideally τ-times repetition of the column vector of a neutral face $\mathbf{y}_n \in \mathbb{R}^d$. Presumably $\mathbf{L} = \left[\mathbf{y}_n|...|\mathbf{y}_n\right]_{d\times\tau}$. As shown in Fig. 2, $\mathbf{X}$ subjects to

$$\mathbf{Y} = \mathbf{D}\mathbf{X} + \mathbf{L},$$

where the dictionary matrix $\mathbf{D}_{d\times n}$ is an arrangement of all sub-matrices $\mathbf{D}_{[j]}$, $j = 1, ..., \lfloor\frac{n}{\tau}\rfloor$. *Only for training*, we have $\lfloor\frac{n}{\tau}\rfloor$ training emotions with neutral faces subtracted. The above constraint of $\mathbf{X}$ characterizes an affine transformation from the latent representation $\mathbf{X}$ to the observation $\mathbf{Y}$. If we write $\mathbf{X}$ and $\mathbf{Y}$ in homogeneous forms [24], then we have

$$\begin{bmatrix} \mathbf{Y}_{d\times\tau} \\ \mathbf{1}_{1\times\tau} \end{bmatrix} = \begin{bmatrix} \mathbf{D}_{d\times n} & \left(\mathbf{y}_n\right)_{d\times 1} \\ \mathbf{0}_{1\times n} & 1 \end{bmatrix} \times \begin{bmatrix} \mathbf{X}_{n\times\tau} \\ \mathbf{1}_{1\times\tau} \end{bmatrix}.$$

In the ideal case with $rank(\mathbf{L}) = 1$, if the neutral face $\mathbf{y}_n$ is pre-obtained [2, 4], it is trival to solve for $\mathbf{X}$. Normally, $\mathbf{y}_n$ is unknown and $\mathbf{L}$ is not with rank 1 due to noises. As $\mathbf{X}$ is supposed to be sparse and $rank(\mathbf{L})$ is expected to be as small as possible (maybe even 1), intuitively our objective is to

$$\min_{\mathbf{X},\mathbf{L}} sparsity(\mathbf{X}) + \lambda_L \cdot rank(\mathbf{L}),$$

where $rank(\mathbf{L})$ can be seen as the sparsity of the vector formed by the singular values of $\mathbf{L}$. Here λ_L is a non-negative weighting parameter we need to tune [25]. When $\lambda_L = 0$, the optimization problem reduces to that in SRC. With both terms relaxed to be ℓ_1 norm, we alternatively solve

$$\min_{\mathbf{X},\mathbf{L}} \|\mathbf{X}\|_1 + \lambda_L\|\mathbf{L}\|_*,$$

where $\|\cdot\|_1$ is the entry-wise ℓ_1 matrix norm, whereas $\|\cdot\|_*$ is the Schatten ℓ_1 matrix norm (nuclear norm, trace norm) which can be seen as applying ℓ_1 norm to the vector of singular values. Now, the proposed joint SLR model is expressed as

$$\min_{\mathbf{X},\mathbf{L}} \|\mathbf{X}\|_1 + \lambda_L\|\mathbf{L}\|_* \qquad s.t. \qquad \mathbf{Y} = \mathbf{D}\mathbf{X} + \mathbf{L} \qquad (1)$$

We solve (1) for matrices $\mathbf{X}$ and $\mathbf{L}$ by the Alternating Direction Method of Multipliers (ADMM) [26] (see Sec. 4).

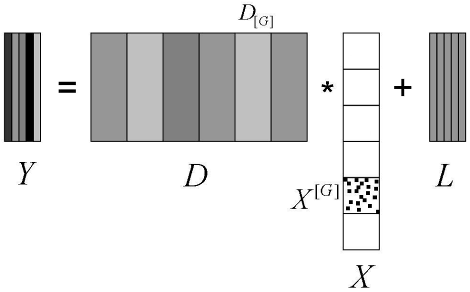

Fig. 3. Pictorial illustration of the constraint in the C-HiSLR.

3.2. C-HiSLR: a Collaborative-Hierarchical SLR model

If there is no low-rank term $\mathbf{L}$, (1) becomes a problem of multi-channel Lasso (Least Absolute Shrinkage and Selection Operator). For a single-channel signal, Group Lasso [27] has explored the group structure for Lasso yet does not enforce sparsity within a group, while Sparse Group Lasso [28] yields an atom-wise sparsity as well as a group sparsity. Then, [8] extends Sparse Group Lasso to multichannel, resulting in a Collaborative-Hierarchical Lasso (C-HiLasso) model. For our problem, we do need $\mathbf{L}$, which induces a Collaborative-Hierarchical Sparse and Low-Rank (C-HiSLR) model:

$$\min_{\mathbf{X},\mathbf{L}} \|\mathbf{X}\|_1 + \lambda_L \|\mathbf{L}\|_* + \lambda_g \sum_{G \in \mathcal{G}} \|\mathbf{X}^{[G]}\|_F \qquad s.t. \quad \mathbf{Y} = \mathbf{D}\mathbf{X} + \mathbf{L} \tag{2}$$

where $\mathbf{X}^{[G]}$ is the sub-matrix formed by all the **rows** indexed by the elements in group $G \subseteq \{1,...,\text{n}\}$. As shown in Fig. 3, given a group G of indices, the sub-dictionary of **columns** indexed by G is denoted as $\mathbf{D}_{[G]}$. $\mathcal{G} = \{G_1, ..., G_K\}$ is a non-overlapping partition of $\{1, ..., n\}$. Here $\|\cdot\|_F$ denotes the Frobenius norm, which is the entry-wise ℓ_2 norm as well as the Schatten ℓ_2 matrix norm and can be seen as a group's magnitude. λ_g is a non-negative weighting parameter for the group regularizer, which is generalized from an ℓ_1 regularizer (consider $\mathcal{G} = \{\{1\}, \{2\}, ..., \{n\}\}$ for singleton groups) [8]. When $\lambda_g = 0$, C-HiSLR degenerates into SLR. When $\lambda_L = 0$, we get back to collaborative Sparse Group Lasso.

3.3. Classification

Following SRC, for each class $c \in \{1, 2, ..., K\}$, let $\mathbf{D}_{[G_c]}$ denote the sub-matrix of $\mathbf{D}$ which consists of all the columns of $\mathbf{D}$ that correspond to emotion class c and similarly for $\mathbf{X}^{[G_c]}$. We classify $\mathbf{Y}$ by assigning it to the class with minimal residual as $c^* = \arg\min_c r_c(\mathbf{Y}) := ||\mathbf{Y} - \mathbf{D}_{[G_c]}\mathbf{X}^{[G_c]} - \mathbf{L}||_F$.

4. OPTIMIZATION

Both SLR and C-HiSLR models can be seen as solving

$$\min_{\mathbf{X},\mathbf{L}} f(\mathbf{X}) + \lambda_L \|\mathbf{L}\|_* \qquad s.t. \quad \mathbf{Y} = \mathbf{D}\mathbf{X} + \mathbf{L} \tag{3}$$

To follow a standard iterative ADMM procedure [26], we write down the augmented Lagrangian function for (3) as

$$\begin{aligned}\mathcal{L}(\mathbf{X}, \mathbf{L}, \mathbf{\Lambda}) &= f(\mathbf{X}) + \lambda_L ||\mathbf{L}||_* \\ &+ \langle \mathbf{\Lambda}, \mathbf{Y} - \mathbf{A}\mathbf{X} - \mathbf{L} \rangle + \frac{\beta}{2} ||\mathbf{Y} - \mathbf{A}\mathbf{X} - \mathbf{L}||_F^2,\end{aligned} \tag{4}$$

where $\mathbf{\Lambda}$ is the matrix of multipliers, $\langle \cdot, \cdot \rangle$ is inner product, and β is a positive weighting parameter for the penalty (augmentation). A single update at the k-th iteration includes

$$\mathbf{L}_{k+1} = \arg\min_{\mathbf{L}} \lambda_L ||\mathbf{L}||_* + \frac{\beta}{2} ||\mathbf{Y} - \mathbf{A}\mathbf{X}_k - \mathbf{L} + \frac{1}{\beta}\mathbf{\Lambda}_k||_F^2 \tag{5}$$

$$\mathbf{X}_{k+1} = \arg\min_{\mathbf{X}} f(\mathbf{X}) + \frac{\beta}{2} ||\mathbf{Y} - \mathbf{A}\mathbf{X} - \mathbf{L}_{k+1} + \frac{1}{\beta}\mathbf{\Lambda}_k||_F^2 \tag{6}$$

$$\mathbf{\Lambda}_{k+1} = \mathbf{\Lambda}_k + \beta(\mathbf{Y} - \mathbf{A}\mathbf{X}_{k+1} - \mathbf{L}_{k+1}). \tag{7}$$

The sub-step of solving (5) has a closed-form solution:

$$\mathbf{L}_{k+1} = \mathcal{D}_{\frac{\lambda_L}{\beta}}(\mathbf{Y} - \mathbf{A}\mathbf{X}_k + \frac{1}{\beta}\mathbf{\Lambda}_k), \tag{8}$$

where $\mathcal{D}$ is the shrinkage thresholding operator. In SLR where $f(\mathbf{X}) = \|\mathbf{X}\|_1$, (6) is a Lasso problem, which we solve by using an existing fast solver [29]. When $f(\mathbf{X})$ follows (2) of C-HiSLR, computing $\mathbf{X}_{k+1}$ needs an approximation based on the Taylor expansion at $\mathbf{X}_k$ [30, 8]. We refer the reader to [8] for the convergence analysis and recovery guarantee.

5. EXPERIMENTAL RESULTS

All experiments are conducted on the CK+ dataset [31] which consists of 321 emotion sequences with labels (angry, contempt[1], disgust, fear, happiness, sadness, surprise) [2] and is randomly divided into a training set (10 sequences per category) and a testing set (5 sequences per category). *For SRC, we assume that the information of neutral face is provided.* We subtract the first frame (a neutral face) from the last frame per sequence for both training and testing. Thus, each emotion is represented as an image. However, *for SLR and C-HiSLR, we assume no prior knowledge of the neutral face.* We form a dictionary by subtracting the first frame from the last τ_{trn} frames per sequence and form a testing unit using the last $(\tau_{tst} - 1)$ frames together with the first frame, which is **not** explicitly known as a neutral face. Thus, each emotion is represented as a video. Here, we set $\tau_{trn} = 4$ or 8, $\tau_{tst} = 8$, $\lambda_L = 10$ and $\lambda_G = 4.5$. Fig. 4 visualizes the recovery results given by C-HiSLR. Facial images are cropped using the Viola-Jones detector [32] and resized to 64×64. As shown in Fig. 5, imperfect alignment may affect the performance.

Firstly, SRC achieves a total recognition rate of **0.80**, against **0.80** for eigenface with nearest subspace classifier and **0.72** for eigenface with nearest neighbor classifier. This verifies that emotion is sparsely representable by training data and SRC can be an alternative to subspace based methods. Secondly, Table 1-3 present the confusion matrix ($\tau_{trn} = 4$) and Table 4 summarizes the true positive rate (*i.e.*, sensitivity). We have anticipated that SLR (**0.70**) performs worse than SRC (**0.80**) since SRC is equipped with neutral faces. However, C-HiSLR's result (**0.80**) is comparable with SRC's. C-HiSLR performs even better in terms of sensitivity, which verifies that the group sparsity indeed boosts the performance.

[1] Contempt is discarded in [2, 4] due to its confusion with other classes.

[2] Please visit http://www.cs.jhu.edu/ xxiang/slr/ for the cropped face data and program of C-HiSLR, SLR, SRC & Eigenface to run the experiments.

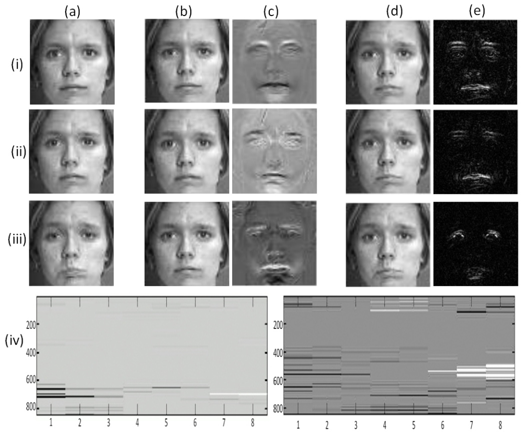

Fig. 4. Effect of group sparsity. $\tau_{trn} = 8$. (a) is the test input $\mathbf{Y}$. (b)(c) are recovered $\mathbf{L}$ and $\mathbf{AX}$, given by **C-HiSLR** which correctly classifies (a) as *contempt*. (d)(e) are recovery results given by **SLR** which mis-classifies (a) as sadness. (i),(ii),(iii) denote results of frame #1, #4, #8 respectively, whereas (iv) displays the recovered $\mathbf{X}$ (left for C-HiSLR and right for SLR). $\mathbf{X}$ given by C-HiSLR is group-sparse as we expected.

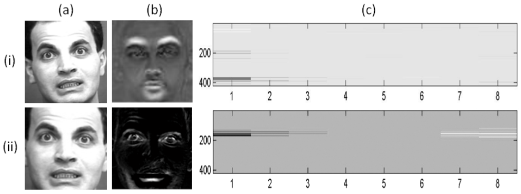

Fig. 5. Effect of alignment. Shown for **C-HiSLR** with $\tau_{trn} = 4$. (a) is the test input (*fear*). (b) and (c) are recovered $\mathbf{AX}$ and $\mathbf{X}$, respectively. (i) is under **imperfect** alignment while (ii) is under **perfect** alignment. $\mathbf{X}$ in (i) is *not* group-sparse.

6. CONCLUSION

We design the C-HiSLR representation model for emotion recognition, unlike [2] requiring neutral faces as inputs and [4] generating labels of identity and emotion as mutual by-products with extra efforts. Our contribution is two-fold. First, we do not recover emotion explicitly. Instead, we treat frames simultaneously and implicitly subtract the low-rank neutral face. Second, we preserve the label consistency by enforcing atom-wise as well as group sparsity. For the CK+ dataset, C-HiSLR's performance on raw data is comparable with SRC given neutral faces, which verifies that emotion is automatically separable from expressive faces as well as sparsely representable. Future works will include handling misalignment [33] and incorporating dictionary learning [12].

ACKNOWLEDGMENTS

This work is supported by US National Science Foundation under Grants CCF-1117545 and CCF-1422995, Army Research Office under Grant 60219-MA, and Office of Naval Research under Grant N00014-12-1-0765. The first author is grateful for the fellowship from China Scholarship Council.

	An	Co	Di	Fe	Ha	Sa	Su
An	**0.77**	0.01	0.09	0.02	0	0.07	0.04
Co	0.08	**0.84**	0	0	0.03	0.04	0
Di	0.05	0	**0.93**	0.01	0.01	0.01	0
Fe	0.09	0.01	0.03	**0.53**	0.12	0.07	0.15
Ha	0.01	0.02	0.01	0.02	**0.93**	0	0.03
Sa	0.19	0.02	0.02	0.05	0	**0.65**	0.07
Su	0	0.02	0	0.02	0	0.02	**0.95**

Table 1. Confusion matrix for **C-HiSLR** on CK+ dataset [31] without explicitly knowing neutral faces. Columns are predictions and rows are ground truths. We randomly choose 15 sequences for training and 10 sequences for testing per class. We let the optimizer run for 600 iterations. Results are averaged over 20 runs and rounded to the nearest. The total recognition rate is **0.80** with a standard deviation of 0.05.

	An	Co	Di	Fe	Ha	Sa	Su
An	**0.51**	0	0.10	0.02	0	0.31	0.06
Co	0.03	**0.63**	0.03	0	0.04	0.26	0.01
Di	0.04	0	**0.74**	0.02	0.01	0.15	0.04
Fe	0.08	0	0.01	**0.51**	0.03	0.19	0.18
Ha	0	0.01	0	0.03	**0.85**	0.08	0.03
Sa	0.09	0	0.04	0.04	0	**0.70**	0.13
Su	0	0.01	0	0.02	0.01	0.02	**0.94**

Table 2. Confusion matrix for **SLR** on CK+ dataset without explicit neutral faces. We randomly choose 15 sequences for training and 10 for testing per class. We let the optimizer run for 100 iterations and Lasso run for 100 iterations. Results are averaged over 20 runs and rounded to the nearest. The total recognition rate is **0.70** with a standard deviation of 0.14.

	An	Co	Di	Fe	Ha	Sa	Su
An	**0.71**	0.01	0.07	0.02	0.01	0.03	0.16
Co	0.07	**0.60**	0.02	0	0.16	0.03	0.12
Di	0.04	0	**0.93**	0.02	0.01	0	0
Fe	0.16	0	0.09	**0.25**	0.25	0	0.26
Ha	0.01	0	0	0.01	**0.96**	0	0.02
Sa	0.22	0	0.13	0.01	0.04	**0.24**	0.35
Su	0	0.01	0	0	0.01	0	**0.98**

Table 3. Confusion matrix for **SRC** [10] with neutral faces explicitly provided, in a similar setting with [2]. We choose half of the dataset for training and the other half for testing per class. The optimizer is OMP and the sparsity level is set to 35%. Results are averaged over 20 runs and rounded to the nearest. The total recognition rate is **0.80** with a standard deviation of 0.05. The rate for *fear* and *sad* are especially low.

Model	An	Co	Di	Fe	Ha	Sa	Su
SRC	0.71	*0.60*	**0.93**	*0.25*	**0.96**	*0.24*	**0.98**
SLR	*0.51*	*0.63*	*0.74*	**0.51**	*0.85*	**0.70**	***0.94***
C-HiSLR	**0.77**	**0.84**	**0.93**	**0.53**	**0.93**	**0.65**	***0.95***

Table 4. Comparison of sensitivity. The **bold** and *italics* denote the highest and lowest respectively. Difference within 0.05 is treated as comparable. C-HiSLR performs the best.

7. REFERENCES

[1] Zhihong Zeng, Maja Pantic, Glenn I. Roisman, and Thomas S. Huang, "A survey of affect recognition methods: Audio, visual, and spontaneous expressions," *IEEE T-PAMI*, vol. 31, no. 1, pp. 39–58, 2009. 1

[2] Stefanos Zafeiriou and Maria Petrou, "Sparse representations for facial expressions recognition via l1 optimization," in *IEEE CVPR Workshop*, 2010. 1, 2, 3, 4

[3] Raymond Ptucha, Grigorios Tsagkatakis, and Andreas Savakis, "Manifold based sparse representation for robust expression recognition without neutral subtraction," in *IEEE ICCV Workshops 2011*, 2013. 1

[4] Sima Taheri, Visha M. Patel, and Rama Chellappa, "Component-based recognition of faces and facial expressions," *IEEE Trans. on Affective Computing*, vol. 4, no. 4, pp. 360–371, 2013. 1, 2, 3, 4

[5] Emmanuel J. Candes, Xiaodong Li, Yi Ma, and John Wright, "Robust principal component analysis?," *Journal of the ACM*, vol. 58, no. 3, pp. 11:1–37. 1

[6] Yonina C. Eldar and Holger Rauhut, "Average case analysis of multichannel sparse recovery using convex relaxation," *IEEE Trans. Inf. Theory*, vol. 56, no. 1, pp. 505–519, 2010. 1

[7] Junzhou Huang and Tong Zhang, "The benefit of group sparsity," *The Annals of Statistics*, vol. 38, no. 4, pp. 1978–2004, 2010. 1

[8] Pablo Sprechmann, Ignacio Ramłrez, Guillermo Sapiro, and Yonina Eldar, "C-HiLasso: A collaborative hierarchical sparse modeling framework," *IEEE Trans. Sig. Proc.*, vol. 59, no. 9, pp. 4183–4198, 2011. 1, 3

[9] Wikipedia, "Tikhonov regularization," http://en.wikipedia.org/wiki/Tikhonov_regularization. 2

[10] John Wright, Allen Y. Yang, Arvind Ganesh, Shankar S Sastry, and Yi Ma, "Robust face recognition via sparse representation," *IEEE T-PAMI*, vol. 31, no. 2, pp. 210–227, 2009. 2, 4

[11] Zhaowen Wang, Jianchao Yang, Nasser Nasrabadi, and Thomas Huang, "A max-margin perspective on sparse representation-based classification," in *IEEE ICCV*, 2013. 2

[12] Yuanming Suo, Minh Dao, Umamahesh Srinivas, Vishal Monga, and Trac D. Tran, "Structured dictionary learning for classification," in *arXiv*, 2014, vol. 1406.1943. 2, 4

[13] Ke Huang and Selin Aviyente, "Sparse representations for signal classification," in *NIPS*, 2006. 2

[14] Lei Zhang, Meng Yang, and Xiangchu Feng, "Sparse representation or collaborative representation: Which helps face recognition?," in *IEEE ICCV*, 2011. 2

[15] John Wright, Arvind Ganesh, Allen Yang, Zihan Zhou, and Yi Ma, "A tutorial on how to apply the models and tools correctly," in *arXiv*, 2011, vol. 1111.1014. 2

[16] Joel A. Tropp and Anna C. Gilbert, "Signal recovery from random measurements via Orthogonal Matching Pursuit," *IEEE Trans. Inf. Theory*, vol. 53, no. 12, pp. 4655–4666, 2007. 2

[17] Emmanuel J. Candes, Justin Romberg, and Terence Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplet frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, 2006. 2

[18] Edoardo Amaldi and Viggo Kann, "On the approximability of minimizing nonzero variables or unsatisfied relations in linear systems," *Theoretical Computer Science*, vol. 209, pp. 237–260, 1998. 2

[19] Emmanuel J. Candes, Justin Romberg, and Terence Tao, "Stable signal recovery from incomplete and inaccurate measurements," *Comm. Pure Appl. Math.*, vol. 59, pp. 1207–1223, 2006. 2

[20] Emmanuel J. Candes and Terence Tao, "Decoding by linear programming," *IEEE Trans. Inf. Theory*, vol. 51, no. 12, pp. 4203–4215, 2005. 2

[21] Guangcan Liu, Zhouchen Liu, Shuicheng Yan, Ju Sun, Yong Yu, and Yi Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE T-PAMI*, vol. 35, no. 1, pp. 171–185, 2013. 2

[22] Ehsan Elhamifar and Rene Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE T-PAMI*, vol. 35, no. 11, pp. 2765–2781, 2013. 2

[23] Meng Yang and Lei Zhang, "Gabor feature based sparse representation for face recognition with gabor occlusion dictionary," in *ECCV*, 2010. 2

[24] Wikipedia, "Homogeneous coordinates," http://en.wikipedia.org/wiki/Homogeneous_coordinates. 2

[25] Raja Giryes, Michael Elad, and Yonina C Eldar, "The projected GSURE for automatic parameter tuning in iterative shrinkage methods," *Appl. Comp. Harm. Anal.*, vol. 30, pp. 407–422, 2010. 2

[26] Stephen P. Boyd, "ADMM," http://web.stanford.edu/~boyd/admm.html. 2, 3

[27] Ming Yuan and Yi Lin, "Model selection and estimation in regression with grouped variables," *J. Royal Statistical Society*, vol. 68, no. 1, pp. 49–67, 2013. 3

[28] Jerome Friedman, Trevor Hastie, and Robert Tibshirani, "A Note on the Group Lasso and a Sparse Group Lasso," in *arXiv*, 2010, vol. 1001.0736. 3

[29] Allen Y. Yang, Arvind Ganesh, Zihan Zhou, Andrew Wagner, Shankar Sastry Victor Shia, and Yi Ma, "Fast l-1 minimization algorithms," http://www.eecs.berkeley.edu/~yang/software/l1benchmark/, 2008. 3

[30] Minh Dao, Nasser M. Nasrabadi, and Trac D. Tran, "Collaborative multi-sensor classification via sparsity-based representation," in *arXiv*, 2014. 3

[31] Patrick Lucey, Jeffrey F. Cohn, Takeo Kanade, Jason Saragih, and Zara Ambadar, "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression," in *IEEE CVPR*, 2010. 3, 4

[32] MathWorks, "Matlab Computer Vision System Toolbox," http://www.mathworks.com/products/computer-vision/. 3

[33] Gregory D. Hager and Peter N. Belhumeur, "Efficient region tracking with parametric models of geometry and illumination," *IEEE T-PAMI*, vol. 20, no. 10, pp. 1025–1039, 1998. 4