Robust and Low-Rank Representation for Fast Face Identification With Occlusions

Michael Iliadis, *Student Member, IEEE*, Haohong Wang, *Member, IEEE*, Rafael Molina, *Member, IEEE*,
and Aggelos K. Katsaggelos, *Fellow, IEEE*

Abstract—In this paper, we propose an iterative method to address the face identification problem with block occlusions. Our approach utilizes a robust representation based on two characteristics in order to model contiguous errors (e.g., block occlusion) effectively. The first fits to the errors a distribution described by a tailored loss function. The second describes the error image as having a specific structure (resulting in low-rank in comparison with image size). We will show that this joint characterization is effective for describing errors with spatial continuity. Our approach is computationally efficient due to the utilization of the alternating direction method of multipliers. A special case of our fast iterative algorithm leads to the robust representation method, which is normally used to handle non-contiguous errors (e.g., pixel corruption). Extensive results on representative face databases (in constrained and unconstrained environments) document the effectiveness of our method over existing robust representation methods with respect to both identification rates and computational time.

Index Terms—Face identification, robust representation, low-rank estimation, iterative reweighted coding.
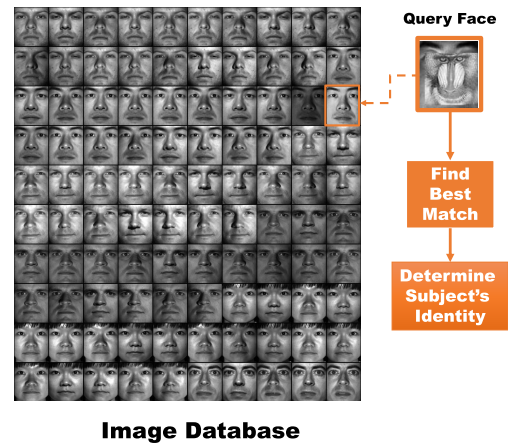
Fig. 1. Overview of the face identification problem investigated in this work. The image database consists of non-occluded faces of different subjects with small illumination changes while the query face exhibits 70% block occlusion that covers most of the informative features of the face.

I. INTRODUCTION

FACE Identification (FI) focuses on deducing a subject's identity through a provided test image and is one of the most popular problems in computer vision [1], [2]. Typically, test images exhibit large variations, such as occlusions. Ideally, if the training set contains the same type of occlusion as the test image then identification becomes a rather straightforward task. In practice, however, there is no guarantee that the collected data would cover all different occlusions for all identities of interest. An example of this problem is presented in Figure 1. The image database consists of non-occluded faces of subjects with intra-class illumination differences while the query face exhibits a 70% random block occlusion that covers

most of the informative features of the face. In applications where prior knowledge such as the region and the object of occlusion is not provided, an appropriate modeling of the error between the test image and the training samples is necessary.

Early works on face identification [3], [4] attempted to deal with illumination variations. The concept of ℓ_1-graph, which is robust to data noises and naturally sparse, was introduced in [5] to encode the overall behavior of the training set in sparse representations. To handle more complex variations such as face disguise and expressions, sparse representation-based classification models were proposed [6]–[10]. The main idea in these approaches is that a subject's face sample can be represented as a linear combination of available images of the same subject. Then, the face class that yields the minimum reconstruction error is selected in order to classify the subject. One recent extension of the sparse representation-based classification model is the class-wise sparse representation [11]. In this method, the number of training classes is minimized to alleviate the problem of representing the query by samples from many different subjects. Another extension is the patch-based classification approach [11], [12]. The patch based approach employs the sparse representation-based classifier to each patch of the face separately and the final decision is made by fusing the patch classification results. Notice that, in patch-based approaches the way to partition the image might be critical for the identification performance, especially when occlusion affects all patches [12].

To address cases with complex occlusion and corruption robust representation models [6], [11], [13]–[17] of the error image[1] were considered, utilizing a non-Gaussian distribution model to minimize the influence of outliers. In these models a Laplacian distribution (sparse error) or more general distributions based on M-Estimators [15], [18] were fitted to the errors. There are, however, two main drawbacks with such approaches. First, the iterative reweighted algorithm used to solve the robust representation problem is computationally expensive when dealing with high-dimensional data [13]–[16]. Second, the performance degrades with over 50% block occlusion. According to [19], this is due to the assumption that error pixels are independent. The robust representation approaches are usually effective in FI cases with non-contiguous[2] errors such as pixel corruption.

In cases of contiguous errors there is a spatial correlation among the error pixels as mentioned in [19]–[21]. To exploit this correlation, the spatial continuity of the error image was integrated into the sparse representation-based classification model [19], [20]. However, these models lack convergence guarantees [21]. To address this issue Qian *et al.* [21] observed that the error image with contiguous occlusion is low-rank and proposed to estimate the error support by solving a nuclear norm minimization problem with the use of ADMM [22]. While the low-rank assumption was well justified the method was effective with up to 50% random block occlusion. A reason might be that only the structure of the error (error support) was exploited and not its distribution (e.g., sparsity).

Low-rank estimation has been considered in [23] where a discriminative low-rank metric learning method was proposed that jointly learns a low-rank linear transformation matrix and a low-rank representation. Authors in [24] developed a graph construction model, with robust similarity metric (low-rank representation, which is robust to noisy data) and balanced property for the application of semi-supervised learning. In [25], a dictionary learning algorithm with low-rank regularization for FI was proposed with Fisher discriminant function to the coding coefficients to make the dictionary more discriminative.

Given corrupted training samples, one can utilize robust PCA (RPCA) [26] or its variants, such as, supervised low-rank (SLR) [27] and double nuclear norm-based matrix decomposition (DNMD) [28] to recover the "clean" data. Since these methods target to recover the original data matrix, they are transductive. Unlike these approaches, inductive methods aim at learning an underlying projection matrix from training data to remove possible corruptions in a new datum. Two popular inductive methods are the inductive RPCA (IRPCA) [29] and inductive DNMD (IDNMD) [28], respectively. In the experiments conducted in [29] and [28] same occlusion and corruption was utilized in both training and testing images (e.g., in [28] the same occlusion object was used in both training and testing samples). The limitation

of these approaches is that the corruption to be handled in the new datum should be similar to corruption present in the training data.

In this work, we propose an iterative method to solve the FI problem with occlusions. We consider the same scenario as in [13], [14], [19], and [21] according to which we are given "clean" frontal aligned views with a block occlusion which appear in any position on the test image but is "unseen" to the training data. When corrupted training data are provided that are not frontally aligned (e.g., scenarios in an unconstrained environment) a tool such as RPCA and SLR is employed to separate outlier pixels and corruptions from the training samples as a pre-processing step. Then, the "clean" frontal aligned faces are used for training data to perform face identification with occluded test images.

As already mentioned, with the robust methods [13], [14], [19], [21] high computational cost is exhibited and identification results significantly degrade with over 50% random block occlusion. Our approach is based on a new iterative method which is efficient in terms of computational cost and robust to block occlusions up to 70%. To describe contiguous errors (e.g., a random block occlusion) the proposed method utilizes a robust representation with two characteristics. The first fits to the errors a distribution based on a tailored loss function. The second models the reduced rank structure of the error in comparison to image size. We will then show that this joint characterization is effective for describing errors with spatial continuity. Our approach is efficient in terms of computational cost. Efficient minimization is performed by reformulating the reweighted coding problem as a constrained ADMM one, thus, avoiding costly matrix inversions.

A special case of our fast iterative algorithm leads to the robust representation method presented in [13] and [14] which is normally used to handle non-contiguous errors (e.g., pixel corruption).

The rest of the paper is organized as follows. In Section II first we present our fast iterative algorithm for FI cases with occlusion. Then, we describe the identification scheme and the weight function used in our method. Finally, a special case of the algorithm for solving efficiently the FI problem with non-contiguous errors is described. Experimental results and discussion on the performance of the proposed algorithms are presented in Section III, and conclusions are drawn in Section IV.

II. ROBUST FACE IDENTIFICATION WITH BLOCK OCCLUSIONS

In this section we propose an iterative method to address the FI problem with block occlusions. Our approach utilizes the robust representation of [13]–[16] with two characteristics and uses a tailored loss function based on M-estimators. The method handles contiguous errors that are considered low-rank in comparison to the size of the image and is efficient in terms of computational cost. A special case of our method is also utilized to solve efficiently the robust representation problem for non-contiguous errors.

Let $\mathbf{y} \in \mathbb{R}^d$ denote the face test sample in a column-wise vectorized form where $j \times k = d$ is the size of the image.

---

[1]Error image is the difference between the occluded test face and the unoccluded training face of the same identity.

[2]We call variations such as block occlusion and face disguise (e.g., scarves) contiguous errors since the error image is zero everywhere except in the region of the occluded object.
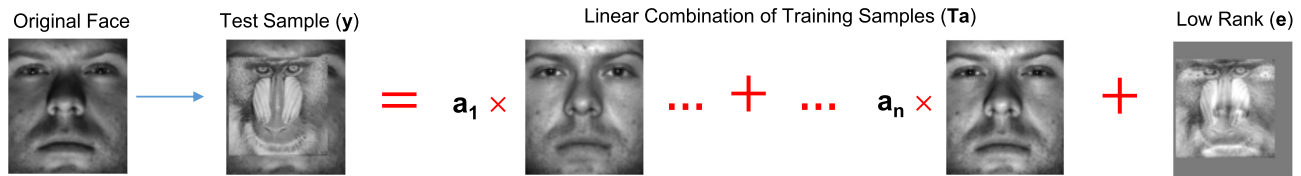
Fig. 2. Degradation Model: the test sample with occlusion can be represented as the linear combination of training samples with some intra-class variations (e.g., lighting) plus the error **e**. The error **e** has two characteristics; it is considered low-rank in comparison of image size and follows a distribution described by a tailored potential loss function.

Let $\mathbf{T} = [T_i, \ldots, T_c] \in \mathbb{R}^{d \times n}$ denote the matrix (dictionary) with the set of samples of c subjects stacked in columns. $T_i \in \mathbb{R}^{d \times n_i}$ denotes the n_i set of samples of the ith subject, such that, $\sum_i n_i = n$.

As illustrated in Figure 2 we can represent the test sample with occlusion as the superposition of training samples and a representation error **e**, thus, the degradation model is,

$$\mathbf{y} = \mathbf{Ta} + \mathbf{e}, \tag{1}$$

where $\mathbf{e} \in \mathbb{R}^d$ is the representation error and $\mathbf{a} \in \mathbb{R}^n$ is the representation vector. Thus, the test sample can be represented as a linear combination of the samples in **T**. The face identity is chosen based on the minimum class residual provided by the estimated coefficients **a** as in [6]. The residual image **e** in our model has two characteristics:

1) It includes mainly the occluded object, and therefore it is considered low-rank in comparison to image size since many of its rows or columns are zero. Please note, that we will be using the term "low-rank" from here onward in this paper to indicate errors that have reduced rank in comparison to the size of the image.

2) It follows a distribution that can be effectively described by a tailored potential loss function (e.g., the logistic function utilized in [14]).

We expect that the calculation of **e** based on the two characteristics mentioned above will lead to an accurate estimation of the **a** coefficients and provide the correct identity. Although the first characteristic has been employed in [21] and the second in robust representation methods such as [14], both of them are necessary to adequately describe the residual image and are used together for the first time in our work. In particular, enforcing only the second characteristic will not necessarily lead to an error image that is structured and low-rank. To the best of our knowledge there is no study that utilizes both the robust representation and the structure of the error using low-rank estimation in a unified framework.

For the problem described above, we propose the following function to be minimized,

$$J(\mathbf{a}) = \sum_{i=1}^{d} \phi((\mathbf{y} - \mathbf{Ta})_i) + \lambda_* \|T_M(\mathbf{y} - \mathbf{Ta})\|_* + \vartheta(\mathbf{a}), \tag{2}$$

where $\phi : \mathbb{R} \to \mathbb{R}$ is a potential loss function which is selected from a variety of M-Estimators [15], [18] and $(\mathbf{y} - \mathbf{Ta})_i$ is the ith component error. The function $\vartheta(\mathbf{a})$ is used as a regularizer of the coefficients **a**, $\lambda_* > 0$ and T_M is an operator that transforms its vector argument into a matrix of appropriate

size such that $T_M(\mathbf{y} - \mathbf{Ta}) \in \mathbb{R}^{j \times k}$. The nuclear norm $\|\cdot\|_*$ is the convex-relaxation of the rank function.

Thus we are looking in minimizing the following problem,

$$\min_{\mathbf{a}} J(\mathbf{a}). \tag{3}$$

Notice, how previous works relate to our model:

1) For $\phi(x) = x^2$, $\lambda_* = 0$ and $\vartheta(\mathbf{a}) = \lambda\|\mathbf{a}\|_1$ with $\lambda > 0$, it is the sparse representation-based classification (SRC) [6] given as,

$$\min_{\mathbf{a}} \|\mathbf{y} - \mathbf{Ta}\|_2^2 + \lambda\|\mathbf{a}\|_1. \tag{4}$$

2) For $\phi(x) = x^2$ and $\vartheta(\mathbf{a}) = \lambda\|\mathbf{a}\|_2^2$, it is the low-rank regularized regression (LR^3) [21] which is formulated as,

$$\min_{\mathbf{a}} \|\mathbf{y} - \mathbf{Ta}\|_2^2 + \lambda_* \|T_M(\mathbf{y} - \mathbf{Ta})\|_* + \lambda \|\mathbf{a}\|_2^2. \tag{5}$$

3) For $\lambda_* = 0$, it is the robust representation problem [13]–[16] formulated as,

$$\min_{\mathbf{a}} \sum_{i=1}^{d} \phi((\mathbf{y} - \mathbf{Ta})_i) + \vartheta(\mathbf{a}). \tag{6}$$

In previous works authors have chosen different functions $\vartheta(\mathbf{a})$ to regularize the coefficients **a**. In the collaborative representation-based classification with regularized least square (CR-RLS) [9] the authors are solving the SRC problem with $\vartheta(\mathbf{a}) = \lambda\|\mathbf{a}\|_2^2$. In [13] and [15] $\vartheta(\mathbf{a}) = \lambda\|\mathbf{a}\|_1$ was used combined with different potential functions while in the regularized robust coding (RRC) [14], $\vartheta(\mathbf{a}) = \lambda\|\mathbf{a}\|_2^2$ was used. In correntropy-based sparse representation (CESR) [16] and structured sparse error coding (SSEC) [19], $\vartheta(\mathbf{a})$ was chosen to be the indicator function of the nonnegative orthant $\mathbb{R}_+^n$, such that a nonnegative $\mathbf{a} \geq 0$ regularization term was enforced.

In this work we choose $\vartheta(\mathbf{a})$ so as the representation coefficients are nonnegative, since it has been shown to provide robust representation for FI in [16] and [19].

The robustness property of $\phi(\cdot)$ in (2), to be described now, in combination with $\vartheta(\cdot)$ will force some of the coefficients of **e** to be zero or very small in magnitude. It will also force **a** to be concentrated in areas of the images that can be represented well by faces in **T**. Furthermore, the use of the nuclear norm will force the residual to be low-rank.

We consider potential loss functions $\phi(\cdot)$ symmetric around zero associated to Super Gaussian (SG) distributions [30]. The function $\phi(\sqrt{x})$ has to be increasing and concave for $x \in (0, \infty)$ [31]. This condition is equivalent to $\phi'(x)/x$ being decreasing on $(0, \infty)$, that is, for $x_1 \geq x_2 \geq 0$,

$\phi'(x_1)/x_1 \le \phi'(x_2)/x_2$. If this condition is satisfied, then $\phi(\cdot)$ can be represented as (using [32, Ch. 12]),

$$\phi(x) = \inf_{\xi>0} \frac{1}{2}\xi x^2 - \phi^*\left(\frac{1}{2}\xi\right), \tag{7}$$

where $\phi^*(\xi)$ is the concave conjugate of $\phi(\sqrt{x})$ and ξ is a variational parameters. The dual relationship to (7) is given by [32],

$$\phi^*\left(\frac{1}{2}\xi\right) = \inf_x \frac{1}{2}\xi x^2 - \phi(x). \tag{8}$$

Equality in (7) is obtained at the optimal values of ξ, which are computed from the dual representation (8) by taking the derivative with respect to x and setting it to zero, which gives $\xi = \phi'(x)/x$. By using (7) we can write the function in (2) as,

$$J(\mathbf{a}) = \min_{\mathbf{w}} \frac{1}{2}\|\sqrt{\mathbf{W}}(\mathbf{y} - \mathbf{Ta})\|_2^2 + \varphi(\mathbf{w}) \\ + \lambda_*\|T_M(\mathbf{y} - \mathbf{Ta})\|_* + \vartheta(\mathbf{a}), \tag{9}$$

where $\mathbf{w} = (\xi_1, \ldots, \xi_d)$ with $\xi_i > 0, i = 1, \ldots, d$, $\mathbf{W} = \text{diag}(\mathbf{w})$, and $\varphi(\mathbf{w}) = \sum_{i=1}^{d} \phi^*(\frac{1}{2}\xi_i)$.

Notice, before proceeding, that the weights in $\mathbf{W}$ are of the form $\phi'(x)/x$ which are large for small values of x for SG potential loss functions, so $\mathbf{a}$ will fit well small values of $\mathbf{e}$ in magnitude.

Let us consider the augmented function by subsistuting (9) into (3),

$$J(\mathbf{a}, \mathbf{w}) = \frac{1}{2}\|\sqrt{\mathbf{W}}(\mathbf{y} - \mathbf{Ta})\|_2^2 + \varphi(\mathbf{w}) \\ + \lambda_*\|T_M(\mathbf{y} - \mathbf{Ta})\|_* + \vartheta(\mathbf{a}). \tag{10}$$

A local minimizer $(\mathbf{a}, \mathbf{w})$ can be calculated by alternating minimization in two steps; in step one, the weights are updated by fixing the representation coefficients $\mathbf{a}$ and in step two the vector $\mathbf{a}$ is updated by fixing the weights in $\mathbf{W}$, i.e.,

$$w_i^{t+1} = \phi'((\mathbf{y} - \mathbf{Ta}^t)_i)/(\mathbf{y} - \mathbf{Ta}^t)_i \tag{11a}$$

$$\mathbf{a}^{t+1} = \underset{\mathbf{a}}{\operatorname{argmin}} \|\sqrt{\mathbf{W}^{t+1}}(\mathbf{y} - \mathbf{Ta})\|_2^2 \\ + \lambda_*\|T_M(\mathbf{y} - \mathbf{Ta})\|_* + \vartheta(\mathbf{a}), \tag{11b}$$

where w_i^{t+1} and $\mathbf{a}^{t+1}$ are the weights and representation coefficients estimated at the tth iteration, respectively. The term $(\mathbf{y} - \mathbf{Ta}^t)_i$ denotes the component error i at the tth iteration. Large weights are assigned to pixels in the residual image with small errors in the previous reconstructed iteration, while small weights are assigned to pixels in the residual image with large errors.

We expect the use of the nuclear norm of the residual to force small weights in $\mathbf{W}$ to be assigned only on the occluded part. Thus, outlier pixels will not contribute much to the reconstruction at the next iteration, and at convergence, the estimated error will mainly consist of those outliers.

Notice that we have,

$$J(\mathbf{a}^t) = J(\mathbf{a}^t, \mathbf{w}^{t+1}) \\ \ge J(\mathbf{a}^{t+1}, \mathbf{w}^{t+1}) \ge J(\mathbf{a}^{t+1}, \mathbf{w}^{t+2}) = J(\mathbf{a}^{t+1}). \tag{12}$$

A. Optimization

Let us now describe the iterative algorithm to find efficiently $\mathbf{a}^{t+1}$ in problem (11b).

In order to solve the proposed problem, first we let $\mathbf{y} - \mathbf{Ta} = \mathbf{e}$ and since we are interested in estimating nonnegative coefficients for the representation vector we also introduce an additional variable $\mathbf{z}$ such that $\mathbf{a} = \mathbf{z}$. Then, the coding step (11b) is reformulated as,

$$\underset{\mathbf{a},\mathbf{z},\mathbf{e}}{\operatorname{minimize}} \|\sqrt{\mathbf{W}^{t+1}}\mathbf{e}\|_2^2 + \lambda_*\|T_M(\mathbf{e})\|_* + \vartheta(\mathbf{z}) \\ \text{subject to } \mathbf{y} - \mathbf{Ta} = \mathbf{e}, \mathbf{a} = \mathbf{z}. \tag{13}$$

Problem (13) is solved efficiently with ADMM which is known for fast convergence to an approximate solution [22]. As in the method of multipliers, the problem takes the form of the augmented Lagrangian,

$$L(\mathbf{e}, \mathbf{a}, \mathbf{z}, \mathbf{u}_1, \mathbf{u}_2, \mathbf{w}^{t+1}) \\ = \|\sqrt{\mathbf{W}^{t+1}}\mathbf{e}\|_2^2 + \lambda_*\|T_M(\mathbf{e})\|_* \\ + \vartheta(\mathbf{z}) + \mathbf{u}_1^T(\mathbf{y} - \mathbf{Ta} - \mathbf{e}) + \frac{\rho_1}{2}\|\mathbf{y} - \mathbf{Ta} - \mathbf{e}\|_2^2 \\ + \mathbf{u}_2^T(\mathbf{a} - \mathbf{z}) + \frac{\rho_2}{2}\|\mathbf{a} - \mathbf{z}\|_2^2, \tag{14}$$

where $\rho_1 > 0$ and $\rho_2 > 0$ are the penalty parameters, and $\mathbf{u}_1$ and $\mathbf{u}_2$ are the dual variables. The ADMM updates can be expressed as,

$$\mathbf{e}_{s+1} = \underset{\mathbf{e}}{\operatorname{argmin}} L(\mathbf{e}, \mathbf{a}_s, \mathbf{u}_{1,s}, \mathbf{w}^{t+1}), \tag{15a}$$

$$\mathbf{z}_{s+1} = \underset{\mathbf{z}}{\operatorname{argmin}} L(\mathbf{z}, \mathbf{a}_s, \mathbf{u}_{2,s}), \tag{15b}$$

$$\mathbf{a}_{s+1} = \underset{\mathbf{a}}{\operatorname{argmin}} L(\mathbf{e}_{s+1}, \mathbf{a}, \mathbf{u}_{1,s}, \mathbf{z}_{s+1}, \mathbf{u}_{2,s}), \tag{15c}$$

$$\mathbf{u}_{1,s+1} = \mathbf{u}_{1,s} + \rho_1(\mathbf{y} - \mathbf{Ta}_{s+1} - \mathbf{e}_{s+1}), \tag{15d}$$

$$\mathbf{u}_{2,s+1} = \mathbf{u}_{2,s} + \rho_2(\mathbf{a}_{s+1} - \mathbf{z}_{s+1}), \tag{15e}$$

where s denotes the ADMM iteration and finally we set $\mathbf{a}^{t+1} = \lim_{s\to+\infty} \mathbf{a}_{s+1}$.

1) Finding $\mathbf{e}_{s+1}$: The update of $\mathbf{e}_{s+1}$ is given by minimizing the following problem,

$$\mathbf{e}_{s+1} = \underset{\mathbf{e}}{\operatorname{argmin}} \frac{1}{2}\|\mathbf{y} - \mathbf{Ta}_s - \mathbf{e}\|_2^2 + \frac{1}{\rho_1}\|\sqrt{\mathbf{W}^{t+1}}\mathbf{e}\|_2^2 \\ + \frac{1}{\rho_1}\mathbf{u}_{1,s}^T(\mathbf{y} - \mathbf{Ta}_s - \mathbf{e}) + \frac{\lambda_*}{\rho_1}\|T_M(\mathbf{e})\|_*. \tag{16}$$

To calculate $\mathbf{e}_{s+1}$ we consider a two-step fast approximation. In step one we solve the weighted norm problem,

$$\tilde{\mathbf{e}} = \underset{\mathbf{e}}{\operatorname{argmin}} \frac{1}{2}\|\mathbf{y} - \mathbf{Ta}_s - \mathbf{e}\|_2^2 + \frac{1}{\rho_1}\|\sqrt{\mathbf{W}^{t+1}}\mathbf{e}\|_2^2 \\ + \frac{1}{\rho_1}\mathbf{u}_{1,s}^T(\mathbf{y} - \mathbf{Ta}_s - \mathbf{e}), \tag{17}$$

and in step two to satisfy the nuclear norm constraint we project the estimated $\tilde{\mathbf{e}}$ to a low-rank space, to obtain,

$$\mathbf{e}_{s+1} = \underset{\mathbf{e}}{\operatorname{argmin}} \frac{1}{2}\left\|T_M\left(\mathbf{e} - \tilde{\mathbf{e}}\right)\right\|_F^2 + \frac{\lambda_*}{\rho_1}\|T_M(\mathbf{e})\|_*. \tag{18}$$

Problem (17) has a closed-form solution given by,

$$\tilde{\mathbf{e}} = \mathbf{C}^{-1}(\mathbf{y} - \mathbf{Ta}_s + \mathbf{u}_{1,s}/\rho_1), \tag{19}$$
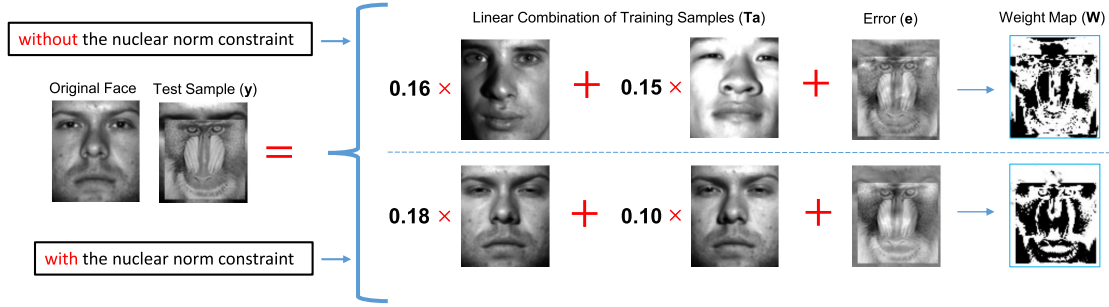
Fig. 3. An example of 70% occluded face. If the low-rank constraint is not used (thus, allowing only the weighting norm in (10) to be utilized) then **y** is reconstructed from faces that belong to the wrong identities. On the other hand, our method which employs both the nuclear and weighting norms to describe the error, is able to reconstruct the occluded face from images that belong to the correct identity.

where $\mathbf{C} = \left(\mathbf{I} + 2\mathbf{W}^{t+1}/\rho_1\right)$ is a diagonal matrix with diagonal entries $c_i = 1 + 2w_i^{t+1}/\rho_1$. Since $\mathbf{C}$ is a diagonal matrix, to update $\tilde{\mathbf{e}}$ we only need to construct a vector with elements equal to $1/c_i$ and perform an element wise-multiplication between the constructed vector and the residual vector $\mathbf{y} - \mathbf{Ta}_s + \mathbf{u}_{1,s}/\rho_1$. Thus, the update of $\tilde{\mathbf{e}}$ can be calculated fast.

The update in (19) is in essence an outlier detector similar to the *soft-thresholding* operator [22]. The values of the residual vector $\mathbf{y} - \mathbf{Ta}_s + \mathbf{u}_{1,s}/\rho_1$ will be weighted according to $\mathbf{C}^{-1}$. A small weight (close to zero) will be given to non-outlier elements (e.g., elements of the residual vector with small values) while a large weight (close to one) will be given to outliers (e.g., elements of the residual vector with large values).

Problem (18) is a nuclear norm minimization problem of the form,

$$\min_{\mathbf{X}} \frac{1}{2}\|\mathbf{X} - \Delta\|_F^2 + \lambda_*\|\mathbf{X}\|_*, \qquad (20)$$

which has a closed-form solution given by [22],

$$\hat{\mathbf{X}} = \mathcal{L}_{\lambda_*}(\Delta) = \mathbf{U}_\Delta \mathcal{S}_{\lambda_*}(\Sigma_\Delta)\mathbf{V}_\Delta^T, \qquad (21)$$

where $\mathcal{S}_{\lambda_*}(\Sigma_\Delta) = \text{sign}(\delta_{ij})\max(0, |\delta_{ij}| - \lambda_*)$ is the soft-thresholding operator, $\mathcal{L}_{\lambda_*}(\cdot)$ is the singular value soft-thresholding operator and $\Delta = \mathbf{U}_\Delta \Sigma_\Delta \mathbf{V}_\Delta^T$ is the SVD of matrix Δ.

Thus, the two-step solution of (16) is given by,

$$\tilde{\mathbf{e}} = \mathbf{C}^{-1}\left(\mathbf{y} - \mathbf{Ta}_s + \mathbf{u}_{1,s}/\rho_1\right) \qquad (22a)$$

$$\mathbf{e}_{s+1} = \mathcal{L}_{\lambda_*/\rho_1}\left(T_M(\tilde{\mathbf{e}})\right). \qquad (22b)$$

The solution in (22) is the low-rank estimation of the weighted error image.[3]

2) Finding $\mathbf{z}_{s+1}$: The update of $\mathbf{z}_{s+1}$ is obtained by solving the following problem,

$$\mathbf{z}_{s+1} = \underset{\mathbf{z}}{\text{argmin}}\, \frac{1}{2}\|\mathbf{a}_s - \mathbf{z}\|_2^2 + \frac{1}{\rho_2}\vartheta(\mathbf{z}) + \frac{1}{\rho_2}\mathbf{u}_{2,s}^T(\mathbf{a}_s - \mathbf{z}). \qquad (23)$$

The solution of (23) is given by,

$$\mathbf{z}_{s+1} = (\mathbf{a}_s + \mathbf{u}_{2,s}/\rho_2)_+, \qquad (24)$$

[3] Note that the solution in (22b) is obtained in a matrix form which is then transformed to a column-wise vectorized form $\mathbf{e}_{s+1}$.

where $(\cdot)_+$ is the function the keeps only the positive coefficients of its argument and set the rest to zero.

Notice that we only need to change the update of $\mathbf{z}$ for different regularization functions $\vartheta(\mathbf{z})$. For example, to solve an iterative reweighted sparse coding (IRSC) problem and regularize the coefficients to be sparse ($\vartheta(\mathbf{z}) = \lambda\|\mathbf{z}\|_1$) we have to substitute (24) with the *soft-thresholding* operator [22]. To solve an iterative reweighted least squares (IRLS) problem ($\vartheta(\mathbf{a}) = \lambda\|\mathbf{a}\|_2^2$) we do not need to introduce and estimate the additional variables, $\mathbf{z}$ and $\mathbf{u}_2$. The coefficients $\mathbf{a}$ can be estimated by solving a regularized least squares problem.

3) Finding $\mathbf{a}_{s+1}$: The update of $\mathbf{a}_{s+1}$ is obtained by solving the following problem,

$$\mathbf{a}_{s+1} = \underset{\mathbf{a}}{\text{argmin}}\, \frac{1}{2}\|\mathbf{y} - \mathbf{Ta} - \mathbf{e}_{s+1}\|_2^2 + \frac{1}{\rho_1}\mathbf{u}_{1,s}^T(\mathbf{y} - \mathbf{Ta} - \mathbf{e}_{s+1})$$
$$+ \frac{1}{\rho_1}\mathbf{u}_{2,s}^T(\mathbf{a} - \mathbf{z}_{s+1}) + \frac{\rho_2}{2\rho_1}\|\mathbf{a} - \mathbf{z}_{s+1}\|_2^2. \qquad (25)$$

Notice that equation (25) is a regularized least squares problem, whose closed-form solution is,

$$\mathbf{a}_{s+1} = \mathbf{P}\Big(\mathbf{T}^T(\mathbf{y} - \mathbf{e}_{s+1} + \mathbf{u}_{1,s}/\rho_1) + (\rho_2/\rho_1)\mathbf{z} - \mathbf{u}_{2,s}/\rho_1\Big), \qquad (26)$$

where $\mathbf{P} = \left(\mathbf{T}^T\mathbf{T} + \frac{\rho_2}{\rho_1}\mathbf{I}\right)^{-1}$ and can be pre-calculated once and cached offline. For large matrices $\mathbf{P}$, iterative algorithms can be used for solving this linear system of equations when matrix inversion is not feasible.

An example of our approach is presented in Figure 3. If the low-rank constraint is not used then **y** is reconstructed using faces that belong to the wrong identities. On the other hand, our approach with the use of the nuclear norm constraint is able to reconstruct the occluded face using images that belong to the correct identity (second row in Figure 3). Also, notice the differences in weight map estimations. In the first case small weights (black values) assigned without any structure to any region of the face. This is not desirable since informative pixels were detected as outliers (e.g., pixels around the occluded object). In addition, many pixels on the occluded object were detected as inliers. In the second case, the error is low-rank and has a spatial continuity around the occluded object (pixels close to zero). In this case, the weight map $\mathbf{W}$ is also enforced to have a spatial continuity (since $\mathbf{W}$ is related to the error) with small weights assigned to the occluded object as desired.

Algorithm 1 Fast & Low-Rank IRNNLS Algorithm

Inputs: $\mathbf{y}$, $\mathbf{T}$, λ_*, ρ_1, ρ_2, ϵ_1, ϵ_2 and ϵ_3.

Initialize $\mathbf{a}^1 = 1/n$, $\mathbf{u}_{1,1} = \mathbf{0}$, $\mathbf{u}_{2,1} = \mathbf{0}$, $t = 0$ and $s = 0$

Repeat

1) $t = t + 1$
2) Estimate the weights,

$$w_i^t = \phi'((\mathbf{y} - \mathbf{Ta}^t)_i)/(\mathbf{y} - \mathbf{Ta}^t)_i, \quad i = 1, \ldots, d$$

    Repeat

        3) $s = s + 1$
        4) Find $\mathbf{e}_s$ using (22) (contiguous errors) or (32)
            (non-contiguous errors)
        5) Find $\mathbf{z}_s$ using (24)
        6) Update $\mathbf{a}_s$ using (26)
        7) Update $\mathbf{u}_{1,s}$ and $\mathbf{u}_{2,s}$ using (15d), (15e)
    Until *converge*

8) Set $\mathbf{a}^t = \mathbf{a}_s$, $\mathbf{u}_{1,1} = \mathbf{0}$, $\mathbf{u}_{2,1} = \mathbf{0}$, $s = 0$

Until *converge*

Output: The final estimates of $\mathbf{a}$ and $\mathbf{w}$.

We would like to mention here a closely related work for block occlusion errors namely robust low-rank regularized regression (RLR[3]) [21]. There are two fundamental differences between our work and RLR[3]. First, the weighted residual $\mathbf{W}(\mathbf{y} - \mathbf{Ta})$ instead of the residual error $\mathbf{y} - \mathbf{Ta}$ was modeled to be low-rank which is different from our method. The model in RLR[3] can handle occlusion of specific objects and size that covers a portion of the face image entirely from left to right (or from top to bottom) such as scarves. Our method handles block occlusions that appear in any size and place in the face. Second, the function to be minimized in RLR[3] is not derived from the duality theorem [32] which raises concerns about its convergence guarantees.

The complete steps of the fast low-rank and iterative reweighted nonnegative least squares (F-LR-IRNNLS) algorithm for contiguous errors are presented in Algorithm 1.

B. Identification Scheme

In SRC [6], the face class that yields the minimum reconstruction error is selected in order to classify or identify the subject. Similarly, in this work the classification is given by computing the residuals e for each class i as,

$$e_i(\mathbf{y}) = \|\sqrt{\mathbf{W}^f}(\mathbf{y} - T_i\hat{\mathbf{a}}_i)\|_2, \tag{27}$$

where $\hat{\mathbf{a}}_i$ is the segment of the final estimated $\mathbf{a}$ associated with class i and $\mathbf{W}^f$ is the final estimated weight matrix from Algorithm 1. Finally, the identity of $\mathbf{y}$ is given as, Identity$(\mathbf{y}) = \text{argmin}_i\{e_i(\mathbf{y})\}$.

C. The Weight Function

Ideally, the weight function should distinguish inliers and outliers given a training dictionary with non-occluded faces and a test sample with occlusion [14]. In particular, given the residual error at any iteration, small weights (close to zero)
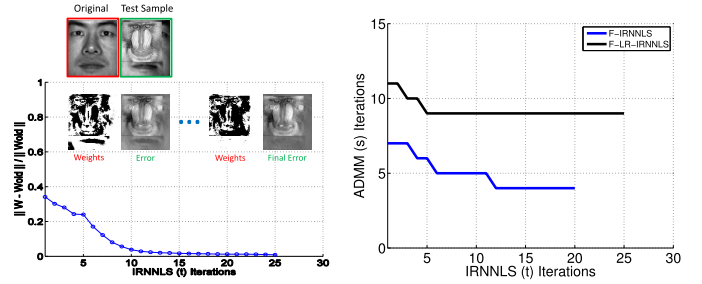

Fig. 4. The left graph shows the convergence of the iterative reweighted and low-rank algorithm (F-LR-IRNNLS). In particular, we present the change of the weights between two consecutive iterations. As $t \to \infty$ the error becomes more sparse and structural, thus, small weights are concentrated on the occluding object. The right graph shows the number of ADMM iterations as a function of the reweighted iterations.

should be assigned to the outlier pixels (large residual error) and larger weight (close to one) to the inlier pixels (small residual error). Although any weight function [15], [18] of the form $w = \phi'(x)/x$ can be used in our framework as long as $\phi'(x)/x$ is decreasing on $(0, \infty)$, in this work we utilize the logistic function proposed in [14]. The logistic function performs particularly well in FI.[4] The weight component w_i as a function of x_i, which is decreasing on $(0, \infty)$, is given by,

$$w_i = \frac{\exp\left(-\mu\left(x_i\right)^2 + \mu\eta\right)}{1 + \exp\left(-\mu\left(x_i\right)^2 + \mu\eta\right)}, \quad i = 1, \ldots, d, \tag{28}$$

where μ and η are positive scalars. As in [14], η denotes the value of the lth largest element of the residual vector $\mathbf{x}$, where $l = \lfloor \gamma d \rfloor$, $\gamma \in (0, 1)$. μ is given as $\frac{\zeta}{\eta}$ with $\zeta = 8$. We also set $\gamma = 0.8$ for the experiments without occlusion and $\gamma = 0.6$ for the experiments with occlusion as in [14].

D. Convergence Criteria

For the purpose of this paper, in order to guarantee convergence of the optimization problem (13) using ADMM, it is sufficient to enforce appropriate termination criteria. As suggested in [22], we enforced the primal residuals to be small such that $\|\mathbf{y} - \mathbf{Ta} - \mathbf{e}\|_2 \leq \epsilon_1$ and $\|\mathbf{a} - \mathbf{z}\|_2 \leq \epsilon_2$, where ϵ_1 and ϵ_2 are small positive numbers.

The termination criterion for the iterative reweighted sequences is $\|\mathbf{W}^t - \mathbf{W}^{t-1}\|_2/\|\mathbf{W}^{t-1}\|_2 < \epsilon_3$, where ϵ_3 is a small positive number. Figure 4 (left) shows the change of the weights between two consecutive iterations for one particular example.

E. Special Case: Robust Representation for Non-Contiguous Errors

A special case of our method leads to the robust representation method [13]–[16] when used for FI problem with non-contiguous errors as given in (6). In the previous robust methods [13]–[16] high computational cost is exhibited due to the reweighted coding step to estimate $\mathbf{a}$. To address this issue the proposed ADMM algorithm described above is utilized

---

[4] For a complete justification of the effectiveness of the logistic function in FI we refer the reader to [14].

here as well to solve this case efficiently. The augmented function of the robust representation problem is formulated as,

$$J(\mathbf{a}, \mathbf{w}) = \frac{1}{2} \|\sqrt{\mathbf{W}}(\mathbf{y} - \mathbf{Ta})\|_2^2 + \varphi(\mathbf{w}) + \vartheta(\mathbf{a}), \qquad (29)$$

which is similar to (10) but without the nuclear norm term. Similarly to (10), a local minimizer $(\mathbf{a}, \mathbf{w})$ can be calculated in two steps,

$$w_i^{t+1} = \phi'((\mathbf{y} - \mathbf{Ta}^t)_i)/(\mathbf{y} - \mathbf{Ta}^t)_i \qquad (30a)$$

$$\hat{\mathbf{a}}^{t+1} = \underset{\mathbf{a}}{\arg\min} \|\sqrt{\mathbf{W}^{t+1}}(\mathbf{y} - \mathbf{Ta})\|_2^2 + \vartheta(\mathbf{a}). \qquad (30b)$$

A major drawback of the iterative reweighted algorithm is that it is computationally expensive [33] due to the coding step (30b). High computational cost is exhibited regardless of the coefficient regularization, for example $\vartheta(\mathbf{a}) = \lambda \|\mathbf{a}\|_2^2$, $\vartheta(\mathbf{a}) = \lambda \|\mathbf{a}\|_1$ or $\vartheta(\mathbf{a})$ is the indicator function of the nonnegative orthant $\mathbb{R}_+^n$ [13]–[16]. The coding step is expensive since in each iteration a new weighted system matrix $\sqrt{\mathbf{W}^{t+1}}\mathbf{T}$ is provided given the updated weights. Moreover, an offline precalculation of the weighted inverse is not possible. Efficient methods to solve problems in the form of (30b) such as conjugate gradient [34] or ℓ_1 algorithms [35], [36] may require several iterations to converge to the desired point [14]. The active set method [37] used for solving the nonnegative least squares problem becomes also slow due to the computation of the pseudoinverse of the system matrix in each iteration.

To solve (30b) efficiently for the $\vartheta(\mathbf{a})$ functions described earlier we utilize the proposed ADMM algorithm described above. Thus, we avoid the explicit calculation of $\mathbf{T}^T\mathbf{W}^{t+1}\mathbf{T}$ inverse during the execution of the algorithm. The idea of accelerating the iterative reweighed scheme using ADMM is also found in [33]. However, in [33] the weights applied to the regularization term $\vartheta(\mathbf{a})$ rather to the residual as in done here. To accelerate (30b), we set $\mathbf{y} - \mathbf{Ta} = \mathbf{e}$, $\mathbf{a} = \mathbf{z}$ and the problem is reformulated as,

$$\underset{\mathbf{a}, \mathbf{z}, \mathbf{e}}{\text{minimize}} \; \|\sqrt{\mathbf{W}^{t+1}}\mathbf{e}\|_2^2 + \vartheta(\mathbf{z})$$
$$\text{subject to } \mathbf{y} - \mathbf{Ta} = \mathbf{e}, \quad \mathbf{a} = \mathbf{z}. \qquad (31)$$

Notice that the $\sqrt{\mathbf{W}^{t+1}}\mathbf{T}$ term is no longer part of the optimization problem which allows us to solve (31) efficiently.

Problem (31) has similar ADMM updates with (13) except $\mathbf{e}_{s+1}$. To find $\mathbf{e}_{s+1}$ we only have to calculate,

$$\mathbf{e}_{s+1} = \mathbf{C}^{-1}\big(\mathbf{y} - \mathbf{Ta}_s + \mathbf{u}_{1,s}/\rho_1\big), \qquad (32)$$

where $\mathbf{C}$ is the diagonal matrix defined in (19) and $\mathbf{C}^{-1}$ can be calculated fast due to the diagonal structure. The updates of $\mathbf{z}_{s+1}$ and $\mathbf{a}_{s+1}$ are similar to (24) and (26) respectively and as explained earlier $\mathbf{P}$ matrix can be pre-calculated once and cached offline.[5] Thus, to solve the robust representation problem we utilize an efficient method since no online inversion of the system matrix is performed in any of the variable updates.

Our fast iterative reweighted nonnegative least squares (F-IRNNLS) algorithm just described solves the

---

[5] As noted earlier, the offline calculation of $\mathbf{P}$ can be utilized when direct inversion is feasible.

robust representation problem efficiently and the steps are also presented in Algorithm 1.

The number of ADMM iterations required for each reweighted iteration is presented in Figure 4 (right) for our method. As expected the number of ADMM iterations for F-LR-IRNNLS is greater than the F-IRNNLS due to the calculation of the nuclear term.

F. Dealing With Corrupted Testing and Training Samples

Until now we have assumed that the training samples represent "clean" frontal aligned views and without large variations of the same identity. However, there might be scenarios such as face identification in an unconstrained environment where *testing* as well as *training* samples are occluded.

To deal with this scenario we incorporate into our framework techniques based on RPCA [26] to separate outlier pixels and occlusions from the training samples. The main idea of RPCA is that the training matrix is decomposed into $\mathbf{T} = \mathbf{A} + \mathbf{E}$, where $\mathbf{A}$ denotes the dictionary with clean faces and $\mathbf{E}$ is the remaining error matrix. Out of the many extensions of RPCA such as in [38] and [28], [39]–[41], in this work we utilize the SLR method in [27] to separate outlier pixels since it is robust to variations such as expression and pose which are very common in an unconstrained environment. One of the main differences between the RPCA [26] and SLR [27] methods is that an additional intra-class variation dictionary $\mathbf{B}$ is estimated in the latter method using $\mathbf{T} = \mathbf{A} + \mathbf{B} + \mathbf{E}$. Thus, having estimated dictionaries $\mathbf{A}$ and $\mathbf{B}$ by [27] we can modify our degradation model (1) to,

$$\mathbf{y} = \mathbf{Aa}_1 + \mathbf{Ba}_2 + \mathbf{e}, \qquad (33)$$

where $\mathbf{A} \in \mathbb{R}^{d \times n}$ denotes a class-specific dictionary (clean images dictionary) and $\mathbf{B} \in \mathbb{R}^{d \times n}$ a non-class specific dictionary (intra-class variation dictionary). Also, $\mathbf{a}_1 \in \mathbb{R}^n$ and $\mathbf{a}_2 \in \mathbb{R}^n$ are the representation vectors for the $\mathbf{A}$ and $\mathbf{B}$ dictionaries respectively and $\mathbf{e}$ is the representation error described earlier in our method. More compactly, the model in (33) can be written as,

$$\mathbf{y} = \mathbf{T}'\mathbf{a}' + \mathbf{e}, \qquad (34)$$

where $\mathbf{T}' = [\mathbf{A} \; \mathbf{B}] \in \mathbb{R}^{d \times 2n}$ and $\mathbf{a}' = [\mathbf{a}_1 \; \mathbf{a}_2] \in \mathbb{R}^{2n}$. Dictionary $\mathbf{B}$ in this case captures variations such as in expressions and pose that cannot be represented by the error term $\mathbf{e}$. On the other hand, $\mathbf{e}$ is utilized to capture occlusions and low-rank variations of the test image, as explained earlier in this work, that cannot be captured by $\mathbf{B}$. For the problem described above the function to be minimized is the same as in (2) where $\mathbf{T}'$ and $\mathbf{a}'$ are used instead of $\mathbf{T}$ and $\mathbf{a}$, respectively. Furthermore, the calculation of $\mathbf{A}$ and $\mathbf{B}$ is described in [27]. In our experiments we denote by F-LR-IRNNLS (SLR), the method when the SLR algorithm [27] is employed as a pre-processing step in our F-LR-IRNNLS method.

III. EXPERIMENTAL RESULTS

In this section we present experiments on four publicly available databases, AR [42], Extended Yale B [43], Multi-PIE [44] and Labeled Faces in the Wild (LFW) [45]

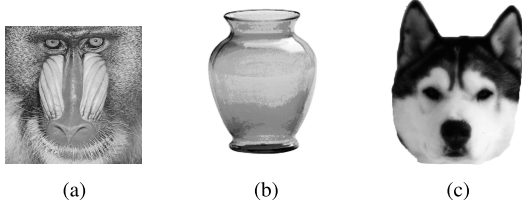(a)   (b)   (c)

Fig. 5. The three artificial images used for the block occlusion experiments. (a) Baboon. (b) Non-square image. (c) Dog.

to show the efficacy of the proposed method. We demonstrate identification and reconstruction results under various artificial and real-world variations. We compare our framework with ten other FI algorithms, SRC [6], CR-RLS [9], LR^3 [21], L12,[6] and the robust algorithms SDR-SLR [27], HQ (additive form) [15], CESR [16], RRC_L1 [14], RRC_L2 [14], SSEC [19]. We consider the following five FI cases:

1) cases with contiguous variations such as random block occlusion with different sizes and objects, face disguise and mixture noise which is a combination of block occlusion and pixel corruption,
2) cases with non-contiguous variations such as illumination variations, pixel corruption, face expressions.
3) cases with random block occlusion and few training samples.
4) cases in unconstrained environment and with corrupted testing and training samples.

For all methods, we used the solvers provided by the authors of the corresponding papers. We chose to solve the ℓ_1 minimization problem in SRC and RRC_L1 with the Homotopy algorithm[7] [46] since it resulted in the highest accuracy in the performance comparison in [35] with reasonable time execution. In our algorithms, we set $\lambda_* = 0.05$, $\rho_1 = 1$ and $\rho_2 = 0.1$. The convergence parameters were set equal to $\epsilon_1 = 10^{-2}$, $\epsilon_2 = 10^{-1}$, $\epsilon_3 = 10^{-2}$. For fair comparisons with respect to execution time and identification rates we set the same ϵ_3 for the RRC algorithm and the same maximum number of iterations ($t = 100$). All face images were normalized to have unit ℓ_2-norm and all variables initialized to zero except for $\mathbf{a}^1 = 1/n$ as in [14].

A. Identification Under Block Occlusions

Experiments with occluded images were conducted on three datasets: Extended Yale B, AR and Multi-PIE.

As in [6], [14], and [21], we chose Subsets 1 and 2 of Extended Yale B for training (in total 719 images) and Subset 3 for testing (in total 455 images). Images were resized to 96×84 pixels. We considered three different artificial objects to occlude the test images as shown in Figure 5. For the first object, block occlusion was tested by placing the square baboon image on each test image. The location of the occlusion was randomly chosen and was unknown during

training. We considered different sizes of the object such that the face is covered with the occluded object from 30% to 90% of its area. Identification rates for the different levels of occlusion are shown in Figure 6(a). For the second and third[8] non-square and smooth (e.g., without textures in it) objects shown in Figure 5(b) and (c) respectively, block occlusion was tested by randomly placing the objects on each test image. Identification rates are shown in Figure 6(b) and 6(c) respectively for each of the non-square object.

Next, we evaluate the performance to block occlusion in the AR database. We chose the 7 non-occluded AR images per subject (each one with a different face expression) from session 1 for training and the 7 non-occluded images per subject from session 2 for testing. This experiment is more challenging since training faces appear with different expressions. In each test image, we replace a random block with the square baboon image. We resized the images to 60×43 pixels. The occlusion ratio increases from 30% to 50% and identification rates are shown in Figure 6(c).

To examine the robustness of our method on a dataset with more subjects we evaluate its performance to block occlusion in the Multi-PIE database [44]. The database contains the images of 337 subjects captured in 4 sessions with simultaneous variations in pose, expression, and illumination. In the experiments we used all 249 subjects in Session 1. We used 6 frontal images with 6 illuminations[9] and neutral expression from Session 1 for training, and 10 frontal images[10] from Session 4 for testing. In this dataset we also examine the robustness of the methods for different percentages of occlusion under the same experiment and same parameters of the methods. In each test image, we replace a random block with the square baboon image. The percentage of the occlusion was randomly chosen from 30% to 60% of the image area. Thus, each test image was chosen to have different percentage of occlusion. Identification rates are shown in Table VI.

From the results with the baboon image we conclude that methods robust to contiguous errors (LR^3, SSEC) performed better than the non-contiguous methods in Yale B. Our F-LR-IRNNLS algorithm outperformed all previous methods overall in all datasets, Yale B, AR and Multi-PIE. SSEC performed well with 80% occlusion in Yale B but it performed poorly in AR and with lower levels of occlusion. As explained in [21], there are no convergence guarantees for SSEC and, perhaps, this explains the unstable results obtained by this method in our experiments. Finally, the performance of non-contiguous error methods, HQ, RRCs and F-IRNNLS significantly dropped by high levels of occlusion. This is due to the fact that these methods cannot handle contiguous variations effectively. In Multi-PIE dataset (Table VI with 6 samples) the F-LR-IRNNLS algorithm performed better than the previous methods. This also suggests that our method performs well under datasets with many subjects and with different occlusion rates.

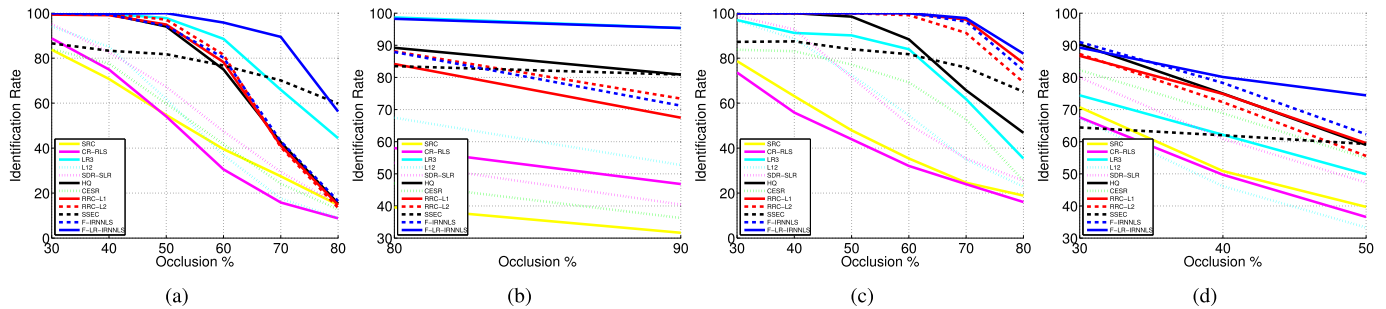From the results with the non-square object and dog image

---

Fig. 6. Identification Rates on the Extended Yale B and AR under Block Occlusion with Baboon, Vase and Dog. (a) Yale B dataset with baboon image. (b) Yale B dataset with vase image. (c) Yale B dataset with dog image. (d) AR dataset with baboon image

TABLE I

AVERAGE RUN TIME PER TEST SAMPLE ON THE EXTENDED YALE B AND AR DATASETS UNDER DIFFERENT VARIATIONS

| Dataset | Yale Baboon Occl. 60% | | AR Baboon Occl. 50% | | Yale Corruption 90% | | AR Corruption 70% | |
	Accuracy	Time	Accuracy	Time	Accuracy	Time	Accuracy	Time
LR^3 [21]	88.57%	0.06s	49.86%	0.02s	14.29%	0.06s	n/a	n/a
HQ [15]	74.95%	3.11s	59.00%	0.99s	42.64%	1.96s	85.54%	1.49s
CESR [16]	41.76%	0.78s	54.86%	0.34s	36.04%	0.99s	85.69%	0.38s
RRC_L1 [14]	78.24%	12.41s	59.57%	3.95s	**84.40%**	15.23s	**91.92%**	29.84s
RRC_L2 [14]	81.54%	10.52s	55.57%	1.96s	76.26%	7.40s	88.38%	10.93s
SSEC [19]	76.70%	1.58s	59.29%	0.80s	4.84%	2.65s	n/a	n/a
Our F-IRNNLS	80.22%	1.52s	62.29%	0.45s	79.78%	2.84s	91.62%	2.25s
Our F-LR-IRNNLS	**95.82%**	2.41s	**74.43%**	0.57s	71.87%	4.30s	n/a	n/a

we found out that all methods performed better than when the baboon was used. We attribute this to the fact that the baboon object exhibits a lot of textures and that it looks like a face. Thus, it is much more difficult for the methods to distinguish inliers and outliers. In particular, for the vase image LR^3 and F-LR-IRNNLS achieved similar identification rates at all levels of occlusion. Perhaps, modeling the error to be low-rank was sufficient for the vase image in the Yale B dataset (the weighted norm was unnecessary). Although the identification rates of SSEC were better than those provided by the RRC the method performed significantly worse than F-LR-IRNNLS. SSEC may not be effective in cases where faces are occluded by non-square objects. Finally, as expected the SDR-SLR method performed significantly better than SRC and CR-RLS in all block occlusion experiments. This is due to the fact that the additional intra-class variation dictionary in SDR-SLR captures more effectively some of the variations of the query.

The time performance for the block occlusion experiments is presented in Table I (columns 3 and 4)[11] for Yale B and AR datasets with the baboon image.[12] A key observation is that while F-IRNNLS achieved identical identification rates with RRC_L1 and RRC_L2 it is computationally more efficient by a magnitude.

B. Identification Under Expressions & Face Disguise

In this experiment we tested our algorithms with face expressions and occlusion with real-world objects in three different scenarios: 1) faces with expressions 2) faces with sunglasses and 3) faces with scarves. The training set consists

---

TABLE II

IDENTIFICATION RATES (%) UNDER FACE DISGUISE ON THE AR DATABASE

Case	Expressions	Sunglasses	Scarves
SRC [6]	82.33%	37.17%	35.17%
CR-RLS [9]	81.33%	33.83%	39.33%
LR^3 [21]	79.50%	80.50%	77.00%
L12	85.17%	40.33%	43.00%
SDR-SLR [27]	92.87%	36.67%	32.83%
HQ [15]	94.33%	66.67%	45.67%
CESR [16]	93.50%	66.50%	17.17%
RRC_L1 [14]	94.33%	83.00%	58.50%
RRC_L2 [14]	93.67%	84.17%	72.50%
SSEC [19]	56.17%	70.67%	75.00%
Our F-IRNNLS	**94.83%**	81.50%	53.67%
Our F-LR-IRNNLS	93.00%	**89.83%**	**78.83%**

of the two neutral images (one from each session) from the AR dataset per subject. For the first scenario (face expressions) the 6 images per subject from sessions 1 and 2 with face expressions (smile, anger and scream) were selected for the testing set. For the second scenario (faces with sunglasses) the testing set consisted of the 6 images per subject with sunglasses from sessions 1 and 2. In the third scenario (faces with scarves) the 6 images per subject with scarves from sessions 1 and 2 were chosen for the testing set. The images were resized to 60×43 pixels. Identification rates for the three scenarios are shown in Table II for the various methods.

For the face expressions experiment all robust algorithms achieved high performance. A key observation for this experiment is that modeling the error as low-rank does not improve the results since face expression errors do not in general form a contiguous area.

For the sunglasses experiment we observed that our F-LR-IRNNLS algorithm outperformed previous methods and was able to detect the outliers effectively. In this case modeling the error to be low-rank was adequate. This is due to the fact that the residual image consisted mainly of the sunglasses that made a contiguous error. SSEC performed poorly, perhaps,

---

[11]In Table I we do not report results from the non-robust methods SRC and CR-RLS since they are both faster than all robust methods. However, they perform poorly in terms of identification rates in all experiments. In this work, our scope is to compare time performance between robust methods.

[12]Similar time results were obtained for the non-square image.

because the method does not capture well contiguous areas that are not square. A similar conclusion was drawn above with the random block occlusion experiment with a non-square image.

Results with the scarves experiment demonstrate that all methods robust to contiguous errors performed well, as expected (since the scarf occlusion is contiguous). Our F-LR-IRNNLS method achieved the best performance with 78.83% identification rate while our non-contiguous F-IRNNLS method achieved only 53.67%. This result emphasizes the fact that exploiting the spatial correlation in contiguous variations, such as scarves, is beneficial. Time performance is not reported here since all methods run very fast (less than a second) due to the fact that the training dictionary in this experiment was relatively small (200 training samples).

C. Identification Under Mixture Noise

In this experiment we evaluate the performance of our algorithm for the case of mixture noise. In this case, both pixel corruption and block occlusion degraded the testing images. An example image with this degradation is shown in Figure 12. This experiment was conducted with two datasets, Extended Yale B and AR. Similarly to the previous Extended Yale B settings, Subsets 1 and 2 of Extended Yale B were used for training and Subset 3 was used for testing. With the AR dataset we chose the 700 non-occluded AR images for training from session 1 and the 700 non-occluded images for testing from session 2. In both datasets, for each testing image a percentage of randomly chosen pixels was corrupted. Corruption was performed by replacing those pixel values with independent and identically distributed samples from a uniform distribution between [0, 255]. Then, we placed the baboon square image on each corrupted test image. In Yale B dataset we performed this experiment with 30% pixel corruption and 60% occlusion. With the AR dataset, experiments were conducted with 20% pixel corruption and 70% occlusion. Identification rates are shown in Table III for the various methods.

F-LR-IRNNLS outperformed all previous methods which indicates that in the mixture noise case, our two error constraints capture the error term effectively. SSEC performed poorly due to the presence of pixel corruption. RRC_L1, RRC_L2 and HQ were robust to pixel corruption, however, their performance remained low since they were not effective on describing the occlusion part. Our F-LR-IRNNLS had a good balance on detecting the corrupted pixels and capturing the occlusion part with the employment of the weighted and nuclear norms. However, although F-LR-IRNNLS achieved significantly higher performance than the previous methods, the actual accuracy was relative low with 63.08% in YaleB and 57.29% in AR. The result may indicate that in mixture of noises further investigation about modeling the error is required.

Execution times in this case are reported in Table III. A key observation is that F-IRNNLS is by an order of magnitude faster than RRC_L1 and RRC_L2. F-LR-IRNNLS was faster than RRCs and slower than LR^3. However, LR^3 achieved significantly lower identification rates.

| Dataset | Yale B | | AR |
	Accuracy	Time	Accuracy
SRC [6]	26.81%	1.04s	27.86%
CR-RLS [9]	14.73%	0.02s	28.29%
LR^3 [21]	44.62%	0.06s	35.43%
SDR-SLR [27]	32.75%	0.50s	40.00%
HQ [15]	42.20%	3.06s	42.00%
CESR [16]	23.96%	0.86s	40.86%
RRC_L1 [14]	43.08%	14.58s	47.00%
RRC_L2 [14]	41.54%	9.65s	38.14%
SSEC [19]	14.95%	1.72s	11.86%
Our F-IRNNLS	45.27%	1.45s	49.00%
Our F-LR-IRNNLS	**63.08**%	6.33s	**57.29**%

| Sessions | Session 2 | Session 3 | Session 4 | |
	Accuracy	Accuracy	Accuracy	Time
SRC [6]	95.48%	92.13%	95.71%	0.63s
CR-RLS [9]	95.30%	90.56%	94.46%	0.02s
LR^3 [21]	90.00%	83.44%	87.89%	0.04s
L12	93.92%	88.69%	92.91%	0.04s
HQ [15]	95.84%	94.63%	97.14%	2.21s
CESR [16]	94.64%	92.94%	96.06%	1.94s
RRC_L1 [14]	95.84%	95.13%	97.14%	6.88s
RRC_L2 [14]	97.11%	94.19%	97.37%	31.11s
SSEC [19]	86.75%	76.13%	83.09%	5.40s
Our F-IRNNLS	**97.29**%	**96.00**%	**98.00**%	1.05s
Our F-LR-IRNNLS	96.87%	94.13%	97.83%	1.83s

D. Identification Under Illumination

Experiments with variations in illumination were conducted on the Multi-PIE dataset. As in the block occlusion experiments, we used all 249 subjects in Session 1. As in [9], we used 14 frontal images with 14 illuminations[13] and neutral expression from Session 1 for training, and 10 frontal images[14] from Session 4 for testing. Identification rates are shown in Table IV for the various methods.

Our first observation is that all methods achieved high identification rates. Simple SRC approaches performed well while robust methods only slightly improved the results. The reason with respect to our method is that for illumination variations modeling the error image as low-rank does not hold in this case. Similar observations can also be deduced from results of the LR^3 method in this case.

With respect to time performance, our algorithm outperforms the previous robust methods. In particular the execution time in our approaches is around 1 second per test image while for RRC_L2 is around 30 seconds. Notice that although this was an experiment with a large training dictionary, our method retains very low running time.

E. Identification Under Pixel Corruptions

Experiments under pixel corruption were conducted on two datasets: Extended Yale B and AR.

As in [6], [14] we used the non-occluded faces of Subsets 1 and 2 of the Extended Yale B for training and Subset 3 for testing. Images were resized to 96 × 84 pixels. In the AR dataset, to make the experiment more challenging

---

[13]Illuminations 0,1,3,4,6,7,8,11,13,14,16,17,18,19.
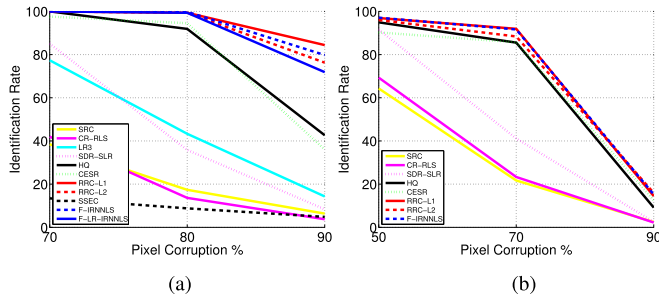[14]Illuminations 0,2,4,6,8,10,12,14,16,18.

(a)

(b)

Fig. 7. Identification Rates on the Extended Yale B and AR datasets under Pixel Corruptions. (a) Yale B dataset. (b) AR dataset.

we chose to use occluded training and testing images. AR has 100 different subjects and for each subject the 13 images from session 1 (7 non-occluded images, 3 images with sunglasses, and 3 images with scarves) were used for training and the 13 images from session 2 for testing. Images were resized to 60×43 pixels.

For each test image in both datasets, a percentage of randomly chosen pixels was corrupted by replacing those pixel values with independent and identically distributed values from a uniform distribution between [0, 255]. The percentage of corrupted pixels was varied between 50 percent and 90 percent. Identification rates are shown in Figure 7 for the various methods.

Figure 7(a) illustrates that the robust non-contiguous methods RRC_L1 and F-IRNNLS achieved the best performance with over 80% accuracy in 90% pixel corruption. Methods able to handle contiguous errors such as SSEC, LR^3 and F-LR-IRNNLS performed poorly. We attribute this to the fact that pixel corruption is not a contiguous variation and modeling the error to have contiguous structure was inadequate. To that extend, with the AR dataset we decided to report results only on methods that handle non-contiguous errors as shown in Figure 7(b). In this dataset the accuracy is low in 90% corruption for all methods. As explained earlier, this was a more challenging experiment with a large number of testing images consisting of faces with pixel corruption on top of occlusion.

With respect to execution time, our F-IRNNLS method outperformed RRC_L1 and RRC_L2 in both Yale B and AR datasets, as shown in Table I (columns 1 and 2). To emphasize the difference in performance, the execution time of F-IRNNLS in AR for 70% pixel corruption was about 2 seconds. In RRC_L1 and RRC_L2 was 29.84 and 10.93 seconds, respectively. Time performance of our algorithms was comparable to CESR and HQ (additive form) and worse than LR^3. However, these methods obtained lower identification rates.

F. Identification Under Unconstrained Environment and Occluded Testing and Training Samples

1) Identification Under Unconstrained Environment: Thus far, we have assumed that training samples are "clean" frontal aligned views and without large variations of the same identity. In an unconstrained environment this assumption does not hold and often face images of the same identity exhibit large variations in pose, illumination, expression and occlusion.
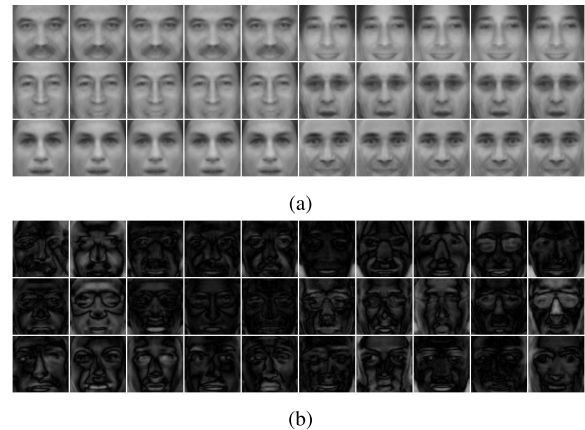

(a)

(b)

Fig. 8. Sample images from the LFW-a dataset and the SDR-SLR decomposition applied to the dataset. (a) Training sample images from the class-specific dictionary **A**. (b) Training sample images from the variation dictionary **B**.

Furthermore, testing images may not contain the same variations and occlusions as the training images.

To examine the robustness of our method on an unconstrained environment we evaluate its performance in the LFW database. The dataset contains images of 5,749 different subjects and in this work we used the LFW-a [47], which is an aligned version of LFW based on commercial face alignment software. We used the subjects that include no less than ten samples and we constructed a dataset with 158 subjects from LFW-a. For each subject, we randomly chose 5 samples for training (resulting in a dictionary of 790 faces) and 5 samples for testing. The images were resized to 90×90.

To deal with such environment we utilize the SDR-SLR algorithm as a pre-processing step as explained in Section II-F. Sample images from dictionaries **A** and **B** estimated on LFW-a dataset are illustrated in Figure 8(a) and (b). More specifically, in Figure 8(b) images cover variations which are not class-specific and are used to represent complex variations of the query. These variations may be represented by the component **B** using images that do not belong to the same identity of the test image. Any remaining variations that cannon be described by **B** are captured by the term **e**.

Identification rates for the LFW dataset are shown in Table V(left column) for the various methods. Our first observation is that the method SDR-SLR achieved better performance than our methods F-IRNNLS and F-LR-IRNNLS. This is expected as in this experiment faces exhibit uncontrolled variations such as pose and expression which are learnt from the training data utilized in SDR-SLR. However, when the method SDR-SLR is combined with our F-LR-IRNNLS method denoted as F-LR-IRNNLS (SLR), performance is improved. This is due to the fact that there are some remaining variations of the query such as occlusion that cannot be learnt from the training data. Our method is able to represent these remaining variations by modeling the representation error as low-rank and fitting to the error a distribution described by a tailored loss function.

2) Identification Under Occluded Testing and Training Samples: In order to further investigate the scenario where we are given corrupted testing and training data in a constrained environment this time,

TABLE V

IDENTIFICATION RATES (%) IN THE UNCONSTRAINED DATASET LFW-a IN
THE LEFT COLUMN. IN THE RIGHT COLUMN WE DEMONSTRATE
IDENTIFICATION RATES (%) IN MULTI-PIE DATASET
UNDER OCCLUDED TRAINING SAMPLES

Dataset	LFW-a	Multi-PIE
SRC [6]	68.35%	14.86%
CR-RLS [9]	64.68%	6.91%
LR^3 [21]	62.53%	6.00%
SDR-SLR [27]	76.20%	39.03%
HQ [15]	71.00%	17.83%
CESR [16]	59.87%	11.43%
RRC_L1 [14]	63.42%	23.20%
RRC_L2 [14]	72.53%	37.66%
SSEC [19]	32.41%	44.14%
Our F-IRNNLS	72.91%	23.89%
Our F-LR-IRNNLS	74.81%	44.06%
Our F-LR-IRNNLS (SLR)	**77.47%**	**47.71%**

We also conducted an experiment on the Multi-PIE dataset
in the scenario that corrupted testing and training data are
provided in a constrained environment. To simulate the cor-
rupted (occluded) training data we considered the same train-
ing and test sets as in Multi-PIE block-occlusion experiment
described above. In particular, we used the 6 frontal images
with 6 illuminations and neutral expression from Session 1 for
training. For each of the 249 subjects we randomly chose
half of the training images to be occluded. In each image
we replace a random block with the square baboon image
with occlusion chosen randomly from 30% to 60%. We chose
the 10 frontal images from Session 4 for testing. In each test
image, we replace a random block with the square baboon
image and the occlusion was randomly chosen from 30% to
60%. Identification rates are shown in Table V(right column).

From the results we observe that as expected SDR-SLR
perform way better than the SRC and CR-RLS. However, not
all block-occlusions are sufficiently covered by **B**. This is due
to the fact that occlusion appears in random places and sizes
in the query as well as in training data. Therefore, it might
be very unlikely that the occlusion on the query and training
images will be of the same type. Thus, when SDR-SLR
method is combined with the F-LR-IRNNLS algorithm the
best accuracy is reported since the proposed modeling of the
error term is robust to handle occlusions of the query. Finally,
we observe that our F-LR-IRNNLS algorithm outperform the
other approaches even when SDR-SLR is not utilized. The
reason may be that our approach chooses the non-occluded
training samples to represent the query since occlusion is
effectively captured by the representation error image **e**.

3) Comparison With Inductive Methods: In this section we
compare our method with two inductive methods, namely
IRPCA [29] and IDNMD [28]. Both methods are able to
handle new data meaning that given a new sample an under-
lying learnt projection matrix **K** can be used to efficiently
remove corruptions and occlusions. A test face is recovered
from occlusions by computing **Ky**. Then, the "clean" test face
Ky is provided as an input to a classifier to identify the subject.
In this work we use the SRC classifier [6] for these methods
in order to make direct comparisons with our method.[15] Also,

[15]We report results with names IRPCA and IDNMD to denote IRPCA combined with SRC and IDNMD combined with SRC classifier, respectively.

in this case we do not perform any pre-processing step such
as SDR-SLR to clean the training data for identification in our
and other methods to make fair comparisons.

There are two main differences between IRPCA, IDNMD
and our method; i) Our method does not require the same
class of occlusions to be present in the training and test
data while inductive methods do to perform well. ii) Our
method describes the error image by using two metrics, namely
weighting and nuclear norms.

To examine the robustness of our method against the induc-
tive methods we conducted the following experiments:

First we investigate the performance on the YaleB dataset.
We chose Subsets 1 and 2 of Extended Yale B for training
and Subset 3 for testing. As in [28], 15% of training images
were occluded with the baboon image in random places.
Three occlusion rates, 30%, 50% and 60% were considered.
For testing images we considered two different scenarios;
1) test images were occluded with the baboon image randomly
placed (different random places than in the training set) 2) test
images were occluded with the dog image (training samples
were still occluded with the baboon image) to explore whether
inductive methods can handle occlusions that are not present
in the training procedure. From the results in Figure 9 (table)
for YaleB dataset we deduce the following findings:

As expected, inductive methods combined with SRC per-
form better than standard SRC methods when the same
occlusion type (baboon) is present in training and testing
images (second and third columns in Figure 9). For, low-
percentage of occlusion (30%) all methods perform well.
However, when occlusion is different in training and testing
images (baboon/dog) as shown in Figure 9 (fourth column),
SRC performs very similarly to inductive methods. This
demonstrates that when a specific occlusion object is involved
in the test but not the training image, IRPCA and IDNMD
may not perform better than SRC.

This is further explained in Figure 9 (figure). First,
we observe from this figure that when a low-percentage
occlusion is present (30%), inductive methods can recover the
test face (see **Ky**) well, and as a consequence can perform well
in identification. However, when the occlusion percentage is
50%, the face recovery by the inductive methods becomes
more noisy. The face becomes a lot more noisy, when the
testing occlusion (dog) is different from the occlusion used
during training (baboon). From the visualizations in Figure 9
we also observe that in our method the occluded object is well
captured in all cases (see weight maps).

IDNMD performs slightly better than IRPCA since it mod-
els each occlusion-induced image as low-rank during training.
Similar results were reported in [28]. For a more effective
comparison, we also report results in Figure 9 for LR^3 [21],
since LR^3 describes the error as low-rank too. It was expected
that IDNMD and LR^3 would have similar performance
when training and testing images have similar occlusions.
However, when occlusions are different in training and testing
images, LR^3 performs better. Our method outperforms signif-
icantly both IDNMD and LR^3 methods.

The second experiment was conducted on the AR database.
We chose the 7 neutral AR images per subject from session 1

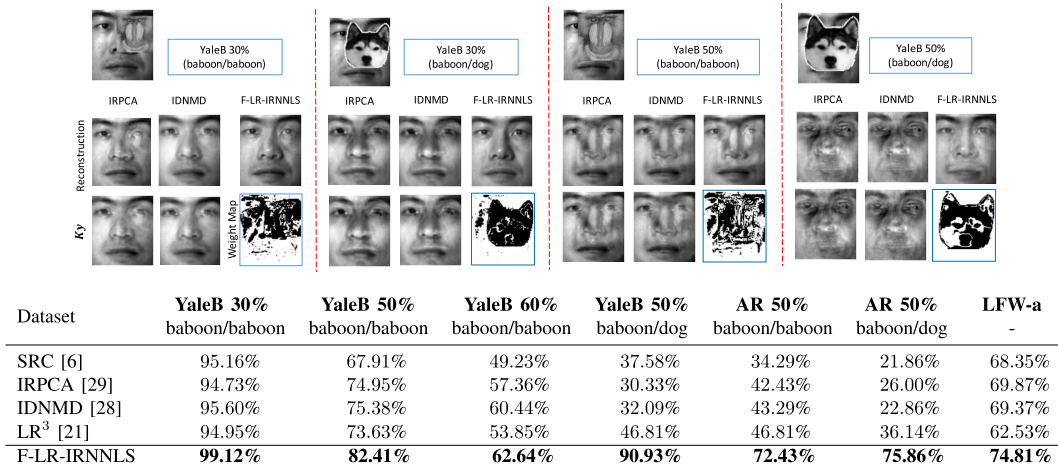| Dataset	YaleB 30% baboon/baboon	YaleB 50% baboon/baboon	YaleB 60% baboon/baboon	YaleB 50% baboon/dog	AR 50% baboon/baboon	AR 50% baboon/dog	LFW-a -
SRC [6]	95.16%	67.91%	49.23%	37.58%	34.29%	21.86%	68.35%
IRPCA [29]	94.73%	74.95%	57.36%	30.33%	42.43%	26.00%	69.87%
IDNMD [28]	95.60%	75.38%	60.44%	32.09%	43.29%	22.86%	69.37%
LR^3 [21]	94.95%	73.63%	53.85%	46.81%	46.81%	36.14%	62.53%
F-LR-IRNNLS	**99.12%**	**82.41%**	**62.64%**	**90.93%**	**72.43%**	**75.86%**	**74.81%**

Fig. 9. Comparisons between our method and inductive approaches. Figure: recovery results of three methods. The first row shows reconstructed faces **Ta** provided by the classifier while the second row visualizes the recovered test sample **Ky** from the inductive methods and weight map estimations of our method. Table: of identification rates of the methods. We denote the different training and testing occlusion scenarios as following; "baboon/baboon" is that training and testing samples are occluded with the baboon image. "baboon/dog" is that training samples are occluded with the baboon image and testing images are occluded with the dog image.

for training, where 15% of those were occluded with the baboon image in random places (occlusion percentage is 50%). The 7 neutral images per subject from session 2 were chosen for testing. In each test image, we replaced a random block with the dog image. Identification rates are shown in Figure 9 (fifth and sixth columns). Results are consistent with the YaleB findings. SRC, IDNMD and IRPCA perform very similarly in fifth column of Figure 9 (since different occlusion is presence in training and testing faces) although the identification rates are really low. Our method again outperforms previous methods significantly.

Finally, we investigate the performance of the algorithms in an unconstrained environment utilizing the LFW database. The experimental settings are similar to the experiment showed in Table V. As with previous experiments, from the results we observe that our proposed method outperforms previous approaches as the error image is described more effectively with the two metrics, weighting and nuclear norms.

G. Identification Under Few Training Samples

To examine the robustness of our method under few training samples per subject we conducted experiments on the Multi-PIE database [44] with uncorrupted training data. As in the experiments under occlusion, we used 6 frontal images with 6 illuminations and neutral expression from Session 1 for training, and 10 frontal images[16] from Session 4 for testing. Then, we randomly selected 2 or 4 samples per subject to perform experiments under fewer training examples. In each test image, we replace a random block with the square baboon image, where each block randomly covered between 30% and 60% of the image area.

From the results in Table VI, we deduce that as expected all methods perform worse when fewer training samples are available. However, our F-LR-IRNNLS algorithm achieved the best performance in 2 and 4 samples. SSEC achieved

[16]Illuminations 0,2,4,6,8,10,12,14,16,18.

TABLE VI
IDENTIFICATION RATES ON THE MULTI-PIE UNDER BLOCK
OCCLUSION AND FEW TRAINING SAMPLES

Samples	2 Samples	4 Samples	6 Samples
SRC [6]	9.37%	13.31%	22.47%
CR-RLS [9]	5.43%	8.40%	18.34%
LR^3 [21]	28.51%	29.71%	37.37%
SDR-SLR [27]	10.17%	21.77%	33.83%
HQ [15]	22.91%	38.23%	49.31%
CESR [16]	11.54%	25.31%	42.11%
RRC_L1 [14]	22.11%	43.66%	64.46%
RRC_L2 [14]	28.86%	44.51%	52.86%
SSEC [19]	28.23%	47.77%	65.43%
Our F-IRNNLS	22.00%	48.86%	64.74%
Our F-LR-IRNNLS	**34.46%**	**52.69%**	**71.09%**

the second best performance while LR^3 performed similarly in 2 samples with the SSEC and RRC_L2. As expected, the SDR-SLR method performed significantly better than SRC and CR-RLS since the additional intra-class variation dictionary alleviates the issue with the limited samples per person. However, since the additional variation dictionary does not capture the unknown occlusion appearing on the test images, SDR-SLR method had lower performance than our F-LR-IRNNLS by a big margin.

Overall our proposed method achieved higher or competitive identification rates across all experiments. In addition, our method incurred lower computational costs than the previous algorithms. In some cases the execution time was lower by an order of magnitude than the second best algorithm overall (e.g., RRC_L1, RRC_L2). Furthermore, our F-LR-IRNNLS and F-LR-IRNNLS (SLR) algorithm outperformed with respect to identification rates all previous robust sparse representation-based methods on images with contiguous errors in all scenarios (uncorrupted training data, occluded testing and training data and unconstrained environment).

H. Weight Map Estimations

Figure 10 shows the estimated weight maps between RRC_L1, RRC_L2 and our F-LR-IRNNLS in experiments with occlusions. Black values (close to zero) represent detected outliers by the various methods. We observe that

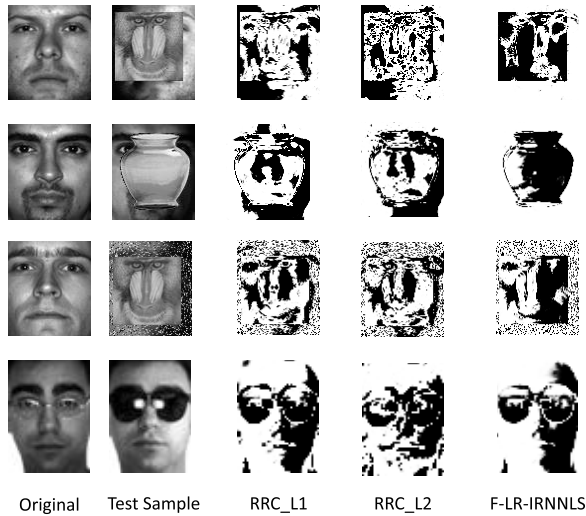Original   Test Sample   RRC_L1   RRC_L2   F-LR-IRNNLS

Fig. 10. Estimated weight maps for three iterative reweighted coding methods. The weight maps of our F-LR-IRNNLS method capture the outlier object of interest more accurately. With the other methods, a number of inlier pixels are detected as outliers.
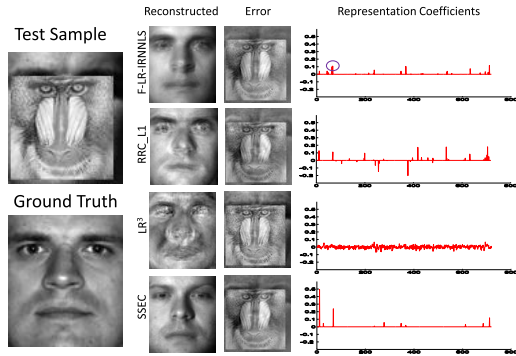

Fig. 11. Reconstruction results in 70% block occlusion.

F-LR-IRNNLS detected the outlier objects more effectively than the other methods. Most of the black regions in the weight maps are concentrated on the occluded area. In particular, we observe in Figure 10 (first row) that for the F-LR-IRNNLS method small weights are only assigned to the occluded (baboon) region as desired. On the other hand, the weight maps of RRC_L1 and RRC_L2 are not as accurate since outliers were detected in important pixels of the face. The reason is that with these methods there is no spatial correlation constraint between the weights. Similar conclusions can be drawn from all other examples in Figure 10.

I. Face Reconstruction Results

Figure 11 illustrates the face reconstruction results and the associated representation coefficients by the four methods. F-LR-IRNNLS and SSEC had the best reconstruction performance. The reconstructed face by LR^3 was poor mainly due to the choice of the regularizer for the representation coefficients (ℓ_2 norm). Similar reconstruction performance for the LR^3 method was encountered in almost all of our conducted experiments.

More reconstruction results for various methods are presented in Figure 12. With mixture noise, our F-LR-IRNNLS achieved the best performance which demonstrates that our modeling was more effective in this case than the other
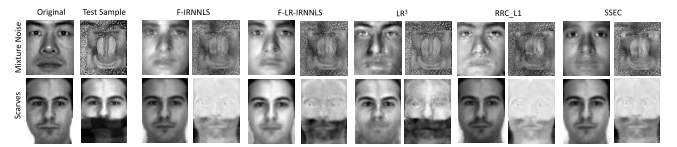

Fig. 12. Reconstruction results for various methods: The left image demonstrates the reconstructed face and the right image shows the estimated error for each of the methods tested.
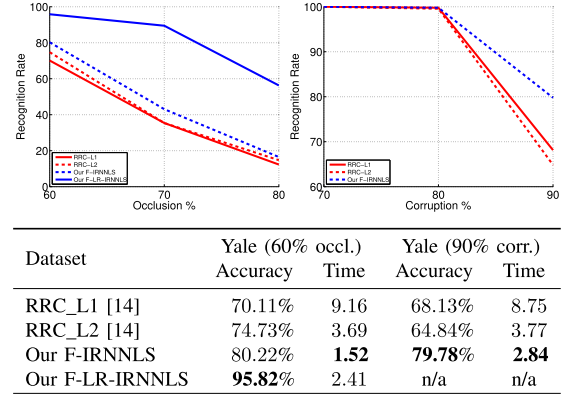

Dataset	Yale (60% occl.)		Yale (90% corr.)	
	Accuracy	Time	Accuracy	Time
RRC_L1 [14]	70.11%	9.16	68.13%	8.75
RRC_L2 [14]	74.73%	3.69	64.84%	3.77
Our F-IRNNLS	80.22%	**1.52**	**79.78**%	**2.84**
Our F-LR-IRNNLS	**95.82**%	2.41	n/a	n/a

Fig. 13. Performance of RRC ($t = 25$ iterations) and our method ($t = 100$ iterations) under Block Occlusion and Pixel Corruption.

methods. Face reconstruction was adequate for the case with scarves occlusion for all methods which validates the identification rates reported in Table II.

J. Time Performance Between Our Method and RRC

In this experiment we evaluate the identification rates in RRC [14] for the case where the maximum reweighted iterations $t = 25$. In other words, we want to investigate the performance degradation of RRC by keeping its execution time similar to our method. In our method we kept $t = 100$. As shown in Figure 13, we observe that the computational time for RRC is now more competitive (although still higher than our method). However, the identification rates dropped significantly in both pixel corruption and block occlusion cases for RRC with $t = 25$.

K. Regularization of the Coefficients

In Table VII we report performance comparisons of our method with different regularizations of the representation coefficients. Our main take away from the results is that sparsity is overall slightly better than the two other regularizers (non-negative and ℓ_2) in terms of identification rates. However, the non-negative regularizer provided a better balance between computational cost and identification rates.

Finally, there is significant difference in time performance between the RRC and our method regardless of the regularization of the coefficients. The efficiency of our method gives rise to robust face recognition systems for which computational time is a critical factor.

IV. CONCLUSIONS

In this work we proposed a method to describe contiguous errors effectively based on two characteristics. The first fits to the errors a distribution described by a tailored loss function. The second describes the error image as structural (low-rank).

TABLE VII
COMPARISON OF IDENTIFICATION RATES AND TIMES BETWEEN RRCs AND OUR ALGORITHMS
UNDER DIFFERENT REGULARIZATIONS OF THE COEFFICIENTS

Dataset	Illumination		Occlusion 60% (Y)		Occlusion 50% (A)		Corruption 70% (A)		Mixture Noise (Y)	
	Accuracy	Time	Accuracy	Time	Accuracy	Time	Accuracy	Time	Accuracy	Time
RRC_L1 [14]	97.14%	6.88s	78.24%	12.41s	59.57%	3.95s	**91.92%**	29.84s	43.08%	14.58s
RRC_L2 [14]	97.37%	31.11s	81.54%	10.52s	55.57%	1.96s	88.38%	10.93s	41.54%	9.65s
F-IRLS	**98.06%**	1.32s	83.96%	1.74s	55.86%	0.39s	87.92%	1.17s	47.91%	1.70s
F-IRSC	97.77%	3.11s	79.78%	1.62s	60.43%	0.51s	90.08%	2.16s	45.05%	1.70s
F-IRNNLS	98.00%	1.05s	80.22%	1.52s	62.29%	0.45s	91.62%	2.25s	45.27%	1.45s
F-LR-IRLS	96.74%	1.36s	**99.12%**	2.81s	71.71%	0.56s	n/a	n/a	57.36%	9.38s
F-LR-IRSC	97.77%	2.79s	98.02%	2.50s	**74.57%**	0.68s	n/a	n/a	**66.81%**	7.96s
F-LR-IRNNLS	97.83%	1.83s	95.82%	2.41s	74.43%	0.57s	n/a	n/a	63.08%	6.33s

Our approach is computationally efficient due to the utilization of ADMM. The extensive experimental results support the claim that the proposed modeling of the error term can be beneficial and more robust than previous state-of-the-art methods to handle occlusions across a multitude of databases and in different scenarios. A special case of our algorithm leads to the robust representation problem which is used to solve cases with non-contiguous errors. We showed that our fast iterative algorithm was in some cases faster by an order of magnitude than the existing approaches.

REFERENCES

[1] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Comput. Surv.*, vol. 35, no. 4, pp. 399–458, Dec. 2003.

[2] Y. Taigman, M. Yang, M. Ranzato, and L. Wolf, "Web-scale training for face identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 2746–2754.

[3] M. Turk and A. Pentland, "Eigenfaces for recognition," *J. Cognit. Neurosci.*, vol. 3, no. 1, pp. 71–86, Jan. 1991.

[4] P. N. Belhumeur, J. P. Hespanha, and D. Kriegman, "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 7, pp. 711–720, Jul. 1997.

[5] B. Cheng, J. Yang, S. Yan, Y. Fu, and T. S. Huang, "Learning with ℓ^1-graph for image analysis," *IEEE Trans. Image Process.*, vol. 19, no. 4, pp. 858–866, Apr. 2010.

[6] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 31, no. 2, pp. 210–227, Feb. 2009.

[7] J. Wright, Y. Ma, J. Mairal, G. Sapiro, T. S. Huang, and S. Yan, "Sparse representation for computer vision and pattern recognition," *Proc. IEEE*, vol. 98, no. 6, pp. 1031–1044, Jun. 2010.

[8] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, H. Mobahi, and Y. Ma, "Toward a practical face recognition system: Robust alignment and illumination by sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 2, pp. 372–386, Feb. 2012.

[9] L. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: Which helps face recognition?" in *Proc. IEEE Int. Conf. Comput. Vis.*, Barcelona, Spain, Nov. 2011, pp. 471–478.

[10] Q. Shi, A. Eriksson, A. van den Hengel, and C. Shen, "Is face recognition really a compressive sensing problem?" in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2011, pp. 553–560.

[11] J. Lai and X. Jiang, "Classwise sparse and collaborative patch representation for face recognition," *IEEE Trans. Image Process.*, vol. 25, no. 7, pp. 3261–3272, Jul. 2016.

[12] J. Lai and X. Jiang, "Modular weighted global sparse representation for robust face recognition," *IEEE Signal Process. Lett.*, vol. 19, no. 9, pp. 571–574, Sep. 2012.

[13] M. Yang, L. Zhang, J. Yang, and D. Zhang, "Robust sparse coding for face recognition," in *Proc. 24th IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2011, pp. 625–632.

[14] M. Yang, L. Zhang, J. Yang, and D. Zhang, "Regularized robust coding for face recognition," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 1753–1766, May 2013.

[15] R. He, W.-S. Zheng, T. Tan, and Z. Sun, "Half-quadratic-based iterative minimization for robust sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 2, pp. 261–275, Feb. 2014.

[16] R. He, W.-S. Zheng, and B.-G. Hu, "Maximum correntropy criterion for robust face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1561–1576, Aug. 2011.

[17] R. He, W.-S. Zheng, B.-G. Hu, and X.-W. Kong, "Two-stage nonnegative sparse representation for large-scale face recognition," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 24, no. 1, pp. 35–46, Jan. 2013.

[18] P. Huber, *Robust Statistics*. New York, NY, USA: Wiley, 1981.

[19] X.-X. Li, D.-Q. Dai, X.-F. Zhang, and C.-X. Ren, "Structured sparse error coding for face recognition with occlusion," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 1889–1900, May 2013.

[20] Z. Zhou, A. Wagner, H. Mobahi, J. Wright, and Y. Ma, "Face recognition with contiguous occlusion using Markov random fields," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Sep./Oct. 2009, pp. 1050–1057.

[21] J. Qian, J. Yang, F. Zhang, and Z. Lin, "Robust low-rank regularized regression for face recognition with occlusion," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2014, pp. 21–26.

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

[23] Z. Ding, S. Suh, J.-J. Han, C. Choi, and Y. Fu, "Discriminative low-rank metric learning for face recognition," in *Proc. 11th IEEE Int. Conf. Workshops Autom. Face Gesture Recognit. (FG)*, vol. 1. May 2015, pp. 1–6.

[24] S. Li and Y. Fu, "Low-rank coding with b-matching constraint for semi-supervised classification," in *Proc. 23rd Int. Joint Conf. Artif. Intell.*, 2013, pp. 1472–1478.

[25] L. Li, S. Li, and Y. Fu, "Learning low-rank and discriminative dictionary for image classification," *Image Vis. Comput.*, vol. 32, no. 10, pp. 814–823, 2014.

[26] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 3, p. 11, May 2011.

[27] X. Jiang and J. Lai, "Sparse and dense hybrid representation via dictionary decomposition for face recognition," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 5, pp. 1067–1079, May 2015.

[28] F. Zhang, J. Yang, Y. Tai, and J. Tang, "Double nuclear norm-based matrix decomposition for occluded image recovery and background modeling," *IEEE Trans. Image Process.*, vol. 24, no. 6, pp. 1956–1966, Jun. 2015.

[29] B.-K. Bao, G. Liu, C. Xu, and S. Yan, "Inductive robust principal component analysis," *IEEE Trans. Image Process.*, vol. 21, no. 8, pp. 3794–3800, Aug. 2012.

[30] S. D. Babacan, R. Molina, M. N. Do, and A. K. Katsaggelos, "Bayesian blind deconvolution with general sparse image priors," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2012, pp. 341–355.

[31] J. A. Palmer, K. Kreutz-Delgado, and S. Makeig, "Strong sub- and super-Gaussianity," in *Latent Variable Analysis and Signal Separation* (Lecture Notes in Computer Science), vol. 6365, V. Vigneron, V. Zarzoso, E. Moreau, R. Gribonval, and E. Vincent, Eds. Berlin, Germany: Springer, 2010, pp. 303–310.

[32] R. T. Rockafellar, *Convex Analysis*. Princeton, NJ, USA: Princeton Univ. Press, 1996.

[33] X. Zhou, R. Molina, F. Zhou, and A. K. Katsaggelos, "Fast iteratively reweighted least squares for l_p regularized image deconvolution and reconstruction," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2014, pp. 1783–1787.

[34] J. R. Shewchuk, "An introduction to the conjugate gradient method without the agonizing pain," School Comput. Sci., Carnegie Mellon Univ., Pittsburgh, PA, USA, Tech. Rep., 1994. [Online]. Available: http://www.ncstrl.org:8900/ncstrl/servlet/search?formname=detail&id=oai%3Ancstrlh%3Acmucs%3ACMU%2F%2FCS-94-125

[35] A. Yang, A. Ganesh, S. Sastry, and Y. Ma, "Fast ℓ_1-minimization algorithms and an application in robust face recognition: A review," Dept. Elect. Eng. Comput. Sci., Univ. California, Berkeley, Berkeley, CA, USA, Tech. Rep. UCB/EECS-2010-13, Feb. 2010.

[36] S. D. Babacan, R. Molina, and A. K. Katsaggelos, "Bayesian compressive sensing using Laplace priors," *IEEE Trans. Image Process.*, vol. 19, no. 1, pp. 53–63, Jan. 2010.

[37] C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems* (Classics in Applied Mathematics), vol. 15. Philadelphia, PA, USA: SIAM, 1995.

[38] C.-F. Chen, C.-P. Wei, and Y.-C. Wang, "Low-rank matrix recovery with structural incoherence for robust face recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2012, pp. 2618–2625.

[39] J. Chen and Z. Yi, "Sparse representation for face recognition by discriminative low-rank matrix recovery," *J. Vis. Commun. Image Represent.*, vol. 25, no. 5, pp. 763–773, Jul. 2014.

[40] Y. Zhang, Z. Jiang, and L. S. Davis, "Learning structured low-rank representations for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 676–683.

[41] G. Liu, Z. Lin, S. Yan, J. Sun, Y. Yu, and Y. Ma, "Robust recovery of subspace structures by low-rank representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 1, pp. 171–184, Jan. 2013.

[42] A. Martínez and R. Benavente, "The AR face database," Comput. Vis. Center, Bellatera, Huntington Beach, CA, USA, Tech. Rep. 24, Jun. 1998.

[43] A. S. Georghiades, P. N. Belhumeur, and D. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 643–660, Jun. 2001.

[44] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-PIE," in *Proc. 8th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, Sep. 2008, pp. 1–8.

[45] G. B. Huang, M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled faces in the wild: A database for studying face recognition in unconstrained environments," Univ. Massachusetts Amherst, Amherst, MA, USA, Tech. Rep. UM-CS-2014-003, Oct. 2007.

[46] D. L. Donoho and Y. Tsaig, "Fast solution of ℓ_1-norm minimization problems when the solution may be sparse," *IEEE Trans. Inf. Theory*, vol. 54, no. 11, pp. 4789–4812, Nov. 2008.

[47] L. Wolf, T. Hassner, and Y. Taigman, "Similarity scores based on background samples," in *Proc. 9th Asian Conf. Comput. Vis. (ACCV)*, Berlin, Germany, 2010, pp. 88–97.

Michael Iliadis received the B.S. degree in digital systems from the University of Piraeus, Greece, in 2008, the M.S. degree in computer science from the University of Bath, U.K., in 2009, and the Ph.D. degree in computer science from Northwestern University, IL, USA, in 2016. He was with TCL Research America as a Research Intern in 2013 and 2014. In 2016, he joined the Sony US Research Center, where he is currently a Research Scientist. He holds five U.S. patents. His current research interests include machine (deep) learning, sparse modeling, compressive sensing, super-resolution, face recognition, and semantic segmentation.

Haohong Wang He received the Ph.D. degree from Northwestern University, Evanston, USA. He held technical and management positions with AT&T, Catapult, Qualcomm, Marvell, and Cisco. He is currently the General Manager of TCL Research America, oversees the Research and Development activities in North America for TCL, a Global Leader in consumer electronics with over 75,000 employees worldwide and ranks the third in TV and the fifth in mobile phone business. He has authored or co-authored five books. He holds over 70 U.S. patents and pending applications. His current research areas include multimedia computing and mobile systems and services.

He was a recipient of the IEEE MMTC Distinguished Service Award. He is a Recognized Leader with Multimedia and Communication Societies. He was the Chair of the IEEE Multimedia Communications Technical Committee, and the IEEE Technical Committee on Human Perception and Multimedia Computing. He was the Chair of the IEEE GLOBECOM 2010, the IEEE ICME 2011, the IEEE VCIP 2014, the IEEE ISM 2015, and the ACM Multimedia 2017. He is the Steering Committee Chair of the IEEE ICME. He has been the Editor-in-Chief of the *Journal of Communications*, and the Vice-President of the *Asia-Pacific Signal and Information Processing* Association.

Rafael Molina received the degree in mathematics (statistics) and the Ph.D. degree in optimal design in linear models from the University of Granada, Granada, Spain, in 1979 and 1983, respectively. He was a Professor of Computer Science and Artificial Intelligence with the University of Granada in 2000. He was the former Dean of the Computer Engineering School with the University of Granada from 1992 to 2002, where he was the Head of the Computer Science and Artificial Intelligence Department from 2005 to 2007. His research interest focuses mainly on using Bayesian modeling and inference in problems, such as image restoration (applications to astronomy and medicine), superresolution of images and video, blind deconvolution, computational photography, source recovery in medicine, compressive sensing, low-rank matrix decomposition, active learning, fusion, and classification.

Dr. Molina was a recipient of an IEEE International Conference on Image Processing Paper Award in 2007 and the ISPA Best Paper Award in 2009. He is a co-author of a paper awarded the runner-Up Prize at Reception for early-stage researchers at the House of Commons. He serves as an Associate Editor of the *Applied Signal Processing* from 2005 to 2007. He has been served as an Associate Editor of the IEEE TRANSACTIONS ON IMAGE PROCESSING since 2010, and the *Progress in Artificial Intelligence* since 2011, and has been an Area Editor of the *Digital Signal Processing* since 2011.

Aggelos K. Katsaggelos (F'98) received the Diploma degree in electrical and mechanical engineering from the Aristotelian University of Thessaloniki, Greece, in 1979, and the M.S. and Ph.D. degrees in electrical engineering from the Georgia Institute of Technology, in 1981 and 1985, respectively. He was the holder of the Ameritech Chair of Information Technology and the AT&T Chair and the Co-Founder and the Director of the Motorola Center for Seamless Communications. In 1985, he joined the Department of Electrical Engineering and Computer Science, Northwestern University, where he is currently a Professor and holder of the Joseph Cummings Chair. He is a member of the Academic Staff, NorthShore University Health System, and a Faculty Member with the Department of Linguistics. He has an appointment with the Argonne National Laboratory. He has authored extensively in the areas of multimedia processing and communications (over 250 journal papers, 500 conference papers and 40 book chapters). He holds 26 international patents. He has authored or co-authored the book *Rate-Distortion Based Video Compression* (Kluwer, 1997), *Super-Resolution for Images and Video* (Claypool, 2007), *Joint Source-Channel Video Transmission* (Claypool, 2007), *Machine Learning Refined: Foundations, Algorithms, and Applications* (Cambridge University Press, 2016), and *The Essentials of Sparse Modeling and Optimization* (Springer, 2017, forthcoming). He has supervised 56 Ph.D. dissertations so far.

Dr. Katsaggelos is a fellow of the SPIE in 2009. He was a recipient of the IEEE Third Millennium Medal in 2000, the IEEE Signal Processing Society Meritorious Service Award in 2001, the IEEE Signal Processing Society Technical Achievement Award in 2010, the IEEE Signal Processing Society Best Paper Award in 2001, the IEEE ICME Paper Award in 2006, the IEEE ICIP Paper Award in 2007, the ISPA Paper Award in 2009, and the EUSIPCO Paper Award in 2013. He was the Editor-in-Chief of the *IEEE Signal Processing Magazine* from 1997 to 2002, a BOG Member of the *IEEE Signal Processing* Society from 1999 to 2001, a member of the Publication Board of the IEEE Proceedings from 2003 to 2007, and he is currently a member of the Awards Board of the Signal Processing Society. He was a Distinguished Lecturer of the IEEE Signal Processing Society from 2007 to 2008.