MULTI-MODEL ROBUST ERROR CORRECTION FOR FACE RECOGNITION

Michael Iliadis[⋆], *Leonidas Spinoulas*[⋆], *Albert S. Berahas*[†], *Haohong Wang*[‡], *Aggelos K. Katsaggelos*[⋆]

[⋆] Dept. of Electrical Engineering and Comp. Sc., Northwestern University, Evanston, IL 60208, USA
[†] Dept. of Eng. Sciences and Appl. Mathematics, Northwestern University, Evanston, IL 60208, USA
[‡] TCL Research America, San Jose, CA 95134, USA

ABSTRACT

In this work we present a general framework for robust error estimation in face recognition. The proposed formulation allows the simultaneous use of various loss functions for modeling the residual in face images, which usually follows non-standard distributions, depending on the image capturing conditions. Our method extends the current vast literature offering flexibility in the selection of the residual modeling characteristics but, at the same time, considering many existing algorithms as special cases. As such, it proves robust for a range of error inducing factors, such as, varying illumination, occlusion, pixel corruption, disguise or their combinations. Extensive simulations document the superiority of selecting multiple models for representing the noise term in face recognition problems, allowing the algorithm to achieve near-optimal performance in most of the tested face databases. Finally, the multi-model residual representation offers useful insights into understanding how different noise types affect face recognition rates.

Index Terms— face recognition, sparse representation, robust representation, error correction

1. INTRODUCTION

Robust error estimation for sparse representation-based classification has been recently investigated in Face Recognition (FR) given frontal views with varying illumination and occlusion as well as disguise [1, 2, 3, 4]. Previous sparse representation-based classifiers solve a regularized regression model, with the coefficients being either sparse or non-sparse, under the assumption that a face can be represented as a linear combination of training faces, as in Sparse Representation-Based Classification (SRC) [5].

Let each face image be of size $j \times k = d$ and $\mathbf{y} \in \mathbb{R}^d$ denote the face test sample. Let $\mathbf{T} = [T_1, \ldots, T_i, \ldots, T_c] \in \mathbb{R}^{d \times n}$ denote a matrix (dictionary) with the set of samples of c subjects stacked in columns. $T_i \in \mathbb{R}^{d \times n_i}$ denotes the set of samples of the i^{th} subject, such that, $\sum_i n_i = n$.

In SRC the regression model is formulated as,

$$\mathbf{y} = \mathbf{T}\mathbf{a} + \mathbf{n}, \tag{1}$$

where $\mathbf{n} \in \mathbb{R}^d$ is the dense error and $\mathbf{a} \in \mathbb{R}^n$ is the coefficient vector to be estimated. In the SRC formulation of model (1), $\mathbf{a}$ is a sparse vector with nonzero elements corresponding to a few samples in $\mathbf{T}$. The coefficients of $\mathbf{a}$ can be found by solving the optimization problem,

$$\min_{\mathbf{a}} \|\mathbf{y} - \mathbf{T}\mathbf{a}\|_2^2 + \lambda \|\mathbf{a}\|_1, \tag{2}$$

where $\lambda > 0$. In order to enforce sparsity, ℓ_1 optimization algorithms [6, 7] can be employed. The final step of the SRC method identifies the subject by selecting the face class that yields the minimum reconstruction error.

Robust error estimation methods [1, 2, 3, 4] have been proposed in order to make regression models more robust to outliers (e.g., occlusions) than the SRC algorithm. These methods introduce a penalty function to the regression model to characterize the residual error which is often not normally distributed [2]. A prior assumption of the distribution of the error is needed in order to estimate it jointly with the representation coefficients. For example, [8] models $\mathbf{y} - \mathbf{T}\mathbf{a}$ as a Laplace distribution using the ℓ_1-norm. In a more general formulation, other methods proposed to use penalty functions based on M-Estimators [3].

Furthermore, researchers, have investigated the use of the whole structure of the residual in order to characterize contiguous occlusions [9, 10]. In [9] the error image is assumed to be low-rank and the residual is modeled using the nuclear norm. In order to jointly handle the pixel-level sparse noise and image-level structural noise, the authors in [10] propose a two norm regression model to characterize the residual.

The regularization of the residual with one prior distribution might not be sufficient to characterize the residual error [10]. Therefore, in this work we propose a robust multi-model representation-based classifier in which we allow the residual to be described by two penalty functions. In contrast to [10] which combined ℓ_1 and nuclear norms, we propose a more general framework, where various residual metrics, proven to be robust [3] in FR, are utilized. We investigate the combination of error distributions, e.g., Huber [11] and Negative Gaussian [3], or Nuclear [9] and Negative Gaussian [3], to model the error residual for different types of occlusions and variations in facial expressions.

2. PROPOSED METHOD

Various existing face recognition schemes have proven robust in modeling the noise characteristics for a range of image corruption types. Nevertheless, most presented approaches are mainly tested on cases where the noise fits their residual model assumptions. For example, the authors in [9] consider a low-rank regularizer for the noise term, hence being successful at recognizing faces under block-occlusion. At the same time, the authors in [10] combine low-rankness and sparsity in order to extend their recognition rates in cases such as random pixel corruptions with block-occlusion. However, when the percentage of corrupted pixels increases significantly, failing to satisfy the low-rank assumption, recognition rates can reduce dramatically.

We aim at presenting a general framework for handling most types of query image corruptions by suggesting a multi-model representation of the residual term. In this way, we relax the modeling

constraints of the residual and enable extra degrees of freedom so that the noise term can be described with non-standard penalty functions in order to handle mixed variations in the captured scene.

Thus, we propose a formulation where two penalty functions are incorporated for better characterizing the residual term $\mathbf{y}-\mathbf{Ta}$. The proposed function to be minimized is written as,

$$J(\mathbf{a}) = \Phi_1(\mathbf{y}-\mathbf{Ta}) + \Phi_2(\mathbf{y}-\mathbf{Ta}) + \lambda_\vartheta \vartheta(\mathbf{a}), \tag{3}$$

where function Φ_k represents a specific potential loss function and is selected from a variety of candidates [3]. The function $\vartheta(\cdot)$, used to regularize the coefficients $\mathbf{a}$, belongs to the class of ℓ_p norms $\left(\|\mathbf{a}^p\|^{\frac{1}{p}}\right)$ and $\lambda_\vartheta > 0$. The formulation above allows for the compact and general representation of many existing algorithms in the face recognition literature. For example:

- For $\Phi_1(\mathbf{x}) = ||\mathbf{x}||_2^2$, $\Phi_2(\mathbf{x}) = \lambda_* ||T_M(\mathbf{s})||_*$ and $\vartheta(\mathbf{a}) = \|\mathbf{a}\|_2^2$ with T_M being an operator that transforms its vector argument into a matrix of appropriate size, $||\cdot||_*$ being the nuclear norm and $\lambda_* > 0$, it is the low-rank regularized regression (LR^3) [9] which is formulated as,

$$\min_{\mathbf{a}} \|\mathbf{y}-\mathbf{Ta}\|_2^2 + \lambda_* \left\|T_M(\mathbf{y}-\mathbf{Ta})\right\|_* + \lambda_\vartheta \|\mathbf{a}\|_2^2. \tag{4}$$

- For $\Phi_1(\mathbf{x}) = ||\mathbf{x}||_1$, $\Phi_2(\mathbf{x}) = \lambda_* ||T_M(\mathbf{s})||_*$ and $\vartheta(\mathbf{a}) = \|\mathbf{a}\|_2^2$, it is the nuclear-$\ell_1$ regression NL_1R [10] which is formulated as,

$$\min_{\mathbf{a}} \|\mathbf{y}-\mathbf{Ta}\|_1 + \lambda_* \left\|T_M(\mathbf{y}-\mathbf{Ta})\right\|_* + \lambda_\vartheta \|\mathbf{a}\|_2^2. \tag{5}$$

- For $\Phi_2(\mathbf{x}) = 0$, it is the robust representation problem [1, 2, 3, 4] formulated as,

$$\min_{\mathbf{a}} \Phi_1(\mathbf{y}-\mathbf{Ta}) + \lambda_\vartheta \vartheta(\mathbf{a}). \tag{6}$$

Although specific cases of the proposed formulation have been presented in the literature our framework generalizes on these results considering such formulations as a specific case while allowing the selection of application specific residual functions.

In previous works authors have chosen different functions $\vartheta(\mathbf{a})$ to regularize the coefficients $\mathbf{a}$. In the collaborative representation-based classification with regularized least squares (CRC-RLS) [12] the authors considered to solve the SRC problem with $\vartheta(\mathbf{a}) = ||\mathbf{a}||_2^2$. In [1, 3] $\vartheta(\mathbf{a}) = ||\mathbf{a}||_1$ was used, combined with different loss functions, while in the regularized robust coding (RRC) [2], $\vartheta(\mathbf{a}) = ||\mathbf{a}||_2^2$ was used. In correntropy-based sparse representation (CESR) [4], $\vartheta(\mathbf{a})$ was chosen to be the indicator function of the non-negative orthant $\mathbb{R}_+^n$, such that a nonnegative $\mathbf{a} \geq 0$ regularization term was enforced.

Recognition rates on human faces with varying illumination and occlusions as well as disguise, indicate that the ℓ_2-norm is more robust than the sparse coding methods presented in [13, 12, 9]. As such we decided to use $\vartheta(\mathbf{a}) = ||\mathbf{a}||_2^2$ as a regularizer for our experiments.

In this work, we consider function $\Phi_k(\mathbf{x})$ to be defined as,

$$\Phi_k(\mathbf{x}) \equiv \sum_{i=1}^{d} \phi_k(x_i) = \min_{\mathbf{e}} \frac{1}{2}\|\mathbf{x}-\mathbf{e}\|_2^2 + \varphi_k(\mathbf{e}). \tag{7}$$

The function $\phi_k : \mathbb{R} \to \mathbb{R}$ is called the Moreau envelope of the penalty function $\varphi_k(\cdot)$. Thus, $\varphi_k(\cdot)$ is the dual function of $\phi_k(\cdot)$ and the proximal mapping of $\varphi_k(\cdot)$ is defined as,

$$\text{prox}_{\varphi_k}(\mathbf{x}) = \underset{\mathbf{e}}{\text{argmin}} \frac{1}{2}\|\mathbf{x}-\mathbf{e}\|_2^2 + \varphi_k(\mathbf{e}). \tag{8}$$

Table 1: Some choices of $\phi_k(\mathbf{x})$ with $\Phi_k(\mathbf{x}) \equiv \sum_{i=1}^{d} \phi_k(x_i)$ and their proximity operators $\text{prox}_{\varphi_k}(x) = x - \phi_k'(x)$. $\sigma > 0$ and $\lambda > 0$.

Estimator	$\phi_k(x)$	$\text{prox}_{\varphi_k}(x)$
Neg. Gaussian	$(\frac{\sigma^2}{2})(1-\exp(-x^2/\sigma^2))$	$x - x\exp(-x^2/\sigma^2)$
Huber	$\begin{cases} x^2/2 & \lvert x\rvert \leq \lambda \\ \lambda\lvert x\rvert - \frac{\lambda^2}{2} & \lvert x\rvert > \lambda \end{cases}$	$\begin{cases} 0 & \lvert x\rvert \leq \lambda \\ x - \lambda\text{sign}(x) & \lvert x\rvert > \lambda \end{cases}$

Table 2: Choice of proximity operators when $\varphi_k(\mathbf{x})$ is a general norm $||\mathbf{x}||$. $\mathcal{L}_{\lambda_*}(\cdot)$ denotes the SVD decomposition of its matrix argument.

	Estimator	Proximity Operator
Nuclear	$\lambda_*\|\|T_M(\mathbf{x})\|\|_*$	$\mathcal{L}_{\lambda_*}(T_M(\mathbf{x})) = \mathbf{U}_X \mathcal{S}_{\lambda_*}(\mathbf{\Sigma}_X)\mathbf{V}_X^T$ $\mathcal{S}_{\lambda_*}(\mathbf{\Sigma}_X) = \text{sign}(x_{ij})\ \max(0, \lvert x_{ij}\rvert - \lambda_*)$
ℓ_1	$\lambda\|\mathbf{x}\|_\mathbf{1}$	$\text{sign}(x_i)\ \max\left(0, \lvert x_i\rvert - \lambda\right)$

Some choices of Φ_k, utilized in equation (7), and their corresponding proximal mappings are presented in Table 1. For the case when Φ_k is a general norm $||\cdot||$, equivalent results are shown in Table 2.

By using (7) we can write the proposed function in (3) as the augmented function,

$$\begin{aligned} J(\mathbf{a},\mathbf{e}) = &\frac{1}{2}\|\mathbf{y}-\mathbf{Ta}-\mathbf{e}\|_2^2 + \varphi_1(\mathbf{e}) \\ &\frac{1}{2}\|\mathbf{y}-\mathbf{Ta}-\mathbf{e}\|_2^2 + \varphi_2(\mathbf{e}) + \lambda_\vartheta \vartheta(\mathbf{a}), \end{aligned} \tag{9}$$

A local minimizer ($\mathbf{a}$,$\mathbf{e}$) can be calculated alternating in two steps; in step one, $\mathbf{e}$ is updated by fixing the coefficient vector $\mathbf{a}$ and in step two the vector $\mathbf{a}$ is updated by fixing $\mathbf{e}$. However, since there are two penalty functions in the above formulation we use a variable splitting technique [14]. Thus, a local minimizer ($\mathbf{a}$,$\mathbf{e}$) of (9) can be approximated by setting $\mathbf{y}-\mathbf{Ta} = \mathbf{e}$ and introducing the auxiliary variable $\mathbf{e} = \mathbf{z}$ such that (9) is reformulated as,

$$\begin{aligned} \underset{\mathbf{a},\mathbf{e},\mathbf{z}}{\text{minimize}} \quad & \varphi_1(\mathbf{e}) + \varphi_2(\mathbf{z}) + \lambda_\vartheta \vartheta(\mathbf{a}) \\ \text{subject to} \quad & \mathbf{y}-\mathbf{Ta} = \mathbf{e},\ \mathbf{e} = \mathbf{z}. \end{aligned} \tag{10}$$

The problem in (10) allows the implicit incorporation of various loss functions $\phi_k(\cdot)$ through their dual potential functions $\varphi_k(\cdot)$.

3. OPTIMIZATION

The problem in (10) can be solved by the Alternating Direction Method of Multipliers (ADMM) [14] which is intended to blend the decomposability of dual ascent with the superior convergence properties of the method of multipliers. In the ADMM formulation of (10), $\mathbf{a}$, $\mathbf{e}$ and $\mathbf{z}$ are updated in an alternating fashion. As in the method of multipliers, it can take the form of augmented Lagrangian,

$$\begin{aligned} L(\mathbf{e},\mathbf{z},\mathbf{a}) = &\varphi_1(\mathbf{e}) + \varphi_2(\mathbf{z}) + \mathbf{u}_1^T\big(\mathbf{y}-\mathbf{Ta}-\mathbf{e}\big) \\ &+ \frac{\rho}{2}\|\mathbf{y}-\mathbf{Ta}-\mathbf{e}\|_2^2 + \mathbf{u}_2^T\big(\mathbf{e}-\mathbf{z}\big) + \frac{\rho}{2}\|\mathbf{e}-\mathbf{z}\|_2^2 + \lambda_\vartheta \vartheta(\mathbf{a}), \end{aligned} \tag{11}$$

where $\rho > 0$ is a penalty parameter, and $\mathbf{u}_1$ and $\mathbf{u}_2$ are the dual variables. Using the scaled dual variables $\mathbf{y_1} = \frac{1}{\rho}\mathbf{u_1}$,$\mathbf{y_2} = \frac{1}{\rho}\mathbf{u_2}$, one can express the ADMM updates as,

$$\mathbf{e}^{t+1} = \underset{\mathbf{e}}{\operatorname{argmin}}\, L(\mathbf{e}, \mathbf{z}^t, \mathbf{a}^t, \mathbf{y_1}^t, \mathbf{y_2}^t), \tag{12a}$$

$$\mathbf{z}^{t+1} = \underset{\mathbf{z}}{\operatorname{argmin}}\, L(\mathbf{e}^{t+1}, \mathbf{z}, \mathbf{y_2}^t), \tag{12b}$$

$$\mathbf{a}^{t+1} = \underset{\mathbf{a}}{\operatorname{argmin}}\, L(\mathbf{e}^{t+1}, \mathbf{a}, \mathbf{y_1}^t), \tag{12c}$$

$$\mathbf{y_1}^{t+1} = \mathbf{y_1}^t + \mathbf{y} - \mathbf{T}\mathbf{a}^{t+1} - \mathbf{e}^{t+1}, \tag{12d}$$

$$\mathbf{y_2}^{t+1} = \mathbf{y_2}^t + \mathbf{e}^{t+1} - \mathbf{z}^{t+1}. \tag{12e}$$

Optimization is conducted in four steps as,

- **Step 1** - Solve (13) to update $\mathbf{e}^{t+1}$,

$$\mathbf{e}^{t+1} = \underset{\mathbf{e}}{\operatorname{argmin}} \frac{1}{2}\|\mathbf{y} - \mathbf{T}\mathbf{a} + \mathbf{y_1} - \mathbf{e}\|_2^2 + \frac{1}{2}\|\mathbf{z} - \mathbf{y_2} - \mathbf{e}\|_2^2 + \frac{1}{\rho}\varphi_1(\mathbf{e})$$

$$\boxed{\mathbf{e}^{t+1} = \frac{1}{2}\text{prox}_{\varphi_1/\rho}(\mathbf{y} - \mathbf{T}\mathbf{a} + \mathbf{y_1} + \mathbf{z} - \mathbf{y_2})}. \tag{13}$$

- **Step 2** - Solve (14) to update $\mathbf{z}^{t+1}$,

$$\mathbf{z}^{t+1} = \underset{\mathbf{z}}{\operatorname{argmin}} \frac{1}{2}\|\mathbf{e} + \mathbf{y_2} - \mathbf{z}\|_2^2 + \frac{1}{\rho}\varphi_2(\mathbf{z})$$

$$\boxed{\mathbf{z}^{t+1} = \text{prox}_{\varphi_2/\rho}(\mathbf{e} + \mathbf{y_2})}. \tag{14}$$

- **Step 3** - Solve (15) to update $\mathbf{a}^{t+1}$,

$$\mathbf{a}^{t+1} = \underset{\mathbf{a}}{\operatorname{argmin}} \frac{1}{2}\|\mathbf{y} - \mathbf{T}\mathbf{a} - \mathbf{e} + \mathbf{y_1}\|_2^2 + \frac{\lambda}{\rho}\vartheta(a)$$

$$\boxed{\mathbf{a}^{t+1} = \text{prox}_{\vartheta\lambda/\rho}(\mathbf{y_1} - \mathbf{e} - \mathbf{y})}. \tag{15}$$

- **Step 4** - Update multipliers $\mathbf{y_1}$ and $\mathbf{y_2}$ using equations (12d) and (12e).

For the purpose of this paper, in order to guarantee convergence of the optimization problem (10) using ADMM, it is sufficient to enforce appropriate termination criteria. In our experiments we used: $\|\mathbf{y} - \mathbf{T}\mathbf{a} - \mathbf{e}\|_\infty \leq \epsilon$ and $\|\mathbf{e} - \mathbf{z}\|_\infty \leq \epsilon$, where $\epsilon = 10^{-7}$.

Furthermore, we adopt the same classification scheme as in SRC (e.g., reconstruction error). The classification is given by computing the residuals e for each class i as,

$$e_i(\mathbf{y}) = \frac{\|\mathbf{y} - \mathbf{e} - T_i\mathbf{a}_i\|_2}{\|\mathbf{a}_i\|_2}, \tag{16}$$

where $\mathbf{a}_i$ is the segment of final estimated $\mathbf{a}$ associated with class i. Finally, the identity of $\mathbf{y}$ is given as,

$$\text{Identity}(\mathbf{y}) = \underset{i}{\operatorname{argmin}}\{e_i\}. \tag{17}$$

4. EXPERIMENTAL RESULTS

In this section we present experimental results on publicly available databases, AR [15], Extended Yale B [16] and and Multi-PIE [17], to show the efficacy of the proposed classifiers. We compare our methods with the non-robust methods SRC [5] and CRC-RLS [12] and with the robust methods HQ-Additive [3][1], HQ-Multiplicative [3], CESR [4], RRC_L1 [2], RRC_L2 [2], LR^3 [9] and NL_1R [10]. We evaluate our method under illumination, random block occlusion, random pixel corruption and an experiment with mixed variations including changes in illumination, expressions and facial disguises (e.g., sunglasses and scarves).

For all methods, we used the solvers and parameters given by the authors in the corresponding papers and source codes. For experiments on datasets that were not originally conducted by the authors, we did a parameter search in order to identify the optimal parameters.

In our robust error correction (REC) framework we use a combination of two metrics to describe the error: Nuclear norm and negative-Gaussian (REC-LG) or Huber and negative-Gaussian (REC-HG). For completeness, we also consider the single metric case where only the negative-Gaussian model is used (REC-G). We investigate the recognition performance of our chosen functions from Table 1 and we report the results. More error functions can be found in the literature [18, 3]; however, the scope of this work is not to report results from an extensive list of error functions, but to build a general framework for their utilization and to thoroughly investigate the ones that proved more robust throughout our experiments.

In all experiments, the input to our classifiers were pixel values, we set $\rho = 1$ and initialized all variables to zero. In the case with the Negative-Gaussian loss function, we set $\sigma^2 = \gamma \times \text{mean}(\|\mathbf{y} - \mathbf{T}\mathbf{a}\|_2^2)$, as in [3], where γ is a tuning parameter between $[0.2, 0.5]$.

4.1. Recognition under Illumination variations

Experiments with illumination variations were conducted on the Multi-Pie dataset. The Multi-PIE database [17] contains images of 337 subjects captured in 4 sessions with simultaneous variations in pose, expression and illumination. In the experiments we used 249 subjects in Session 1, and a subset of that in Sessions 2 to 4. We followed the experimental setup of [12], and used 14 frontal images[2] per subject with neutral expressions from Session 1 for training, and used 10 frontal images[3] per subject from Sessions 2 to 4 for testing. We used cropped face images with dimensions 50×41 pixels. Recognition rates are shown in Table 3 for various methods.

We observe that our proposed methods REC-HG and REC-G achieve the best results. This indicates that the negative-Gaussian error metric combined with the ℓ_2 regularization on $\mathbf{a}$, models illumination variations adequately. Another interesting observation is the fact that when low-rankness is combined with the negative-Gaussian error metric (REC-LG), it performs better than when low-rankness is combined with sparsity (NL_1R).

4.2. Recognition under Block Occlusions

As in [2, 5, 9, 10], we chose Subsets 1 and 2 of Extended Yale B for training and Subset 3 for testing in order to evaluate the performance of the algorithms on occluded images. We resized the images to 96×84 pixels. Block occlusion was tested by placing an unrelated square block image on each test image. The location of the occlusion was randomly chosen for each image and was unknown during training.

[1]The Welsch function was used as it achieved the best performance. In this work we call the Welsch function negative-Gaussian.

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

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

Table 3: Recognition Rates under Illumination variations.

Sessions	**Session 2** Accuracy	**Session 3** Accuracy	**Session 4** Accuracy
SRC [5]	95.48%	92.13%	95.71%
CRC-RLS [12]	94.16%	87.63%	91.83%
HQ-Additive [3]	95.84%	94.63%	97.14%
HQ-Multiplicative [3]	96.51%	94.94%	97.54%
CESR [4]	95.00%	92.38%	95.94%
RRC_L1 [2]	96.51%	94.50%	97.31%
RRC_L2 [2]	95.78%	90.94%	96.51%
LR^3 [9]	94.82%	90.44%	94.80%
NL_1R [10]	94.54%	89.56%	94.63%
REC-HG	96.93%	95.81%	**98.51%**
REC-LG	96.27%	90.56%	94.97%
REC-G	**96.99%**	**96.19%**	98.40%

Recognition rates for 50% block ooclusion are shown in Table 4 (first column) for the various methods.

One of the highlights of our proposed method is that we achieved almost 100% accuracy (REC-HG) on images with 50% block occlusion on the Yale B database.

4.3. Recognition under Pixel Corruption

For the pixel corruption experiments we used Subsets 1 and 2 of the Extended Yale B for training and Subset 3 for testing, and we resized the images to 96×84 pixels as in [2, 5]. A percentage of randomly chosen pixels from each of the test images was corrupted by replacing those pixel values with independent and identically distributed samples from a uniform distribution between [0, 255]. The percentage of corrupted pixels was 90 percent. Recognition rates are shown in Table 4 (second column) for the various methods.

It is clear from the results that this was the hardest experiment that we ran. Methods that use low-rank as an error metric, perform poorly. We attribute this to the fact that when images are 90% corrupt, modeling the error as low-rank is severely inadequate. Comparing the results for RRC_L1, REC-HG and REC-G, it is not clear if the choice of the error metrics or the regularizer of the coefficients is more important, thus these types of experiments need further investigation.

4.4. Recognition under Mixed Variations

This experiment is a reproduction of the experiment in section 5 of [13]. The AR database consists of over 3000 frontal images of 126 individuals (26 images per individual). Each individual participated in two sessions separated by two weeks, i.e., 13 images were taken at each session. The faces in AR contain variations such as changes in illumination, expressions and facial disguises (e.g., sunglasses or scarves). In our experiments, 100 subjects were chosen randomly (50 male and 50 female). For each subject, we randomly permuted the 26 images, and then took the first half for training and the rest for testing. Thus, we had 1300 training and 1300 testing samples. For statistical stability, we generated 10 different training and testing dataset pairs. The images were cropped to have dimensions 60×43 pixels and converted to gray-scale. Recognition rates are shown in Table 4 (third column) for the various methods.

The fact that images are corrupted with various forms of noise suggests that a two metric model would better capture the error. This however, is not the case for any two metrics. The results verify this claim as the two metric method REC-HG performs better than REC-G in this case, but the two metric method REC-LG performs on par with LR^3.

Overall our proposed methods achieved high recognition rates across all the experiments that we conducted. Figure 1 shows challenging cases in which our algorithm reports superior performance.

Table 4: Recognition Rates and Time Performance under 50% occlusion, 90% pixel corruption and mixed variations.

Variation	**50% Occlusion** Accuracy	Time	**90% Corruption** Accuracy	**Mixed Variations** Accuracy
SRC [5]	54.72%	1.00s	7.25%	97.80% ±0.31
CRC-RLS [12]	47.03%	0.02s	44.40%	96.72% ±0.32
HQ-Additive [3]	94.07%	2.70s	42.64%	96.13% ±0.71
HQ-Multiplicative [3]	95.60%	10.00s	51.21%	96.80% ±0.54
CESR [4]	57.40%	0.70s	41.50%	71.43% ±1.68
RRC_L1 [2]	95.82%	10.30s	83.74%	96.26% ±0.51
RRC_L2 [2]	95.16%	8.80s	55.38%	98.61% ±0.23
LR^3 [9]	96.48%	4.10s	7.69%	98.34%±0.41
NL_1R [10]	93.63%	4.30s	7.47%	98.35% ±0.40
REC-HG	**99.56%**	3.01s	87.69%	**98.68%±0.27**
REC-LG	97.36%	4.40s	7.03%	98.34% ±0.41
REC-G	**99.56%**	3.00s	**87.91%**	98.43% ±0.34

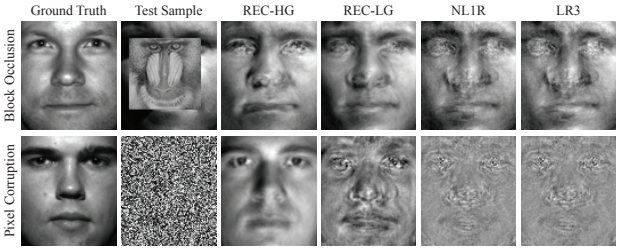

Fig. 1: Example face reconstructions for challenging cases of test samples for our proposed approaches REC-HG and REC-LG as well as the compared error correction algorithms NL_1R [10] and LR^3 [9].

4.5. Computational Cost

Table 4 (first column) summarizes the time experiments we conducted on the extended Yale B database with 50% block occlusion. A first observation is that RRC_L1, RRC_L2 and HQ-Multiplicative are a lot more computationally expensive than our approaches. Methods that modeled error as low rank (NL_1R and REC-LG) were more expensive than methods that used other error metrics, due to the fact that each iteration required an SVD decomposition. The cheapest method by far was CRC-RLS, however, it also achieved the lowest recognition rate. As is clear from Table 4 our proposed methods REC-HG, REC-LG and REC-G strike a good balance between recognition rate and computational cost.

5. CONCLUSIONS

In this work we presented a general framework for incorporating multi-model representation of the residual in face recognition. A vast number of existing methods are special sub cases of the proposed approach for specific choices of loss functions. The experimental results support the claim that the multi-modeling of the residual term combined with the ℓ_2 regularization of the coefficient vector can be beneficial and more robust across a multitude of databases. We believe that this framework will extend and ease further research in face recognition algorithms.

6. REFERENCES

[1] M. Yang, D. Zhang, J. Yang, and D. Zhang, "Robust sparse coding for face recognition," in *IEEE Conf. Comp. Vision Pattern Recognition*, Jun. 2011, pp. 625–632.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] E. J. Candès, J. Romberg, and T. Tao, "Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information," *IEEE Trans. Inf. Theory*, vol. 52, no. 2, pp. 489–509, 2006.

[7] Y. Tsaig and D. L. Donoho, "Extensions of compressed sensing," *Signal Process.*, vol. 86, no. 3, pp. 549–571, Mar. 2006.

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

[9] J. Qian, J. Yang, F. Zhang, and Z. Lin, "Robust low-rank regularized regression for face recognition with occlusion," in *IEEE Conf. Comp. Vision Pattern Recognition Workshops*, Jun. 2014, pp. 21–26.

[10] L. Luo, J. Yang, J. Qian, J. Chen, and Y. Tai, "Nuclear-L1 norm joint regression for face reconstruction and recognition," in *Asian Conf. Computer Vision*, Nov. 2014.

[11] P. J. Huber and M. R. Elvezio, *Robust statistics*, Wiley, New York, Feb. 2009.

[12] D. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: Which helps face recognition?," in *IEEE Int. Conf. Comp. Vision*, Nov. 2011, pp. 471–478.

[13] Q. Shi, A. Eriksson, A. Van den Hengel, and C. Shen, "Is face recognition really a compressive sensing problem?," *IEEE Conf. Comp. Vision Pattern Recognition*, vol. 0, pp. 553–560, 2011.

[14] S. Boyd, N. Parikh, E. Chu, Bo. 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.

[15] A. Martínez and R. Benavente, "The AR face database," Tech. Rep. 24, Computer Vision Center, Bellatera, Jun. 1998.

[16] 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.

[17] R. Gross, I. Matthews, J. Cohn, T. Kanade, and S. Baker, "Multi-PIE," *Image Vision Comput.*, vol. 28, no. 5, pp. 807–813, May 2010.

[18] J. Idier, "Convex half-quadratic criteria and interacting auxiliary variables for image restoration," *IEEE Trans. Image Process.*, vol. 10, no. 7, pp. 1001–1009, Jul. 2001.