Generalized Representation-based Classification by Lp-norm for Face Recognition

Jing Wang, Xiao Xie, Li Zhang, Xusheng Chen, Hongzhou Yue and Huaping Guo

***Abstract*—Representation-based Classification (RC) algorithms have been extensively applied in the domain of face recognition. This paper introduces a novel approach by incorporating arbitrary norms into both the objective and constraint functions of conventional RC algorithms, resulting in the development of the Generalized RC (GRC) algorithm. The primary goal of this enhancement is to fully leverage the unique advantages offered by various norms. Within the majorization-minimization framework, an iterative procedure was designed to solve the optimization problem of GRC. This approach ensures that a closed-form solution is obtained in each iteration and a locally optimal solution can be found upon convergence. Experimental results on four benchmark face databases demonstrated that GRC generally outperforms competing RC algorithms in face recognition. Data and source code of this study are publicly available on the GitHub repository at https://github.com/yuzhounh/GRC.**

***Index Terms*—representation-based classification, sparse, face recognition, Lp-norm, majorization-minimization**

I. INTRODUCTION

Representation-based Classification (RC) algorithms [1, 2] have been widely utilized in face recognition. These algorithms serve as generalizations of popular classification algorithms such as Nearest Neighbor (NN) [3], Nearest Feature Line (NFL) [4], and Nearest Subspace (NS) [5]. NN represents a test sample by a training sample, NFL represents a test sample by a linear combination of a pair of training samples in each class, and NS represents a test sample by a linear combination of all training samples in each class. Then the test sample is assigned to the class with the smallest representation error. The Linear Regression-based Classification (LRC) algorithm [6] can be categorized as an NS method. Recently, Li et al. [7] proposed a new algorithm aimed at selecting an optimal neighborhood size k for k Nearest Neighbor (KNN). Wang and Yang [8] proposed a Nearest Neighbor with Double Neighborhoods (NNDN) algorithm for imbalanced data classification. Impressively, the NNDN algorithm achieved superior classification performance when compared to the most advanced algorithms within the KNN family.

As the foremost RC algorithm, Sparse Representation-based Classification (SRC) [9] incorporates the theory of sparse representation and compressed sensing into the realm of face recognition. The fundamental concept underlying SRC is to represent a test sample by a sparse combination of all training samples and assign it to the class with the smallest representation error. Due to the great success in face recognition, SRC has witnessed widespread adoption across various fields, including but not limited to hyperspectral image classification [10, 11], motor imagery EEG pattern recognition [12, 13], automatic identification of epileptic seizures [14], protein-protein interactions prediction [15].

The importance of sparse representation received considerable attention in SRC and related studies [1, 9] while the importance of collaborative representation was somewhat overlooked. To address this problem, Collaborative Representation-based Classification (CRC) [16] was proposed. CRC achieves very competitive classification results with significantly reduced computational complexity compared to SRC [16]. Therefore, it stands as a good alternative to SRC [2, 16]. Furthermore, these two algorithms can be synergistically combined to enhance classification performance [17].

The optimization problems of SRC and CRC are very similar. They share an identical objective function, with the primary distinction being that SRC employs the L1-norm in its constraint function, whereas CRC opts for the L2-norm. Considering that L1-norm and L2-norm represent two specific instances of the more general Lp-norm, it is natural to replace them with Lp-norm to leverage the inherent sparsity and robustness associated with various norms. Drawing inspiration from this idea, we present an innovative RC algorithm termed Generalized Representation-based Classification (GRC), In GRC, Lp-norm is applied in both the objective function and the constraint function of traditional RC algorithms, thereby providing a more versatile and adaptable norm-based approach.

Lp-norm has demonstrated its effectiveness in extended versions of Penalized Least Squares (PLS) [18], Principal Component Analysis (PCA) [19, 20], two-dimensional PCA

Manuscript received June 28, 2023; revised November 30, 2023.

This work was supported in part by the National Natural Science Foundation of China under Grant 31900710 and Grant 31600862; in part by the Youth Scientific Research Fund Project of Xinyang Normal University under Grant 2021-QN-036; and in part by the Nanhu Scholars Program for Young Scholars of Xinyang Normal University.

Jing Wang is a lecturer at School of Computer and Information Technology, Xinyang Normal University, Xinyang, 464000, China. (corresponding author to provide phone: 86-0376-6390765; fax: 86-0376-6390765; e-mail: wangjing@xynu.edu.cn).

Xiao Xie is a teacher at School of Computer and Information Technology, Xinyang Normal University, Xinyang, 464000, China. (e-mail: xiexiao@xynu.edu.cn).

Li Zhang is an associate professor at Preschool Teachers College, Nanjing Xiaozhuang University, Jiangsu, 211171, China (e-mail: lily613@foxmail.com).

Xusheng Chen is a lecturer at School of Computer and Information Technology, Xinyang Normal University, Xinyang, 464000, China. (e-mail: 471877096@qq.com).

Hongzhou Yue is a lecturer at School of Computer and Information Technology, Xinyang Normal University, Xinyang, 464000, China. (e-mail: yuehz@xynu.edu.cn).

Huaping Guo is an associate professor at School of Computer and Information Technology, Xinyang Normal University, Xinyang, 464000, China. (e-mail: hpguo@xynu.edu.cn).

(2DPCA) [21, 22], Linear Discriminant Analysis (LDA) [23], two-dimensional LDA (2DLDA) [24], Common Spatial Patterns (CSP) [25, 26], etc. These generalized algorithms consistently outperform their traditional counterparts and find their applications in various fields such as matrix completion, image classification, image reconstruction, brain signal classification, etc. Therefore, it is reasonable to anticipate that extending classical RC algorithms to GRC can lead to performance enhancements.

Among the existing studies, the research target of the Generalized Iterated Shrinkage Algorithm (GISA) [27] is the closest to that of GRC. GISA aims to solve a problem known as Lp-norm non-convex sparse coding. Except for GISA, several alternative algorithms [27] were proposed to solve the same problem, including Iteratively Reweighted Least Squares (IRLS), Iteratively Reweighted L1-minimization (IRL1), Iteratively Thresholding Method (ITM-Lp), and Look-up Table (LUT). These algorithms had been successfully applied in image classification [28-30] and image inpainting [31, 32]. It was demonstrated that GISA is superior to the competing algorithms [27]. Nevertheless, these studies primarily focused on the Lp-norm minimization problem within the range of $0 \leq p < 1$. None of them provides a complete solution to the GRC problem proposed in this paper.

The major differences between GISA and GRC exist in three aspects. Firstly, GISA exclusively applies Lp-norm in the constraint function of the sparse coding problem, whereas GRC applies Lp-norm in both the objective function and the constraint function of RC algorithms. Secondly, GISA only deals with the cases when $0 \leq p < 1$ while GRC deals with the cases when $0 < s \leq 2$ and $0 < p \leq 2$. Here, s denotes the value of the arbitrary norm applied in the objective function. Thirdly, GISA designs the solution based on a Generalized Soft-thresholding (GST) function and converges to the optimal solution only in several special cases while GRC designs the solution within the Majorization-Minimization (MM) framework [33], which guarantees that the solution always converges to a locally optimal one.

While Lp-norm in the constraint function of RC algorithms has been abundantly studied [27], Lp-norm in the objective function is yet to be investigated. The most relevant studies are the robust extensions of SRC and CRC [34-36], which apply L1-norm in the objective function of RC algorithms. That is, the parameter s is fixed to 1 in the Robust RC (RRC) algorithms. These RRC algorithms can be effectively solved through methods such as Alternating Direction Method (ADM) and IRLS. In this paper, we investigate GRC with $0 < s \leq 2$ and $0 < p \leq 2$, which substantially extends RRC.

In addition to the aforementioned approaches, Majumdar and Ward [37] introduced the concept of group lasso [38] into sparse coding to promote group sparsity. Huang et al. [39] provided an IRLS-based algorithm to solve the group sparse coding problem. The group sparsity finds wide-ranging applications in tasks like multi-task classification [40] and multi-modal classification [41]. Wang et al. [34] introduced a correlation regularization technique that is formulated by employing the trace norm into RRC to promote adaptive sparsity. The correlation regularization effectively strikes a balance between the L2-norm and L1-norm regularizations. Yang et al. [42] replaced L2-norm with nuclear norm in the objective function of CRC to emphasize the low-rank property of the residual image. Xiao [43, 44] introduced a L1-L2 blur regularization and a double L0-regularization for blind image restoration. Ding et al. [45] introduced Schatten-2/3 and Schatten-1/2 quasi-norms for low rank tensor completion. The variants of norms utilized in these studies significantly differ from Lp-norm, which is the major focus of this study.

The contributions of this study can be summarized in the following points: (1) A Generalized Representation-based Classification algorithm (GRC) was proposed for face recognition by applying arbitrary norms in both the objective function and the constraint function of conventional RC algorithms. (2) An iterative algorithm was designed to solve the optimization problem of GRC within the MM framework. (3) The classification performance of GRC was thoroughly evaluated on four well-established benchmark face databases. Experimental results demonstrated that GRC, when configured with optimal parameters, outperforms competing algorithms.

The remainder of this paper is organized as follows. Section II reviews the relevant studies and proposes the GRC approach. Section III presents the techniques that will be used to solve the GRC problem. Section IV presents the solutions derived for GRC. Section V conducts experiments to evaluate the classification performance of GRC. Section VI presents and analyzes the experimental results. Section VII makes discussions on this study and related works. Finally, Section VIII presents the conclusions and summarizes the findings.

II. RELATED WORKS

The notations in this paper are described as follows. Lowercase letters denote scalars, boldface lowercase letters denote vectors, boldface uppercase letters denote matrices; $\mathrm{sign}(\cdot)$ denotes the sign function; $|\cdot|$ denotes the absolute value function; $|\boldsymbol{w}|^p$ denotes the element-wise power of the absolute value of a vector; $\boldsymbol{w}^{(k)}$ denotes a vector in the kth iteration; $\mathrm{diag}(\boldsymbol{w})$ denotes a square and diagonal matrix formed by placing the elements of vector $\boldsymbol{w}$ on the main diagonal; $\boldsymbol{w} \circ \boldsymbol{v}$ denotes the Hadamard product, i.e., the element-wise multiplication of two vectors; $\|\cdot\|_1$, $\|\cdot\|_2$, and $\|\cdot\|_p$ denote L1-norm, L2-norm, and Lp-norm, respectively. It's important to note that the sign function and the absolute value function can be applied either to a scalar or to a vector in an element-wise manner.

We first review two pertinent algorithms, i.e., SRC and CRC, in a sequential manner. Then, the proposed GRC algorithm is described and its relationships with both SRC and CRC are discussed.

A. SRC

Suppose there are n training image samples for c classes $\boldsymbol{X} = [\boldsymbol{X}_1, \boldsymbol{X}_2, \dots, \boldsymbol{X}_c] \in \mathbb{R}^{d \times n}$ where each column of $\boldsymbol{X}$ is a training image sample obtained by stacking the columns of an image and each row of $\boldsymbol{X}$ is a feature. $\boldsymbol{y} \in \mathbb{R}^d$ is a test image sample. SRC [9] finds its representation coefficient $\boldsymbol{\alpha} \in \mathbb{R}^n$ by solving the following optimization problem:

$$\min_{\boldsymbol{\alpha}} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\alpha}\|_2^2 + \lambda \|\boldsymbol{\alpha}\|_1 \quad (1)$$

where λ is a tuning parameter. To enhance clarity, we refer to $\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\alpha}\|_2^2$ the objective function and $\|\boldsymbol{\alpha}\|_1$ as either the constraint function or the regularization term. Several representative approaches for solving this L1-norm minimization problem are reviewed in [46]. Let $\hat{\boldsymbol{\alpha}}_i$ denote the representation coefficient associated with class i. The residuals corresponding to different classes are computed as follows:

$$\|\boldsymbol{y}-\boldsymbol{X}_i\hat{\boldsymbol{\alpha}}_i\|_2, \ i=1,2,\dots,c. \tag{2}$$

Since the L1-norm sparsity $\|\boldsymbol{\alpha}\|_1$ can augment discriminatory information for classification [16], we modified the residuals by incorporating the L1-norm sparsity. The modified residuals are computed as follows:

$$r_i=\frac{\|\boldsymbol{y}-\boldsymbol{X}_i\hat{\boldsymbol{\alpha}}_i\|_2}{\|\hat{\boldsymbol{\alpha}}_i\|_1}, \ i=1,2,\dots,c. \tag{3}$$

This modification can improve the classification performance of SRC in practice. After calculating the residuals, the identity of the test sample $\boldsymbol{y}$ is assigned to $\arg min_i\{r_i\}$.

B. CRC

CRC [16] can be formulated by replacing the L1-norm in the constraint function of SRC with the L2-norm. That is, CRC determines its representation coefficient by solving the following optimization problem:

$$\min_{\boldsymbol{\alpha}}\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\alpha}\|_2^2+\lambda\|\boldsymbol{\alpha}\|_2^2. \tag{4}$$

The solution of GRC can be analytically derived as

$$\boldsymbol{\alpha}=(\boldsymbol{X}^T\boldsymbol{X}+\lambda\boldsymbol{I})^{-1}\boldsymbol{X}^T\boldsymbol{y}. \tag{5}$$

After obtaining the representation coefficient $\boldsymbol{\alpha}$ for CRC, the residuals are computed as

$$r_i=\frac{\|\boldsymbol{y}-\boldsymbol{X}_i\hat{\boldsymbol{\alpha}}_i\|_2}{\|\hat{\boldsymbol{\alpha}}_i\|_2}, \ i=1,2,\dots,c. \tag{6}$$

Then the identity of $\boldsymbol{y}$ is assigned to $\arg min_i\{r_i\}$.

C. GRC

Inspired by the above two RC algorithms, i.e., SRC and CRC, we formulate the optimization problem of GRC by applying arbitrary norms in both the objective function and the constraint function of conventional RC algorithms as follows:

$$\min_{\boldsymbol{\alpha}}\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\alpha}\|_s^s+\lambda\|\boldsymbol{\alpha}\|_p^p \tag{7}$$

where $0<s\le 2$ and $0<p\le 2$. After obtaining the representation coefficient $\boldsymbol{\alpha}$ for GRC, the residuals are computed as:

$$r_i=\frac{\|\boldsymbol{y}-\boldsymbol{X}_i\hat{\boldsymbol{\alpha}}_i\|_s}{\|\hat{\boldsymbol{\alpha}}_i\|_p}, \ i=1,2,\dots,c. \tag{8}$$

Then the identity of $\boldsymbol{y}$ is assigned to $\arg min_i\{r_i\}$.

GRC serves as a natural extension of SRC and CRC. When $s=2$ and $p=1$, GRC reduces to SRC. When $s=2$ and $p=2$, GRC reduces to CRC. Consequently, GRC inherits the strengths of both SRC and CRC. Furthermore, GRC can leverage diverse norms to explore additional potentialities such as sparsity and robustness. As a result, GRC is expected to outperform other RC algorithms in terms of classification performance.

III. METHODOLOGY

Solving the optimization problem of GRC is a nontrivial task due to its inherent nonsmooth nature. To address this challenge, this paper introduces the MM framework. The MM framework effectively transforms a nonsmooth problem into a smooth one that is much easier to handle.

A. MM Framework

Suppose $f(\boldsymbol{\alpha})$ is the objective function to be minimized. Within the MM framework, if there exists a surrogate function $g\left(\boldsymbol{\alpha}\middle|\boldsymbol{\alpha}^{(k)}\right)$ that satisfies two key conditions:

$$\begin{aligned} f\left(\boldsymbol{\alpha}^{(k)}\right)&=g\left(\boldsymbol{\alpha}^{(k)}\middle|\boldsymbol{\alpha}^{(k)}\right)\\ f(\boldsymbol{\alpha})&\le g\left(\boldsymbol{\alpha}\middle|\boldsymbol{\alpha}^{(k)}\right) \ \text{ for all } \boldsymbol{\alpha}\end{aligned} \tag{9}$$

where $\boldsymbol{\alpha}^{(k)}$ is $\boldsymbol{\alpha}$ at the kth step in the iteration procedure. Then $f(\boldsymbol{\alpha})$ could be optimized by iteratively minimizing the surrogate function $g\left(\boldsymbol{\alpha}\middle|\boldsymbol{\alpha}^{(k)}\right)$ as follows:

$$\boldsymbol{\alpha}^{(k+1)}=\arg\min_{\boldsymbol{\alpha}} g\left(\boldsymbol{\alpha}\middle|\boldsymbol{\alpha}^{(k)}\right). \tag{10}$$

One could observe that

$$\begin{aligned} f\left(\boldsymbol{\alpha}^{(k+1)}\right)&=f\left(\boldsymbol{\alpha}^{(k+1)}\right)-g\left(\boldsymbol{\alpha}^{(k+1)}\middle|\boldsymbol{\alpha}^{(k)}\right)+g\left(\boldsymbol{\alpha}^{(k+1)}\middle|\boldsymbol{\alpha}^{(k)}\right)\\ &\le f\left(\boldsymbol{\alpha}^{(k)}\right)-g\left(\boldsymbol{\alpha}^{(k)}\middle|\boldsymbol{\alpha}^{(k)}\right)+g\left(\boldsymbol{\alpha}^{(k+1)}\middle|\boldsymbol{\alpha}^{(k)}\right)\\ &\le f\left(\boldsymbol{\alpha}^{(k)}\right)-g\left(\boldsymbol{\alpha}^{(k)}\middle|\boldsymbol{\alpha}^{(k)}\right)+g\left(\boldsymbol{\alpha}^{(k)}\middle|\boldsymbol{\alpha}^{(k)}\right)\\ &=f\left(\boldsymbol{\alpha}^{(k)}\right)\end{aligned} \tag{11}$$

where the first inequality holds because $f(\boldsymbol{\alpha})-g\left(\boldsymbol{\alpha}\middle|\boldsymbol{\alpha}^{(k)}\right)$ reaches its maximum at $\boldsymbol{\alpha}=\boldsymbol{\alpha}^{(k)}$ as the result of the two key conditions, while the second inequality holds because $g\left(\boldsymbol{\alpha}\middle|\boldsymbol{\alpha}^{(k)}\right)$ reaches its minimum at $\boldsymbol{\alpha}=\boldsymbol{\alpha}^{(k+1)}$ as a result of the update rule. Therefore, the value of the objective function monotonically decreases during the iteration procedure and will converge to a local optimum.

B. Inequalities

A critical point of the MM framework is to find a surrogate function that can be optimized directly. This is realized by introducing appropriate inequalities. According to [21], we have the following two inequalities.

Lemma 1: Let $\boldsymbol{\alpha}\in\mathbb{R}^n$, $\boldsymbol{\beta}\in\mathbb{R}^n$, and $0<p<1$, then

$$\|\boldsymbol{\alpha}\|_p^p\le p[|\boldsymbol{\beta}|^{p-1}\circ\operatorname{sign}(\boldsymbol{\beta})]^T\boldsymbol{\alpha}+(1-p)\|\boldsymbol{\beta}\|_p^p \tag{12}$$

holds and the inequality becomes equality when $\boldsymbol{\alpha}=\boldsymbol{\beta}$.

Lemma 2: Let $\boldsymbol{\alpha}\in\mathbb{R}^n$, $\boldsymbol{\beta}\in\mathbb{R}^n$, $\boldsymbol{\beta}$ has no zero elements, and $0<p<2$, then

$$\|\boldsymbol{\alpha}\|_p^p\le\frac{p}{2}\boldsymbol{\alpha}^T\operatorname{diag}(|\boldsymbol{\beta}|^{p-2})\boldsymbol{\alpha}+\left(1-\frac{p}{2}\right)\|\boldsymbol{\beta}\|_p^p \tag{13}$$

holds and the inequality becomes equality when $\boldsymbol{\alpha}=\boldsymbol{\beta}$.

Lemma 1 establishes a linear function of $\boldsymbol{\alpha}$ as the upper bound of $\|\boldsymbol{\alpha}\|_p^p$ while Lemma 2 establishes a quadratic function of $\boldsymbol{\alpha}$ as the upper bound of $\|\boldsymbol{\alpha}\|_p^p$. The two inequalities will be used to solve the optimization problem of GRC.

C. Linear Optimization Problem with Lp-Norm Constraint

Another lemma that will be used to solve the GRC problem is the solution of a linear optimization problem with Lp-norm constraint.

Lemma 3: Let $\boldsymbol{\alpha} \in \mathbb{R}^n$, $\boldsymbol{\beta} \in \mathbb{R}^n$, $\boldsymbol{\beta} \neq \mathbf{0}$. Let $p, q \in [1, \infty]$ be two scalars that satisfies $1/p + 1/q = 1$. The optimization problem

$$\max_{\boldsymbol{\alpha}} \boldsymbol{\beta}^T \boldsymbol{\alpha}, \qquad \text{s.t. } \|\boldsymbol{a}\|_p^p = 1 \tag{14}$$

has a closed-form solution

$$\boldsymbol{\alpha} = \frac{|\boldsymbol{\beta}|^{q-1} \circ \text{sign}(\boldsymbol{\beta})}{\|\boldsymbol{\beta}\|_q^{q-1}}. \tag{15}$$

This lemma is also demonstrated in [21].

IV. SOLUTIONS OF GRC

With the above techniques, we proceed to solve the GRC problem. GRC aims to optimize the following problem:

$$\min_{\boldsymbol{\alpha}} \|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\alpha}\|_s^s + \lambda \|\boldsymbol{\alpha}\|_p^p \tag{16}$$

where $0 < s \leq 2$ and $0 < p \leq 2$. Let $\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\alpha}\|_s^s + \lambda \|\boldsymbol{\alpha}\|_p^p$ be denoted as $f(\boldsymbol{\alpha})$, where $\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\alpha}\|_s^s$ is the objective function and $\|\boldsymbol{\alpha}\|_p^p$ is the regularization term. We can obtain four different solutions for the GRC problem by relaxing $f(\boldsymbol{\alpha})$ in various ways, depending on the ranges of s and p.

A. Case 1

When $0 < s < 1$ and $1 \leq p \leq 2$, by relaxing the objective function in $f(\boldsymbol{\alpha})$ according to Lemma 1, we have

$$\begin{aligned} &\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\alpha}\|_s^s + \lambda \|\boldsymbol{\alpha}\|_p^p \\ &\leq s\left[\left|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\alpha}^{(k)}\right|^{s-1} \circ \text{sign}\left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\alpha}^{(k)}\right)\right]^T (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\alpha}) \\ &\quad + (1-s)\left\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\alpha}^{(k)}\right\|_s^s + \lambda \|\boldsymbol{\alpha}\|_p^p. \end{aligned} \tag{17}$$

Let the relaxed function be denoted as $g\left(\boldsymbol{\alpha}|\boldsymbol{\alpha}^{(k)}\right)$, i.e.,

$$\begin{aligned} &g\left(\boldsymbol{\alpha}|\boldsymbol{\alpha}^{(k)}\right) \\ &= s\left[\left|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\alpha}^{(k)}\right|^{s-1} \circ \text{sign}\left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\alpha}^{(k)}\right)\right]^T (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\alpha}) \\ &\quad + (1-s)\left\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\alpha}^{(k)}\right\|_s^s + \lambda \|\boldsymbol{\alpha}\|_p^p. \end{aligned} \tag{18}$$

We have $f\left(\boldsymbol{\alpha}^{(k)}\right) = g\left(\boldsymbol{\alpha}^{(k)}|\boldsymbol{\alpha}^{(k)}\right)$ and $f(\boldsymbol{\alpha}) \leq g\left(\boldsymbol{\alpha}|\boldsymbol{\alpha}^{(k)}\right)$ for all $\boldsymbol{\alpha}$, satisfying the two key conditions of the MM framework. Therefore, $g\left(\boldsymbol{\alpha}|\boldsymbol{\alpha}^{(k)}\right)$ is a feasible surrogate function of $f(\boldsymbol{\alpha})$. According to the MM framework, the GRC problem can be turned into iteratively minimizing the surrogate function as follows:

$$\boldsymbol{\alpha}^{(k+1)} = \arg\min_{\boldsymbol{\alpha}} g\left(\boldsymbol{\alpha}|\boldsymbol{\alpha}^{(k)}\right). \tag{19}$$

By ignoring the items irrelevant to $\boldsymbol{\alpha}$, minimizing the surrogate function leads to the following problem:

$$\boldsymbol{\alpha}^{(k+1)} = \arg\min_{\boldsymbol{\alpha}} -s\boldsymbol{z}^T\boldsymbol{\alpha} + \lambda \|\boldsymbol{\alpha}\|_p^p \tag{20}$$

where $\boldsymbol{z} = \boldsymbol{X}^T\left[\left|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\alpha}^{(k)}\right|^{s-1} \circ \text{sign}\left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\alpha}^{(k)}\right)\right]$. According to Lemma 3, its solution is

$$\boldsymbol{\alpha}^{(k+1)} = \frac{|\boldsymbol{z}|^{q-1} \circ \text{sign}(\boldsymbol{z})}{\|\boldsymbol{z}\|_q^{q-1}} \tag{21}$$

where $q = p/(p-1)$. By applying this update rule, we can obtain a locally optimal solution for GRC. This completes the solution of case 1.

Note that this solution also can be used to solve GRC with $0 < s < 1$ and $p > 2$ because the optimization problem in Equation (20) is solved by Lemma 3, which is applicable for $p \geq 1$. However, for clarity, we mainly focus on GRC with $0 < s \leq 2$ and $0 < p \leq 2$, and neglect the case when $0 < s < 1$ and $p > 2$. On the other hand, the classification accuracies of GRC with $0 < s < 1$ are very low in practice. Therefore, we pay little attention to GRC with $0 < s < 1$ and $p > 2$ in our study.

B. Case 2

When $0 < s < 1$ and $0 < p \leq 2$, by relaxing the objective function in $f(\boldsymbol{\alpha})$ according to Lemma 1 and relaxing the regularization in $f(\boldsymbol{\alpha})$ according to Lemma 2, we have

$$\begin{aligned} &\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\alpha}\|_s^s + \lambda \|\boldsymbol{a}\|_p^p \\ &\leq s\left[\left|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\alpha}^{(k)}\right|^{s-1} \circ \text{sign}\left(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\alpha}^{(k)}\right)\right]^T (\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\alpha}) \\ &\quad + (1-s)\left\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\alpha}^{(k)}\right\|_s^s + \lambda \frac{p}{2} \boldsymbol{\alpha}^T \text{diag}\left(\left|\boldsymbol{\alpha}^{(k)}\right|^{p-2}\right)\boldsymbol{\alpha} \\ &\quad + \lambda\left(1 - \frac{p}{2}\right)\left\|\boldsymbol{\alpha}^{(k)}\right\|_p^p. \end{aligned} \tag{22}$$

Let the relaxed function be denoted as $g\left(\boldsymbol{\alpha}|\boldsymbol{\alpha}^{(k)}\right)$. Again, we have $f\left(\boldsymbol{\alpha}^{(k)}\right) = g\left(\boldsymbol{\alpha}^{(k)}|\boldsymbol{\alpha}^{(k)}\right)$ and $f(\boldsymbol{\alpha}) \leq g\left(\boldsymbol{\alpha}|\boldsymbol{\alpha}^{(k)}\right)$ for all $\boldsymbol{\alpha}$, satisfying the two key conditions of the MM framework. Therefore, $g\left(\boldsymbol{\alpha}|\boldsymbol{\alpha}^{(k)}\right)$ is a feasible surrogate function of $f(\boldsymbol{\alpha})$. According to the MM framework, the GRC problem can be turned into iteratively minimizing the surrogate function. Its solution is

$$\begin{aligned} \boldsymbol{\alpha}^{(k+1)} &= \frac{s}{\lambda p} \text{diag}\left(\left|\boldsymbol{\alpha}^{(k)}\right|^{2-p}\right)\boldsymbol{z} \\ &= \frac{s}{\lambda p}\left(\left|\boldsymbol{\alpha}^{(k)}\right|^{2-p} \circ \boldsymbol{z}\right). \end{aligned} \tag{23}$$

This completes the solution of case 2.

C. Case 3

When $0 < s \leq 2$ and $0 < p < 1$, by relaxing the objective function in $f(\boldsymbol{\alpha})$ according to Lemma 2 and relaxing the regularization in $f(\boldsymbol{\alpha})$ according to Lemma 1, we have

$$\begin{aligned} &\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\alpha}\|_s^s + \lambda \|\boldsymbol{a}\|_p^p \\ &\leq \frac{s}{2}(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\alpha})^T \text{diag}\left(\left|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\alpha}^{(k)}\right|^{s-2}\right)(\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\alpha}) \\ &\quad + \left(1 - \frac{s}{2}\right)\left\|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\alpha}^{(k)}\right\|_s^s \\ &\quad + \lambda p\left[\left|\boldsymbol{\alpha}^{(k)}\right|^{p-1} \circ \text{sign}\left(\boldsymbol{\alpha}^{(k)}\right)\right]^T \boldsymbol{\alpha} \\ &\quad + \lambda(1-p)\left\|\boldsymbol{\alpha}^{(k)}\right\|_p^p. \end{aligned} \tag{24}$$

Let the relaxed function be denoted as $g\left(\boldsymbol{\alpha}|\boldsymbol{\alpha}^{(k)}\right)$. Similarly, $g\left(\boldsymbol{\alpha}|\boldsymbol{\alpha}^{(k)}\right)$ is a feasible surrogate function of $f(\boldsymbol{\alpha})$. Therefore, the GRC problem can be turned into iteratively minimizing the surrogate function. Its solution is

$$\boldsymbol{\alpha}^{(k+1)} = \boldsymbol{U}^{-1}\left[\boldsymbol{v} - \frac{\lambda p}{s}\left|\boldsymbol{\alpha}^{(k)}\right|^{p-1} \circ \text{sign}\left(\boldsymbol{\alpha}^{(k)}\right)\right] \tag{25}$$

where

$$\boldsymbol{U} = \boldsymbol{X}^T \text{diag}\left(\left|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\alpha}^{(k)}\right|^{s-2}\right)\boldsymbol{X} \tag{26}$$

and

$$\boldsymbol{v} = \boldsymbol{X}^T \text{diag}\left(\left|\boldsymbol{y} - \boldsymbol{X}\boldsymbol{\alpha}^{(k)}\right|^{s-2}\right)\boldsymbol{y}. \tag{27}$$

This completes the solution of case 3.

D. Case 4

When $0 < s \leq 2$ and $0 < p \leq 2$, by relaxing the objective function and the regularization in $f(\boldsymbol{\alpha})$ according to Lemma 2, we have

$$\begin{aligned} &\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\alpha}\|_s^s+\lambda\|\boldsymbol{\alpha}\|_p^p \\ &\leq \frac{s}{2}(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\alpha})^T \operatorname{diag}\left(\left|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\alpha}^{(k)}\right|^{s-2}\right)(\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\alpha}) \\ &+\left(1-\frac{s}{2}\right)\left\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\alpha}^{(k)}\right\|_s^s+\lambda\frac{p}{2}\boldsymbol{\alpha}^T \operatorname{diag}\left(\left|\boldsymbol{\alpha}^{(k)}\right|^{p-2}\right)\boldsymbol{\alpha} \\ &+\lambda\left(1-\frac{p}{2}\right)\left\|\boldsymbol{\alpha}^{(k)}\right\|_p^p. \end{aligned} \tag{28}$$

Let the relaxed function be denoted as $g\left(\boldsymbol{\alpha}|\boldsymbol{\alpha}^{(k)}\right)$. Similarly, $g\left(\boldsymbol{\alpha}|\boldsymbol{\alpha}^{(k)}\right)$ is a feasible surrogate function of $f(\boldsymbol{\alpha})$. Therefore, the GRC problem can be turned into iteratively minimizing the surrogate function. Its solution is

$$\boldsymbol{\alpha}^{(k+1)}=\left[\boldsymbol{U}+\frac{\lambda p}{s}\operatorname{diag}\left(\left|\boldsymbol{\alpha}^{(k)}\right|^{p-2}\right)\right]^{-1}\boldsymbol{v}. \tag{29}$$

This completes the solution of case 4.

The above completes solving the GRC problem with $0 < s \leq 2$ and $0 < p \leq 2$. Importantly, in all cases, a closed-form solution is obtained in each iteration. The solution avoids some common problems in solving Lp-norm-based algorithms such as learning rates [20] and zero-finding problems [27]. Furthermore, a locally optimal solution can be guaranteed due to the MM framework.

Fig. 1 illustrates the solution domains of the four cases. Excluding case 1 with $p > 2$ where the classification accuracies tend to be very low in practice, the solution domain of case 4 covers those of cases 1 to 3. For simplicity, we apply the solution in case 4 to solve the GRC problem, as summarized in Algorithm 1.

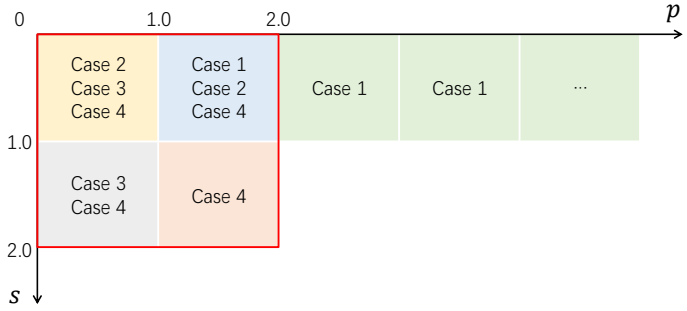

Fig. 1. The solution domains of the four cases.

Algorithm 1 The Algorithm Procedure of GRC

Input: n training image samples $\boldsymbol{X} \in \mathbb{R}^{d\times n}$ and a test image sample $\boldsymbol{y} \in \mathbb{R}^d$, $s \in (0,2]$, $p \in (0,2]$.

Output: the representation coefficient $\boldsymbol{\alpha}$.

Initialize $\boldsymbol{\alpha}^{(0)} = (\boldsymbol{X}^T\boldsymbol{X}+\lambda\boldsymbol{I})^{-1}\boldsymbol{X}^T\boldsymbol{y}$.

while $\boldsymbol{\alpha}$ is not converged **or** the maximal iteration number is not reached **do**

$\boldsymbol{U} = \boldsymbol{X}^T \operatorname{diag}\left(\left|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\alpha}^{(k)}\right|^{s-2}\right)\boldsymbol{X}$

$\boldsymbol{v} = \boldsymbol{X}^T \operatorname{diag}\left(\left|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\alpha}^{(k)}\right|^{s-2}\right)\boldsymbol{y}$

$\boldsymbol{\alpha}^{(k+1)} = \left[\boldsymbol{U}+\frac{\lambda p}{s}\operatorname{diag}(|\boldsymbol{\alpha}^{(k)}|^{p-2})\right]^{-1}\boldsymbol{v}$

end while

V. EXPERIMENTS

A. Face databases

Four benchmark face databases, i.e., the AR face database [47], the FEI face database [48], the FERET face database [49], and the UMIST face database [50] were used in our experiments to evaluate the classification performance of GRC.

The AR face database contains 3120 images from 120 subjects, 26 images per subject. The images were taken with different facial expressions and illumination conditions. Some of the images are occluded with sunglasses or scarves. All images have been cropped and scaled to 50 by 40. This database can be downloaded from http://www2.ece.ohio-state.edu/~aleix/ARdatabase.html.

The FEI face database contains 2800 images from 200 subjects, 14 images per subject. The images were taken with different view angles, scales, facial expressions, and illumination conditions. All images have been cropped and scaled to 48 by 64. This database can be downloaded from https://fei.edu.br/~cet/facedatabase.html.

The FERET face database contains 1400 images from 200 subjects, 7 images per subject. The images were taken with different facial expressions and view angles. All images have been cropped and scaled to 80 by 80. This database can be downloaded from https://www.nist.gov/itl/products-and-services/color-feret-database.

The UMIST face database contains 380 images from 20 subjects, 19 images per subject. The images were taken with different view angles. All images have been cropped and scaled to 112 by 92. This database can be downloaded from http://eprints.lincoln.ac.uk/id/eprint/16081/.

Fig. 2 shows some sample images from the four face databases. Table I shows the statistics of these databases.

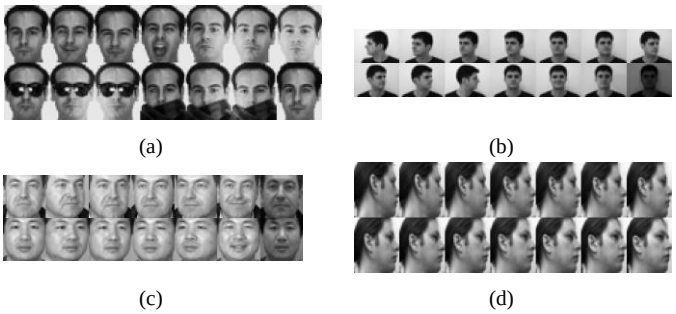

Fig. 2. Fourteen sample images from (a) the AR face database, (b) the FEI face database, (c) the FERET face database, and (d) The UMIST face database.

TABLE I
THE STATISTICS OF THE FOUR DATABASES

Face Database	Image size	Total number of images	Number of images per subject	Number of subjects
AR	50*40	3120	26	120
FEI	48*64	2800	14	200
FERET	80*80	1400	7	200
UMIST	112*92	380	19	20

B. Parameter settings

To identify the optimal parameter pair (s,p) for GRC, we conducted a search within the range of $s = [0.1{:}\,0.1{:}\,2.0]$ and $p = [0.1{:}\,0.1{:}\,2.0]$. Since GRC reduces to SRC when $(s,p) =$

$(2,1)$ and reduces to CRC when $(s,p)=(2,2)$, we could make a direct comparison between GRC and the two classical RC algorithms, i.e., SRC and CRC.

For the parameter λ, it was reported that SRC and CRC can achieve good results when λ is assigned a small positive value, typically ranging from 10^{-6} to 10^{-1} [16]. Our experimental results showed that this observation holds true for GRC. Therefore, we set λ to 10^{-3}.

C. Experiment 1

In the first experiment, we initially extracted principal components from the raw face data. Subsequently, we applied different representation-based classification algorithms to perform face recognition. The optimal parameters of GRC were determined in this procedure. Finally, we made a comprehensive comparison between GRC and its related algorithms.

For each subject in the FERET face database, we randomly chose three images for training and two images for testing. For each subject in the remaining three face databases, we randomly chose six images for training and four images for testing. The training images were normalized by z-score so that each feature was centered to have a mean of zero and scaled to have a standard deviation of one. Then the testing images were normalized by applying the same parameters. After that, we applied Principal Component Analysis (PCA) [51, 52] to reduce the dimension of the normalized data. Suppose there are n training image samples. $\lambda_1, \lambda_2, \ldots, \lambda_m$ represent m eigenvalues obtained through eigenvalue decomposition of the covariance matrix derived from the normalized training data. Without loss of generality, these eigenvalues are sorted in descending order. The percentage of total variance explained by these principal components is defined as

$$\frac{\sum_{i=1}^{m}\lambda_i}{\sum_{i=1}^{n}\lambda_i}. \tag{30}$$

The maximal number of principal components that explain no more than 98% of total variance were calculated. Then the normalized data were projected by these principal components for dimensionality reduction. After that, GRC was employed to perform classification. The above procedure was repeated ten times and the average classification accuracies were calculated to evaluate the performance of GRC. The overall procedure of applying GRC in classification is summarized in Algorithm 2.

Algorithm 2 The Procedure of Applying GRC in Classification

Input: n training image samples $\boldsymbol{X}\in\mathbb{R}^{d\times n}$ and a test image sample $\boldsymbol{y}\in\mathbb{R}^{d}$, $s\in(0,2]$, $p\in(0,2]$.

Output: the identity of $\boldsymbol{y}$.

1. Preprocess $\boldsymbol{X}$ and $\boldsymbol{y}$ by z-score and PCA in order.
2. Solve the GRC problem (Algorithm 1):
$$\min_{\boldsymbol{\alpha}}\|\boldsymbol{y}-\boldsymbol{X}\boldsymbol{\alpha}\|_s^s+\lambda\|\boldsymbol{\alpha}\|_p^p$$
where $\boldsymbol{\alpha}$ is the representation coefficient.
3. Compute the residuals
$$r_i=\frac{\|\boldsymbol{y}-\boldsymbol{X}_i\widehat{\boldsymbol{\alpha}}_i\|_s}{\|\widehat{\boldsymbol{\alpha}}_i\|_p},\ i=1,2,\ldots,c$$
where $\widehat{\boldsymbol{\alpha}}_i$ is the representation coefficient associated with class i.
4. Output the identity of $\boldsymbol{y}$ as $\arg\min_i\{r_i\}$.

To evaluate the influence of each parameter on the classification performance of GRC, we changed one parameter at a time while keeping another parameter fixed.

Fig. 3 shows the classification accuracies of GRC with $s=[0.1:0.1:2.0]$ and $p=2.0$ on the four face databases. When $s\le 1.0$, the classification accuracies are very close to zero. When $s\ge 1.5$, the classification accuracies approach their peak values and remain stable across various s values. Therefore, to ensure a valid classification result, it is essential to configure the parameter s with a value no less than 1.5.

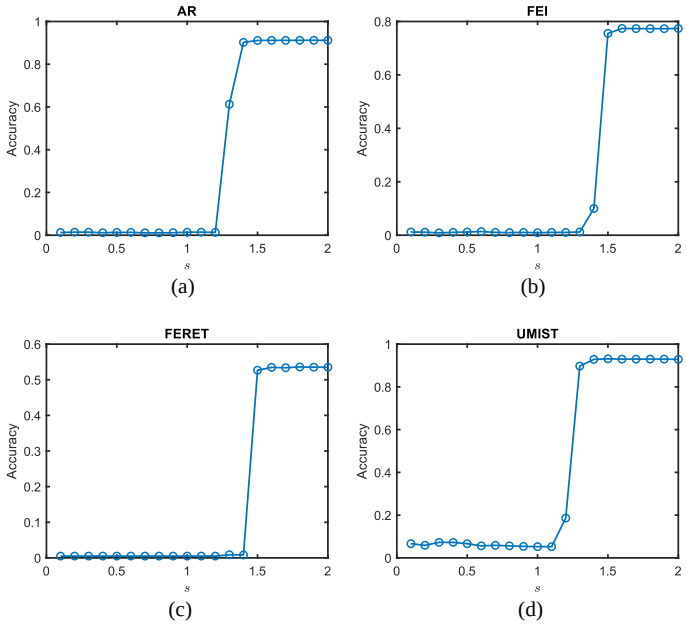

Fig. 3. Classification accuracies of GRC with $s=[0.1:0.1:2.0]$ and $p=2.0$ on (a) the AR face database, (b) the FEI face database, (c) the FERET face database, and (d) The UMIST face database.

To be specific, on the AR face database, the highest classification accuracy is 0.9121, obtained when s=1.9. On the FEI face database, the highest classification accuracy is 0.7739, obtained when s=1.6. On the FERET face database, the highest classification accuracy is 0.5360, obtained when s=1.8. On the UMIST face database, the highest classification accuracy is 0.9313, obtained when s=1.5. The optimal s value falls within the range of 1.5 to 1.9 across the four face databases.

Fig. 4 shows the classification accuracies of GRC with $s=2.0$ and $p=[0.1:0.1:2.0]$ on the four face databases. On the AR face database, the highest classification accuracy is 0.9185, obtained when p=1.6. On the FEI face database, the highest classification accuracy is 0.8111, obtained when p=1.1. On the FERET face database, the highest classification accuracy is 0.5622, obtained when p=1.3. On the UMIST face database, the highest classification accuracy is 0.9412, obtained when p=1.1. The optimal p value falls within the range of 1.1 to 1.6 across the four face databases.

To determine the optimal parameter pair (s,p) for GRC, we simultaneously fine-tuned both parameters. Fig. 5 shows the classification accuracies of GRC with $s=[0.1:0.1:2.0]$ and $p=[0.1:0.1:2.0]$ on the four face databases. When $1.5\le$

$s \leq 2.0$ and $0.9 \leq p \leq 2.0$, the classification accuracies are generally higher than the results in other cases. It is worth noting that most of these high-performance areas are within the solution domain of case 4, as illustrated in Fig. 1. This further solidifies our decision to utilize the solution from case 4 for addressing the GRC problem. Additionally, we had conducted experiments with the solutions from cases 1 to 3 within their applicable solution domains and found that the results closely resemble those shown in Fig. 5.

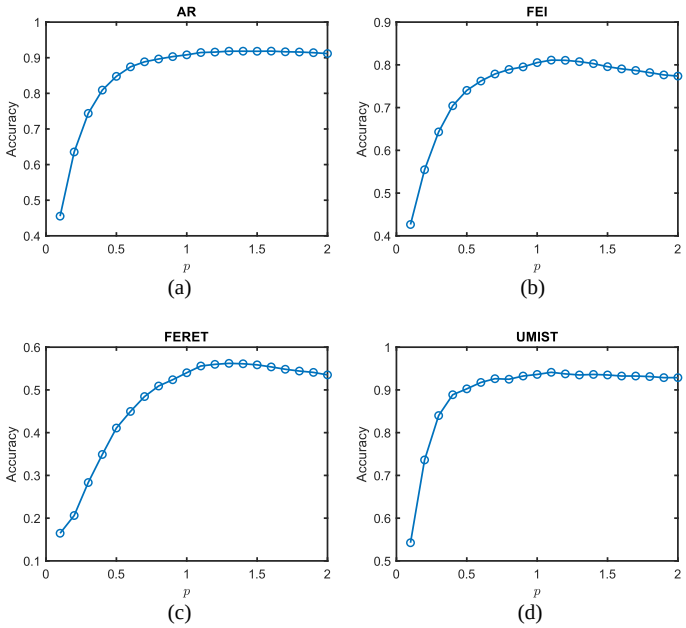

Fig. 4. Classification accuracies of GRC with $s = 2.0$ and $p = [0.1{:}\,0.1{:}\,2.0]$ on (a) the AR face database, (b) the FEI face database, (c) the FERET face database, and (d) The UMIST face database.

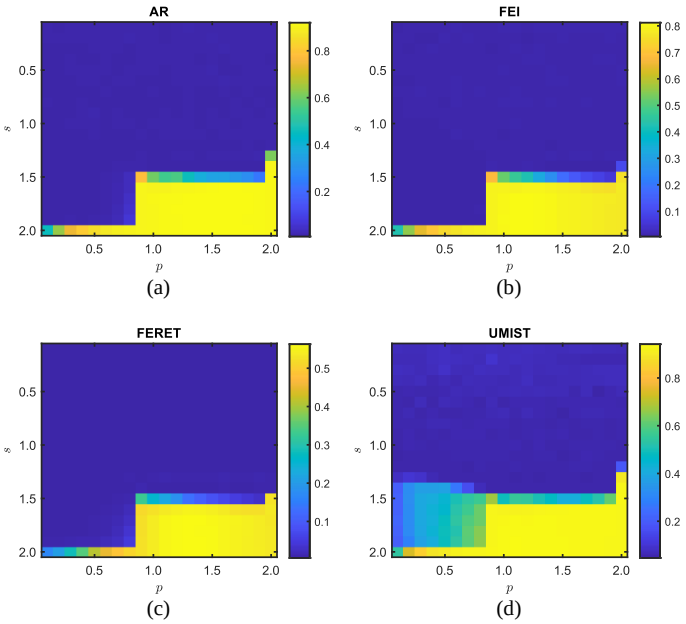

Fig. 5. Classification accuracies of GRC with $s = [0.1{:}\,0.1{:}\,2.0]$ and $p = [0.1{:}\,0.1{:}\,2.0]$ on (a) the AR face database, (b) the FEI face database, (c) the FERET face database, and (d) The UMIST face database.

On the AR face database, the highest classification accuracy of GRC is 0.9196, obtained when the parameter pair (s, p) equals (1.7, 1.6). When $s = 2.0$ and $p = 2.0$, GRC reduces to CRC, and the corresponding accuracy on this database is 0.9117. When $s = 2.0$ and $p = 1.0$, GRC reduces to SRC, and the corresponding accuracy on this database is 0.9081.

On the FEI face database, the highest classification accuracy of GRC is 0.8116, obtained when (s, p) equals (1.8, 1.2). The accuracies of CRC and SRC on this database are 0.7737 and 0.8053, respectively.

On the FERET face database, the highest classification accuracy of GRC is 0.5627, obtained when (s, p) equals (1.8, 1.3). The accuracies of CRC and SRC on this database are 0.5352 and 0.5402, respectively.

On the UMIST face database, the highest classification accuracy of GRC is 0.9412 obtained when (s, p) equals (2.0, 1.1). The accuracies of CRC and SRC on the UMIST face database are 0.9287 and 0.9363, respectively.

Table II lists the classification accuracies obtained by the three RC algorithms on the four face databases. The optimal parameter pairs for GRC are also presented in the table. In all cases, GRC with optimal parameters outperforms the other two algorithms. The results demonstrate that applying Lp-norm in both the objective and constraint functions of the two RC algorithms can improve their classification performances.

TABLE II
THE CLASSIFICATION ACCURACIES OF FOUR ALGORITHMS ON FOUR FACE DATABASES

Face database	CRC	SRC	GRC (s, p)
AR	0.9117	0.9081	**0.9196** (1.7, 1.6)
FEI	0.7737	0.8053	**0.8116** (1.8, 1.2)
FERET	0.5352	0.5402	**0.5627** (1.8, 1.3)
UMIST	0.9287	0.9363	**0.9412** (2.0, 1.1)

To further compare GRC with competing algorithms, we set the parameter pair (s, p) to the optimal one for each face database and changed the dimensionality reduced by PCA. To be specific, the optimal parameter pair (s, p) for GRC was chosen based on Table II. When applying PCA to reduce the dimensionality of normalized data, we reserved the principal components that explain about 90%, 95%, or 98% of total variance. Then, CRC, SRC, and GRC with the optimal parameter pair were applied to perform classification. As a representative NS approach, LRC [6] was also included for comparison. This procedure was repeated only three times. Finally, the average classification accuracies were calculated to evaluate the classification performances of the four RC algorithms.

Table III shows the classification accuracies of the four RC algorithms on the four face databases when principal components with different percentages of explained variances are extracted. The classification accuracies increase with the percentage of explained variance in most cases. On the FEI face database, when the principal components that explain 90% of total variance are extracted, LRC achieves the highest classification accuracy among the four competing algorithms. Similarly, on the FERET face database, when the principal components that explain 90% of total variance are extracted, SRC achieves the highest classification accuracy among the

four competing algorithms. Except for the two cases, GRC outperforms the other three algorithms. These results provide further evidence that introducing Lp-norm to the two RC algorithms improves their classification performances.

TABLE III
THE CLASSIFICATION ACCURACIES OF FOUR ALGORITHMS ON FOUR FACE DATABASES

Face database	Percentage of explained variance	LRC	CRC	SRC	GRC
AR	90%	0.6742	0.8131	0.8650	**0.8665**
	95%	0.7027	0.9050	0.9156	**0.9173**
	98%	0.7163	0.9215	0.9160	**0.9260**
FEI	90%	**0.7291**	0.4809	0.6932	0.6924
	95%	0.7325	0.6441	0.7529	**0.7679**
	98%	0.7420	0.7431	0.7985	**0.8135**
FERET	90%	0.2838	0.2767	**0.3440**	0.3180
	95%	0.3505	0.4422	0.4962	**0.4968**
	98%	0.3652	0.4968	0.5212	**0.5353**
UMIST	90%	0.9300	0.9013	0.9288	**0.9325**
	95%	0.9275	0.9213	0.9325	**0.9425**
	98%	0.9300	0.9325	0.9263	**0.9400**

It is worth noting that the optimal parameter pairs listed in Table II may not necessarily be optimal for the experiments in Table III. This indicates that there is still room for improving the classification performance of GRC. However, it should be acknowledged that tuning the parameters for GRC is a time-consuming process, particularly when striving to identify the optimal parameter pair through nested cross-validation. This represents a major limitation of GRC when contrasted with competing algorithms.

D. Experiment 2

To further demonstrate the superiority of GRC, we fixed the number of principal components and applied various RC algorithms to perform classification on the AR and FERET face databases.

For each subject in the AR face database, we randomly chose six images for training and four images for testing. Following this, we normalized the images by z-score and applied PCA on the normalized data to reduce its dimension to 54, 120, 200, and 300, as previously done in [16]. Finally, GRC with $s = [0.1{:}\,0.1{:}\,2.0]$ and $p = [0.1{:}\,0.1{:}\,2.0]$ was applied to perform classification. The procedure was repeated three times and the average classification accuracies were calculated to assess the performance of GRC.

Fig. 6 shows the classification accuracies across various parameter settings and numbers of principal components. When $1.5 \le s \le 2.0$ and $0.9 \le p \le 2.0$, the classification accuracies are significantly higher than those in other cases. The patterns appeared in Fig. 6 are consistent with those in Fig. 5.

Table IV lists the classification accuracies of the four algorithms on the AR face database when the dimension of images is reduced to 54, 120, 200, and 300. For GRC, only the highest classification accuracies are reported. Again, GRC greatly outperforms the other three algorithms. The optimal (s, p) pairs are (1.9, 1.1), (1.9, 1.2), (1.9, 1.4), and (1.8, 1.3) when the image dimensions are 54, 120, 200, and 300, respectively.

TABLE IV
THE CLASSIFICATION ACCURACIES OF FOUR ALGORITHMS ON THE AR FACE DATABASE

Dimension	54	120	200	300
LRC	0.6597	0.6854	0.6917	0.6951
CRC	0.8035	0.9028	0.9188	0.9160
SRC	0.8632	0.9139	0.9278	0.9139
GRC	**0.8701**	**0.9236**	**0.9313**	**0.9257**

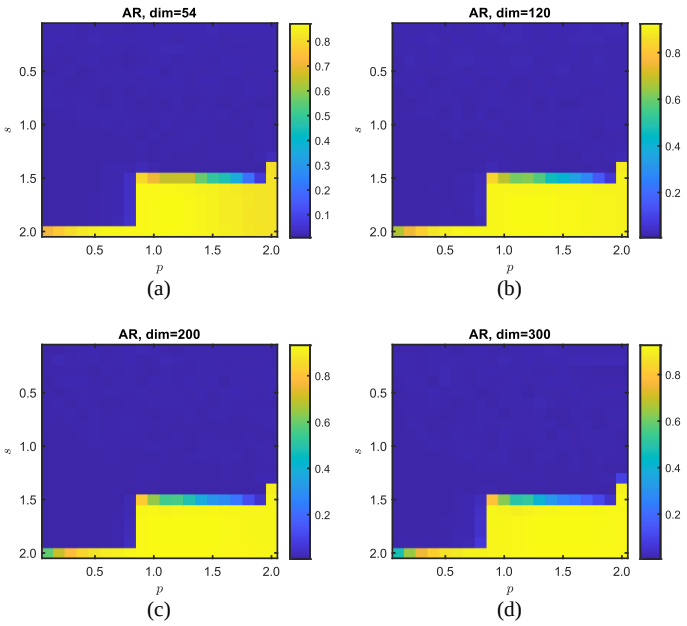

Fig. 6. Classification accuracies of GRC with s = [0.1:0.1:2.0] and p = [0.1:0.1:2.0] on the AR face database. The four subfigures correspond to the results when the dimension of images is reduced to (a) 54, (b) 120, (c) 200, and (d) 300 by PCA in order.

Fig. 7 shows the classification accuracies of the four algorithms when the parameter pair (s, p) is set to (1.8, 1.2) for GRC and the image dimension varies from 10 to 300. GRC and SRC generally outperform CRC. LRC obtains the worst classification performance among the four algorithms. When comparing GRC and SRC, GRC is slightly better. Since the optimal parameter pair for GRC varies with the dimension of images, changing the (s, p) values may yield better accuracy results on this database. In summary, introducing Lp-norm into existing RC algorithms improves their classification performances.

Then we proceeded to conduct a similar experiment on the FERET face database. For each subject in the FERET face database, we randomly chose four images for training and use the remaining three images for testing. The training images were normalized by z-score so that each feature was centered to have a mean of zero and scaled to have a standard deviation of one. Then the testing images were normalized by applying the same parameters. Following this, we applied PCA to reduce the dimension of normalized data to 54, 120, 200, and 300. Finally, GRC with $s = [0.1{:}\,0.1{:}\,2.0]$ and $p = [0.1{:}\,0.1{:}\,2.0]$ was applied to perform classification. This procedure was repeated three times and the average classification accuracies were calculated to assess the performance of GRC.

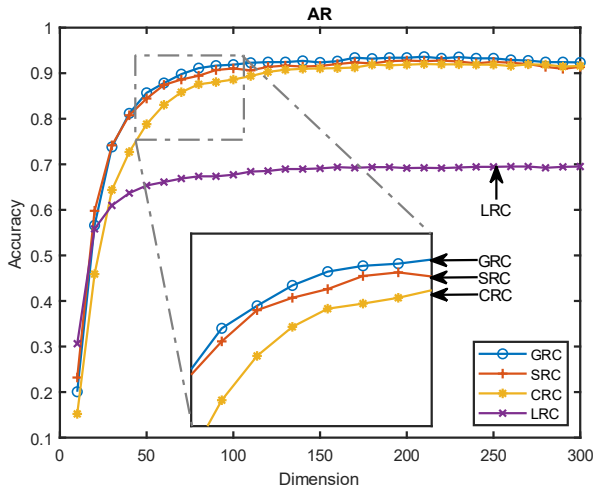

Fig. 7. Classification accuracies of the four algorithms on the AR face database when the parameter pair (s, p) is set to (1.8, 1.2) for GRC and the dimension of images varies from 10 to 300.

Fig. 8 shows the classification accuracies across various parameter settings and numbers of principal components. When $1.5 \leq s \leq 2.0$ and $0.9 \leq p \leq 2.0$, the classification accuracies are significantly higher than those in other cases. These results closely resemble the results obtained on the AR face database, as shown in Fig. 6.

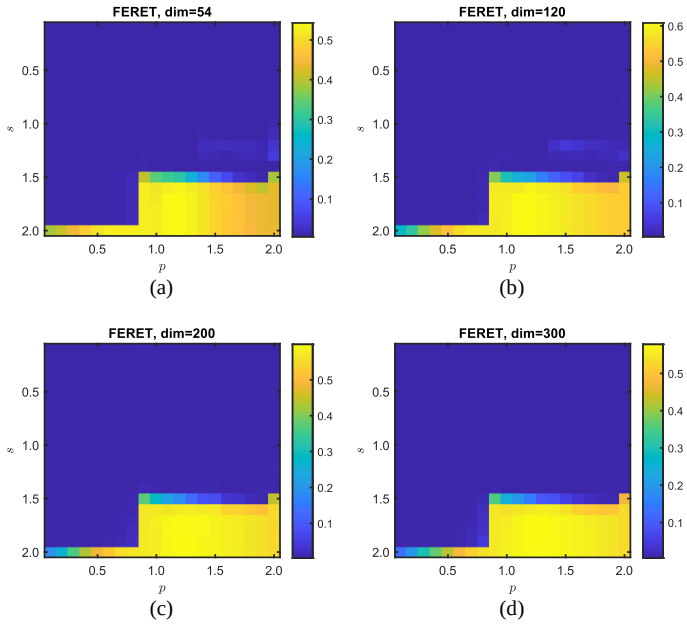

Fig. 8. Classification accuracies of GRC with s = [0.1:0.1:2.0] and p = [0.1:0.1:2.0] on the FERET face database. The four subfigures correspond to the results when the dimension of images is reduced to (a) 54, (b) 120, (c) 200, and (d) 300 by PCA in order.

Table V lists the classification accuracies of the four algorithms on the FERET face database when the dimension of images is reduced to 54, 120, 200, and 300. For GRC, only the highest classification accuracies are reported. GRC consistently outperforms the other three algorithms. The optimal (s, p) pairs are (1.8, 1.2) in all four cases.

TABLE V
THE CLASSIFICATION ACCURACIES OF FOUR ALGORITHMS ON THE FERET FACE DATABASE

Dimension	54	120	200	300
LRC	0.3828	0.4122	0.4211	0.4217
CRC	0.4411	0.5356	0.5467	0.5294
SRC	0.5289	0.5889	0.5767	0.5606
GRC	**0.5428**	**0.6083**	**0.5989**	**0.5778**

Fig. 9 shows the classification accuracies of the four algorithms when the parameter pair (s, p) of GRC is set to (1.8, 1.2) and the image dimension varies from 10 to 300. GRC generally outperforms the other three algorithms. SRC is slightly worse than GRC. Both GRC and SRC outperform CRC. LRC obtains the worst classification performance among the four algorithms. The results further demonstrate that introducing Lp-norm into the RC algorithms improves their classification performances.

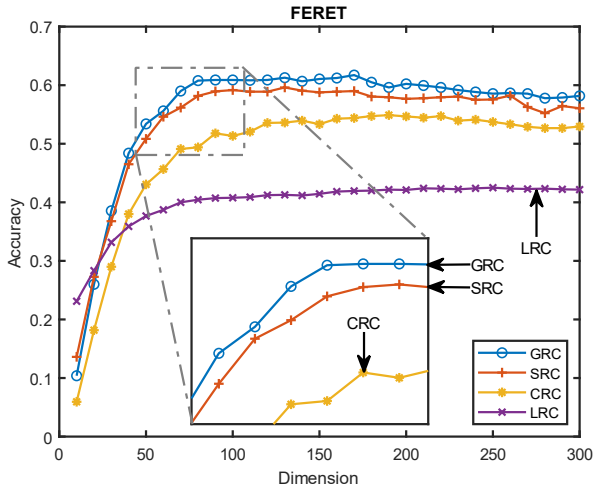

Fig. 9. Classification accuracies of the four algorithms on the FERET face database when the parameter pair (s, p) is set to (1.8, 1.2) for GRC and the dimension of images varies from 10 to 300.

Data and source code of this study are publicly available on the GitHub repository at https://github.com/yuzhounh/GRC.

VI. DISCUSSION

The proposed algorithm offers a novel solution for the well-known L1-norm minimization problem [9], which is different from existing solutions [46] such as interior-point methods, homotopy methods, the Fast Iterative Soft-thresholding algorithm (FISTA), proximal-point methods, Parallel Coordinate Descent (PCD), Approximate Message Passing (AMP), Templates for Convex Cone Solvers (TFOCS), etc. Furthermore, the scope of the Lp-norm minimization problem addressed in this paper is more extensive than that of the L1-norm minimization problem, thus carrying substantial theoretical significance.

Two major aspects of this study remain uncertain. Firstly, an unresolved challenge is to design a solution for GRC beyond the solution domains illustrated in Fig. 1. Secondly, how to determine the theoretically optimal (s, p) pair remains unclear. In practice, the optimal parameter settings vary with

the specific face database and image dimension. Moreover, the computational resources required for identifying the optimal parameter pair through nested cross-validation are notably intensive. Hence, it is imperative to develop a method that can theoretically determine the optimal parameters for GRC in the future work.

VII. CONCLUSION

This paper proposes the Generalized Representation-based Classification (GRC) algorithm, which extends traditional RC algorithms by incorporating Lp-norm in both the objective function and the constraint function. To solve the optimization problem of GRC, an iterative procedure was designed within the MM framework, ensuring that a closed-form solution is obtained in each iteration and a locally optimal solution can be achieved upon convergence. To assess the effectiveness of GRC, we compared it with other prevalent RC algorithms on four benchmark face databases. The experimental results consistently demonstrated that GRC with optimal parameters outperforms competing algorithms. Therefore, incorporating Lp-norm into traditional RC algorithms can improve their classification performances.

REFERENCES

[1] Z. Zhang, Y. Xu, J. Yang, X. Li, and D. D. Zhang, "A Survey of Sparse Representation: Algorithms and Applications," *IEEE Access*, vol. 3, pp. 490-530, 2015.

[2] J. Zhou, S. Zeng, and B. Zhang, "Linear Representation-Based Methods for Image Classification: A Survey," *IEEE Access*, vol. 8, pp. 216645-216670, 2020.

[3] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. Wiley, 2001.

[4] S. Z. Li and J. Lu, "Face recognition using the nearest feature line method," *IEEE Transactions on Neural Networks*, vol. 10, no. 2, pp. 439-43, 1999.

[5] M. Dyrholm and L. C. Parra, "Smooth bilinear classification of EEG," in *2006 International Conference of the IEEE Engineering in Medicine and Biology Society*, 2006: IEEE, pp. 4249-4252.

[6] I. Naseem, R. Togneri, and M. Bennamoun, "Linear Regression for Face Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 11, pp. 2106-2112, 2010.

[7] Y. Li, Y. Yang, J. Che, and L. Zhang, "Predicting the number of nearest neighbor for kNN classifier," *IAENG International Journal of Computer Science*, vol. 46, no. 4, pp. 662-669, 2019.

[8] C. Wang and Y. Yang, "Nearest Neighbor with Double Neighborhoods Algorithm for Imbalanced Classification," *IAENG International Journal of Applied Mathematics*, vol. 50, no. 1, pp. 147-159, 2020.

[9] J. Wright, A. Y. Yang, A. Ganesh, S. Sastry, S., and Y. Ma, "Robust Face Recognition via Sparse Representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210-227, 2009.

[10] J. Peng *et al.*, "Low-Rank and Sparse Representation for Hyperspectral Image Processing: A review," *IEEE Geoscience and Remote Sensing Magazine*, vol. 10, pp. 10-43, 2022.

[11] H. Wang and T. Çelik, "Sparse representation-based hyperspectral image classification," *Signal, Image and Video Processing*, vol. 12, pp. 1009 - 1017, 2018.

[12] M. Meng, X. Yin, Q. She, Y. Gao, W. Kong, and Z. Luo, "Sparse representation-based classification with two-dimensional dictionary optimization for motor imagery EEG pattern recognition," *Journal of Neuroscience Methods*, vol. 361, 2021.

[13] S. R. Sreeja and D. Samanta, "Dictionary reduction in sparse representation-based classification of motor imagery EEG signals," *Multimedia Tools and Applications*, 2023.

[14] S. Sheykhivand, T. Y. Rezaii, Z. Mousavi, A. Delpak, and A. Farzamnia, "Automatic Identification of Epileptic Seizures From EEG Signals Using Sparse Representation-Based Classification," *IEEE Access*, vol. 8, pp. 138834-138845, 2019.

[15] M. Kong, Y. Zhang, D. Xu, W. Chen, and M. Dehmer, "FCTP-WSRC: Protein–Protein Interactions Prediction via Weighted Sparse Representation Based Classification," *Frontiers in Genetics*, vol. 11, 2020.

[16] L. Zhang, D. Meng, and X. Feng, "Sparse Representation or Collaborative Representation: Which Helps Face Recognition?," in *International Conference on Computer Vision*, 2012, pp. 471-478.

[17] L. Xun, J. Zhang, D. Cao, J. Wang, S. Zhang, and F. Yao, "Mapping cotton cultivated area combining remote sensing with a fused representation-based classification algorithm," *Comput. Electron. Agric.*, vol. 181, p. 105940, 2021.

[18] G. Marjanovic and V. Solo, "On Lq Optimization and Matrix Completion," *IEEE Transactions on Signal Processing*, vol. 60, pp. 5714-5724, 2012.

[19] Z. Liang, S. Xia, Y. Zhou, L. Zhang, and Y. Li, "Feature extraction based on Lp-norm generalized principal component analysis," *Pattern Recognit. Lett.*, vol. 34, pp. 1037-1045, 2013.

[20] N. Kwak, "Principal component analysis by Lp-norm maximization," *IEEE Transactions on Cybernetics*, vol. 44, no. 5, pp. 594-609, 2014.

[21] J. Wang, "Generalized 2-D Principal Component Analysis by Lp-Norm for Image Analysis," *IEEE Transactions on Cybernetics*, vol. 46, no. 3, pp. 792-803, 2016.

[22] H. Wang and J. Wang, "2DPCA with L1-norm for simultaneously robust and sparse modelling," *Neural Networks*, vol. 46, pp. 190-198, 2013.

[23] Q. Ye, L. Fu, Z. Zhao, H. Zhao, and N. Meem, "Lp- and Ls-Norm Distance Based Robust Linear Discriminant Analysis," *Neural Networks*, vol. 105, pp. 393-404, 2018.

[24] B. Cnla, A. Yhs, W. C. Zhen, and D. Nyd, "Robust bilateral Lp-norm two-dimensional linear discriminant analysis," *Information Sciences*, vol. 500, pp. 274-297, 2019.

[25] N. Fang and H. Wang, "Generalization of Local Temporal Correlation Common Spatial Patterns Using Lp-norm ($0 < p < 2$)," 2017, in Neural Information Processing, pp. 769-777.

[26] Q. Cai, W. Gong, Y. Deng, and H. Wang, "Single-Trial EEG Classification via Common Spatial Patterns with Mixed Lp- and Lq-Norms," *Mathematical Problems in Engineering*, vol. 2021, no. 6645322, 2021.

[27] W. Zuo, D. Meng, L. Zhang, X. Feng, and D. Zhang, "A Generalized Iterated Shrinkage Algorithm for Non-convex Sparse Coding," in *International Conference on Computer Vision*, 2013, pp. 217-224.

[28] Q. Zhu, N. Xu, S. J. Huang, J. Qian, and D. Zhang, "Adaptive feature weighting for robust Lp-norm sparse representation with application to biometric image classification," *International journal of machine learning and cybernetics*, vol. 11, no. 2, pp. 463-474, 2020.

[29] K. A. Toh, G. Molteni, and Z. Lin, "Deterministic Bridge Regression for Compressive Classification," 2022.

[30] S. Wang, C. Fang, J. Gu, and J. Fang, "Cancer classification using collaborative representation classifier based on non-convex Lp-norm and novel decision rule," in *Seventh International Conference on Advanced Computational Intelligence*, 2015.

[31] R. Li, L. Tang, Y. Bai, Q. Wang, X. Zhang, and M. Liu, "Group-Based Sparse Representation Based on Lp-Norm Minimization for Image Inpainting," *IEEE Access*, vol. 8, pp. 60515-60525, 2020.

[32] R. Li, Y. Bai, X. Zhang, T. Lan, and Q. Wang, "Group-Based Sparse Representation Based on Lp-norm Minimization for Compressive Sensing," 2020.

[33] D. R. Hunter and K. Lange, "A Tutorial on MM Algorithms," *American Statistician*, vol. 58, no. 1, pp. 30-37, 2004.

[34] J. Wang, C. Lu, M. Wang, P. Li, S. Yan, and X. Hu, "Robust Face Recognition via Adaptive Sparse Representation," *IEEE Transactions on Cybernetics*, vol. 44, pp. 2368-2378, 2014.

[35] S. Cai, L. Zhang, W. Zuo, and X. Feng, "A Probabilistic Collaborative Representation Based Approach for Pattern Classification," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2950-2959, 2016.

[36] J. Gou, B. Hou, W. Ou, Q.-r. Mao, H. Yang, and Y. Liu, "Several robust extensions of collaborative representation for image classification," *Neurocomputing*, vol. 348, pp. 120-133, 2019.

[37] A. Majumdar and R. K. Ward, "Classification via group sparsity promoting regularization," *2009 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 861-864, 2009.

[38] M. Yuan and Y. Lin, "Model selection and estimation in regression with grouped variables," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 68, 2006.

[39] J. Huang, F. Nie, H. Huang, and C. Ding, "Supervised and Projected Sparse Coding for Image Classification," *Proceedings of the AAAI Conference on Artificial Intelligence*, 2013.
[40] X.-T. Yuan and S. Yan, "Visual classification with multi-task joint sparse representation," *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 3493-3500, 2010.
[41] S. Shekhar, V. M. Patel, N. M. Nasrabadi, and R. Chellappa, "Joint Sparse Representation for Robust Multimodal Biometrics Recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, pp. 113-126, 2014.
[42] J. Yang, L. Luo, J. Qian, Y. Tai, F. Zhang, and Y. Xu, "Nuclear Norm Based Matrix Regression with Applications to Face Recognition with Occlusion and Illumination Changes," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, pp. 156-171, 2014.
[43] S. Xiao, "Blind Image Restoration Based on l_1-l_2 Blur Regularization," *Engineering Letters*, vol. 28, no. 1, pp. 148-154, 2020.
[44] S. Xiao, "Image Restoration Algorithm Based on Double L0-Regularization and ALM," *Engineering Letters*, vol. 31, no. 1, pp. 52-65, 2023.
[45] X. Ding, T. Wei, Y. Wang, and J. Li, "Low Rank Tensor Completion via Schatten-2/3 Norm and Schatten-1/2 Norm," *IAENG International Journal of Applied Mathematics*, vol. 53, no. 1, pp. 317-328, 2023.
[46] A. Y. Yang, Z. Zhou, A. Ganesh, S. S. Sastry, and Y. Ma, "Fast L1-Minimization Algorithms For Robust Face Recognition," *IEEE Transactions on Image Processing*, vol. 22, no. 8, pp. 3234-3246, 2010.
[47] A. Martínez and R. Benavente, "The AR Face Database," *CVC Technical Report*, vol. 24, pp. 1-8, 1998.
[48] C. E. Thomaz and G. A. Giraldi, "A new ranking method for principal components analysis and its application to face image analysis," *Image Vis. Comput.*, vol. 28, pp. 902-913, 2010.
[49] P. J. Phillips and H. Moon, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090-1104, 2000.
[50] D. B. Graham and N. M. Allinson, "Characterising Virtual Eigensignatures for General Purpose Face Recognition," 1998.
[51] J. Shlens, "A Tutorial on Principal Component Analysis," *ArXiv*, vol. abs/1404.1100, 2014.
[52] I. T. Jolliffe and J. Cadima, "Principal component analysis: a review and recent developments," *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, vol. 374, 2016.