

Research Article

Linear Discriminative Learning for Image Classification

Rab Nawaz Jadoon ^{1,2}, Waqas Jadoon,¹ Ahmad Khan,¹ Zia ur Rehman,¹ Sajid Shah,¹ Iftikhar Ahmed Khan ¹ and WuYang Zhou ²

¹Department of Computer Science, COMSATS University Islamabad, Abbottabad Campus, KPK, Pakistan

²School of Information Science and Technology, University of Science and Technology of China, Hefei, China

Correspondence should be addressed to WuYang Zhou; wyzhou@ustc.edu.cn

Received 27 May 2019; Revised 26 August 2019; Accepted 20 September 2019; Published 20 October 2019

Guest Editor: Marco Perez-Cisneros

Copyright © 2019 Rab Nawaz Jadoon et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

In this paper, we propose a linear discriminative learning model called adaptive locality-based weighted collaborative representation (ALWCR) that formulates the image classification task as an optimization problem to reduce the reconstruction error between the query sample and its computed linear representation. The optimal linear representation for a query image is obtained by using the weighted regularized linear regression approach which incorporates intrinsic locality structure and feature variance between data into representation. The resultant representation increases the discrimination ability for correct classification. The proposed ALWCR method can be considered an extension of the collaborative representation- (CR-) based classification approach which is an alternative to the sparse representation- (SR-) based classification method. ALWCR improved the discriminant ability for classification as compared with CR original formulation and overcomes the limitations that arose due to a small training sample size and low feature dimension. Experimental results obtained using various feature dimensions on well-known publicly available face and digit datasets have verified the competitiveness of the proposed method against competing image classification methods.

1. Introduction

The variations in face images such as illumination change, random noise, orientation, and different facial expressions make the face recognition (FR) difficult and challenging. Recognition in the original image space degrades the accuracy and leads to computational complexity due to the curse of dimensionality. Different approaches such as eigenface [1], fisherface, and nearest subspace [2] have been devised to reduce the dimensionality and perform the classification in reduced feature space. In these methods, it is assumed that the same class of faces can be linearly mapped to a low-dimensional subspace [3]. However, the illumination change, noise, shadows, orientation, and depth of the face images often violate this subspace assumption. Conventional methods like principal component analysis (PCA) fail to achieve the desired results due to a gross error of arbitrary magnitude

exhibited in real-world face images [4]. In recent years, pattern classification based on a sparse representation (SR) has attracted interest due to the superior results of sparse signal coding in the field of image reconstruction [3, 5]. Generally, the sparse representation is computed by representing the query sample as the linear combination of bases (involving sparse samples) over training dictionary, by employing sparsity constraint computed by l_1 minimization. It has been claimed that the robustness and accuracy for the classification is considerably improved due to sparsity constraint for computing representation [3, 5].

An image-based general classification algorithm called the sparse representation for classification (SRC) is proposed by Wright et al. [3]. The sparsest linear representation is recovered via l_1 minimization. The resulting representation automatically discriminates among various classes by using minimum reconstruction error. The

robustness of SRC to change in illumination, occlusion, and noise has diverted the research towards image classification in the sparse domain. However, several recent studies [6, 7] questioned the role of SR in robust classification, and the working mechanism is not fully explained yet. Most of the sparse representation techniques focus on the sparsity constraints of sample representation computed via l_1 minimization and their aim is to improve the sparsest representation for the query sample. Weighted sparse representation-based classification (WSRC), an extension of SRC is proposed by Lu et al. [8] to improve discriminative ability for classification. The method overcomes the limitations that arise due to the linearity structure in low-dimensional subspace in traditional SRC formulation by employing data locality into the sparse representation. Gao et al. [9] have used locality preserving projection for linear approximation of high-dimensional data residing on a manifold and then used the sparse representation for hyperspectral image classification, to address the imbalance between high-dimensionality and limited training samples. A feature-based structural measure (FSM) is proposed by Shnain et al. [10]. The measure estimates a reliable similarity between similar human face images and discrimination among dissimilar images by incorporating statistical structure and edges as structural features. Iqbal et al. [11] have proposed a collaborative representation-based classification method for image classification. The method utilizes virtual samples and computed representation on local image patches to address the limited sample face recognition problem. Jiang et al. [12] have introduced a low-dimensional data representation model by integrating LLE and PCA with a closed-form solution. The resulting representation incorporates both local neighborhood structure and global structure and is robust to outliers. An improved principal component analysis (IPCA) method for face recognition was proposed by Zhu et al. [13]. The approach extracts useful features from the original feature space (original face images) through dimensionality reduction and then employed the least square regression approach for classification by considering the minimum reconstruction error. The experimental results show promising results compared with some state-of-the-art methods.

Many recent studies have questioned the role of l_1 minimization-based sparsity constraint in the classification and showed that the data generally used in various SR-based classification methods are unable to hold the sparsity assumption presumed [7, 14]. Zhang et al. [7] have argued the crucial role of representing the query sample as the linear combination by using all samples over the training dictionary hence called collaborative representation (CR). A simple l_2 minimization-based least square CR representation achieves competitive classification with less computational complexities as compared with sparse representation approaches. Solving the SR problem to learn large representation is considered computationally expensive with unknown algebraic solutions [6].

In dimensionality reduction, classification and clustering the proximity relationship (intrinsic locality structure) among data elements is important [3, 15]. The existing CR methods encode a new sample as a linear combination of previously learned features irrespective of data point proximity relationship which leads to poor classification [16].

In this work, a collaborative representation method is proposed to formulate the image classification problem to compute optimal linear representation (computed via weighted variant of l_2 minimization problem which having a closed form of algebraic solution) against the query image. The resulting representation scheme incorporates proximity relationship and distinctive data features into linear representation.

2. Related Work

Linear regression-based classification techniques have gained a lot of interest in recent years. The classification problem is formulated to find the coefficients that minimize the reconstruction error between a probe image and its linear reconstruction from an over-complete basis.

For a classification problem, suppose there are n number of samples chosen from known classes such that the matrix $A_k = [a_{k1}, \dots, a_{kn}] \in \mathfrak{R}^{m \times n_k}$ is formed by using training samples from the k th class where a_{kj} represents the column vector formed from the j th sample (image) of dimension m . The training matrix $A = [A_1, \dots, A_C] \in \mathfrak{R}^{m \times n}$ is formed by concatenating samples from C known classes having m number of rows and n columns where $n = \sum_{k=1}^C n_k$ shows the number of samples in training dictionary.

The nearest subspace (NS) [17] classifier looks for the minimal distance between sample Y and its principal projection. The assignment of the sample to the k th class is performed on the basis of minimal reconstruction error r_k between Y and subspace spanned by data points of class A_k , that is,

$$r_k = \arg \min_k \|Y - A_k x_{Is}\|_2^2, \quad (1)$$

where x_{Is} is a coefficient vector which is estimated using the least square regression method, that is, $x_{Is} = (A_k^T A_k)^{-1} A_k^T Y$, $k = 1, 2, \dots, c$

In the sparse representation-based classification (SRC) [3] technique, which is a generalization of the NS method, each probe image is represented as the linear combination of all the training samples from an over-complete dictionary resulting in an ill-condition inverse problem. The solution is obtained by harnessing the sparsity of coefficient's vector. That is, only a small number of solution vector elements have nonzero values. The resulting sparse representation (SR) scheme has shown very interesting robust face recognition performance.

The concept of collaborative representation (CR) has been introduced more recently in various studies [7, 17–19]. The role of sparse representation in robust face recognition and pattern classification has raised many questions and there is still no clear explanation why sparsity constraints on

coefficient vector produce robust face recognition results. However, an over-complete dictionary leads to infinite representations for the query sample. Since the query sample representation is formed from the collaboration of all samples in the dictionary, hence, the representation is referred as collaborative representations (CRs). Furthermore, it has been shown that the adoption of CRs has been more crucial for robust classification than the sparsity constraints using l_1 -regularization on coefficients. Adoption of CRs based on regularized least square (l_2 -regularization) to compute representation will lead to similar performance with much lower computational complexity in the classification.

In collaborative representation- (CR-) based classification [7, 14, 17] framework, the representation of a query sample Y is obtained by coding it collaboratively over the dictionary formed by using all training samples. The classification is considered as an optimization problem that aims to reduce the reconstruction cost between a test sample $Y \in \mathfrak{R}^m$ and the corresponding representation. The sparse representation-based classification (SRC) [3] approach is one example of CR-based classifiers. SRC assumed that the sparsest representation of the query image Y over fixed sized training dataset A will correctly classify the query sample Y . The sparse representation of A is obtained by l_1 minimization:

$$x = \arg \min_x \|Y - Ax\|_2^2 + \lambda \|x\|_1, \quad (2)$$

where $\|\cdot\|_1$ represent the sparsity constraints of coefficient vector x while λ is a small positive constant used to control the sparsity level. The resulting optimization problem is solved by using linear programming with unknown algebraic solution. Computing the sparse representation x , using equation (2), then the SRC will assign the query image to the class having the minimum reconstruction error $r_k(Y)$ among all classes:

$$r_k(Y) = \|Y - A\sigma_k(x)\|_2, \quad (3)$$

where $\sigma_k(x) \in \mathfrak{R}^n$ is an n -dimensional vector. The nonzero coefficients of $\sigma_k(x)$ corresponds to the k th training class. The classification is performed subject to the following condition:

$$\text{identity}(Y) = \arg \min_x r_k(Y). \quad (4)$$

Liu et al. [8] have pointed out the performance degradation of SRC in a lower dimensional feature subspace is caused by limited discriminative information due to the linearity structure of data. To overcome this limitation, they proposed the weighted sparse representation-based classification (WSRC) method to integrate data locality and sparsity into formulation, that is,

$$(\text{weighted } l_1) : x = \arg \min_x \|Wx\|_1 \text{ subject to } Y = Ax, \quad (5)$$

where W is a block diagonal matrix that is used to characterize the similarity structure between test sample and training data and is calculated as follows:

$$\text{Diag}(W) = [\text{dist}(Y, a_1^1), \dots, \text{dist}(Y, a_n^C)]^T, \quad (6)$$

where $\text{dist}(y, a_i^C) = \text{norm}\|Y - a_i^C\|^s$, s is the locality parameter. After obtaining more discriminate sparse code of the test sample representation, the recognition is performed using minimum reconstruction error.

Another approach for CR is to utilize the regularized least squares (CRC-RLS) [7] method. A query sample is encoded from all the training samples of dictionary A using l_2 minimization. CRC handles the classification as an optimization problem and reduces the reconstruction cost between query image Y and its estimated representation. The coding model of CRC irrespective of outlier robustness is given by:

$$x = \arg \min_x \|Y - Ax\|_2^2 + \lambda \|x\|_2^2. \quad (7)$$

Equation (7) has the following closed-form solution:

$$x = (A^T A + \lambda I)^{-1} A^T Y, \quad (8)$$

where λ is the regulatory parameter and $I \in \mathfrak{R}^{n \times n}$ represents the identity matrix. The classification decision is based on the minimum regularized reconstruction error computed as follows:

$$r_k(Y) = \frac{\|Y - A_k \sigma_k(x)\|_2}{\|\sigma_k(x)\|_2}. \quad (9)$$

The classification decision is taken as in equation (4).

For classification, the CRC performs best, in terms of computational efficiency, compared to SR-based approaches, see [14] for detail. It is the collaborative representation (CR) which makes CRC robust to classification and enhances its discrimination power. In the classification decision of CR, the minimum distance between collaborative components and the projection within the class is considered:

$$r_k = \|\tilde{Y} - A_k x\|_2, \quad (10)$$

where $\tilde{Y} = \sum_{k=1}^c A_k x_k$ represents the perpendicular projection of the query sample on subspace span by training dictionary. In query sample classification decision, first CR checks the angle between the perpendicular projection and the principal signal (i.e., representation using coefficients from the correct class). This angle should be smaller among all other classes. Secondly, CR also checks the angle between the principal signal and the rest of the classes (other than correct class) should be larger. Equation (10) is the actual amount that plays an important role in classification. This double checking mechanism makes CR very effective and robust for classification problems.

3. Proposed Methodology

It has been observed in various research works [15, 20, 21] that the locality structure among data elements plays an important role for classification. In low-dimensional subspace, a reconstructed data point and its nearest

neighbors for the nonlinear manifold will preserve the local geometry. Hence, the robust classification of observations (data points) can be achieved by modeling the nonlinear geometry of the manifolds. Furthermore, the discriminate representation can be enhanced by exploiting the distinctiveness of various features among data elements for accurate classification [15].

Motivated by the aforementioned observations, this paper integrates locality structures and feature distinctiveness among data points to compute linear representation. The aim is to learn discriminant representation by exploiting the distinctiveness of various features to reduce variations among data elements of similar class and using locality information to distinguish data elements between dissimilar classes. We propose an image classification method called adaptive locality-based weighted collaborative representation (ALWCR) with the ability to reconstruct the query image from optimal neighbors to further discriminate the classification. The proposed model is formulated as a unified weighted least square regularization:

$$x_{\text{ALWCR}} = \arg \min_x \left\{ \frac{1}{2} \| (Y - Ax)\Omega \|_2^2 + \frac{1}{2} \lambda \sum_{i=1}^n x_i^2 Q_i \right\}, \quad (11)$$

where $A = [a_1, \dots, a_n]$ represent the samples of the training dictionary A ; $Q_i = \|Y - a_i\|_2^2$, $i = 1, \dots, n$, represent the locality structure similarities based on Euclidean distance between query Y and the i th sample in dictionary A ; and $\Omega \in \mathfrak{R}^{m \times m}$ is a diagonal matrix which describes the weight calculated for the m th feature (Y_m) to mitigate the importance in representation. Intuitively, the solution coefficients for the computed representation x_{ALWCR} vary if Y_m is dissimilar to other features and suppress the coefficients of similar features; λ is a small positive constant and act as a regularization parameter.

The proposed problem formulation, i.e., equation (11), has an algebraic solution using fixed Ω values [22]. If there are sufficient training samples available, then the matrix Ω is precalculated using variance among training samples. The coefficients of solution vector x_{ALWCR} represent the contribution of samples chosen from the training dictionary during reconstruction of the query sample. The higher coefficient values represent the neighboring samples which span the same linear subspace as span by query sample Y and mitigate the importance of feature dimension. Intuitively, the first term is used to reduce the variance among data element of similar classes, and the weight matrix (Ω) is introduced for variance estimation. These weights are prelearned from training data by using the weighted least square regression (WLS) approach. For correct noise estimation, each training class should have enough samples to model various variations such as pose, illumination, and outliers among query samples. Therefore, if the training dictionary covers enough variations among testing and training samples, then noise is correctly estimated using weights which improves discrimination for classification. The second term penalizes those

training samples which lie on different subspaces than the test sample. Therefore, training sample contribution in the linear representation of test samples (the training samples that lie on different subspaces as compared with the test sample) is reduced, which decreases the overall reconstruction error and improves classification results. In term of training samples, the test sample Y can be linearly represented as $Y = Ax + \xi$, where ξ represents the noise term. According to weighted least square (WSL) regression [23] for noise estimation,

$$x_{wls} = \arg \min_x (Y - Ax)^T \Omega_{wls} (Y - Ax), \quad (12)$$

where Ω_{wls} represents the diagonal matrix that is proportional to covariance matrix ξ . The algebraic solution of equation (12) is given as

$$x_{wls} = (A^T \Omega_{wls} A)^{-1} A^T \Omega_{wls} Y. \quad (13)$$

An iterative procedure [24] can be used to estimate Ω_{wls} . First, a least square optimization is used to compute the residual r_{ls} while Ω_{wls} (in (13)) can be initialized as a diagonal matrix to $\sqrt{|r_{ls}|}$,

$$\begin{cases} r_{ls} = Y - Ax_{ls}, \\ x_{ls} = (A^T A)^{-1} A^T Y, \\ \Omega_{wls} = \text{diag}(\sqrt{|r_{ls}|}). \end{cases} \quad (14)$$

In the second iteration, the coefficients x_{wls} are estimated using equation (13), then the weighted least square residual r_{wls} is computed, and the weight matrix Ω_{wls} is updated to Ω_{wls1} :

$$\begin{cases} r_{wls} = Y - Ax_{wls}, \\ \Omega_{wls1} = \text{diag}(\sqrt{|r_{wls}|}). \end{cases} \quad (15)$$

Similarly, the next iteration leads to

$$\begin{cases} r_{wls1} = Y - Ax_{wls1}, \\ \Omega_{wls2} = \text{diag}(\sqrt{|r_{wls1}|}), \end{cases} \quad (16)$$

where r_{wls} represents the weighted least square residual at the 3rd iteration of the procedure and x_{wls1} is the coefficient vector computed using equation (13) by substituting Ω_{wls1} . Finally, Ω_{wls2} represents the updated weight matrix at the end of the current iteration. Note that Ω_{wls} will converge after some iterations.

3.1. Optimization Algorithm. First, the weights Ω are learned, with respect to query sample Y , using (13)–(16). The algebraic solution of equation (11) can be determined by its partial differentiation w.r.t x and equating to zero:

$$x_{\text{ALWCR}} = -A^T \Omega (Y - Ax) + \lambda Q \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = 0, \quad (17)$$

where

$$\Omega = \begin{pmatrix} \sigma_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_m \end{pmatrix} \in \mathfrak{R}^{m \times m}, \quad (18)$$

$$Q = \begin{pmatrix} \|Y - a_1\|_2^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \|Y - a_n\|_2^2 \end{pmatrix} \in \mathfrak{R}^{n \times n}. \quad (19)$$

The term x_i in equations (11) and (17) represents a coefficient vector having positive values. A reasonable solution of equation (17) can be estimated by replacing coefficients x_i with the corresponding solution vector (x_{ALWCR}) elements as

$$x_{\text{ALWCR}} = (A^T \Omega A + \lambda Q)^{-1} A^T \Omega Y, \quad (20)$$

where $Q \in \mathfrak{R}^{n \times n}$ is a diagonal matrix which represents the similarities between the test sample and the corresponding training samples drawn from dictionary A . To maintain an effective similarity between the query sample and training samples, the matrix Q can be normalized as $Q = Q / (1 + \max(Q))$. Thus, the matrix Q mitigates the neighborhood contributions in the solution space. Note that the diagonal matrix $\Omega \in \mathfrak{R}^{m \times m}$ modulates the feature distinctiveness into solution.

The resultant representation vector x_{ALWCR} expresses a test sample Y as the linear combination of optimal neighbors that spanned in the subspace of Y . The computed representation encodes the distinctive features and improves the final image classification. The proposed method performs the classification on the basis of regularized residual error of each class. The coding error for the k th class and query image Y is computed as

$$r_k(Y) = \frac{\|Y - A\sigma_k(x_{\text{ALWCR}})\|_2}{\|\sigma_k(x_{\text{ALWCR}})\|_2}, \quad (21)$$

where $\sigma_k(x_{\text{ALWCR}}) \in \mathfrak{R}^n$ is the coefficient vector, whose nonzero coefficient is chosen from x_{ALWCR} which corresponds to the k th class. The final assignment of a query sample to a class is given via

$$\text{identify}(Y) = \arg \min_k r_k(Y). \quad (22)$$

Figure 1 shows the working of the proposed solution for classification of a query sample, and Algorithm 1 summarizes the proposed method.

4. Experimental Results

Two real-world face databases of AR [25] and ORL [24] are used in experimentation to assess the proposed ALWCR method for face recognition. Both databases contain images of different poses, expressions, and illuminations. Further experiments are conducted on MNIST handwritten digit database [26] to test the effectiveness of the proposed method for image classification in general.

The conventional dimensionality reduction techniques, eigenfaces [1] and randomfaces [7], are used to map the original image space into various low-dimensional feature spaces. The performance of the proposed method is compared against several well-known classification algorithms, i.e., nearest neighbor (NN) [27] (as the baseline method), nearest subspace (NS) [2], collaborative representation-based classification (CRC) [7], sparse representation-based classification (SRC), and a newly proposed sparse representation-based method of WSRC [8].

The CVX [28] package is used to implement SRC and WSRC to solve l_1 minimization problem. In case of NN classifier [27], the neighborhood size is set to 1. For WSRC [8], the regularization parameter λ is set to $1e-4$ and $s = 1.5$, respectively. The regularization parameter λ for CRC and the proposed ALWCR method is set to $1e-3$ for all experiments unless specified.

4.1. Face Representation and Recognition

4.1.1. ORL Database. The ORL database contains 400 images of 40 different subjects and each subject has 10 images. The variations in images include pose changes and expression changes like smiling, open and close eyes, and facial details that include wearing glasses or without glasses. We have cropped and resized all images to a resolution of size 64×64 for this experiment, and the intensity values of all images are scaled to unit length, i.e., l_2 -norm. The well-known two-set experimental protocol is followed for the given experimentation. A subset with $n_k = (4, 5)$ images per class is randomly chosen to be a training set, while rest of the database is treated as a test set. Figure 2 shows few train and test samples from the ORL database. Figure 3 demonstrates the recognition results of the proposed method and other state-of-the-art methods for different dimensions generated by eigen and random projections.

The recognition rates on the ORL 5 training set versus dimensions are shown in Figures 3(a) and 3(b) using eigen and random projections, respectively. In the case of eigen projections (Figure 3(a)), the proposed ALWCR outperforms NN, NS, and original CRC methods, while it performs better than WSRC for many feature dimensions. For a random feature dimension (Figure 3(b)), the performance of the proposed ALWCR is far better as compared to WSRC, CRC, and NS. Figures 3(c) and 3(d) demonstrate the recognition results for the ORL 4-train dataset. It can be observed that ALWCR beats the other state-of-the-art methods in the eigen feature space (Figure 3(c)), while for random feature dimensions, its performance is comparable to the NS method.

The performance of linear representation techniques such as CRC and SRC is not dependent to the choice of an "optimal" feature selection used for dimension reduction. Simply a random feature transformation should lead to comparable performance as any other carefully engineered features such as eigen or fisher transformation

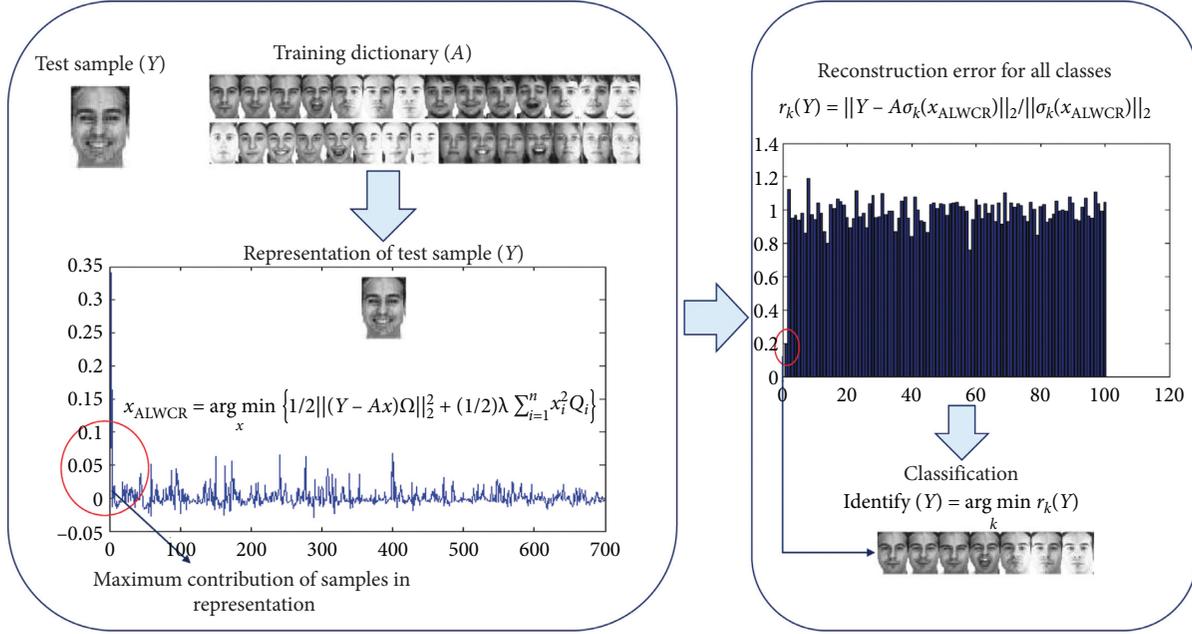


FIGURE 1: Classification steps of the proposed solution.

Input: A -unit l_2 -normalized data matrix; unit l_2 -normalized query samples matrix $\hat{Y} = [y_1, \dots, y_n]$ and $\lambda = 1e^{-3}$, regularity parameter

Output: return labels of the query samples $y_i, i = 1, \dots, n$

for $i \leftarrow 1$ to n do

Use equation (19) and determine the similarity matrix Q ; estimate weighted matrix
$$\begin{cases} r_{i_s} = y_i - Ax_{i_s} \\ x_{i_s} = (A^T A)^{-1} A^T y_i \\ \hat{\Omega} = \text{diag}(\sqrt{|r_{i_s}|}) \end{cases}$$

Compute coefficients \hat{x}_{ALWCR} using equation (20), to estimate the residual \hat{r}_{ALWCR} and update $\hat{\Omega}$ as

$$\begin{cases} \hat{x}_{ALWCR} = (A^T \hat{\Omega} A + \lambda Q)^{-1} A^T \hat{\Omega} y_i \\ \hat{r}_{ALWCR} = y_i - A \hat{x}_{ALWCR} \\ \hat{\Omega} = \text{diag}(\sqrt{|\hat{r}_{ALWCR}|}) \end{cases}$$

Update $\Omega \leftarrow \hat{\Omega}$ and solve the objective function in equation (11) to obtain linear representation x_{ALWCR} ; compute coding error $r_k(y_i)$ against each class using equation (21); assign the class label to query sample y_i according to equation (22)
End.

ALGORITHM 1: Proposed method.



FIGURE 2: Some training and testing samples from ORL database.

provided sufficiently large number of linear measurements. However, in low-dimensional feature space, the discrimination information from linearity structure to date is not enough to correctly classify the data and severely affects the performance of linear representation approaches. The results from Figure 4 show that the ALWCR method outperforms competing methods

including original CRC and WSRC approach across different feature dimensions using both eigen and random features. The performance improvement becomes more significant in low feature dimension because the linearity structure alone is not enough to correctly separate data among different classes, and encoding data locality and feature variance into representation help improves the



FIGURE 3: Classification accuracy on ORL dataset as a function of various dimensions.

discriminative power for classification. Furthermore, these results demonstrate the consistent performance of the proposed method across various feature dimensions compared to that of competing methods, especially WSRC and CRC.

4.1.2. AR Database. There are 4000 color images of 126 individual subjects in the AR database. These frontal view face images have wide variations in term of occlusions, facial expression, and illumination. We have used a subset of the AR database with similar settings as reported in [3, 7]. The subset contains 100 individual subjects (50 male and 50 female subjects) having 14 images per subject taken in two sessions. The sessions were separated over two weeks having 7 images containing different illuminations and expressions and are resized to 66×48 .

In session 1, seven images are selected for each subject as a training set, while the other seven images are used as a testing set from session 2. Figure 4 demonstrates some instances of sample images taken from training and testing sets. Tables 1 and 2 show the recognition results for the AR dataset against different feature dimensions obtained by eigen and random projections, respectively.

It is clear from the results (Tables 1 and 2) that the proposed ALWCR method outperforms other methods and specially WSRC in random feature space. In a nutshell, the proposed ALWCR is stable across different dimensions and perfectly handles the limited train samples compared with other mentioned methods.

The proposed method is able to handle the limitations of original CRC which faces the recognition degradation due to the small sample size. Furthermore, the proposed method is computationally efficient than the SR-based methods due to its closed form solution.

4.2. Regularization Parameters and Data Dimensionality. To investigate how the regulatory parameter versus data dimensionality affects the performance of various methods, we used the AR database with the same settings as mentioned above. The images were cropped to 60×43 pixel size. Training and testing data are l_2 -normalized prior to applying

all considered methods. We use eigen and random projections to obtain low- and high-dimensional embeddings to compare CRC, SRC, and the proposed ALWCR methods in various feature dimensional spaces. For the proposed ALWCR method, we estimated the covariance matrix Ω_{wcr} from the training data by using (14) for the first iteration, and then estimating Ω using (20), we used three iterations in all experiments. To solve the singularity issue in the training sample matrix A , we used the singular value decomposition (SVD) method to compute matrix inverse, i.e., (14) (for estimate Ω_{wcr}) without specifying parameter λ .

From Figure 5, it can be seen that performance of the CRC method, i.e., l_2 -regularization in low-dimensional feature space is inferior than SRC, i.e., l_1 -regularization, which is much more effective in low-dimensional feature space due to the sparsity constraints on coefficients that improve the discrimination in classification [1]. CRC performance improves as the feature dimension increases, and in high-dimensional feature space, CRC outperforms SRC by including data variance and data locality structure into l_2 -regularization formulation, i.e., the proposed ALWCR method improves the results over those of the flat formulation, i.e., CRC in low-dimensional feature space.

The proposed approach outperforms CRC and SRC methods in term of classification accuracy in low-dimensional feature space and demonstrates competing results as compared with CRC and better than SRC in high-dimensional feature space. The effect of choosing a regulatory parameter has less influence on ALWCR compared to other methods on recognition performance. Figure 6 provides an overview of classification accuracy versus different feature dimensions on the AR dataset for different classifiers. The regulatory parameter is fixed to $\lambda = 1e^{-01}$ for all methods except NS. It can be seen that ALWCR outperforms all other considered methods in low-dimensional feature space and have competitive classification accuracy in high-dimensional feature space as well.

4.3. Digit Classification. To evaluate the proposed approach for general pattern classification performance

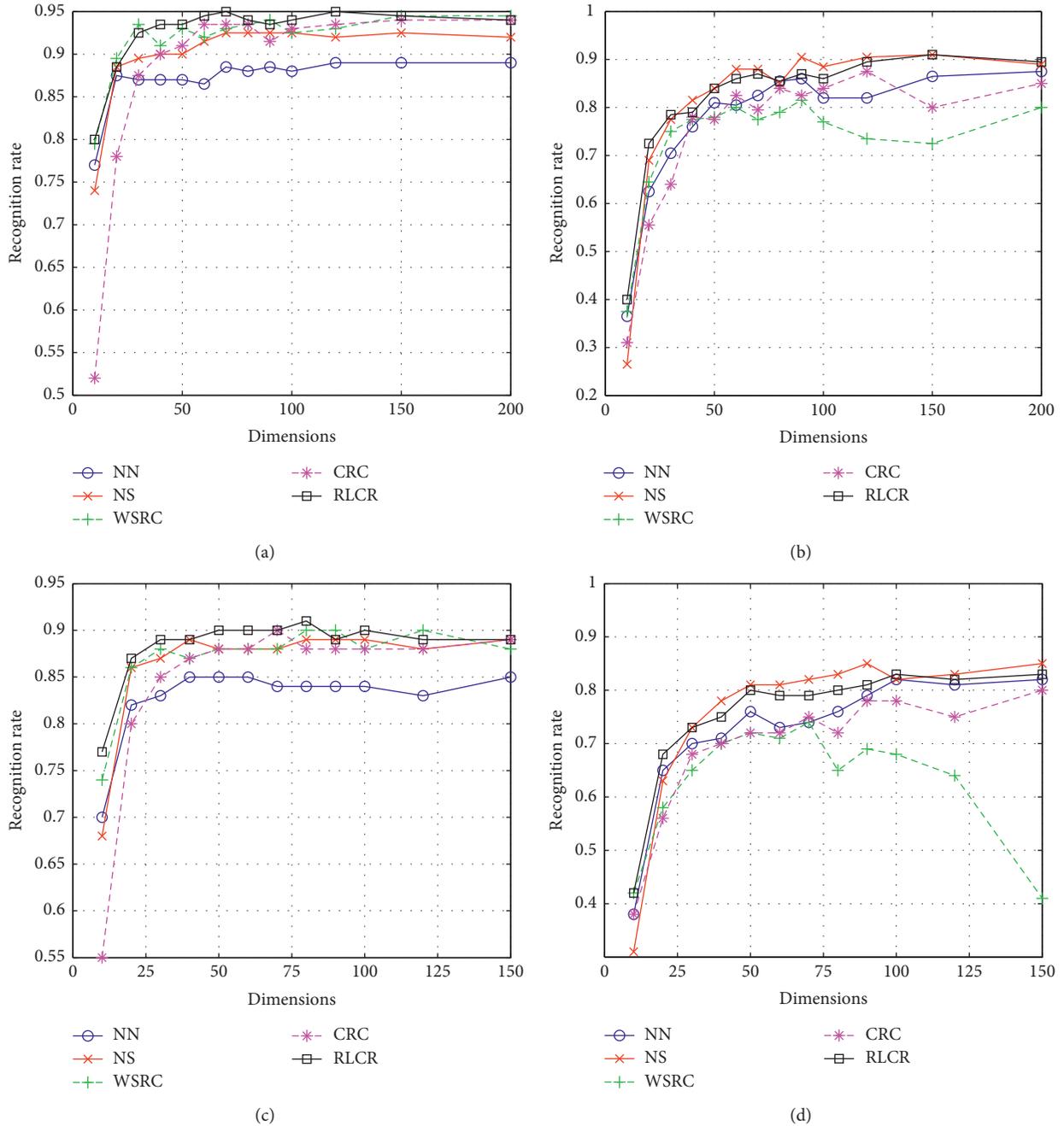


FIGURE 4: Some sample images from training and testing set from AR database; top two rows show the sample from the training set while remaining rows show samples from the testing set. (a) 5-train dataset using eigen features, (b) 5-train dataset using Random features, (c) 4-train dataset using eigen features, and (d) random features on 4-train dataset.

against the several well-known methods, i.e., NS, CRC, and SRC, we have used MNIST handwritten digit dataset [27]. The training set consists of 60,000 samples of handwritten digits (from 0 to 9) while the testing set consists of 10,000 samples, where each sample is represented as an 8-bit gray scale image of size 28×28 . We have chosen the same experimental setting as described in detail previously [14].

The training set was composed by randomly selecting 50 samples from each class to form a training dictionary of 500

samples. The testing set is formed by selecting 700 samples randomly from the original test set. Eigen and random projections are used to obtain various feature space dimensions. Figure 7 shows the classification results on the MNIST dataset against various feature dimensions computed using eigen and random embedding.

It can be seen from Figure 7 that the proposed ALWCR method outperforms all other considered methods in various feature space dimensions (see Table 3 for detail comparison of classification results on various

TABLE 1: The classification results (%) comparison on the AR database for different eigen feature dimensions.

Dim	10	30	40	50	60	70	80	90	100	120	150
1-NN	44.85%	62.57%	65.42%	67.71%	68.57%	68.85%	69.28%	69.28%	69.85%	70.14%	70.28%
NS	18.85%	66.71%	69.00%	70.28%	71.00%	72.42%	72.71%	73.71%	74.28%	75.42%	75.57%
SRC	36.53%	68.42%	73.64%	77.58%	79.59%	80.39%	82.24%	82.02%	83.43%	83.89%	84.67%
WSRC	47.42%	74.71%	79.28%	83.00%	85.57%	87.00%	88.14%	88.85%	88.28%	89.71%	90.28%
CRC	19.14%	62.42%	73.71%	79.42%	83.00%	84.28%	86.85%	87.28%	89.00%	90.00%	91.00%
ALWCR	48.00%	72.57%	78.42%	82.42%	85.57%	85.57%	87.57%	88.00%	89.00%	90.00%	89.85%

The bold face shows the best classification results

TABLE 2: The classification results (%) comparison on the AR database for different random feature dimensions.

Dim	10	30	40	50	60	70	80	90	100	120	150
1-NN	25.00%	57.42%	62.00%	65.00%	61.85%	63.00%	66.00%	66.14%	66.42%	68.57%	68.14%
NS	6.57%	57.71%	63.28%	65.85%	68.71%	68.85%	67.42%	69.14%	71.71%	71.00%	72.71%
SRC	15.00%	49.75%	59.34%	64.52%	66.73%	66.48%	67.78%	71.64%	72.00%	74.51%	76.84%
WSRC	30.71%	58.57%	68.14%	71.28%	72.57%	72.14%	74.85%	79.28%	79.00%	81.71%	81.57%
CRC	12.85%	44.71%	60.00%	65.71%	71.71%	75.71%	57.57%	78.28%	83.00%	84.57%	85.85%
ALWCR	31.14%	62.42%	68.14%	75.57%	76.42%	82.85%	78.28%	82.00%	85.14%	85.71%	86.57%

The bold face shows the best classification results.

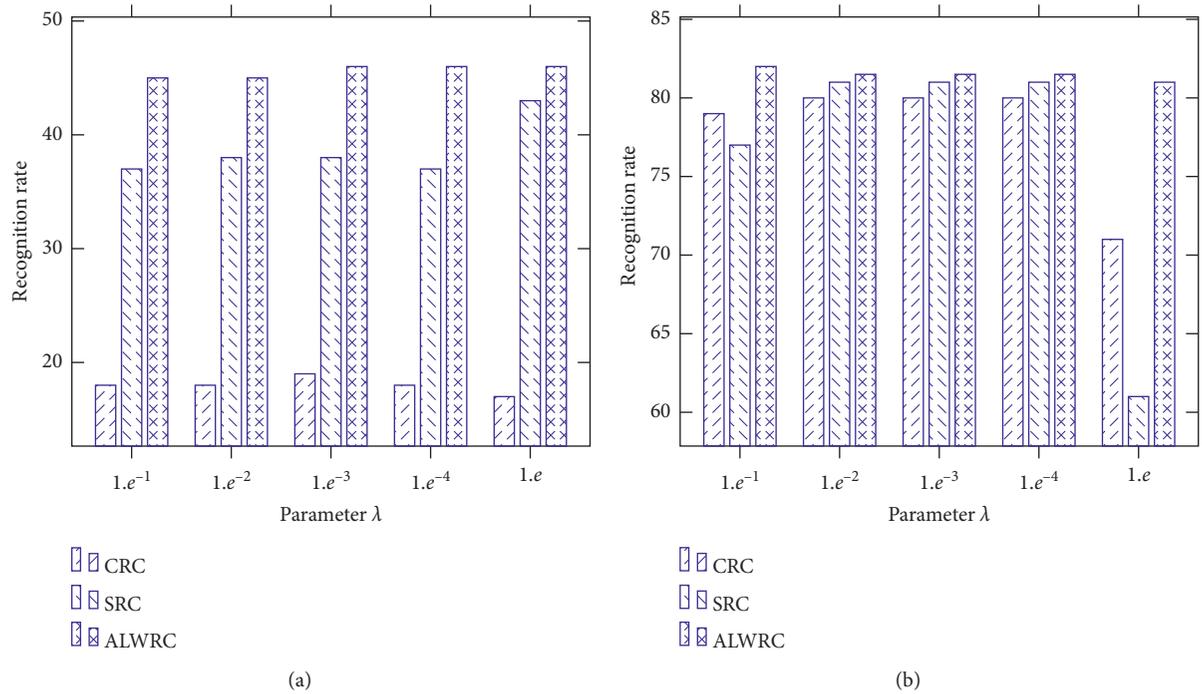


FIGURE 5: Classification results for low-dimensional feature space eigen embedding versus regulatory parameters for CRC, SRC, and ALWCR methods on AR database. (a) Feature dimensions 10. (b) Feature dimensions 54.

feature dimensions). The ALWCR method generally outperforms the CRC method but has increased time complexity than CRC due to estimating variance among samples and incorporating locality structure. However,

ALWCR is significantly faster than the SRC method and outperforms SRC in terms of speed and accuracy. See Table 4 for running time comparison of SRC and ALWCR.

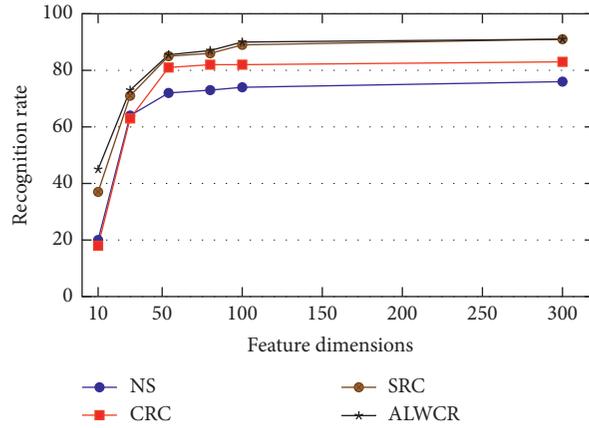


FIGURE 6: Recognition rate versus different feature dimensions; dimensionality is reduced using eigen projection on the AR database. Parameter $\lambda = 1e^{-01}$ is fixed for all methods.

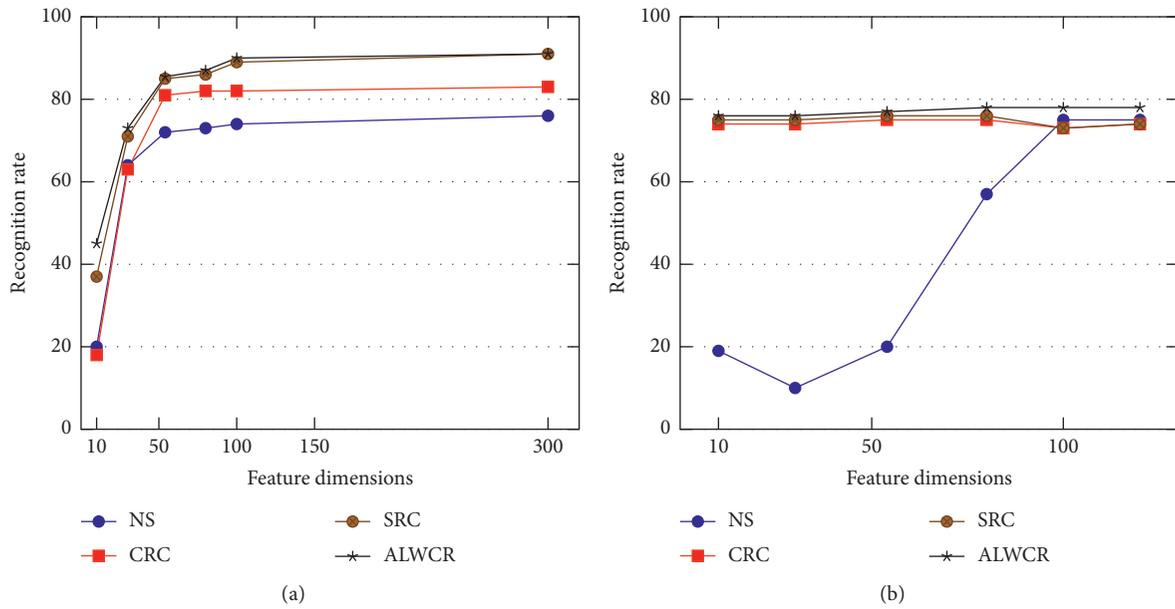


FIGURE 7: Recognition rate versus different feature space dimensions on the MINISET database. Regulatory parameter value is fixed for all methods except NS. (a) Eigen projection. (b) Random projection.

TABLE 3: Image classification results on MNIST dataset. The parameter value $\lambda = 1e^{-03}$ is fixed for all methods. Eigen and random projection is used to get different feature spaces.

Dimension		150	200	250	300	350
Eigenfaces	CRC	75.71%	74%	72.57%	70.42%	70.42%
	SRC	74.85%	73.28%	70.57%	70.28%	66.85%
	ALWCR	77.28%	75.14%	74%	71.71%	74.05%
Randomfaces	CRC	73%	71.14%	70.28%	68.28%	66%
	SRC	70%	68.42%	69.28%	63.71%	59.14%
	ALWCR	75.71%	73.28%	72.28%	69%	67.14%

TABLE 4: Running time comparison on the MNIST dataset for various feature dimensions using eigen and random projections.

Method	Dimension											
	Eigen-20		Eigen-54		Eigen-120		Random-20		Random-54		Random-120	
	Recog.	Time										
CRC	71.61%	0.73	73.00%	0.84	73.89%	1.05	53.49%	0.54	66.61%	0.61	71%	0.80
SRC	76.71%	4.63×10^2	77.28%	5.06×10^2	74.85%	8.73×10^2	59.42%	2.33×10^2	70.71%	3.15×10^2	72.85%	5.83×10^2
ALWCR	77.71%	3.37×10^1	78.28%	4.26×10^1	76.85%	4.86×10^1	63.14%	2.57×10^1	74.28%	2.80×10^1	75.28%	3.16×10^1

Time = running time in seconds for all testing samples in the dataset. The bold numbers are the highest recognition rates and lowest running time for experiment.

5. Conclusions

In this paper, an adaptive locality-based weighted collaborative representation (ALWCR) is presented. The proposed (ALWCR) method effectively encodes the local geometry and feature variances among data elements and performs robust classification. It is an extension of original CRC formulation which significantly improves classification discrimination power. Performance of the proposed method is evaluated against standard benchmarks and state-of-the-art methods. Experimental results confirm that the proposed ALWCR is stable across different feature dimensions for different datasets and outperforms other methods in terms of recognition and computational cost.

Data Availability

The data used to support this study will be made available upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

Acknowledgments

We are thankful to COMSATS University, Islamabad (Abbottabad Campus), and the School of Information Science and Technology, USTC Hefei, China, which fully supported us by providing all key resources during the implementation and all afterward phases of this project. We would also like to personally thank Prof. WuYang Zhou and Dr. Waqas Jadoon for their continuous encouragement and massive support both academically and socially during this project. This work was sponsored by the key program of National Natural Science Foundation of China (Grant no. 61631018).

References

- [1] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Proceedings of the 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 586–591, IEEE, Maui, HI, USA, June 1991.
- [2] K.-C. Lee, J. Ho, and D. J. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 5, pp. 684–698, 2005.
- [3] J. Wright and A. Y. Yang, S. S. Sastry, S. Yi Ma, 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.
- [4] B. Wang, W. Li, Z. Li, and Q. Liao, "Adaptive linear regression for single-sample face recognition," *Neurocomputing*, vol. 115, pp. 186–191, 2013.
- [5] A. Y. Yang, J. Wright, Y. Ma, and S. S. Sastry, "Feature selection in face recognition: a sparse representation perspective," *IEEE Transactions Pattern Analysis and Machine Intelligence*, no. 2, 2007.
- [6] Q. Shi, A. Eriksson, A. van Den Hengel, and C. Shen, "Is face recognition really a compressive sensing problem?," in *Proceedings of the 2011 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 553–560, IEEE, Colorado, CO, USA, June 2011.
- [7] L. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: which helps face recognition?," in *Proceedings of the 2011 IEEE International Conference on Computer Vision (ICCV)*, pp. 471–478, IEEE, Barcelona, Spain, December 2011.
- [8] C.-Y. Lu, H. Min, J. Gui, L. Zhu, and Y.-K. Lei, "Face recognition via weighted sparse representation," *Journal of Visual Communication and Image Representation*, vol. 24, no. 2, pp. 111–116, 2013.
- [9] H. Yu, L. Gao, W. Li, Q. Du, and B. Zhang, "Locality sensitive discriminant analysis for group sparse representation-based hyperspectral imagery classification," *IEEE Geoscience and Remote Sensing Letters*, vol. 14, no. 8, pp. 1358–1362, 2017.
- [10] N. A. Shnain, Z. M. Hussain, and S. F. Lu, "A feature-based structural measure: an image similarity measure for face recognition," *Applied Sciences*, vol. 7, no. 8, p. 786, 2017.
- [11] O. Iqbal, J. Waqas, F. G. Khan, B. Nazir, I. A. Khan et al., "Robust face recognition under limited training sample scenario using linear representation," *KSII Transactions on Internet & Information Systems*, vol. 12, no. 7, 2018.
- [12] B. Jiang, C. Ding, and B. Luo, "Robust data representation using locally linear embedding guided PCA," *Neurocomputing*, vol. 275, pp. 523–532, 2018.
- [13] Y. Zhu, C. Zhu, and X. Li, "Improved principal component analysis and linear regression classification for face recognition," *Signal Processing*, vol. 145, pp. 175–182, 2018.
- [14] J. Waqas, Z. Yi, and L. Zhang, "Collaborative neighbor representation based classification using-minimization approach," *Pattern Recognition Letters*, vol. 34, no. 2, pp. 201–208, 2013.
- [15] X. He, D. Cai, S. Yan, and H.-J. Zhang, "Neighborhood preserving embedding," in *Proceedings of the Tenth IEEE International Conference on Computer Vision*, vol. 1, pp. 1208–1213, IEEE, Beijing, China, October 2005.
- [16] W. Li and D. Qian, "A survey on representation-based classification and detection in hyperspectral remote sensing imagery," *Pattern Recognition Letters*, vol. 83, pp. 115–123, 2016.

- [17] Y. Chi and F. Porikli, "Connecting the dots in multi-class classification: from nearest subspace to collaborative representation," in *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3602–3609, IEEE, Portland, OR, USA, June 2012.
- [18] J. Waqas, Y. Zhang, and L. Zhang, "Graph-based features extraction via datum adaptive weighted collaborative representation for face recognition," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 28, no. 2, p. 1451003, 2014.
- [19] M. Yang, L. Zhang, D. Zhang, and S. Wang, "Relaxed collaborative representation for pattern classification," in *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2224–2231, IEEE, Madison, WI, USA, June 2012.
- [20] X. He and P. Niyogi, "Locality preserving projections," in *Advances in neural information processing systems*, pp. 153–160, MIT Press, Cambridge, MA, USA, 2004.
- [21] S. T. Roweis and K. Lawrence, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, 2000.
- [22] P. Meer, D. Mintz, A. Rosenfeld, and D. Y. Kim, "Robust recognition methods for computer vision: a review," *International Journal of Computer Vision*, vol. 6, no. 1, pp. 59–70, 1991.
- [23] D. B. Rubin and R. J. A. Little, *Statistical Analysis with Missing Data*, Vol. 793, John Wiley & Sons, Hoboken, NJ, USA, 2019.
- [24] F. S. Samaria and A. C. Harter, "Parameterisation of a stochastic model for human face identification," in *Proceedings of the 1994 IEEE Workshop on Applications of Computer Vision*, pp. 138–142, IEEE, Sarasota, FL, USA, December 1994.
- [25] A. M. Martinez and R. Benavente, "The ar face database," Technical Report 24, Springer, Berlin, Germany, 1998.
- [26] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [27] T. Cover and P. Hart, "Nearest neighbor pattern classification," *IEEE Transactions on Information Theory*, vol. 13, no. 1, pp. 21–27, 1967.
- [28] M. Grant, S. Boyd, and Y. Ye, *CVX: Matlab Software for Disciplined Convex Programming*, 2008.

