This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Sparse Supervised Representation-Based Classifier for Uncontrolled and Imbalanced Classification

Ting Shu, *Student Member, IEEE*, Bob Zhang, *Member, IEEE*, and Yuan Yan Tang, *Life Fellow, IEEE*

***Abstract*—The sparse representation-based classification (SRC) has been utilized in many applications and is an effective algorithm in machine learning. However, the performance of SRC highly depends on the data distribution. Some existing works proved that SRC could not obtain satisfactory results on uncontrolled data sets. Except the uncontrolled data sets, SRC cannot deal with imbalanced classification either. In this paper, we proposed a model named sparse supervised representation classifier (SSRC) to solve the above-mentioned issues. The SSRC involves the class label information during the test sample representation phase to deal with the uncontrolled data sets. In SSRC, each class has the opportunity to linearly represent the test sample in its subspace, which can decrease the influences of the uncontrolled data distribution. In order to classify imbalanced data sets, a class weight learning model is proposed and added to SSRC. Each class weight is learned from its corresponding training samples. The experimental results based on the AR face database (uncontrolled) and 15 KEEL data sets (imbalanced) with an imbalanced rate ranging from 1.48 to 61.18 prove SSRC can effectively classify uncontrolled and imbalanced data sets.**

***Index Terms*—Data driven, face recognition, imbalanced classification, spare supervised representation-based classifier (SSRC), sparse representation-based classification (SRC).**

I. Introduction

THE sparse representation-based classification (SRC) was first proposed by Wright *et al.* [1]. The fundamental assumption of the SRC is that every class subspace is independent of each other. Based on this assumption, the SRC represents a test sample through a linear combination of the training samples and the coefficients are sparse. Some of the coefficients of the training samples from the same class with the test sample have significant values, while the others are close to zero. The SRC reconstructs the test sample via the training samples of each class and its corresponding coefficients, the sparse coefficients can make the residual of the same class to be the minimum among all class residuals.

In the recent years, the SRC has been proven to be effective and efficient in many applications, such as: face recognition [2], denoising [3], superresolution [4], hyperspectral image classification [5], visual tracking [6], and so on. The sparse constraint is always used in other algorithms, such as subspace learning [7], discriminant manifold learning [8], [9], and so on. Recently, many variants of the SRC have been developed, such as: kernel SRC [10]–[12] combining the kernel trick with the SRC, collaborative representation-based classifier (CRC) [13] replacing the l_1-norm with the l_2-norm, weighted SRC (WSRC) [14], [15] adding the distance information as the weight in the objective function.

Wright *et al.* [16] stated in the review paper: *The sparse representation-based face recognition assumes that the training images have been carefully controlled and that the number of samples per class is sufficiently large. Outside these operating conditions, it should not be expected to perform well.* In other words, the performance of the SRC relied on the quality and quantity of the training samples. Deng *et al.* [17], [18] explored the influences of the data structure to SRC in face recognition. Deng *et al.* [18] revealed that the discriminant nature of the collaborative representation (including the SRC) is determined by the class separability of the data dictionary. The theoretical analysis and experimental results in [18] proved that the SRC cannot classify uncontrolled data sets very well. In [17], a superposed linear representation that encodes the test sample as a superposition of the prototype and variation dictionaries was proposed to address this limitation. The proposed method in [17] and [18] only focused on the uncontrolled face recognition data sets (such as various illuminations).

Imbalanced data sets can be found in many applications such as biometrics, e-business, medical diagnosis, information security, and so on. Traditionally, conventional classifiers treat all the samples from different subjects equally. Hence, these types of methods cannot classify the minority class (also named the positive class) samples correctly, but instead identify them into the majority class (also known as the negative class). However, in many applications, the minority class is more important than the majority class. For example, in medical diagnosis, the cost that a patient (belonging to the minority class) is misclassified to be healthy is much higher than that of the reverse situation.

In the past years, the methods dealing with imbalanced data sets can be categorized into three main categories: 1) data-preprocessing methods [19]–[21]; 2) cost-sensitive methods [22], [23]; and 3) algorithmic-level methods [24], [25]. The data-preprocessing methods process the imbalanced data to make them balanced before the classification scheme through oversampling (increasing the sample number in the minority class), undersampling (decrease

Manuscript received June 28, 2018; revised September 5, 2018; accepted November 28, 2018. This work was supported in part by the National Natural Science Foundation of China under Grant 61602540 and Grant 61273244 and in part by the Research Grants of University of Macau under Grant MYRG2018-00053-FST. *(Corresponding author: Bob Zhang.)*

The authors are with the Department of Computer and Information Science, University of Macau, Macau 999078, China (e-mail: yb57406@um.edu.mo; bobzhang@um.edu.mo; yytang@um.edu.mo).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TNNLS.2018.2884444

the sample number in the majority class), or both. However, oversampling may bring noises to the training data set and undersampling could delete crucial information from the majority class. The cost-sensitive methods deal with the imbalanced problem using a higher penalty for misclassifying minority class data during the decision phase. The algorithm-level methods modify the classifier objective functions to process the imbalanced data sets. To this day, imbalanced classification is still a challenging problem.

To solve the above-mentioned issues (uncontrolled and imbalanced data sets classification), we propose a data-driven algorithm which is a variant of the SRC. Our proposed method named sparse supervised representation-based classifier (SSRC) involves the training class label information during the test sample representation phase. SSRC linearly represents the test sample in each class subspace. To further classify imbalanced data sets, a weight learning model is proposed and utilized in SSRC. We tested our proposed algorithm in the public AR [26] and KEEL [27], [28] data sets and the experimental results proved that SSRC can correctly classify the uncontrolled and imbalanced data sets.

There are three main contributions in our paper as follows.

1) A data-driven algorithm is proposed to deal with uncontrolled and imbalanced training sets.
2) The reconstruction term is split (for the first time) into more than one term depending on the applications.
3) SSRC can correctly classify the test sample based on the imbalanced training set without any data preprocessing.

Throughout this paper, $X = [X_1, X_2, \dots, X_K] \in \mathbb{R}^{d\times N}$ denotes the whole training samples, where $X_i \in \mathbb{R}^{d\times n_i}$ signifies the ith subject, d is the dimensionality of the training data, and $N = n_1 + n_2 + \dots + n_K$ represents the size of the training data set. $X_{i,j}$ is the jth training sample from the ith class. The test sample is described as $y \in \mathbb{R}^d$. $\alpha = [\alpha_1^T, \alpha_2^T, \dots, \alpha_K^T]^T \in \mathbb{R}^N$ indicates the coefficient of X representing y, where $\alpha_i \in \mathbb{R}^{n_i}$ stands for the coefficient of X_i reconstructing y. λ and θ are the scales controlling the tradeoff among different terms in the objective function. I represents the identity matrix. The weight vector is denoted as $w = [w_1, \dots, w_k]$ and w_i describes its ith element.

For the following content, the related works including the imbalanced classification and the SRC algorithm are first introduced in Section II. Section III gives the details about SSRC followed by the results of the experiments tested on the AR and KEEL data sets in Section IV. Finally, some conclusions will be represented in Section V.

II. Related Works

In this section, the conventional SRC and its variants are first introduced. Next, uncontrolled data sets classification via the SRC is described. The works about imbalanced classification are finally given.

A. SRC and Its Variants

The conventional SRC was first proposed in [1] and its objective function is defined as

$$\min_{\alpha} \frac{1}{2}\|y - X\alpha\|_2^2 + \lambda\|\alpha\|_1. \tag{1}$$

The main principle of the SRC is: a linear combination of the training samples can reconstruct a test sample, where only the coefficients from the same class have significant values and the others are close to zero. The l_1-norm regularization term in (1) can make the coefficients to be sparse and ensure the main assumption of the SRC. For example, when the test sample y is from the ith class, its coefficient α solved from (1) will be $[0, \dots, \alpha_i, \dots, 0]$.

Zhang *et al.* [13] explored the relationships between the discriminant ability and the sparse coefficients of the SRC. The study in [13] revealed that the collaborative representation of the SRC plays a more important role than the sparse coefficients in face recognition. Based on the finding, a CRC was proposed via replacing the l_1-norm in (1) with a l_2-norm on the coefficients

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

Different from the conventional SRC, the objective function of the CRC is convex and smooth. Hence, the coefficients of (2) can be easily and analytically solved as

$$\hat{\alpha} = (X^T X + \lambda I)^{-1} X^T y. \tag{3}$$

The probabilistic theory is always used to avoid the overfitting and underfitting problem of algorithms [29]. Cai *et al.* [30] proposed a probabilistic collaborative representation framework, where the probability that a test sample belongs to the collaborative subspace of all classes can be well defined and computed. A probabilistic CRC (ProCRC) was proposed in [30] and it is defined as

$$\min_{\alpha} \left\{ \|y - X\alpha\|_2^2 + \lambda\|\alpha\|_2 + \frac{\theta}{K}\sum_{k=1}^{K} \|X\alpha - X_i\alpha_i\|_2^2 \right\} \tag{4}$$

where the third term $(\theta/K)\sum_{k=1}^{K} \|X\alpha - X_i\alpha_i\|_2^2$, which is different from the objective function [refer to (2)] of the CRC considers the probability that a test sample belongs to the collaborative subspace of all classes.

As the objective function of the ProCRC is convex and smooth, it also has a closed-form solution. The solution can be easily solved by the iterative reweighted least square technique

$$\hat{\alpha} = \left(X^T X + \frac{\theta}{K}\sum_{k=1}^{K} \left(\overline{X}_k'\right)^T \overline{X}_k' + \lambda I \right)^{-1} X^T y \tag{5}$$

where $\overline{X}_k' = X - X_k'$ and $X_k' = [0, \dots, X_k, \dots, 0]$ have the same size as X.

B. Uncontrolled Data Set Classification via the SRC

Deng *et al.* [18] discussed the sparsity of the conventional SRC model and found that the performance of the traditional SRC algorithm relied on the data set. In order to solve this problem, they proposed a novel SRC model named superposed linear representation based classification (SLRC). The SLRC handled the uncontrolled data sets through decomposing the training samples into two different dictionaries. The two dictionaries are prototype (class centroids) and variation (sample-to-centroid difference) dictionaries, respectively. They also proposed a superposed linear representation that encodes the test sample as a superposition of the prototype and variation dictionaries. The class centroids (prototype dictionary) are average images of each class and the variation images separate out the uncontrolled factors, such as lighting and sunglasses. With different regularization terms, they developed two classifiers (SLRC-l_1 and SLRC-l_2). The objective function of the SLRC is defined as

$$\min_{\alpha,\beta} \left\| [P, V] \begin{bmatrix} \alpha \\ \beta \end{bmatrix} - y \right\|_2^2 + \lambda \left\| \begin{bmatrix} \alpha \\ \beta \end{bmatrix} \right\|_l \tag{6}$$

where $P = [c_1, \ldots, c_i, \ldots, c_K] \in \mathbb{R}^{d\times K}$ represents the prototype dictionary and $c_i = (1/n_i)\sum_{j=1}^{n_i} X_{i,j}$ is the mean of the i^{th} class training samples. $V = [X_1 - c_1 e_1^T, \ldots, X_K - c_K e_K^T] \in \mathbb{R}^{d\times N}$ is the variation dictionary and $e_i^T = [1, \ldots, 1] \in \mathbb{R}^{1\times n_i}$ is a vector only having 1s. The norm of the coefficients can be $l \in \{1, 2\}$ and its corresponding algorithms are SLRC-l_1 and SLRC-l_2, respectively.

Zou *et al.* [31] developed a data-driven method named sparse representation-based over-sampling technique (SROT) to apply the SRC in oversampling to deal with imbalanced classification. SROT first calculates the sparse coefficients of the current minority sample based on the remaining minority samples. Then, it randomly selects a subset from the nonzero items of the coefficients. Afterward, it adds a random value [ranging from $(0, 1)$] to each selected nonzero coefficient. Finally, it generates a new sample through multiplying the new coefficients to the remaining minority samples.

C. Imbalanced Classification

Most of the previous works processing imbalanced classification focused on data-driven methods. The simplest way to oversample the minority class is through randomly replicating the existing minority data [32]. In 2002 [33], the synthetic minority oversampling technique (SMOTE) was proposed. For SMOTE, a start point is first randomly selected from the minority class as the considered point (x_i) and its nearest neighbors were generated based on the Euclidean distances. Among these neighbors, one point (x_j) is then chosen. Finally, SMOTE created a new minority class sample $(\hat{x})$ by selecting a random point on the line from the considered point (x_i) to one (x_j) of its nearest neighbors. The data are generated using

$$\hat{x} = x_i + a \circ (x_j - x_i) \tag{7}$$

where a with the same size of $\hat{x}$ is a random vector and each value ranges from $[0, 1]$, $\circ$ means the Hadamard product (also named elementwise product) of two vectors.

Mullick *et al.* [34] modified the k-nearest neighbor (k-NN) classifier for imbalanced classification. Mullick *et al.* used two methods to select the k value based on the data structure. One method learns the k value with the help of an artificial neural network and it is called the Adaptive k-NN (Ada-k−NN). The another selects the k value using information about its neighboring training points and is named Ada-k-NN2. Furthermore, the authors proposed a class-based global weighting scheme (global imbalance handling scheme or GIHS) to class imbalance. Lu *et al.* [35] proposed an incremental learning method called dynamic weighted majority for imbalance learning to deal with imbalanced data sets.

III. Proposed Method

This section first introduces the proposed method. Section III-B discusses the difference between the conventional SRC model and our proposed model (SSRC) for imbalanced classification. How to solve the model is given in Section III-C followed by the representation of the method for classification in Section III-D.

A. Sparse Supervised Representation-Based Classifier

When using the conventional SRC model [refer to (1)], the training samples are used as an intact dictionary to linearly represent the test sample without any class information. In other words, the representation phase of the conventional SRC belongs to an unsupervised case. After the coefficients are solved from the unsupervised phase, each class residual is calculated and the label of the test sample is predicted as the class with the minimum residual. However, during the unsupervised representation phase, the sparse coefficients would always be biased to the majority class. Given this reason, the conventional SRC cannot obtain satisfactory performances for most of the uncontrolled or imbalanced data sets. The experimental results in Section IV can prove this fact in practice.

In order to solve the above-mentioned problems, we proposed a data-driven SRC model named SSRC, involving the class label information during the representation phase. In SSRC, the training samples from every class are used to linearly represent the test sample independently and a l_1-norm is added to make different class coefficients compete with each others. The classification phase of SSRC is similar with that of the conventional SRC model. The objective function of SSRC is defined as

$$(\hat{\alpha}) = \arg\min_{\alpha} \sum_{i=1}^{K} \frac{w_i}{2} \|y - X_i \alpha_i\|_2^2 + \lambda \|\alpha\|_1 \tag{8}$$

where K is the class number of the training samples and w_i represents the ith class weight.

In SSRC, the test sample is represented by a linear combination of the training samples in each subject. According to the subspace learning theory [36], the matrix (training data set) from the same subspace (class) is enough to linearly represent the vector (test sample) and two or more than two matrices from different classes are needed to linearly represent the vector. In other words, a linear combination of the training samples from other classes can also represent the test sample, while these training samples come from the above-mentioned two classes. Hence, using all the training samples to linearly represent the test sample (such as SRC, CRC, and so on) is hard to handle uncontrolled data sets. For example, a test image shows a face with sunglasses, the training samples from the same class have no face with sunglasses, and some face with sunglasses belong to other training classes. In this situation, using the training samples as a whole matrix to linearly represent the test sample, the coefficients of the other

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

classes including face with sunglasses can have significant values. This could make SRC-based algorithms misclassify the test sample. Therefore, using each class training samples to linearly represent the test sample is helpful for our algorithm to handle the uncontrolled data sets.

The class weights are added to handle the training data sets with a high imbalanced rate (IR). For uncontrolled training sets, each class weight (w_i) is set to be 1. For an imbalanced data set, the minority class reconstruction term will be given a larger weight than the other terms. The larger weight makes the biased case caused by the imbalanced data set numbers to be more fair to the minority class in the classification phase. In order to distinguish the two cases, the SSRC model without weights is named as SSRC-NW and defined as

$$(\hat{\alpha}) = \arg\min_{\alpha} \frac{1}{2}\sum_{i=1}^{K} \|y - X_i\alpha_i\|_2^2 + \lambda\|\alpha\|_1. \quad (9)$$

The model with different class weights is called as SSRC-W and defined as

$$(\hat{\alpha}) = \arg\min_{\alpha} \sum_{i=1}^{K} \frac{w_i}{2}\|y - X_i\alpha_i\|_2^2 + \lambda\|\alpha\|_1$$
$$\text{s.t } \exists\ w_i \neq 1, \quad \forall i = 1, \ldots, K. \quad (10)$$

Algorithm 1 Learning Class Weights

1: **Input:** $X = [X_1, \ldots, X_K]$
2: **Initialization:** $w_i = 1, \forall i = 1, \ldots, K$
3: *# Calculate the training data matrix variance of each class:*
4: $v_i = var(X_i)$
5: *# Find the class with the maximum variance:*
6: $j = \arg\max_i v_i$
7: *# Compute the mean of the j^{th} class:*
8: $M_j = mean(X_j)$
9: *# Learn the class weights except j:*
10: **for** $i = 1 : j-1,\ j+1 : K$
11: *# Calculate the current class mean:*
12: $M_i = mean(X_i)$
13: *# Compute the correlation coefficient between M_i and M_j*
14: $c_i = corrcoef(M_i, M_j)$
15: *# Assign value to the i^{th} class weight:*
16: $w_i = 1 + 0.001 * c_i$
17: **end**
18: **Output:** $w = [w_1, \ldots, w_K]$

How each class weight is learned is described in Algorithm 1. Variance is the measure of the width of a distribution [37]. The class with the maximum variance means it has the biggest data distribution width among those of all classes. Hence, the class with the maximum variance is selected as the start class and its weight is assigned 1 (Lines 3–6). Other class weights are learned via comparing with this start class. Covariance is the measure of how one variable varies with another and correlation is the normalized way of the covariance [38]. Therefore, other class weights are decided by the correlation coefficients between them and the start class. For the ith class, its weight is calculated via $w_i = 1 + 0.001 * c_i$, where c_i is the correlation coefficient between the ith class and the start class. According to our experiments, we found that the highest results were always achieved when the weights vary from $[1.0001, 1.01]$. More details about the effect of the class weights are described in Section IV-E. The variances and correlation coefficients are computed through the inner functions of the MATLAB.

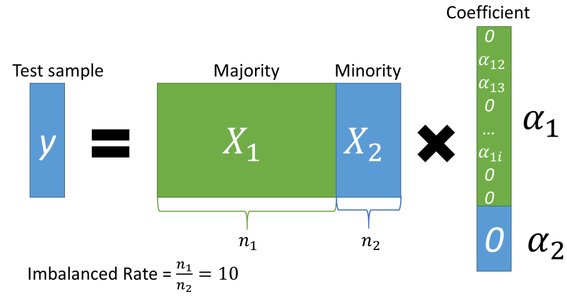

Fig. 1. Example of the SRC model diagram.

B. Difference Between SSRC and SRC for Imbalanced Classification

For convenience, we used a binary-class data set with an IR of 10 as the example to discuss the differences between SSRC and SRC. The majority class denoted as X_1 has n_1 samples and X_2 is the minority class with n_2 samples. The IR of 10 means $(n_1/n_2) = 10$.

The key assumption of the SRC is that data samples in the same class lie in a linear subspace [1]. However, the SRC requires that the sampling of each training class data is sufficient to make sure that the test sample can be classified correctly according to the sparse coefficient. For imbalanced data sets, the minority class X_2 does not have enough samples to linearly represent the test sample from the same subject better than the majority class X_1 which has 10 times the number of data. When classifying a test sample y belonging to the minority class, X_1 has a quantitative advantage compared with X_2. Hence, most of the significant values of the sparse coefficient α would exist in the majority class coefficient α_1, even when α_2 has only zero entries. Fig. 1 illustrates the example using the SRC [refer to (1)] to classify y based on X_1 and X_2, where green signifies the data related with X_1 and the data related with X_2 is represented in blue. Therefore, the SRC can not deal with the imbalanced data sets in classifying the samples from the minority class.

The example of SSRC [refer to (8)] classifying the test sample y is displayed in Fig. 2. In SSRC, the samples of each class are used to linearly represent the test sample, which means that the minority class has an equal opportunity to classify the test sample. However, for some highly imbalanced data sets, splitting the reconstruction term is not enough to help the minority class remove the influences of the majority class. Hence, adding a weight to each subject's reconstruction term in the objective function is necessary for these highly imbalanced data sets.

The regularization term using l_1-norm in the SSRC objective function (8) ensures the coefficient to be sparse and to have

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

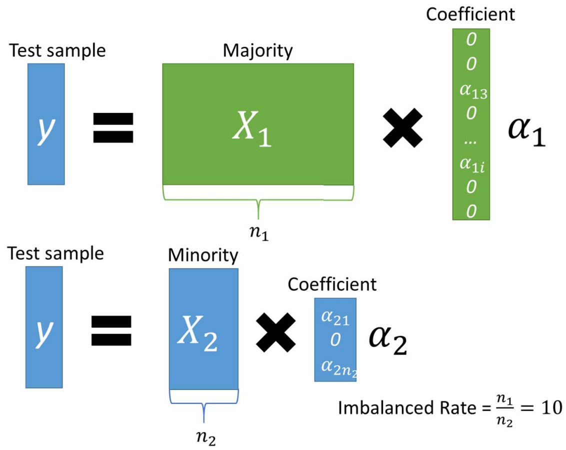

Fig. 2. Example of reconstruction term splitting in SSRC.

the ability to classify the test sample correctly. This guarantees that the residual of the test sample with the reconstruction term from the same class is minimal among all the reconstruction terms.

C. Optimization

In order to solve (8), the different class reconstruction terms are combined together

$$(\hat{\alpha}) = \arg\min_{\alpha} \frac{1}{2}\|\tilde{y} - \tilde{X}\alpha\|_2^2 + \lambda\|\alpha\|_1 \tag{11}$$

where $\tilde{y}$ and $\tilde{X}$ are described in (12) and (13)

$$\tilde{y} = \begin{bmatrix} \sqrt{w_1}y \\ \sqrt{w_2}y \\ \vdots \\ \sqrt{w_K}y \end{bmatrix} \in \mathbb{R}^{Kd\times 1} \tag{12}$$

$$\tilde{X} = \begin{bmatrix} \sqrt{w_1}X_1 & 0 & \cdots & 0 \\ 0 & \sqrt{w_2}X_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{w_K}X_K \end{bmatrix} \in \mathbb{R}^{Kd\times N}. \tag{13}$$

The combined SSRC model (11) is very similar with the traditional SRC model (1). Therefore, we can use the same solution with the conventional SRC to solve $\hat{\alpha}$ in (11), which is equivalent to (8) sharing the same minimizer (α). Hence, the coefficient solved from (11) can be used directly to minimize (8).

For our proposed model, which is a variant of the Lasso model, we use a fast iterative shrinkage–thresholding algorithm (FISTA) [39]–[41] to solve the coefficient of SSRC. How to use FISTA to solve the coefficient is presented in Algorithm 2. In this algorithm, the objective function of (11) is denoted as $f(x) = (1/2)\|\tilde{y} - \tilde{X}\alpha\|_2^2 + \lambda\|\alpha\|_1$. At line 8 of this algorithm, $\tau_s(\alpha)_j = (|\alpha_j| - s)_+ \mathrm{sign}(\alpha_j)$ is a proximal operator—soft thresholding [42] and defined in (14). More details about FISTA can be found in [39]

$$\tau_s(\alpha)_j = \begin{cases} \alpha_j - s & \alpha_j > s \\ \alpha_j + s & \alpha_j < -s \\ 0 & \text{otherwise.} \end{cases} \tag{14}$$

D. Classification Scheme

After solving the coefficient of the test sample y via (11) and Algorithm 2, each subject's residual is calculated by

$$r_i(y) = \|y - X_i\hat{\alpha}_i\|_2^2 \tag{15}$$

Algorithm 2 Using FISTA to Solve (11)

1: **Input:** $\tilde{y}$, $\tilde{X}$, and λ
2: **Initialization:** $\beta_1 = \alpha_0 \in \mathbb{R}^N$, $t_1 = 1$, and Lipschitz constant $L = 2\Lambda_{\max}(\tilde{X}^T\tilde{X})$, and $\xi_0 = 0$
3: **for** i = 1 : max iterative number
4: *Calculate the medium-term sequences:*
5: $\xi_i = \frac{1+\sqrt{1+4\xi_{i-1}^2}}{2}$
6: $\mu_i = \frac{1-\xi_{i-1}}{\xi_i}$
7: *Update the extrapolated point (the $\tau_s(\alpha)_j$ definition is given in Equation (14)):*
8: $\beta_i = \tau_{\lambda t_i}(\alpha_{i-1} - t_k\nabla f(\alpha_{i-1}))$
9: *Update the coefficient:*
10: $\alpha_i = (1-\mu_i)\beta_i + \mu_i\beta_{i-1}$
11: **end**
12: **Output:** $\hat{\alpha}$

where $r_i(y)$ indicates the residual of reconstructing the test sample. The test sample y is assigned the label of the class where its residual is the minimum and is defined as

$$\text{Label}(y) = \arg\min_i r_i(y). \tag{16}$$

To summarize, SSRC establishes a "new" test sample and a "new" training data set through combining different class reconstruction terms with its corresponding weights [refers to (8)–(13)]. The SSRC steps are given in Algorithm 3.

Algorithm 3 SSRC

Input: a test sample y, training samples $X = [X_1, \ldots, X_K]$ in K classes, a tradeoff scale λ, and K weights $w_1, \ldots, w_K$.
Normalize the columns of X to have unit l_2-norm.
Constitute a new test sample $\tilde{y}$ through Equation (12).
Constitute a new training data set $\tilde{X}$ through Equation (13).
With $\tilde{y}$ and $\tilde{X}$, solve the coefficient $\hat{\alpha}$ by Algorithm 2.
Compute each class residual $(r_i(y))$ through Equation (15).
Label the test sample (y) by Equation (16).
Output: Label(y).

IV. Experimental Results

Experimental results of our proposed methods and the comparison classifiers on the uncontrolled and imbalanced data sets are given in this section. Face recognition with uncontrolled data set is first presented in Section IV-A followed by the imbalanced classification settings in Section IV-B. Binary-class and multiclass imbalanced data set results are discussed in Sections IV-C and IV-D, respectively. The effects of the parameters of SSRC are described in Section IV-E.

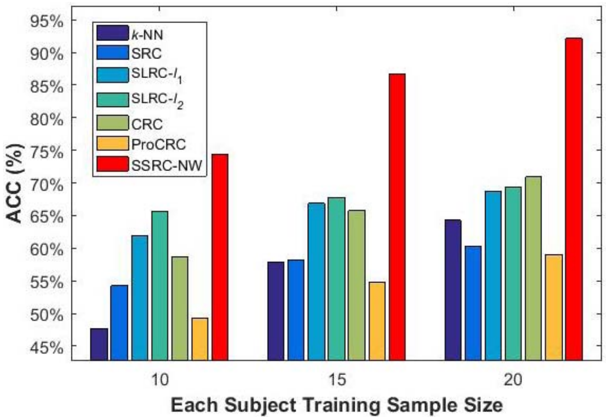

Fig. 3. ACC (%) of SSRC-NW and the six comparison methods for AR database.

A. Face Recognition With Uncontrolled Data Set

The ability of our proposed method (SSRC-NW) to classify uncontrolled data set was tested on the public AR face database [26]. The AR database contains over 4000 color images of 126 people. For each individual, 26 images were captured in two different days and 13 images each day were taken under different facial expressions, illumination conditions, and occlusions.

In the following experiments, 120 subjects consisting of 65 males and 55 females were randomly selected from the AR database. We extracted texture features from each image via a 2-D Gabor filter bank including $5 \times 8 = 40$ 2-D Gabor filters [43]. The 40 2-D filters include 5 σ values $\times 8\theta$ values with a size of 16×16. The feature vector of each facial image is a concatenation of the mean values of 40 filter response. More details about how to extract the Gabor features from the images can be found in [44] and [45].

For each subject, the training size varies from 10, 15, and 20, where the remaining samples are utilized as the testing. In order to decrease the influences of the training and testing data sets splitting, five random iterations with the same setting were applied. The final result is the mean of the five random experimental accuracies, where the accuracy (ACC) is the ratio of the correctly classified sample number divided by the whole test data set size.

In order to show the performance of SSRC-NW, six comparison classifiers were applied consisting of one traditional classifier (k-NN) [46], the SRC, the SLRC-l_1, the SLRC-l_2, the CRC, and the ProCRC. All the methods were tested under the same experimental settings. For the following experiments, $k = 1$ for k-NN; $\lambda = 0.001$ for the SRC, SLRC-l_1, SLRC-l_2, and CRC; $\lambda = 0.001$ and $\theta = 0.001$ for the ProCRC; and $\lambda = 0.0001$ for SSRC-NW. All the parameter(s) of the seven classifiers were selected as the optimal based on the experiments. The λ selection of SSRC-NW is presented in Section IV-E.

Fig. 3 shows the results of SSRC-NW and the six comparison classifiers for the AR database with different training sample sizes. In the bar chart, the results of SSRC-NW are illustrated in the red bar. The x-axis is the training sample number of each subject and it varies from 10, 15, and 20. The y-axis is the ACC in a percentage format. According to the figure, it is obvious that our proposed method (SSRC-NW) always obtained the highest accuracies with various training sample sizes among the seven classifiers. The SRC depicted in blue performed poorly and even got a lower ACC than that of k-NN when the training sample number is 20. The numerical results of SSRC-NW and the six comparison methods can be found in Table I.

TABLE I

NUMERICAL RESULTS OF SSRC-NW AND THE SIX COMPARISON METHODS FOR THE AR DATABASE

Classifier \ Number	10	15	20
k-NN	47.73%	57.88%	64.28%
SRC	54.28%	58.27%	60.33%
SLRC-l_1	61.93%	66.88%	68.67%
SLRC-l_2	65.70%	67.76%	69.36%
CRC	58.67%	65.79%	70.92%
ProCRC	49.28%	54.85%	59.00%
SSRC-NW	**74.35%**	**86.73%**	**92.14%**

According to Table I, SSRC-NW obtained 8.65%, 18.97%, and 21.22% higher accuracies than the second highest results with various training sample sizes, respectively. The mean difference between the highest (obtained by SSRC-NW) and the second highest accuracies (the first and second ones were achieved via SLRC-l_2 for 10 and 15 training samples and the last one was attained through CRC) was 16.28%. It is well known that the images from the AR database are captured under different facial expressions, illumination conditions, and occlusions. Based on the results, it is worth noting that SSRC-NW can recognize the face with uncontrolled data sets effectively and efficiently, while the conventional SRC cannot.

B. Imbalanced Classification Setting

Different from SSRC-NW, SSRC-W was developed to deal with imbalanced data sets through learning a specific weight for each class reconstruction in its objective function. The weights are learned based on the data distribution and the learning procedure can be found in Algorithm 1. The following experiments were applied on 15 KEEL data sets [27] including 7 binary-class and 8 multiclass data sets. The IRs of the 15 data sets vary from 1.48 to 61.18.

For all of the following experiments, fivefold cross validation [47] was applied. During fivefold cross validation, the data set is first split into five equal parts. Each part is used as the test data set and the remaining parts are the training data sets, where the final result is the mean of the five rounds.

As the distribution of the imbalanced data, ACC is not the suitable measurement for imbalanced classification. For example, there are two classes and the size of the majority class is 95 and the other is 5. When the classifier predicts all the samples to be the majority class, the ACC is still very high (95/100 = 95%). However, the classifier misclassified all the minority class samples with a high ACC. Therefore, ACC cannot display the actual performance in the imbalanced classification.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE II

LIST OF THE SEVEN BINARY-CLASS IMBALANCED DATA SETS AND ITS CORRESPONDING λ VALUES FOR SSRC-W

ID	Dim.	Maj No.	Min No.	IR	λ
BC-1	4	288	24	12	0.7
BC-2	10	2572	164	15.68	0.7
BC-3	10	2692	44	61.18	0.7
BC-4	34	112	63	1.78	0.1
BC-5	5	2289	130	17.61	0.1
BC-6	47	900	100	9	0.4
BC-7	9	42	10	4.2	0.6

In this paper, balanced ACC (BACC) and G-means [48] are utilized to measure the performances of the classifiers in the imbalanced classification. BACC and G-means separately consider the classifier's performance on each of the classes. For a C-class classification problem, the number of the ith class sample is denoted as n_i and the number of the predicted ith class sample is assigned as n_i'. BACC and G-means are calculated as BACC$=(1/C)\sum_{i=1}^{C}(n_i'/n_i)$ and GM$=(\prod_{i=1}^{C}((n_i'/n_i)))^{(1/C)}$, respectively.

Along with the above six comparison classifiers (refer to Section IV-A), eight additional methods were compared with our proposed algorithm (SSRC-W). The eight methods consist of the above six comparison classifiers applied on the preprocessing data sets through SMOTE and two algorithms (Ada-K-NN+GIHS and Ada-K-NN2+GIHS [34]) specially designed for the imbalanced classification. The parameters of the first 12 comparison classifiers utilized the same setting as in Section IV-A and the final two comparison methods belong to the nonparametric model.

C. Binary-Class Imbalanced Data Sets

Table II lists the seven binary-class imbalanced data sets used in this paper. The first column gives the data set ID, Dim. represents the dimensionality, Maj No. and Min No. are short for the number of the majority and minority class samples, respectively, and IR denotes the IR. The last column of the table gives the corresponding λ values for SSRC-W. The optimal parameters of SSRC-W for different data sets were chosen according to the best results.

The results of SSRC-W and the 14 comparison classifiers for the 7 binary-class imbalanced data sets with various IRs are illustrated in Fig. 4. The BACC of SSRC-W is depicted in the red bar. Table III presents the numerical mean results (BACC and GM) of SSRC-W and the 14 comparison classifiers in binary-class imbalanced classification. According to Fig. 4 and Table III, it is obvious that SSRC-W can classify binary-class imbalanced data sets with various IRs effectively and efficiently without any data preprocessing.

D. Multiclass Data Sets

Information of the eight multiclass imbalanced data sets is given in Table IV. Compared with Table II, this table has one more column (Class No.) showing the class numbers of the multiclass data sets. Similar with binary-class imbalanced classification settings, the parameter (λ) of SSRC-W was selected according to the experiments.

TABLE III

NUMERICAL MEAN RESULTS OF SSRC-W AND THE 14 COMPARISON CLASSIFIERS FOR THE 7 BINARY-CLASS IMBALANCED DATA SETS

Classifier	BACC	GM
k-NN	81.72%	0.7354
SRC	79.52%	0.6356
SLRC-l_1	75.13%	0.6825
SLRC-l_2	65.89%	0.3988
CRC	73.82%	0.5681
ProCRC	57.37%	0.2052
k-NN+SMOTE	87.26%	0.8670
SRC+SMOTE	89.23%	0.8822
SLRC-l_1+SMOTE	75.92%	0.6758
SLRC-l_2+SMOTE	54.82%	0.1568
CRC+SMOTE	85.54%	0.8153
ProCRC+SMOTE	80.93%	0.7448
Ada-kNN+GIHS	84.84%	0.8237
Ada-kNN2+GIH	85.41%	0.8415
SSRC-W	**92.39%**	**0.9212**

TABLE IV

LIST OF THE EIGHT MULTICLASS IMBALANCED DATA SETS AND ITS CORRESPONDING λ VALUES FOR SSRC-W

ID	Dim.	Class No.	Maj No.	Min No.	IR	λ
MC-1	13	3	71	48	1.48	0.3
MC-2	16	10	115	105	1.1	0.1
MC-3	5	3	150	30	5	0.2
MC-4	34	6	112	20	5.6	0.01
MC-5	4	3	288	49	5.88	0.1
MC-6	9	5	76	17	4.47	0.4
MC-7	25	5	48	16	3	0.1
MC-8	9	3	1706	131	13.02	0.1

TABLE V

NUMERICAL MEAN RESULTS OF SSRC-W AND THE 14 COMPARISON CLASSIFIERS FOR THE 8 MULTICLASS IMBALANCED DATA SETS

Classifier	BACC	GM
k-NN	79.02%	0.6859
SRC	74.84%	0.5648
SLRC-l_1	74.58%	0.6729
SLRC-l_2	62.06%	0.4061
CRC	71.08%	0.5229
ProCRC	66.36%	0.4642
k-NN+SMOTE	80.87%	0.7658
SRC+SMOTE	80.81%	0.7330
SLRC-l_1+SMOTE	74.27%	0.6721
SLRC-l_2+SMOTE	56.02%	0.2968
CRC+SMOTE	75.89%	0.6071
ProCRC+SMOTE	79.83%	0.7072
Ada-kNN+GIHS	79.67%	0.7687
Ada-kNN2+GIHS	80.06%	0.7709
SSRC-W	**86.54%**	**0.8093**

Fig. 5 illustrates the results of SSRC-W and the 14 comparison classifiers for the 8 multiclass imbalanced data sets with various IRs. The BACC of SSRC-W is depicted in the red bar. Table V presents the numerical mean results (BACC and GM) of SSRC-W and the 14 comparison classifiers in multiclass imbalanced classification. According to Fig. 5 and Table V, it is obvious that SSRC-W can classify multiclass imbalanced

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

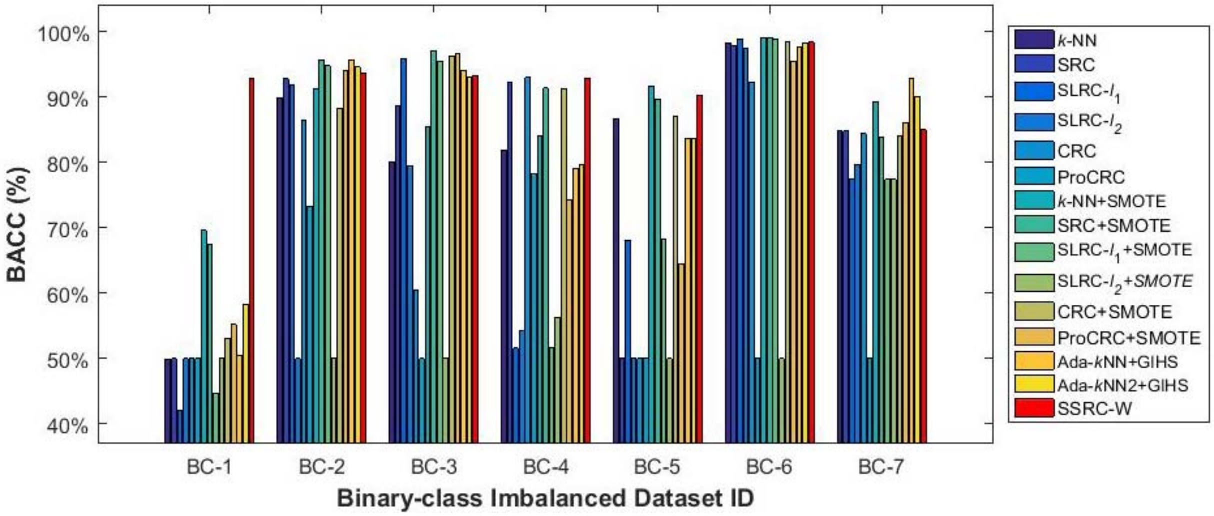

Fig. 4. BACC (%) of SSRC-W and the 14 comparison methods for binary-class imbalanced data sets.

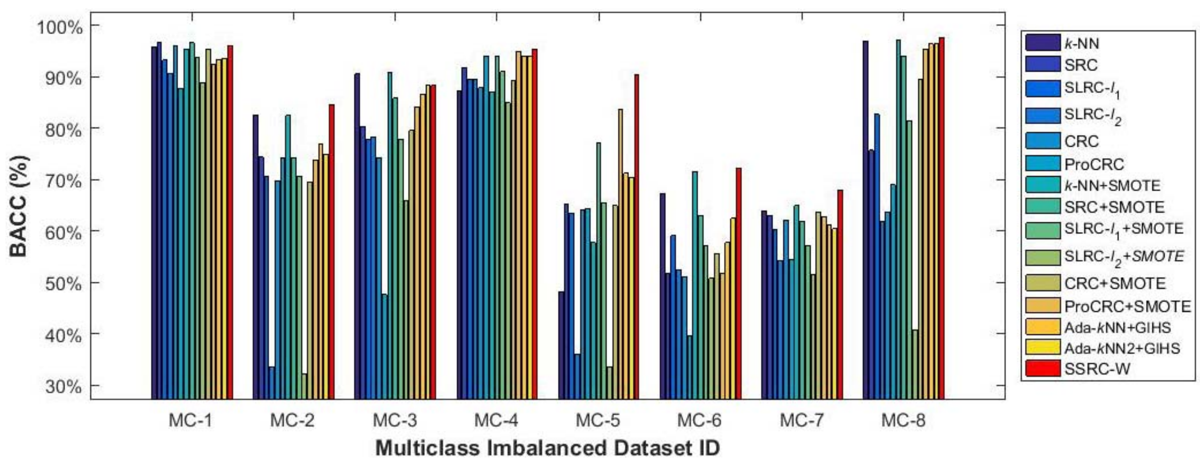

Fig. 5. BACC (%) of SSRC-W and the 14 comparison methods for multiclass imbalanced data sets.

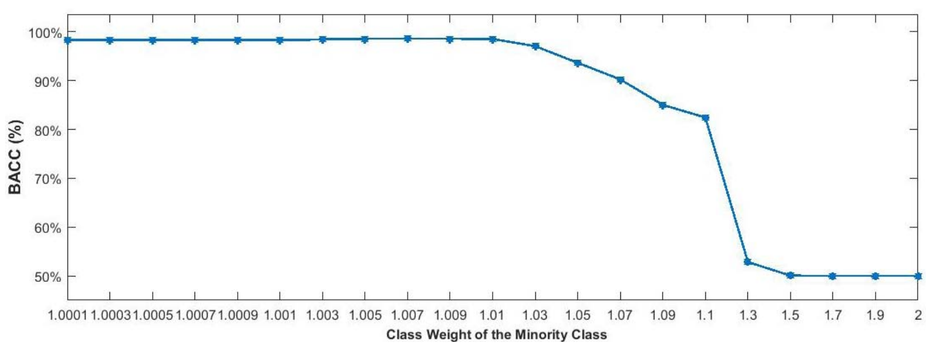

Fig. 6. BACC(%) of SSRC-W on the BC-6 data set with various class weights.

data sets with various IRs effectively and efficiently without any data preprocessing.

E. SSRC Parameters Analysis

According to (8), there are two types of parameters (class weights and tradeoff scale) in SSRC. The class weights are automatically learned from the training samples (refers to Algorithm 1). This section discusses the effects of the two types of parameters for SSRC.

Fig. 6 depicts the results of the BC-6 data set with various weights. There are two classes in this data set: majority and minority subjects. The x-axis gives the weight of the minority class and the majority class weight is set to be 1. The class weight varies from [1.00012]. The BACCs had very small changes when the weight is in [1.00011.01]. After the weight

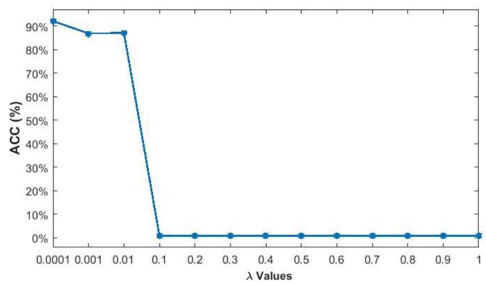

Fig. 7. ACC(%) of SSRC-NW on the AR data set with various λ values.

equals to 1.01, the results decreased very quickly. Hence, the optimal weight varies from [1.00011.01].

Fig. 7 gives an example, where the experimental setting is the AR data set with 20 training data samples in each class, where it has various λ values. It is obvious that 0.0001 is the optimal value for λ in our experiments.

V. Conclusion

SRC is a popular and effective classification algorithm. However, its results are closely related with the distribution of the data. Specifically, it has been shown to perform poorly in the classification of uncontrolled data sets. Furthermore, when it comes to pattern classification using imbalanced data sets, the minority class samples are always misclassified in SRC, where these samples can be more important than the majority class. In order to resolve the above-mentioned issues of SRC, we proposed an SSRC via the use of the class label information during the test sample reconstruction phase. Different from SRC, each training class samples linearly represent the test sample in its corresponding subspace. In SSRC, the reconstruction term is split into different parts, where one part is represented through one subject. In the objective function, a l_1-norm term on the coefficient makes the same class reconstruction term to be close to the test sample. A specially designed class weights learning model is utilized in SSRC to classify imbalanced data sets. To the best of our knowledge, SSRC is the first algorithmic-level method to deal with imbalanced data sets based on SRC. In this paper, two forms of our proposed method were utilized in uncontrolled face recognition (SSRC-NW) and imbalanced classification (SSRC-W), respectively. Through the comparison of the six classifiers using the AR face database (uncontrolled data set) with various training sample sizes, the results showed that SSRC-NW always obtained the best accuracies. For the imbalanced classification, SSRC-W and 14 other methods were tested on 15 public KEEL data sets with various IRs. The proposed method (SSRC-W) also attained the highest results most of the time in classifying both binary-class and multiclass imbalanced data sets.

As part of our future work, we will use the deep learning technology to improve the SRC in uncontrolled data set classification.

References

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

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

[3] M. Elad and M. Aharon, "Image denoising via sparse and redundant representations over learned dictionaries," *IEEE Trans. Image Process.*, vol. 15, no. 12, pp. 3736–3745, Dec. 2006.

[4] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process.*, vol. 19, no. 11, pp. 2861–2873, Nov. 2010.

[5] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification using dictionary-based sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 49, no. 10, pp. 3973–3985, Oct. 2011.

[6] X. Mei and H. Ling, "Robust visual tracking and vehicle classification via sparse representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 11, pp. 2259–2272, Nov. 2011.

[7] S. Yi, Z. He, Y.-M. Cheung, and W.-S. Chen, "Unified sparse subspace learning via self-contained regression," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 28, no. 10, pp. 2537–2550, Oct. 2017.

[8] P. Zhang, X. You, W. Ou, C. L. P. Chen, and Y.-M. Cheung, "Sparse discriminative multi-manifold embedding for one-sample face identification," *Pattern Recognit.*, vol. 52, pp. 249–259, Apr. 2016.

[9] M. Pang, B. Wang, Y.-M. Cheung, and C. Lin, "Discriminant manifold learning via sparse coding for robust feature extraction," *IEEE Access*, vol. 5, pp. 13978–13991, 2017.

[10] S. Gao, I. W.-H. Tsang, and L.-T. Chia, "Kernel sparse representation for image classification and face recognition,' in *Proc. Eur. Conf. Comput. Vis. (ECCV)*. Berlin: Springer, 2010, pp. 1–14.

[11] L. Zhang *et al.*, "Kernel sparse representation-based classifier," *IEEE Trans. Signal Process.*, vol. 60, no. 4, pp. 1684–1695, Apr. 2012.

[12] Y. Chen, N. M. Nasrabadi, and T. D. Tran, "Hyperspectral image classification via kernel sparse representation," *IEEE Trans. Geosci. Remote Sens.*, vol. 51, no. 1, pp. 217–231, Jan. 2013.

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

[14] C. Y. Lu, H. Min, J. Gui, L. Zhu, and Y. K. Lei, "Face recognition via weighted sparse representation," *J. Vis. Commun. Image Represent.*, vol. 24, no. 2, pp. 111–116, Feb. 2013.

[15] Z. Fan, M. Ni, Q. Zhu, and E. Liu, "Weighted sparse representation for face recognition," *Neurocomputing*, vol. 151, pp. 304–309, Mar. 2015.

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

[17] W. Deng, J. Hu, and J. Guo, "In defense of sparsity based face recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2013, pp. 399–406.

[18] W. Deng, J. Hu, and J. Guo, "Face recognition via collaborative representation: Its discriminant nature and superposed representation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 10, pp. 2513–2521, Oct. 2017.

[19] P. Chujai, K. Chomboon, K. Chaiyakhan, K. Kerdprasop, and N. Kerdprasop, "A cluster based classification of imbalanced data with overlapping regions between classes," in *Proc. Int. MultiConf. Eng. Comput. Sci.*, 2017, Hong Kong, Mar. 2017, pp. 353–358.

[20] Q. Kang, L. Shi, M. Zhou, X. Wang, Q. Wu, and Z. Wei, "A distance-based weighted undersampling scheme for support vector machines and its application to imbalanced classification," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 9, pp. 4152–4165, Sep. 2017.

[21] C. Huang, Y. Li, C. C. Loy, and X. Tang, "Learning deep representation for imbalanced classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 5375–5384.

[22] C. Huang, C. C. Loy, and X. Tang, "Discriminative sparse neighbor approximation for imbalanced learning," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 5, pp. 1503–1513, May 2017.

[23] P. Cao, D. Zhao, and O. Zaiane, "An optimized cost-sensitive SVM for imbalanced data learning," in *Proc. Pacific–Asia Conf. Knowl. Discovery Data Mining, (PAKDD)*. Berlin, Germany: Springer, 2013, pp. 280–292.

[24] Y. Zhu, Z. Wang, and D. Gao, "Gravitational fixed radius nearest neighbor for imbalanced problem," *Knowl.-Based Syst.*, vol. 90, pp. 224–238, Dec. 2015.

[25] B. Nikpour, M. Shabani, and H. Nezamabadi-pour, "Proposing new method to improve gravitational fixed nearest neighbor algorithm for imbalanced data classification," in *Proc. 2nd Conf. Swarm Intell. Evol. Comput. (CSIEC)*, Mar. 2017, pp. 6–11.

[26] A. M. Martinez and R. Benavente, “The AR face database,” CVC, New Delhi, India, Tech. Rep. 24, Jun. 1998.

[27] A. Fernández, S. García, M. J. del Jesus, and F. Herrera, “A study of the behaviour of linguistic fuzzy rule based classification systems in the framework of imbalanced data-sets,” *Fuzzy Sets Syst.*, vol. 159, no. 18, pp. 2378–2398, 2008.

[28] A. Fernández, M. J. del Jesus, and F. Herrera, “Hierarchical fuzzy rule based classification systems with genetic rule selection for imbalanced data-sets,” *Int. J. Approx. Reasoning*, vol. 50, no. 3, pp. 561–577, 2009.

[29] Z. Ma, Y. Lai, W. B. Kleijn, Y.-Z. Song, L. Wang, and J. Guo, “Variational Bayesian learning for Dirichlet process mixture of inverted Dirichlet distributions in non-Gaussian image feature modeling,” *IEEE Trans. Neural Netw. Learn. Syst.*, to be published.

[30] S. Cai, L. Zhang, W. Zuo, and X. Feng, “A probabilistic collaborative representation based approach for pattern classification,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 2950–2959.

[31] X. Zou, Y. Feng, H. Li, and S. Jiang, “SROT: Sparse representation-based over-sampling technique for classification of imbalanced dataset,” in *Proc. IOP Conf. Ser. Earth Environ. Sci.*, vol. 81, no. 1, 2017, p. 012201.

[32] N. Japkowicz, “The class imbalance problem: Significance and strategies,” in *Proc. Int. Conf. Artif. Intell., (ICAI)*, 2000, pp. 111–117.

[33] N. V. Chawla, K. W. Bowyer, L. O. Hall, and W. P. Kegelmeyer, “SMOTE: Synthetic minority over-sampling technique,” *J. Artif. Intell. Res.*, vol. 16, no. 1, pp. 321–357, 2002.

[34] S. S. Mullick, S. Datta, and S. Das, “Adaptive learning-based k-nearest neighbor classifiers with resilience to class imbalance,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 11, pp. 5713–5725, Nov. 2018.

[35] Y. Lu, Y.-M. Cheung, and Y. Y. Tang, “Dynamic weighted majority for incremental learning of imbalanced data streams with concept drift,” in *Proc. 26th Int. Joint Conf. Artif. Intell.*, 2017, pp. 2393–2399.

[36] R. A. Horn, R. A. Horn, and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1990.

[37] W. J. Dixon and F. J. Massey, Jr., *Introduction to Statistical Analysis*, 2nd ed. New York, NY, USA: McGraw-Hill, 1957, pp. 191–195.

[38] *Understanding Variance, Covariance, and Correlation*. Accessed: Jun. 14, 2018. [Online]. Available: https://www.countbayesie.com/blog/2015/2/21/variance-co-variance-and-correlation

[39] A. Beck and M. Teboulle, “A fast iterative shrinkage-thresholding algorithm for linear inverse problems,” *SIAM J. Imag. Sci.*, vol. 2, no. 1, pp. 183–202, 2009.

[40] P. R. Johnstone and P. Moulin. (2015). “A Lyapunov analysis of FISTA with local linear convergence for sparse optimization.” [Online]. Available: https://arxiv.org/abs/1502.02281v1

[41] S. Tao, D. Boley, and S. Zhang, “Local linear convergence of ISTA and FISTA on the lasso problem,” *SIAM J. Optim.*, vol. 26, no. 1, pp. 313–336, 2016.

[42] N. Parikh and S. Boyd, “Proximal algorithms,” *Found. Trends Optim.*, vol. 1, no. 3, pp. 127–239, Jan. 2014.

[43] J.-K. Kamarainen, V. Kyrki, and H. Kalviainen, “Invariance properties of Gabor filter-based features-overview and applications,” *IEEE Trans. Image Process.*, vol. 15, no. 5, pp. 1088–1099, May 2006.

[44] S. Ting and B. Zhang, “Diabetes mellitus detection based on facial block texture features using the Gabor filter,” in *Proc. IEEE 17th Int. Conf. Comput. Sci. Eng. (CSE)*, Dec. 2014, pp. 1–6.

[45] T. Shu, B. Zhang, and Y. Y. Tang, “An extensive analysis of various texture feature extractors to detect diabetes mellitus using facial specific regions,” *Comput. Biol. Med.*, vol. 83, pp. 69–83, Apr. 2017.

[46] L. E. Peterson, “K-nearest neighbor,” *Scholarpedia*, vol. 4, no. 2, p. 1883, 2009.

[47] R. Kohavi *et al.*, “A study of cross-validation and bootstrap for accuracy estimation and model selection,” in *Proc. Int. Joint Conf. AI*, vol. 14, no. 2, Aug. 1995, pp. 1137–1145.

[48] M. Kubat and S. Matwin, “Addressing the curse of imbalanced training sets: One-sided selection,” in *Proc. 4th Int. Conf. Mach. Learn.*, Nashville, TN, USA, vol. 97, Jul. 1997, pp. 179–186.

Ting Shu (S’14) received the B.A. degree in software engineering from Chongqing University, Chongqing, China, in 2013, and the M.S. degree in software engineering from the University of Macau, Macau, China, in 2015, where she is currently pursuing the Ph.D. degree.

Her current research interests include machine learning, pattern recognition, medical biometrics, image processing, and multiview learning.

Bob Zhang (M’11) received the B.A. degree in computer science from York University, Toronto, ON, Canada, in 2006, the M.A.Sc. degree in information systems security from Concordia University, Montreal, QC, Canada, in 2007, and the Ph.D. degree in electrical and computer engineering from the University of Waterloo, Waterloo, ON, Canada, in 2011.

He was with the Center for Pattern Recognition and Machine Intelligence, Carnegie Mellon University, Pittsburgh, PA, USA, where he was a Post-Doctoral Researcher with the Department of Electrical and Computer Engineering. He is currently an Assistant Professor with the Department of Computer and Information Science, University of Macau, Macau, China. His current research interests include medical biometrics, pattern recognition, and image processing.

Yuan Yan Tang (S’88–M’88–SM’96–F’04–LF’16) is currently a Chair Professor with the Faculty of Science and Technology, University of Macau, Macau, China. He is also a Professor with Chongqing University, Chongqing, China, an Adjunct Professor with Concordia University, Montreal, QC, Canada, and an Honorary Professor with Hong Kong Baptist University, Hong Kong. He has authored or co-authored over 400 academic papers, over 25 monographs, books, and book chapters. His current research interests include wavelets, pattern recognition, and image processing.

Mr. Tang is an IAPR Fellow. He was the General Chair, the Program Chair, and a Committee Member for many international conferences. He is the Founder and the Editor-in-Chief of the *International Journal on Wavelets, Multiresolution, and Information Processing*. He is the Founder and the Chair of the Pattern Recognition Committee in the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS. He is the Founder and the General Chair of the Series International Conferences on Wavelets Analysis and Pattern Recognition. He is the Founder and the Chair of the Macau Branch of International Association of Pattern Recognition. He is an Associate Editor of several international journals.