ABSTRACT Recently, collaborative representation based classifiers (CRC) have shown outstanding performances in recognition tasks. The key to success of most CRC algorithms states that the testing samples can be coded well by a suitable dictionary globally, while the local knowledge between samples has not been fully considered. We observe that the representations of similar samples have a high degree of similarity. In order to take advantage of this important similarity information, this paper proposes a new local knowledge-based collaborative representation model for image classification. Specifically, certain adjacent training samples of the testing image should be determined firstly, and then the representations of these neighborhoods can be applied to guide the coefficients of the testing samples to be more discriminative. Further, we derive a robust version of the proposed method to treat the face recognition with occlusions or corruptions. Extensive experiments are carried out to show the superiority of the proposed method over other state-of-the-art classifiers on various image recognition tasks.

INDEX TERMS Collaborative representation, local consistency, robustness, image recognition, supervised learning.

I. INTRODUCTION
Image recognition, which aims to determine the labels of those query samples, has attracted the attention of many scholars in the machine learning community [1], [2]. In many computer vision applications, ranging from hyperspectral image analysis [3], image denoising [4] to face recognition [5], image recognition plays an important and fundamental role. Due to the uncontrollable factors such as illumination and occlusion which commonly occur in images, designing effective and efficient recognition methods is still a challenging and urgent topic [6].

In the past few years, a series of classifiers have been proposed for image recognition, which could be roughly divided into two types [7]: parametric-dependent algorithms and non-parametric algorithms. For parametric-dependent algorithms [8], there are a large number of weight parameters needed to be learnt in the training process, which not only consumes too much computational time, but also can easily make the algorithms over-fitting. To avoid these shortcomings, non-parametric algorithms focus on how to represent the query samples with respect to certain training samples directly, and the representations can be applied for classification further. Typical representative non-parametric methods include nearest neighborhood (NN) and its variants (NC, NFL). Compared with parametric-dependent algorithms, recent works also have found the advantages of non-parametric methods in dealing with small sample data [9].

The core idea in non-parametric classifiers is to find the discriminative representation of the testing sample. In [10], Wright et al. advocated the query sample should be coded with respect to all the training samples, and the coefficient can be constrained by \( \ell_1 \)-norm to reflect the contributions of different samples. This so-called sparse representation based
classifier (SRC) further utilizes the approximation residuals within each class for classification. With the success of SRC, a number of variants have been proposed. In [11], the authors proposed a structured SRC (SSRC) by exploiting the intra-block structure in all the training data. Based on the motivation that the image can be represented biased towards their respective categories, Huang et al. [12] presented a class specific SRC method to make those class groups compete for coding the query image. Further, Mi et al. [13] proposed an adaptive class preserving SRC to balance the sparsity with block lasso regression adaptively. On the other hand, some recent works concentrate on studying the internal working mechanism of SRC. Xu et al. [14] gave an efficient two-phase test sample sparse representation to get sparsity and also provided a probability interpretation. From a different perspective, Zhang et al. [15] pointed out that it is the collaborative representation principle between the entire training set plays a more critical role than the sparsity induced by the \(\ell_1\)-norm constrained. Motivated by this observation, a simple but efficient collaborative representation based classifier (CRC) was proposed, in which the \(\ell_1\) regularizer was replaced by the \(\ell_2\)-norm regularization. Naturally, this ridge regression-based CRC method can be solved efficiently by solving a ridge regression problem. Yang et al. [16] further gave a comparative analysis of the two methods and concluded that sparsity is necessary when there is enough training sample, if not, CRC without sparsity can get higher accuracy than SRC. In [7], the authors tried to reveal the intrinsic mechanism of CRC from the probabilistic view, and proposed a new probabilistic CRC (ProCRC) by incorporating the relationship between training data of each object class and the testing sample. Though SRC and CRC have shown their respective advantages in different recognition tasks, how to get a more discriminative representation of the query sample efficiently and further improve the performances of SRC or CRC by incorporating another appropriate knowledge are still open problems [17].

Most existing CRC methods use the linear combination of all the training samples, while neglect the local property within data. Here the local knowledge means the similar sample may share similar properties [5]. In this paper, based on the motivation that the representations of the similar images should have a high degree of similarity, we propose a local knowledge-based CRC (LKCR) method. In particular, the neighboring training samples are searched firstly, then their representations with respect to the entire training set can be obtained precisely. Further, these representations are utilized to constrain the calculation of the coefficient of the testing sample. By this way, the obtained representation can be discriminative and share nearly the same sparsity like SRC. Nevertheless, the proposed LKCR is based on the efficient \(\ell_2\) regularization, thus there exists an efficient analytical solution. Moreover, by utilizing the \(\ell_1\)-based error model, the proposed method can be extended to a robust version to treat the recognition problems with noise. An early short version of this work can be found in [18], and here a more detailed analysis and evaluations of the proposed method will be given. In summary, the contributions of this paper can be listed as follows.

1) A new local prior knowledge is proposed and incorporated into the framework of CRC for image recognition.
2) The proposed model can be solved efficiently and achieve the similar sparsity without the \(\ell_1\) constraint which consumes too much computation time.
3) Extensive experimental results on various image databases can show the outstanding performances of the proposed method.

The rest of this paper is organized as follows. In Section II, we give a brief review of related works. Section III introduces the proposed method in details. Experimental results are presented in Section IV. Finally, Section V concludes this paper.

II. RELATED WORKS

In this section, we briefly introduce some typical non-parametric classifiers. Firstly, the notations used throughout this paper should be provided. Given \(C\) types of samples, and there will be a total of \(N (N = nC)\) training samples, where \(n\) means the number of training samples in each class. Each two-dimensional image used in this paper is converted to a column vector. And the entire training set can be denoted as \(\mathbf{D} = \{\mathbf{D}_1, \mathbf{D}_2, \ldots, \mathbf{D}_C\}\), where \(\mathbf{D}_c\) is the samples from the \(c\)th category. The test sample is \(\mathbf{y}\). Let \(\mathbf{l}_D\) be the label space of the training set. And the target of a classification system is to determine the label of every test sample \(\mathbf{y}\), i.e., \(\hat{\mathbf{l}}_y\).

A. NN AND NC

Among non-parametric classifiers, the nearest neighborhood (NN) classifier [19] is the most simple yet popular one. For a test sample \(\mathbf{y}\), NN calculates its distance between each training sample one-by-one, then find the sample which is closest to \(\mathbf{y}\), as follows,

\[
\hat{\mathbf{x}} = \arg\min_{\mathbf{x}\in \mathbf{D}} \text{dist}(\mathbf{y} - \mathbf{x})
\]

where \(\text{dist}(\cdot)\) means the distance between \(\mathbf{y}\) and \(\mathbf{x}\), such as the Euclidean distance. Then the decision of \(\hat{\mathbf{l}}_y\) is according to the label of \(\hat{\mathbf{x}}\), i.e., \(\hat{\mathbf{l}}_y = \hat{\mathbf{l}}_{\hat{\mathbf{x}}}\). We can see there is no training process in NN classifier, thus it can be considered as a lazy algorithm. The biggest shortcoming of NN lies that NN is very sensitive to the noise.

Nearest centroid (NC) classifier is a representative variant of NN. In NC, the centroid of each class is calculated to represent this kind of samples. Generally, the mean value is chosen as the centroid. Let \(\mathbf{x}_c\) denote the centroid of the \(c\)th class \(\mathbf{D}_c\), NC determines the identity of \(\mathbf{y}\) as follows,

\[
\hat{\mathbf{l}}_y = \arg\min_{c} \text{dist}(\mathbf{y} - \mathbf{x}_c).
\]  

It is obvious that the centroid \(\mathbf{x}_c\) is not so easy to be affected by the noise or outliers in training samples. Hence, NC classifier can be more robust than NN. Here we should note that the robustness and performance of NN or NC can be improved
by choosing other distance functions, such as Euclidean distance, Cosine distance and Cityblock distance and so on [20].

**B. LRC**

Different from NN and NC using only one sample to represent a certain class, linear regression classifier (LRC) [21] utilizes all the training samples from some class to approximate the test sample \( y \). This process can be modeled as

\[
\bar{a}_c = \arg \min_{a_c} \| y - D_c a_c \|_2^2.
\]

(3)

where \( D_c \) means the training samples in the \( c \)th class, and \( a_c \) is the representation. Then \( y \) can be classified into the class which has the smallest reconstruction error, as follows,

\[
l_y = \arg \min_c r_c.
\]

(4)

where \( r_c = \| y - D_c \bar{a}_c \|_2 \) is the approximation residuals according to certain class.

**C. SRC**

As mentioned before, SRC method assumes the test sample \( y \) can be represented by the linear combination of all the training samples, while different samples have different contributions. The \( \ell_1 \)-norm is incorporated to make the representation sparse. The SRC can be modeled in mathematics as

\[
\bar{a} = \arg \min_a \| y - Da \|_2^2 + \lambda \|a\|_1,
\]

(5)

where \( \lambda \) denotes the regularization parameter. The core idea of SRC is that a test sample should be a linear combination of all training set, but only the coefficients from the same object have significant values while the others should be close to zero. The classification rule is also based on the reconstruction error between \( y \) and each class. That is

\[
r_c(y) = \| y - D_c \bar{a}_c \|_2,
\]

(6)

where \( \bar{a}_c \) is the coefficients associated with the \( c \)th class in \( \bar{a} \). Then \( l_y \) can be determined by

\[
l_y = \arg \min_c r_c(y).
\]

(7)

**D. CRC AND ProCRC**

Zhang et al. [15] explored the internal working mechanism of CRC, and pointed out that the collaborative representation in CRC was modeled based on \( \ell_1 \)-norm, the complexity was very high. According to these motivations, a new collaborative representation classifier (CRC) was proposed which constrained the coefficient with a smooth \( \ell_2 \)-norm regularization. Its objective function is as follows,

\[
\bar{a} = \arg \min_a \| y - D_a \|_2^2 + \lambda \|a\|_2^2.
\]

(8)

Different from SRC, the solution of (8) can be obtained directly by solving a ridge regression problem, just like,

\[
\bar{a} = (D^T D + \lambda I)^{-1} D^T y.
\]

(9)

To further interpret the working mechanism of CRC methods, Cai et al. [7] proposed to explain CRC from the probabilistic theory. Consequently, a probabilistic collaborative representation (ProCRC) was proposed, which is modeled as follows,

\[
\bar{a} = \arg \min_a \| y - Da \|_2^2 + \lambda \|a\|_2 + \gamma \sum_{c=1}^C \| D_c a_c - D_c \bar{a}_c \|_2^2,
\]

(10)

where the third term emphasizes the probability which \( y \) belongs to all object classes. The same as CRC, the solution of ProCRC can be got analytically as

\[
\bar{a} = \left( D^T D + \gamma \sum_{c=1}^C \left( \bar{D}_c^T \bar{D}_c + \lambda I \right)^{-1} \right)^{-1} D^T y.
\]

(11)

where \( \bar{D}_c = D - D'_c \), and \( D'_c = [0, \ldots, D_c, \ldots, 0] \).

**III. THE PROPOSED METHOD**

In this section, the proposed LKCR model will be presented in details. First, we will explain the motivation to propose this new classifier. Then, the modeling process of LKCR will be presented. Finally, we will extend the proposed LKCR to a robust version to treat the face recognition with noise.

**A. MOTIVATION**

The existing representation based classification methods can be divided into two kinds roughly. The first kind follows \( \ell_1 \)-norm constraint, whose typical representative is SRC [10]. The other one is the collaborative representative-based algorithms, such as CRC [15]. Generally speaking, \( \ell_1 \) regularizer can introduce sparsity into the representations, thereby achieving good classification rate. Nevertheless, we know that SRC-based classifiers consume too much computation time in optimization. CRC is based on the efficient \( \ell_2 \) regularization technology. One advantage of this kind of methods is there is a closed-form solution. But the classification accuracy sometimes decreases with the reduction of sparsity [22].

Now we turn to see the representation based classifiers from the viewpoint of manifold learning. The representation based classifiers try to encode each testing sample with a linear combination of certain training samples, but they all ignore the local consistence information within samples. Here we use an example to illustrate this new perspective. In Fig. 1, the first two samples A and B are from the same subject, and the third image C is selected from other subject. Obviously, the similarity degree between A and B is higher than A (or B) and C. The second and third column in Fig. 1 is the coefficients computed by CRC and SRC, respectively. By observation and contrast, we can not only find the difference in sparsity between SRC and CRC, but also the consistent relationship between the coefficients and the samples. That means, if the samples are similar to each other, their representations based on the dictionary should also be similar. In this work, we aim at incorporating this prior knowledge into the modeling process of CRC, thus the representations...
of query samples can be more discriminative. What is more, the obtained coding can maintain a certain sparsity with an efficient way.

### B. LKCR MODEL

As a basis to propose a new model, we first take a general analysis on the representation based classifiers. The objective function of the methods aforementioned can be reformulated into a unified framework as follows,

$$\mathcal{J} = \|y - \Theta a\|^2_2 + \lambda \|a\|^q_p + \gamma E_a,$$

(12)

where the first term denotes the error fidelity, and $\Theta a$ is the representative vector to approximate the test sample $y$. $\|a\|^q_p$ is the regularizer, $E_a$ can be seen the prior knowledge of different models derived for various purposes, and both $\lambda$ and $\gamma$ are regularization parameters. The specific settings of different models are summarized in Table 1, in which $\sqrt{\cdot}$ and $\times$ denote there exists the terms or not. The key work to build a new model is how to set these terms.

NN as well as its variant NC only utilizes one vector to approximate the testing sample. In LRC, the samples from the same category are combined linearly to approximate the test sample. The classification rule is based on the shortest distance from a query sample to each combination. All these three classifiers have no any constraints and prior knowledge on the coefficient. For the SRC and CRC, they both encode the test sample over the entire training samples. The difference between them is SRC emphasizes the sparsity by $\ell_1$-norm on $a$, while CRC emphasizes the collaboration representation of all the samples, and use the efficient $\ell_2$-norm to constrain the coefficient. These two classifiers have advantages in different scenarios, but the same shortcoming is that they do not incorporate any other prior knowledge to improve the recognition performance further. Taking advantage of some useful prior knowledge is an important strategy to build new classification model. For example, by giving a probabilistic interpretation of CRC, ProCRC further considers the correlation between $Da$ and $D_a c$ via the inner product, and achieved outstanding performance by incorporating this prior knowledge. In [20], Peng et al. incorporated the locality preservation property into the CRC framework (LCCR). In LCCR, the local consistency was preserved by incorporating the regularization term as follows

$$E_a = \sum_{x_i \in X} \|x_i - Da\|^2_2,$$

(13)

where $X$ is $y$‘s neighboring samples from the training space.
From the analysis above, we can find that the prior knowledge plays an important role in improving the performances of classifiers. In this work, we will propose a new local knowledge which will be incorporated into the framework of CRC. The new local prior knowledge can be formulated as

$$E_a = \sum_{i=1}^{K} w_i |a_i - a|^2_2,$$  

where \( \{a_i\}, i = 1, \dots, K \) are the representations of \( y \)'s neighborhoods in the training space. Thus, the first step in our proposed LKCR is to determine the neighborhoods of every query sample. Specifically, given test sample \( y \), we can obtain its \( K \)-neighboring samples \( X \) from the entire training set. Then we should calculate the representations of \( X \) by the following:

$$A = \arg\min_{A} \|X - DA\|_2^2 + \lambda_1 \|A\|_2^2.$$  

where \( A = [a_1, a_2, \ldots, a_K] \). Obviously, these training representations \( A \) can be calculated precisely.

The prior term \( E_a \) establishes the relationships between \( a \) and \( A \). To guarantee the effectiveness of the proposed model, here we utilize the strategy, e.g. set \( p = q = 2 \). We further replace (14) into Eq. (12), and the proposed LKCR model can be expressed as,

$$\bar{a} = \arg\min_{a} \|y - Da\|_2^2 + \lambda |a|_2^2 + \gamma \sum_{i=1}^{K} w_i |a_i - a|^2_2.$$  

We utilize \( w_i \) as the locality adaptor to further measure the importance of each adjacent training samples to represent the testing sample, here it is determined as [23]

$$w_i = \exp \left( \frac{-\|y - x_i\|_2^2}{\sigma} \right).$$  

where \( \sigma \) is the bandwidth parameter and used for adjusting the weight decay speed. And the \( w_i \) usually can be normalized to be between \((0, 1]\) by subtracting \( \min_{i} \|y - x_i\|_2^2 \) in the inner molecule of (17) [24]. The existence of \( w_i \) indicates that two close samples are set with a larger weight, so that their representations can be more similar.

As LKCR model is based on the \( \ell_2 \)-norm regularization, the optimization is very simple and efficient. By taking the the derivation of Eq. (16) with respect to \( a \) to be 0, we can get the solution of LKCR efficiently as

$$\bar{a} = \left(D^T D + \lambda I + \gamma \Sigma \Sigma w_i I\right)^{-1} \left(D^T y + \gamma \Sigma w_i a_i\right).$$  

We remark that the neighboring training samples of the test sample can be obtained by many methods, like \( \epsilon \)-ball or \( k \)NN method [20]. For simplicity, in this paper the \( \epsilon \)-ball method is utilized. The classification rules follows the same strategy shown in (6) and (7).

In summary, the procedure of our proposed model can be depicted in Algorithm 1.

**C. ROBUST LKCR MODEL**

In this part, the proposed LKCR model will be extended to its robust version (R-LKCR) to treat the face recognition in uncontrolled environments. In face recognition field, the classification performances of many existing classifiers (like CRC) can be affected seriously by the occlusions and corruptions in the images. As pointed out in [25], the measure used in the model’s error fidelity plays an important role in the robustness. To reduce the interference of occlusions on our classifier, here we enforce the \( \ell_2 \)-norm to measure the error fidelity. And the novel R-LKCR model can be formulated as follows,

$$\bar{a} = \arg\min_{a} \|y - Da\|_1 + \lambda |a|_2^2 + \gamma \sum_{i=1}^{K} w_i |a_i - a|^2_2.$$  

We can see the \( \ell_2 \)-norm is replaced by the \( \ell_1 \)-norm in the robust model, which cannot be solved directly by solving a ridge regression. Here we turn to utilize the iterative reweighted least square (IRLS) to optimize (22) effectively. In iterative process of IRLS, we should first calculate a diagonal weight matrix \( W \) whose diagonal element can be calculated as,

$$W(i, i) = \frac{1}{|D(i, :)a - y_i|},$$  

where \( D(i,:) \) represents the \( i \)th row of \( D \), and \( y_i \) denotes the \( i \)th element of testing sample \( y \). Following the optimization strategy of IRLS, the Eq. (22) can be rewritten as

$$\bar{a} = \arg\min_{a} \|y - Da\|^T W(y - Da)$$

$$+ \lambda |a|_2^2 + \gamma \sum_{i=1}^{K} w_i |a_i - a|^2_2.$$  

\[\text{(24)}\]
Thus in each iterative step, the representation \( a \) can be solved by:

\[
\bar{a} = (D^T WD + \lambda I + \gamma \Sigma w I)^{-1}(D^T Wy + \gamma \Sigma w I a). 
\]  
(25)

In the entire iterative process, the representation \( a \) and weighting matrices \( W \) should be updated alternatively until satisfying the convergence conditions. A detailed theoretical analysis has been given in [26], [27] to guarantee the iteration process in IRLS method can convergence to the global minima, here we do not repeat it.

D. RELATIONSHIP WITH OTHER METHODS

From the descriptions above, we can see there are two stages in the proposed LKCR model. Indeed, the related two-stages strategy has also been applied in SRC-KNS [28] and CFFR [29]. But it is different from ours. In SRC-KNS or CFFR, some candidate training classes are determined in the first stage, whose labels will be utilized as the entire label space in the second stage. Though nearest subspace can be chosen and reduce the computational time, the real labels of testing samples may be missing due to the inaccuracy of the first stage. In LKCR, the representations of the nearest training samples are calculated to induce the coefficient which will be obtained in the second stage to be more discriminative, and the candidate label space has not changed. For the locality constraint, the LLC [24] applied this prior knowledge from another viewpoint. In LLC, the locality is just utilized as a series of weights which can measure the similarity between the testing sample and all the training images. By comparison, the locality prior knowledge is utilized more fully and rationally in our proposed method. The advantages of our proposed method will be further verified in the experimental results.

IV. EXPERIMENTAL RESULTS

In this section, several experiments will be carried to show the outstanding performance of our proposed method over many state-of-the-art classifiers. Subsection IV-A will depict the performance of LKCR in treating face recognition tasks. And in subsection IV-B, we will evaluate its recognition ability in action recognition with two complicated datasets \( KTH \) [30] and \( HMDB51 \) [31]). Further, we will verify the robustness of LKCR model in subsection IV-C. The source codes of the competitive classifiers are gotten from the corresponding authors, in which the hyper-parameters are tuned carefully to achieve their respective best performance in each experiment. All experiments involved in this section are conducted on MATLAB 2016a with a 3.40GHz and 15.90 GB RAM machine.

A. RESULTS ON FACE RECOGNITION

Face recognition has been an active research point in many computer vision and artificial intelligence fields, such as smart transportation, security monitor and financial security. In this part, two popular face datasets: LFW [32] and AR [33] are utilized to evaluate the performances of the proposed method. Their detailed descriptions are listed as follows:

- **LFW.** This database contains 13,123 facial images from 5,749 subjects. The images are pictured in uncontrollable environments with large variations in occlusions, misalignment, expressions, illumination and so on. Some selected samples are shown in Fig 2(a). Here we select a subset which includes 143 classes with more than 11 samples per class for the experiments in this subsection.

- **AR.** This database contains over 4000 facial images from 126 persons (56 female and 70 male). There are a great variations in these images like expression, illumination and disguise (wearing scarves or sunglasses). In this subsection, we take a subset which includes 50 male and 50 female with changing in illuminations and expressions for experiments. Fig 2(b) shows several example images. And the images used here are all resized to \( 60 \times 43 \).

![FIGURE 2. Some samples from two face database. (a) LFW database, (b) AR database.](image-url)
There are three hyper-parameters in the proposed LKCR model, two regularization parameter $\lambda$, $\gamma$ and the number of neighboring samples $K$. In this experiment, we set $\lambda = 0.1$ empirically, while $\gamma$ and $K$ are selected by appreciate grid search strategy. Several other popular classifiers, include SVM [34], SRC, LRC, CESR [35], CRC, ProCRC and LCCR are compared with our proposed LKCR. To conduct the comparison fairly, we take 10 trials with randomly combining the training and testing samples, and the mean results are reported.

In the experiment on LFW, we randomly choose 10 images from each class as training set, and the rest are used for testing. Table 2 shows the classification rates of different classifiers. Here we note that four popular features: Gray, LBP, FFT and Gabor [20] are extracted from the images for the recognition in this experiment. The best results for each feature is marked in bold in Table 2. From the results, it can be seen that the classifiers’ performance can be influenced heavily by different kinds of features. In all these cases, SVM performances the worst, LRC and CRC can achieve the similar classification accuracy. The performances of SRC and CESR can be improved slightly with the help of $\ell_1$-norm regularization. By enhancing the collaborative learning from each class-specific, ProCRC can have a further improvements compared with SRC and CRC classifier. Obviously, the proposed LKCR method can perform the best in most cases, which confirms the rationality of this new local prior knowledge.

In the experiment on AR, 7 samples in each category are randomly selected for training and the rest for testing. To comprehensively show the performance of different classifiers on AR, we project the images into different dimensions with PCA, i.e., $\dim = 54, 120, 300, 2580$. In each case, 10 trials are carried out, then the average results are reported. Table 3 shows the performances of different classifiers with diverse feature dimensions. From Table 3, it can been that with the increasing of feature dimensions, the classification performances of all classifiers can be higher. Among these classifiers, SVM, CESR and LRC can always achieve similar rates in different cases. SRC can perform a little better than these three classifiers. And CRC classifier can achieve a higher classification rates than SRC in this databases, especially for the high dimension case. And our proposed method can still have great advantages over other competitors in most cases, which further verify the effectiveness of our method in treating the image recognition problems.

Now we further evaluate the computation time of our proposed method and the competitive classifiers on the LFW database with FFT features. The running time of different algorithms are listed in Table 4. Since a extra projection $P$ and $a_i$ should be calculated first in the first stage of LKCR, we can see the computation time is slower than the CRC and ProCRC, which just has one stage with the analytical solution. But our method is still much faster than the SRC and CESR which is constrained by the $\ell_1$-norm. Comprehensively considering the efficiency and accuracy, the advantage of our proposed LKCR still can be realized.

### TABLE 2. Classification rates (%) on LFW of different classifiers.

| Algorithm | Gray | FFT | Gabor | LBP |
|-----------|------|-----|-------|-----|
| SVM [34]  | 3.1  | 5.8 | 42.4  | 18.5|
| SRC [10]  | 7.3  | 33.6| 68.7  | 61.6|
| CESR [35] | 9.6  | 37.6| 65.7  | 63.1|
| LRC [21]  | 7.1  | 13.9| 25.4  | 26.3|
| CRC [15]  | 7.3  | 14.0| 25.4  | 26.3|
| LCCR [20] | 12.3 | 39.5| 72.2  | 67.8|
| ProCRC [7]| 13.1 | 38.2| 71.3  | 67.9|
| LKCR      | 17.7 | 39.7| 73.7  | 69.0|

### TABLE 3. Classification rates (%) on AR of different classifiers.

| Algorithm | 54   | 120  | 300  | 2580 |
|-----------|------|------|------|------|
| SVM [34]  | 73.4 | 81.0 | 82.0 | 83.1 |
| SRC [10]  | 81.7 | 88.7 | 90.3 | —    |
| CESR [35] | 74.0 | 81.4 | 84.6 | 84.6 |
| LRC [21]  | 80.6 | 90.1 | 93.6 | 82.3 |
| CRC [15]  | 80.6 | 90.4 | 94.0 | 94.4 |
| LCCR [20] | 84.7 | 90.7 | 94.1 | 94.4 |
| ProCRC [7]| 81.1 | 89.0 | 92.6 | 94.4 |
| LKCR      | 80.7 | 90.7 | 94.3 | 95.0 |

### B. RESULTS ON ACTION RECOGNITION

In this subsection, two complicated databases: KTH and HMDB51 are utilized to test the performances of LKCR model on action recognition problem. In computer vision community, action recognition has attracted lots of attention, as automatic inference of actions needs so much activity information. Several other existing algorithms are used as the competitors, such as SRC, NFLS [36], LCCR, DU [37] and RDU [40]. SVM is also applied as a baseline method. The hyper-parameters in each competitive algorithm is tuned to make the classifier perform the best. Fig. 3 depicts some example images from the two databases, and the following are their detailed descriptions:

- **KTH.** This database contains six kinds of actions, including running, hand waving, walking, boxing, jogging and hand clapping. These actions are acted lots of times by twenty-five persons in 4 scenarios: indoors, outdoors with different clothes, outdoors with scale variations and outdoors. In our experiment, 100 videos for each kind of actions are chosen from the entire database to do the training and testing.

- **HMDB51.** There are simple facial actions, general body movements, and human interactions and so on, which is collected from many sources, ranging from digitized movies to YouTube videos. Following the same experimental configuration in [40], 100 video sequence for each kind of action are selected from the original database to conduct our experiment in this part.
In the experiment on *KTH*, the methods provided by [41] is used to represent each video data. In order to evaluate the recognition accuracy of various classifiers comprehensively, 5, 10 and 15 video features of each action are for training, and the remaining data can be used for testing. The performances of various classifiers in different scenarios are reported in Table 5. By analyzing these results, it can be found the accuracy of different classifiers can increase with the increasing of training number. For all the classifiers, SVM performs the worst, especially for the small training sample. By utilizing the kernel trick into DU [37], the recognition accuracy of RDU is better than most classifiers. However, our proposed method performs the best in nearly all the cases.

**TABLE 5.** Classification rate (%) on KTH of different classifiers.

| Algorithm | 5 train | 10 train | 15 train |
|-----------|---------|----------|----------|
| SVM [34]  | 81.40   | 88.89    | 93.83    |
| SRC [10]  | 91.40   | 90.56    | 94.90    |
| SSRC [38] | 92.11   | 90.00    | 96.28    |
| NFLS [36] | 90.53   | 89.63    | 93.14    |
| DPL [39]  | 91.08   | 91.30    | 96.00    |
| DU [37]   | 90.70   | 90.37    | 95.29    |
| LCCR [20] | 92.98   | 93.89    | 94.90    |
| RDU [40]  | 92.63   | 91.48    | 96.28    |
| LKCR      | 95.09   | 95.37    | 96.86    |

We further test the method’s ability in action recognition with a more complicated *HMDB51* database. To compare the performance of different classifiers fairly, we use the representations of videos provided by [41]. We select different number of training samples, e.g., 5, 10 and 15 video features of each action for training, and the remaining video data are for testing. Table 6 lists the comparison results of different classifiers. From this table, we can see the LCCR and LKCR method have a great advantage than other algorithms. It can conclude that local consistency knowledge is an important prior for action recognition. Obviously, our LKCR model can perform better than LCCR in all scenarios. As the *HMDB51* is quite complex, we can see the recognition rates in Table 6 are all quite low. But here we should note that the aim of this paper is to propose a suitable classifier, which plays an important role in the recognition system. And the performances can be further improved if more discriminative features can be incorporated, like deep features [42]. This will also be one of our future research directions.

**TABLE 6.** Classification rate (%) on HMDB51 of different classifiers.

| Algorithm | 5 train | 10 train | 15 train |
|-----------|---------|----------|----------|
| SVM [34]  | 4.87    | 6.36     | 8.88     |
| SRC [10]  | 9.72    | 12.64    | 14.97    |
| SSRC [38] | 9.45    | 12.18    | 14.90    |
| NFLS [36] | 8.90    | 10.68    | 12.78    |
| DPL [39]  | 9.40    | 13.70    | 17.40    |
| DU [37]   | 9.91    | 14.21    | 16.80    |
| LCCR [20] | 17.83   | 22.70    | 26.27    |
| RDU [40]  | 9.85    | 14.44    | 17.07    |
| LKCR      | 18.16   | 23.88    | 27.96    |

**C. EVALUATIONS ON ROBUST FACE RECOGNITION**

In this subsection, the robustness performance of our proposed R-LKCR model will be investigated. In practice,
robustness is one of the key indicators of an algorithm. We will test the R-LKCR on the Extended Yale B [43] face database with different outliers, and compare it with the relative SRC, CRC, LKCR, R-DCNR [44] and R-SRC classifiers [45].

Firstly, we conduct the experiments on the Extended Yale B with random block occlusion. For each individual in this dataset, we randomly select nearly a half images as the training data, and the rest can be used for testing. To evaluate the robustness, a local region in each image is selected and replaced with an unrelated monkey image. The occlusion levels are set as 10%, 20%, 30% and 40%. Fig. 5 shows some images with block occlusions. For each case, 10 trails are conducted and the average results are reported in Table 7. From Table 7, we can see the classification accuracy of classifiers can be influence seriously by block occlusions, especially for the $\ell_2$ error fidelity based methods, e.g., SRC, CRC and LKCR. As the robust version of the corresponding classifiers, R-DCNR and R-SRC can show great improvements in noise environment. Nevertheless, our proposed R-LKCR can still have advantages in comparison with these competitors. In nearly all the cases, our method can achieve the best recognition rates.

| Algorithm | 10% | 20% | 30% | 40% |
|-----------|-----|-----|-----|-----|
| SRC [10]  | 94.59 | 80.62 | 67.92 | 58.26 |
| CRC [15]  | 94.37 | 80.87 | 67.63 | 59.35 |
| LKCR      | 95.48 | 83.21 | 68.58 | 61.72 |
| R-DCNR [44] | 98.25 | 97.48 | 89.97 | 87.36 |
| R-SRC [45] | 97.49 | 95.60 | 90.19 | 83.47 |
| R-LKCR    | 98.59 | 97.33 | 90.50 | 88.21 |

D. PARAMETER SELECTION

In our proposed LKCR classifier, there are two important hyperparameters, i.e., $\lambda$ and $\gamma$ to be tuned. They are used to balance the importance of the corresponding constraint terms. $\lambda$ is used to avoid the trivial solution of coefficient $a$, and the second term is utilized to preserve the local similarity of the training data. The sensitivity of these two parameters will be analyzed in this subsection.

We conduct this experiment on the KTH action database with different training number. And $\lambda$ and $\gamma$ are both chosen from the candidate set $\{2^{-3}, 2^{-2}, \ldots, 2^2\}$, then we perform the LKCR method with different combinations of these parameters. Fig. 7 shows the relationships of classification accuracy with the different combinations of parameters. It is obvious that our method are not so sensitivity to the selections of parameters. And satisfactory accuracy can be achieved when parameters $\lambda$ and $\gamma$ locate in the range of $[2^{-3}, 2^{-1}]$ and $[2^{-2}, 2^{-1}]$.  

Table 8. Robust face recognition results (%) with random pixel corruption.
V. CONCLUSION
In this paper, by following the strategy of collaborative representation, a new local knowledge-based collaborative representation classifier (LKCR) was proposed for image recognition. The key idea is to incorporate the local consistency prior information into the framework of collaborative representation, then make representations of query sample to be more discriminative for image recognition. To make our proposed method be robust in the noise environment, a robust model based on the $\ell_1$ error fidelity was presented. Extensive experiments on diverse image databases were conducted to demonstrate the superiority of our proposed method in comparison with other classifiers. An unsolved problem of LKCR is the search strategy of query sample’s neighboring training samples. For real-world large training data, it is time-consuming to determine the adjacent training data. In the future, we will try to combine the multidimensional indexing method [47] to speed up the exact search.

REFERENCES
[1] C. L. P. Chen and Z. Liu, “Broad learning system: An effective and efficient incremental learning system without the need for deep architecture,” IEEE Trans. Neural Netw. Learn. Syst., vol. 29, no. 1, pp. 10–24, Jan. 2018.
[2] S. Yang, J.-Z. Guo, and J.-W. Jin, “An improved Id3 algorithm for medical data classification,” Comput. Electr. Eng., vol. 65, pp. 474–487, Jan. 2018.
[3] Y. Kong, X. Wang, Y. Cheng, and C. Chen, “Hyperspectral imagery classification based on semi-supervised broad learning system,” Remote Sens., vol. 10, no. 5, p. 685, 2018. [Online]. Available: https://www.mdpi.com/2072-4292/10/5/685
[4] L. Liu, L. Chen, C. L. P. Chen, Y. Y. Tang, and C. M. Pun, “Weighted joint sparse representation for removing mixed noise in image,” IEEE Trans. Cybern., vol. 47, no. 3, pp. 600–611, Mar. 2017.
[5] J. Jin, Z. Liu, and C. L. P. Chen, “Discriminative graph regularized broad learning system for image recognition,” Sci. China Inf. Sci., vol. 61, no. 11, Nov. 2018.
[6] Z. Jiang, Z. Lin, and L. S. Davis, “Label consistent K-SVD: Learning a discriminative dictionary for recognition,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 11, pp. 2651–2664, Nov. 2013.
[7] 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. (CVPR), Jun. 2016, pp. 2950–2959.
[8] J.-W. Jin and C. L. Philip Chen, “Regularized robust broad learning system for uncertain data modeling,” Neurocomputing, vol. 322, pp. 58–69, Dec. 2018.
[9] M. Yang, X. Wang, G. Zeng, and L. Shen, “Joint and collaborative representation with local adaptive convolution feature for face recognition with single sample per person,” Pattern Recognit., vol. 66, pp. 117–128, Jun. 2017.
[10] H. Zhang, Y. Zhang, and T. S. Huang, “Pose-robust face recognition via sparse representation,” Pattern Recognit., vol. 46, no. 5, pp. 1511–1521, May 2013.
[11] E. Elhamifar and R. Vidal, “Block-sparse recovery via convex optimization,” IEEE Trans. Signal Process., vol. 60, no. 8, pp. 4094–4107, Aug. 2012.
[12] S. Huang, Y. Yang, D. Yang, L. Huangfu, and X. Zhang, “Class specific sparse representation for classification,” Signal Process., vol. 116, pp. 38–42, Nov. 2015. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S016516841500153X
[13] J.-X. Mi, Q. Fu, and W. Li, “Adaptive class preserving representation for image classification,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 7427–7435.
[14] Y. Xu, D. Zhang, J. Yang, and J.-Y. Yang, “A two-phase test sample sparse representation method for use with face recognition,” IEEE Trans. Circuits Syst. Video Technol., vol. 21, no. 9, pp. 1255–1262, Sep. 2011.
[15] L. Zhang, M. Yang, and X. Feng, “Sparse representation or collaborative representation: Which helps face recognition?” in Proc. Int. Conf. Comput. Vis., Nov. 2011, pp. 471–478.
[16] J. Yang, L. Zhang, Y. Xu, and J. Y. Yang, “Beyond sparsity: The role of $\ell_1$-optimizer in pattern classification,” Pattern Recognit., vol. 45, no. 3, pp. 1104–1118, 2012.
[17] H. Yuan, X. Li, F. Xu, Y. Wang, L. L. Lai, and Y. Y. Tang, “A collaborative-competitive representation based classifier model,” Neurocomputing, vol. 275, pp. 627–635, Jan. 2018.
[18] J. Jin, C. L. Philip Chen, and J. Zhou, “Image recognition using manifold constrained collaborative representation,” in Proc. Int. Conf. Secur., Pattern Anal., Cybern. (SPAC), Dec. 2018, pp. 52–56.
[19] T. Cover and P. Hart, “Nearest neighbor pattern classification,” IEEE Trans. Inf. Theory, vol. 13, no. 1, pp. 21–27, Jan. 1967.
[20] X. Peng, L. Zhang, Z. Yi, and K. K. Tan, “Learning locality-constrained collaborative representation for robust face recognition,” Pattern Recognit., vol. 47, no. 9, pp. 2794–2806, Sep. 2014.
[21] I. Naseem, R. Togneri, and M. Bennamoun, “Linear regression for face recognition,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 11, pp. 2106–2112, Nov. 2010.
[22] Y. Xu, Z. Zhong, J. Yang, J. You, and D. Zhang, “A new discriminative sparse representation method for robust face recognition via l2 regularization,” IEEE Trans. Neural Netw. Learn. Syst., vol. 28, no. 10, pp. 2233–2242, Feb. 2017.
[23] X. Tang, G. Feng, and J. Cai, “Weighted group sparse representation for undersampled face recognition,” Neurocomputing, vol. 145, pp. 402–415, Dec. 2014.
[24] J. Wang, J. Yang, K. Yu, F. Lv, T. Huang, and Y. Gong, “Locality-constrained linear sparse representation method for use with face recognition,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 32, no. 11, pp. 2106–2112, Nov. 2010.
[25] Y. Naseem, R. Togneri, and M. Bennamoun, “Linear regression for face recognition,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 13, no. 1, pp. 21–27, Jan. 1967.
[26] H. Yuan, X. Li, F. Xu, Y. Wang, L. L. Lai, and Y. Y. Tang, “A collaborative-competitive representation based classifier model,” Neurocomputing, vol. 275, pp. 627–635, Jan. 2018.
[27] D. Ba, B. Babadi, P. L. Purdon, and E. N. Brown, “Convergence and stability of iteratively re-weighted least squares algorithms,” IEEE Trans. Signal Process., vol. 62, no. 1, pp. 183–195, Jan. 2014.

FIGURE 7. Analysis of the parameter sensitivity of LKCR on KTH with (a) 5 train; (b) 10 train; (c) 15 train.
JUNWEI JIN received the B.S. degree from Ningxia University, Ningxia, China, in 2013, and the M.S. and Ph.D. degrees from the University of Macau, Macau, in 2015 and 2019, respectively. Since May 2020, he has been an Assistant Professor with the School of Computer and Communication Engineering, Henan University of Technology. His current research interests include machine learning, computer vision, and neural networks.

YANTING LI received the B.S. and M.S. degrees from Henan University, China, in 2012 and 2015, respectively, and the Ph.D. degree from the University of Macau, Macau, in 2019. She has been an Assistant Professor with the School of Computer and Communication Engineering, Zhengzhou University of Light Industry. Her current research interests include function approximation theory, image processing, sparse coding, and machine learning.

LIJUN SUN received the B.S. degree from Xidian University, China, in 1989, the M.S. degree from the Hefei University of Technology, China, in 2001, and the Ph.D. degree from Northwestern Polytechnical University, China, in 2005. She is currently a Professor with the College of Information Science and Engineering, Henan University of Technology, Zhengzhou, China. Her research interests include artificial intelligence, wireless sensor networks, computational intelligence, image processing, and robot application.

JIANYU MIAO received the M.S. degree in applied mathematics from Zhengzhou University, Zhengzhou, China, in 2014, and the Ph.D. degree in operational research and cybernetics from the University of Chinese Academy of Sciences, Beijing, China, in 2018. Since 2018, he has been on the faculty of the College of Information Science and Engineering, Henan University of Technology, Zhengzhou, China, where he is currently a Lecturer. His current research interests include optimization and machine learning.

C. L. PHILIP CHEN (Fellow, IEEE) is a Chair Professor with the Department of Computer and Information Science, Faculty of Science and Technology, University of Macau, Macau. The University of Macau Engineering and Computer Science Programs receiving the Hong Kong Institute of Engineers’ (HKIE) Accreditation and the Washington/Seoul Accord is his utmost contribution in engineering/computer science education for Macau as the former Dean of the Faculty. His current research interests include optimization and machine learning.

He is a Fellow of AAAS, IAPR, CAA, and HKIE. He is a Foreign Member of the Academy of Europe and a member of the European Academy of Sciences and Arts. He was the IEEE SMC Society President, from 2012 to 2013, and is currently the Vice President of the Chinese Association of Automation (CAA). He was the Chair of the TC 9.1 Economic and Business Systems of International Federation of Automatic Control, from 2015 to 2017, and also a Program Evaluator of the Accreditation Board of Engineering and Technology Education (ABET) of the U.S. for computer engineering, electrical engineering, and software engineering programs. He received the 2016 Outstanding Electrical and Computer Engineers Award from his alma mater, Purdue University after he graduated from the University of Michigan, Ann Arbor, MI, USA. He is the Editor-in-Chief of the IEEE TRANSACTIONS ON CYBERNETICS and an Associate Editor of several IEEE Transactions.