Score level fusion in representation-based classification method for face recognition

Zi-Qi Li¹, Jun Sun¹,², Xiao-Jun Wu¹,² and He-Feng Yin¹

Abstract
Recent years have witnessed the success of representation-based classification method (RBCM) in the domain of face recognition. Collaborative representation-based classification (CRC) and linear regression-based classification (LRC) are two representative approaches. CRC is a global representation method which uses all training samples to represent test samples and utilizes representation residuals to perform classification, whereas LRC is a local representation method which exploits training samples from each class to represent test samples. Related researches indicate that the combination of LRC and CRC can fully exploit the representation residuals produced by them, thus improving the performance of RBCM. However, the representation coefficients obtained by CRC usually contain negative values which may result in overfitting problem. Therefore, to solve this problem to some extent, the combination of LRC and non-negative least square-based classification (NNLSC) is proposed in this paper. Experimental results on benchmark face datasets show that the proposed method is superior to the combination of LRC and CRC and other state-of-the-art RBCMs. The source code of our proposed method is available at https://github.com/li-zi-qi/score-level-fusion-of-NNLS-and-LRC.

Keywords
Representation-based classification method, collaborative representation-based classification, linear regression-based classification, non-negative least square-based classification, score level fusion, face recognition

Received 16 June 2019; revised 31 January 2020; accepted 8 April 2020

Introduction
Currently, and in the past few years, representation-based classification method (RBCM),¹,² which can be applied in a wide variety of fields – e.g. face recognition³–⁶ and hyperspectral imagery classification – has received more attention.⁷–⁹ RBCM represents a test sample as a linear combination of training samples and then employs the representation results to classify the test sample. Generally, RBCM can be divided into two categories: global representation methods and local representation methods.

The global representation methods utilize all the training samples to represent test samples, while the local representation methods use class-specific training samples. Actually, both of them employ their representation residuals for the final classification process. In 2008, Wright et al.¹⁰ proposed a sparse representation-based classification (SRC) method, which achieves impressive recognition results in the community of face recognition.¹ SRC first obtains the sparse coefficient of a test sample using the dictionary formed by all the training samples, then reconstructs the test sample using the sparse coefficient and each class of training samples, and finally designates the test sample into the category which yields the minimum reconstruction error. And related experiments have proved that SRC is robust to image occlusion and pixel corruption,

¹School of Internet of Things Engineering, Jiangnan University, Wuxi, China
²Jiangsu Provincial Laboratory of Pattern Recognition and Computational Intelligence, Jiangnan University, Wuxi, China

Corresponding author:
Jun Sun, School of Internet of Things Engineering, Jiangnan University, Wuxi 214122, China.
Email: junsun@jiangnan.edu.cn
pointing out a new direction for face recognition research. However, the original SRC method needs to solve the \( \ell_1 \)-norm optimization problem. It is worth noting that when the number of training samples increases, the sparse decomposition process becomes very slow. In order to overcome this problem, Zhang et al.\(^{1,2,13}\) proposed to use the \( \ell_2 \)-norm to replace the original \( \ell_1 \)-norm, namely the collaborative representation classification (CRC) algorithm, which can achieve comparable performance to SRC, and the calculation speed is faster, making it more suitable for practical applications. Vo et al.\(^{14}\) and Xu et al.\(^{15}\) proposed a non-negative least square classification (NNLSC) algorithm, which is also a global representation method. NNLSC represents a test sample as a non-negative linear combination of all training samples, then calculates the residuals using each class of training samples and the corresponding non-negative coefficients, and finally classifies the test sample into the category corresponds to the minimum residual. Experimental results show that the performance of NNLSC is better than traditional classifiers, e.g. nearest neighbor (NN), nearest centroid (NC), and nearest subspace (NS).

In contrast to the global representation methods, the local representation methods use class-specific training samples to represent test samples and exploit the local information between training samples to improve performance. Naseem et al.\(^{16}\) presented a linear regression classification (LRC) method, which selects \( k \) training samples adjacent to the current test sample by \( k \)NN to form the dictionary. Xu et al.\(^{17}\) proposed a local sparse representation (LSRC) method, which selects \( k \) training samples adjacent to the current test sample by \( k \)NN to form the dictionary. The first stage represents a test sample as a linear combination of training samples and uses the training samples to determine \( M \) adjacent training samples of the current test sample. The second stage employs the CRC method to classify the test sample on the \( M \) adjacent samples.

Since both the global representation methods and the local representation methods use residuals for classification, only the training samples used to represent the current test sample are different. Additionally, these two types of methods have various advantages on different datasets. Therefore, Zhang et al.\(^{19}\) proposed to combine CRC and LRC and used the residuals obtained by CRC and LRC to perform score level fusion. The experimental results showed that the performance of the fusion method is better than that of the single algorithm. However, the representation coefficient obtained by CRC usually contains some negative values, which may cause overfitting problem. To tackle this problem, we propose a new score level fusion method which is based on LRC and NNLSC. To be specific, residuals of LRC and NNLRS are first produced, then these residuals are normalized and fused in a weighted sum manner; finally, the test sample is classified to the class that leads to the minimum residual.

Here, we clearly show the advantages of our proposed scheme:

1. Non-negative constraint of NNLSC can avoid overfitting problem.
2. Score level fusion strategy of LRC and NNLSC is presented, in which the complementary information of residuals produced by local and global representation approaches can be fully exploited.

This paper is structured as follows: “Related work” section reviews several related methods which contain LRC, CRC, and NNLSC. Score level fusion on the combination of NNLSC and LRC and the analysis of the proposed method is presented in the following section. “Experimental results and analysis” section shows the results of our experiments on publicly available face databases and analyzes the parameter in our proposed algorithm. Conclusion and future works are presented in the final section.

**Related work**

Suppose there are \( n \) training samples for \( C \) classes, the \( i \)th class contains \( n_i \) training samples (\( i = 1, 2, \ldots , C \)), \( \sum_{i=1}^{C} n_i = n \). Stack one training sample into a column vector, then all the column vectors form the training data matrix \( \mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \ldots , \mathbf{X}_C] \in \mathbb{R}^{d \times n} \) and \( \mathbf{X}_i = [\mathbf{x}_{i1}, \mathbf{x}_{i2}, \ldots , \mathbf{x}_{in}] \in \mathbb{R}^{d \times n_i} \), where \( d \) represents the dimension of samples.

**Linear regression classification**

The linear regression classification method is based on linear subspace hypothesis. Suppose a test sample \( \mathbf{y} \in \mathbb{R}^{d \times 1} \) belongs to the \( i \)th class, then it can be represented as a linear combination of all the training samples belonging to the \( i \)th class, i.e.

\[
\mathbf{y} = \beta_1 \mathbf{x}_{i1} + \beta_2 \mathbf{x}_{i2} + \cdots + \beta_{n_i} \mathbf{x}_{in_i} \quad (1)
\]

Equation (1) can be written in the following matrix form

\[
\mathbf{y} = \mathbf{X}_i \beta_i \quad (2)
\]

where \( \beta_i = [\beta_{1i}, \beta_{2i}, \ldots , \beta_{n_i}]^T \in \mathbb{R}^{n_i \times 1} \) is the coefficient vector, which can be obtained by least square estimation

\[
\hat{\beta}_i = (\mathbf{X}_i^T \mathbf{X}_i)^{-1} \mathbf{X}_i^T \mathbf{y} \quad (3)
\]
Using the $i$th class samples $X_i$ and the corresponding coefficient vector $\hat{\beta}_i$, we can get a reconstructed sample

$$\tilde{y}_i = X_i\hat{\beta}_i = X_i(X_i^T X_i)^{-1}X_i^T y$$  \hspace{1cm} (4)

If the test sample belongs to the $i$th class, then $\tilde{y}_i$ is the closest to $y$, so by calculating the distance between $y$ and $\tilde{y}_i$

$$d_i(y) = \|y - \tilde{y}_i\|_2$$  \hspace{1cm} (5)

We can classify $y$ into the category corresponding to the minimum distance

$$\text{identity}(y) = \arg\min_i d_i(y), \quad i = 1, 2, \ldots, C$$  \hspace{1cm} (6)

**Collaborative representation-based classification**

LRC encodes a test sample on each class of training samples, whereas CRC represents the test sample on all training samples, i.e.

$$y = z_{1,1}x_{1,1} + \cdots + z_{1,n_1}x_{1,n_1} + \cdots + z_{C,1}x_{C,1} + \cdots + z_{C,n_C}x_{C,n_C}$$  \hspace{1cm} (7)

which can be reformulated as

$$y = Xz$$  \hspace{1cm} (8)

where $z = [z_1; z_2; \ldots; z_C] \in \mathbb{R}^{n \times 1}$ and $z_i = [z_{i,1}, z_{i,2}, \ldots, z_{i,n_C}]^T \in \mathbb{R}^{n \times 1}$ are representation coefficients. The objective function of CRC is formulated as

$$\hat{z} = \arg\min_z \left\{ \|y - Xz\|_2^2 + \lambda \|z\|_2^2 \right\}$$  \hspace{1cm} (9)

where $\lambda$ is a regularization parameter which balances the minimization between the terms of residual and regularization. Equation (9) has the following closed-form solution

$$\hat{z} = (X^T X + \lambda I)^{-1}X^T y$$  \hspace{1cm} (10)

Besides the class-specific reconstruction error, Zhang et al.\textsuperscript{1,2} found that the representation coefficient $\hat{z}_i$ contains some discriminative information for classification, so the classification rule can be formulated as

$$\text{identity}(y) = \arg\min_i \frac{\|y - X_i\hat{z}_i\|_2}{\|\hat{z}_i\|_2}$$  \hspace{1cm} (11)

**Non-negative least square-based classification**

Inspired by non-negative matrix factorization (NMF)\textsuperscript{20} and Xu et al.,\textsuperscript{15} we represent a test sample $y$ as a non-negative linear combination of training samples. The non-negative constraint can yield more discriminative coefficient vectors, which ultimately promote the performance of classification.

The least square classification method usually solves the following problem

$$\min_{z} \|y - Xz\|_2^2$$  \hspace{1cm} (12)

The coefficient will contain negative values, but in many practical problems, the coefficient is often required to be non-negative. Therefore, by introducing the non-negative constraint into equation (12), we can obtain the following non-negative least square model

$$\min_{z} \|y - Xz\|_2^2, \text{s.t.} \ z \geq 0$$  \hspace{1cm} (13)

It is worth noting that in Xu et al.,\textsuperscript{15} they solved equation (13) under the framework of alternating direction method of multipliers (ADMM).\textsuperscript{21} By introducing an auxiliary variable $\gamma$, equation (13) can be reformulated as

$$\min_{z} \|y - Xz\|_2^2, \text{s.t.} \ z = \gamma, \gamma \geq 0$$  \hspace{1cm} (14)

**Score level fusion method based on NNLSC and LRC**

Combine NNLSC and LRC

Zhang et al.\textsuperscript{19} proposed the combination of CRC and LRC, because they both use residuals to classify an input test sample. The rationality of this combination has twofold. First, the residuals of them have little correlation, and CRC is a global representation method which uses all training samples to represent test samples, whereas LRC is a local representation method which only exploits training samples from each class. Second, the superiority of CRC and LRC is different on various datasets, which can enhance the stability of performance to some extent under combination.

Zhang et al.\textsuperscript{1,2} found that CRC can achieve comparable performance to SRC, and the speed is faster. However, the goal of CRC is to make training samples represent the test sample as good as possible, that is, the reconstruction error is the smallest; thus,
the coefficient will contain negative values, which will lead to overfitting problem. In order to tackle this problem, we employ NNLSC to get the global representation coefficient. The experimental results show that the performance by fusing NNLSC and LRC is better than that of by combining CRC and LRC.

Suppose \( p = [p_1, p_2, \ldots, p_C] \) and \( q = [q_1, q_2, \ldots, q_C] \) are the residuals obtained by NNLSC and LRC, respectively.

\[
p = \|y - X\hat{\mathbf{a}}\|_2 \tag{15}
\]

\[
q = \|y - X\hat{\mathbf{b}}\|_2 \tag{16}
\]

Then, the above two residuals are normalized according to the following formulations

\[
p'_i = \frac{p_i - p_{\text{min}}}{p_{\text{max}} - p_{\text{min}}} \tag{17}
\]

\[
q'_i = \frac{q_i - q_{\text{min}}}{q_{\text{max}} - q_{\text{min}}} \tag{18}
\]

where \( p_{\text{max}} \) and \( p_{\text{min}} \) are the largest and smallest residuals by NNLSC, and \( q_{\text{max}} \) and \( q_{\text{min}} \) are the largest and smallest residuals by LRC. The fused residuals on the \( i \)th class can be formulated as

\[
r_i = (1 - \omega)p'_i + \omega q'_i \tag{19}
\]

where \( \omega > 0 \) is a balancing parameter.

Finally, we can classify the test sample \( y \) into the category that yields the least residual

\[
\text{identity \( (y) = \arg \min_i r_i, \quad i = 1, 2, \ldots, C \)} \tag{20}
\]

The complete procedures of our proposed method are outlined in Algorithm 1.
Algorithm 1. Score level fusion of NNLSC and LRC

Input: training data matrix $X$, test data $y$, parameter $\omega$ for fusion.

Output: the identity of $y$.

1. Compute the coefficient $\hat{\alpha}_i$ of NNLSC by equation (13) and the corresponding residual $p$ by equation (15);
2. Obtain the coefficient $\hat{\beta}_i$ of LRC by equation (3) and the corresponding residual $q$ by equation (16);
3. Normalize $p$ and $q$ by equations (17) and (18);
4. Compute the fused residual $r_i$ by equation (19);
5. Classify $y$ according to equation (20).

Analysis of the proposed method

In order to demonstrate the effectiveness of our proposed algorithm, we here present an example on the ORL database. Figure 1 is a test image from the fifth class, whose size is $56 \times 46$. The training images are the first three samples per subject. Figure 2 shows the residuals obtained by CRC. There are 40 residuals corresponding to 40 subjects. We can see that the residual from the 21st class is a little smaller than that from the fifth class, whereas in Figure 3, the fifth class has the least residual obtained by NNLSC, which can correctly classify the test image.

Figure 4 is a test image from the third class on the ORL database. Figure 5 displays the residuals obtained by CRC+LRC, we can observe that the 26th class has the least residual; thus, the test sample is misclassified into the 26th class. The residuals of our proposed method are plotted in Figure 6, one can see that the third class leads to the minimum residual, which means that the test sample is correctly classified.

Experimental results and analysis

In this section, we report the performance of the score level fusion based on NNLSC and LRC under three publicly available datasets, i.e. FERET\textsuperscript{23}, GT\textsuperscript{24} and XM2VTS\textsuperscript{25} databases, the details of which are listed in Table 1. To illustrate the superiority of our proposed method, SRC\textsuperscript{10}, CRC\textsuperscript{1,2}, CCRC\textsuperscript{26}, NCRC\textsuperscript{27}, LRC\textsuperscript{16}, NNLSC\textsuperscript{15} and LRC+CRC\textsuperscript{19}.

![Figure 4](image) A test image in the ORL database (the third class).

![Figure 3](image) The residual obtained by NNLSC, and the fifth class has the least residual.
approaches are compared. The parameters $\lambda_{CRC}$ in CRC and CRC+LRC, $\lambda_1$ and $\lambda_2$ in CCRC, and $\lambda_{NCRC}$ in NCRC are selected from the set \{10^{-6}, 10^{-5}, \ldots, 10^{-1}, 0.25, 0.5, 1, 10, 100\}. To solve SRC optimization problem, SolveFISTA.m\textsuperscript{28} is employed in this paper. Another parameter $\omega$ in our proposed method is selected from the set \{0.1, 0.2, \ldots, 0.9\}. Finally, we select the optimal values.

**Figure 5.** The residual obtained by CRC+LRC, and the 26th class has the least residual.

**Figure 6.** The residual obtained by NNLSC+LRC, and the third class has the least residual.
by experiments, the details of which on each database are summarized in Table 2. All experiments are conducted with MATLAB R2018b under Windows 10 on a 3.30 GHz CPU and 16 GB RAM machine.

**FERET database**

The FERET face database includes 1400 face images of 200 subjects, each having seven images with the variation in pose, facial expressions, and lighting conditions. Some typical face images are shown in Figure 7. All face images are resized as 40 × 40 pixels in our experiments, and the first one to four face images per subject are treated as training samples and the remaining as testing samples.

The classification results of various competing methods under different number of training samples are summarized in Table 3. It can be seen that our proposed method outperforms other algorithms in most cases expect when the number of training samples is four per subject. By combining CRC and LRC, it can obtain better classification accuracy than those, respectively. The parameter $\lambda_{CRC}$ in CRC and CRC + LRC is selected as $1 \times 10^{-2}$, and the contribution of LRC in CRC + LRC is 0.9. By introducing the constraint of non-negative, NNLSC outperforms CRC. The parameter $\lambda$ in our proposed method is 0.9. When the number of training samples per class is three, the results of all algorithms drop. The reason behind this phenomenon may be that we select the first three images fixed over all subjects. The reasonable way is to randomly choose the training face images in each class.

**GT database**

The GT database contains 750 face images from 50 subjects. For per individual, there are 15 face images with variation in frontal and/or leaned face images under different scales, facial expressions, and lighting conditions. In our experiments, each image is cropped and resized to 60 × 50 pixels, and some example images are shown in Figure 8. The first one to six face images of each class are selected as training samples and the rest as testing samples.
Table 4 details the recognition results of various methods. We can see that all the algorithms increase consistently when the number of training samples increases, and our proposed method can achieve better recognition results when the parameter \(\omega = 0.5\) or 0.6. By integrating CRC and LRC, it can acquire better performance than respective method (i.e. CRC and LRC), and we set the parameter \(\lambda_{CRC} = 0.25\) in the related methods, and the weight of LRC in CRC + LRC is 0.1. One can see that NNLSC is superior to CRC by introducing the non-negative constraint.

Table 3. Classification accuracy (%) of competing methods on the FERET database.

| Methods | 1  | 2  | 3  | 4  | 5  | 6  |
|---------|----|----|----|----|----|----|
| SRC     | 45.83 | 57.20 | 52.38 | 68.00 |
| CRC     | 41.08 | 57.20 | 48.75 | 54.33 |
| CCRC    | 42.08 | 57.20 | 48.25 | 53.83 |
| NCRC    | 41.33 | 57.00 | 49.00 | 54.67 |
| LRC     | 44.50 | 63.10 | 58.88 | **76.33** |
| NNLSC   | 49.67 | 63.10 | 57.63 | 72.67 |
| CRC + LRC | 47.92 | 65.00 | 59.00 | 76.17 |
| Proposed (\(\omega = 0.9\)) | **49.92** | **66.10** | **60.50** | **76.17** |

Bold values signify the highest recognition rate accuracy.

Table 4. Classification accuracy (%) of various methods on the GT database.

| Methods | 1  | 2  | 3  | 4  | 5  | 6  |
|---------|----|----|----|----|----|----|
| SRC     | 38.71 | 47.38 | 49.17 | 52.18 | 53.60 | 62.44 |
| CRC     | 36.29 | 45.54 | 50.00 | 54.18 | 59.00 | 66.22 |
| CCRC    | 33.43 | 44.46 | 48.67 | 51.64 | 55.40 | 61.33 |
| NCRC    | 36.29 | 45.54 | 50.17 | 53.82 | 59.00 | 66.89 |
| LRC     | 36.14 | 46.62 | 51.33 | 55.64 | 59.60 | 67.33 |
| NNLSC   | 37.71 | 50.46 | 52.50 | 59.09 | 62.60 | 69.56 |
| CRC + LRC | 36.71 | 47.38 | 52.50 | 56.55 | 60.20 | 68.89 |
| Proposed (\(\omega = 0.5\)) | 38.71 | **50.92** | 52.83 | **60.00** | 62.80 | 70.22 |
| Proposed (\(\omega = 0.6\)) | **38.86** | 50.77 | **53.33** | 59.82 | **63.20** | **70.67** |

Bold values signify the highest recognition accuracy.
XM2VTS database

There are 2360 face images (295 subjects and each has 8 face images) in the XM2VTS database. These face images were captured from four recording sessions with two face images per subject, which had no expression changes under the same conditions. Figure 9 shows some face images from this database. In our experiments, we treat the first one to four images as training samples, and the remaining as testing samples. These images are of size $55 \times 51$ pixels.

Table 5 lists the experimental results with various algorithms on the XM2VTS database. It can be seen that with the increase of the number of training samples in each individual, the performance of all competing methods improves steadily, and our proposed method outperforms other compared approaches except when the first two images are used for training. We choose $\omega = 0.3, 0.4, \text{ and } 0.5$ in score level fusion. NNLSC consistently outperforms CRC. For CRC + LRC, we set the parameter $\lambda_{CRC} = 0.25$, and the occupation of LRC is 70%.

Parameter analysis

In our proposed method, the parameter $\omega$ used for score level fusion needs to be determined. To examine how $\omega$ influences the performance of our proposed method, we conduct experiments on the FERET database, and the first two images per subject are used for

| Methods          | 1     | 2     | 3     | 4     |
|------------------|-------|-------|-------|-------|
| SRC              | 61.26 | 64.75 | 74.10 | 76.69 |
| CRC              | 67.99 | 78.25 | 86.51 | 89.92 |
| CCRC             | 69.83 | 78.87 | 86.98 | 89.66 |
| NCRC             | 68.09 | 78.19 | 86.64 | 89.92 |
| LRC              | 60.97 | 71.19 | 82.85 | 86.27 |
| NNLSC            | 70.31 | 79.60 | 88.68 | 91.02 |
| CRC + LRC        | 69.78 | 79.60 | 88.14 | 90.51 |
| Proposed ($\omega = 0.3$) | 70.65 | 79.38 | 88.81 | 91.10 |
| Proposed ($\omega = 0.4$) | 70.61 | 79.44 | 88.81 | 91.10 |
| Proposed ($\omega = 0.5$) | **70.75** | **79.44** | **88.75** | **91.36** |

Note: bold text represent statistical significance i.e. $p$ value < 0.05.
training and the remaining for testing. \( \omega \) is selected from the candidate set \( \{0.1, 0.2, \ldots, 0.9\} \). Figure 10 displays the effect of parameter \( \omega \) on the classification accuracy. We can observe that when \( \omega = 0.9 \), our method achieves the best result.

**Conclusions**

RBCM can be roughly divided into two categories: global representation and local representation, they both employ the residuals to classify an input test sample. Related researches indicate that the residuals of these two types of methods have low correlation, and their performance varies on different datasets. Therefore, a method which fuses the residuals of CRC and LRC was proposed by other scholars. Nevertheless, the coefficient of CRC often contains negative values, which may lead to overfitting problem. To address this problem, we obtain the global residual by NNLSC which introduces the non-negative constraint into the model of least square regression. Experimental results demonstrate that by fusing NNLSC and LRC in score level, our proposed method can achieve promising performance. In future, we will solve some potential problems when the number of training samples is insufficient in LRC by adding some local constraints.

**Acknowledgements**

We would like to thank Prof Lei Zhang from Hong Kong Polytechnic University for releasing the source code of CRC at http://www4.comp.polyu.edu.hk/~cslzhang/papers.htm.

**Declaration of Conflicting Interests**

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

**Funding**

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: National Natural Science Foundation of China (Projects Numbers: 61673194, 61672263, 61672265, and 61876072), and in part by the national first-class discipline program of Light Industry Technology and Engineering (Project Number: LITE2018-25).

**ORCID iDs**

Zi-Qi Li https://orcid.org/0000-0003-4694-4456
He-Feng Yin https://orcid.org/0000-0001-5831-1475

**References**

1. Li C and Wu X-J. Neighborhood preserving sparse representation based on Nyström method for image set classification on symmetric positive definite matrices.
2. Shrivastava A, Patel VM and Chellappa R. Multiple kernel learning for sparse representation-based classification. *IEEE Trans Image Process* 2019; 23: 3013.

3. Peng Y, Li L, Liu S, et al. Virtual samples and sparse representation-based classification algorithm for face recognition. *IET Comput Vis* 2018; 13: 172–177.

4. Peng Y, Li L, Liu S, et al. Extended sparse representation-based classification method for face recognition. *Mach Vis Appl* 2018; 29: 991–1007.

5. Yin J, Lai W, Miao S, et al. Optimized projection for collaborative representation based classification and its applications to face recognition. *Pattern Recogn Lett* 2016; 73: 83–90.

6. Du HS, Hu QP, Qiao DF, et al. Robust face recognition via low-rank sparse representation-based classification. *Int J Auton Comput* 2015; 12: 579–587.

7. Li W, Tramel EW, Prasad S, et al. Nearest regularized subspace for hyperspectral classification. *IEEE Trans Geosci Remote Sens* 2013; 52: 477–489.

8. Jia S, Yao X, Tang G, et al. Spatial-spectral-combined sparse representation-based classification for hyperspectral imagery, soft computing – a fusion of foundations. *Soft Comput* 2016; 20: 4659–4668.

9. Zhou G, Zheng Z, Iftekharuddin K, et al. Sparse coding based dense feature representation model for hyperspectral image classification. *J Electron Imag* 2015; 24: 063009.

10. Wright J, Yang AY, Ganesh A, et al. Robust face recognition via sparse representation. *IEEE Trans Pattern Anal Mach Intell* 2008; 31: 210–227.

11. Yin J, Liu Z, Zhong J, et al. Kernel sparse representation based classification. *Neurocomputing* 2012; 77: 120–128.

12. Zhang L, Yang M and Feng X. Sparse representation or collaborative representation: Which helps face recognition? In: *2011 international conference on computer vision*, 2011, pp.471–478. Piscataway, NJ: IEEE.

13. Zhang L, Yang M, Feng X, et al. Collaborative representation based classification for face recognition. *Comput Vis Pattern Recogn* arXiv:1204.2358.

14. Vo N, Moran B and Challa S. Nonnegative-least-square classifier for face recognition. In: *International symposium on neural networks*, 2009, pp.449–456. Berlin, Heidelberg: Springer.

15. Xu J, An W, Zhang L, et al. Sparse, collaborative, or nonnegative representation: which helps pattern classification?. *Pattern Recogn* 2019; 88: 679–688.

16. Naseem I, Togneri R and Bennamoun M. Linear regression for face recognition. *IEEE Trans Pattern Anal Mach Intell* 2010; 32: 2106–2112.

17. Li CG, Guo J and Zhang HG. (2010) Local sparse representation based classification. In: *20th international conference on pattern recognition*, 2010, pp.649–652. Piscataway, NJ: IEEE.

18. Xu Y, Zhang D, Yang J, et al. A two-phase test sample sparse representation method for use with face recognition. *IEEE Trans Circuit Syst Video Technol* 2011; 21: 1255–1262.

19. Zhang H, Wang F, Chen Y, et al. Combination of linear regression classification and collaborative representation classification. *Neural Comput Appl* 2014; 25: 833–838.

20. Lee DD and Seung HS. Learning the parts of objects by non-negative matrix factorization. *Nature* 1999; 401: 788–791.

21. Boyd S, Parikh N, Chu E, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers. *FNT Mach Learn* 2010; 3: 1–122.

22. Samaria FS and Harter AC. Parameterisation of a stochastic model for human face identification. In: *Proceedings of 1994 IEEE workshop on applications of computer vision*, 1994, pp.138–142. Piscataway, NJ: IEEE.

23. Phillips PJ, Hyeonjoon Moon Rizvi SA, et al. The FERET evaluation methodology for face-recognition algorithms. *IEEE Trans Pattern Anal Machine Intell*. 2000, 22(10): 1090–1104.

24. Chen L, Man H, Nefian A V, et al. Face recognition based on multi-class mapping of Fisher scores[J]. *Pattern Recognition*, 2005, 38(6): 799-811.

25. Kittler J and Nixon MS. Audio-and video-based biometric person authentication. In: *4th international conference, AVBPA*, Guilford, UK, 9–11 June 2003.

26. Yuan H, Li X, Xu F, et al. A collaborative-competitive representation based classifier model. *Neurocomputing* 2018; 275: 627–635.

27. Zhou J and Zhang B. Collaborative representation using non-negative samples for image classification. *Sensors* 2019; 19: 2609.

28. Yang AY, Sastry SS, Ganesh A, et al. Fast l_1-minimization algorithms and an application in robust face recognition: a review. In: *2010 IEEE international conference on image processing*, 2010, pp.1849–1852. Piscataway, NJ: IEEE.