IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received May 25, 2021, accepted June 3, 2021, date of publication June 16, 2021, date of current version July 2, 2021.

Digital Object Identifier 10.1109/ACCESS.2021.3089836

Extended JSSL for Multi-Feature Face Recognition via Intra-Class Variant Dictionary

GUOJUN LIN[1,2], QINRUI ZHANG[2], SHUNYONG ZHOU[1,2], XINGGUO JIANG[2], HAO WU[1,2], HAIRONG YOU[3], ZUXIN LI[4], PING HE[2], AND HENG LI[5]

[1]Artificial Intelligence Key Laboratory of Sichuan Province, Sichuan University of Science and Engineering, Zigong 643000, China
[2]School of Automation and Information Engineering, Sichuan University of Science and Engineering, Zigong 643000, China
[3]School of Information Science and Engineering, Northeastern University, Shenyang 110819, China
[4]School of Engineering, Huzhou University, Huzhou 313000, China
[5]Smart Construction Laboratory (BRE), The Hong Kong Polytechnic University, Hong Kong

Corresponding author: Ping He (pinghecn@qq.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 11705122; in part by the Scientific Research Foundation of Sichuan University of Science and Engineering under Grant 2019RC11 and Grant 2019RC12; in part by the Sichuan Science and Technology Program of China under Grant 2019YJ0477, Grant 2020YFSY0027, and Grant 2020YFH0124; in part by the Guangdong Basic and Applied Basic Research Foundation under Grant 2021A1515011342; in part by the Open Foundation of Artificial Intelligence Key Laboratory of Sichuan Province under Grant 2019RZJ03 and Grant 2020RZY02; and in part by the Applied Basic Research Programs of Science and Technology Department of Zigong under Grant 2019YYJC29 and Grant 2020YGJC01.

ABSTRACT This paper focuses on how to represent the testing face images for multi-feature face recognition. The choice of feature is critical for face recognition. The different features of the sample contribute differently to face recognition. The joint similar and specific learning (JSSL) has been effectively applied in multi-feature face recognition. In the JSSL, although the representation coefficient is divided into the similar coefficient and the specific coefficient, there is the disadvantage that the training images cannot represent the testing images well, because there are probable expressions, illuminations and disguises in the testing images. We think that the intra-class variations of one person can be linearly represented by those of other people. In order to solve well the disadvantage of JSSL, in the paper, we extend JSSL and propose the extended joint similar and specific learning (EJSSL) for multi-feature face recognition. EJSSL constructs the intra-class variant dictionary to represent the probable variation between the training images and the testing images. EJSSL uses the training images and the intra-class variant dictionary to effectively represent the testing images. The proposed EJSSL method is perfectly experimented on some available face databases, and its performance is superior to many current face recognition methods.

INDEX TERMS Sparse representation, image classification, multi-feature, face recognition.

I. INTRODUCTION

Face recognition is a very important biometric technology and also the application field of computer vision and pattern recognition. In the last decades, many scholars have proposed a lot of methods of face recognition such as Eigenfaces, Fisherfaces and Laplacianfaces and made great progress. According to the sparse representation principle of human visual system, an image or a signal can be linearly represented by representation bases that are atoms of the dictionary. In the recent years, many research results have been achieved by applying the sparse representation method to the various applications which include image classification [1]–[3], image denoising [4], [5] and face recognition [6], [7]. The sparse representation based classification (SRC) [6] method has already been used to make robust face recognition successful and has achieved the amazing performance in face recognition. The research of the sparse representation based pattern classification is promoted, because SRC is the successful method. Some scholars have proposed many variants of SRC. For example, the weighted SRC [8]–[10] methods were proposed for face recognition. Locality-sensitive discriminative dictionary learning for SRC [11]–[13] was proposed to improve image classification via the constraints on the representation coefficients. The Gabor filters can be used to effectively extract the local features from the face images.

The associate editor coordinating the review of this manuscript and approving it for publication was Jiachen Yang.

SRC can be successfully combined with the Gabor local features. The Gabor-feature based SRC (GSRC) [14] was proposed and achieved higher recognition rates than SRC. The inter-class sparsity based discriminative least square regression (ICS_DLSR) [15] was proposed for image classification.

Although SRC has already been very successfully used in face recognition and image classification, some scholars queried the role of l_1-norm sparsity in SRC and thought that it was the collaborative representation that played the important role in SRC and made SRC effective. The collaborative representation based classification (CRC) [16] was proposed by using the l_2-norm to constraint the representation coefficients. The computational complexity of CRC is remarkably lower than that of SRC, but CRC has similar face recognition results to SRC. Many modifications of CRC have been proposed for face recognition. The regularized robust coding (RRC) [17] was used for face recognition with expression changes, illuminations and disguises. Based on the two-stage classification, the hierarchical collaborative representation [18] was proposed for robust face recognition. As the intrinsic classification mechanism of CRC remained unclear, the probabilistic collaborative representation based classification (ProCRC) [19] was proposed. ProCRC maximizes the likelihood that the testing image belongs to each class. By combining NSC and SRC/CRC, the collaborative representation optimized classifier (CROC) [20], [21] was proposed for multi-class classification. The collaborative neighbor representation-based classification (CNRC) [22] was proposed for face recognition by using l_2-norm to regularize the representation coefficients. Based on CNRC and DSRC, the discriminative collaborative neighbor representation (DCNR) [23] method was proposed for robust face recognition.

Thinking that the intra-class variations of one person can be linearly represented by those of other people, SRC was extended and the extended sparse representation based classifier (ESRC) [24], [25] was proposed for undersampled face recognition with expression changes, illuminations and disguises. ESRC constructs the intra-class variant dictionary to represent the probable variation between the training images and the testing images. ESRC has superior performance to SRC. According to the advantage that the computational complexity of CRC is remarkably lower than that of SRC, enlightened from ESRC, the extended collaborative representation based classification (ECRC) [26] was proposed for undersampled face recognition. ECRC performs better than CRC.

In face recognition, the choice of feature is critical. The excellent feature can improve the recognition performance. The different features of the sample contribute differently to face recognition. The multi-task joint sparse representation based classification (MTJSRC) [27] that combined various types of features was proposed for multi-feature face recognition. The multi-view discriminant dictionary learning via learning view-specific and shared structured dictionaries (MDVSD) [28] was proposed for image classification. The multi-view synthesis and analysis dictionaries learning (MSADL) [29] was proposed for image classification. The robust algorithm that combined finite rate of innovation theory and multimodal dictionary learning was proposed for depth image super resolution [30]. The relaxed collaborative representation (RCR) [31] was proposed to effectively make use of the similarity and distinctiveness of the different features. Based on RCR, the joint similar and specific learning (JSSL) [32] method was proposed by combining the different features, which were face, tongue and sublingual vein. Different from RCR, in the JSSL, the representation coefficient was divided into the similar coefficient and the specific coefficient.

In order to solve well the disadvantage that the training images cannot effectively represent the testing images in RCR and JSSL, because there are probable expression changes, illuminations and disguises in the testing images, and thinking that the intra-class variations of one person can be linearly represented by those of other people, in the paper, we extend JSSL and propose the extended joint similar and specific learning (EJSSL) method for multi-feature face recognition. The main contributions of the paper are summarized in the following two points:

1) Different from many existing discriminative dictionary learning methods and RCR, in the JSSL, the representation coefficient was divided into the similar coefficient and the specific coefficient. Based on JSSL, the novelty of the paper is that the proposed EJSSL is the extended JSSL and constructs the intra-class variant dictionary to effectively represent the probable variation between the training images and the testing images. EJSSL uses the training images and the intra-class variant dictionary to represent the testing images. EJSSL has better performance than JSSL. Figure 1 demonstrates the flowchart of the proposed EJSSL method.

2) Compared with many current face recognition methods, the experimental results prove that the proposed EJSSL method can achieve better performance.

The remainder of the paper is organized as follows. The related method is briefly introduced in Section II. The proposed EJSSL method is presented in Section III. The optimization procedure of EJSSL is described in Section IV. The classification scheme of EJSSL is presented in Section V. The analysis of computational complexity is shown in Section VI. The experimental results and discussion are shown in Section VII. Finally, the paper is concluded in Section VIII.

II. RELATED METHOD

The joint similar and specific learning (JSSL) [32] method was proposed by combining the different features which were face, tongue and sublingual vein. There are K types of features for a face database. The training sample set of the k-th feature are denoted by $\boldsymbol{D}_k = \left[\boldsymbol{D}_k^1, \boldsymbol{D}_k^2, \ldots, \boldsymbol{D}_k^n\right] \in R^{m_k \times M_k}$ $\left(k = 1, 2, \ldots, K\right)$, where n is the number of classes and $\boldsymbol{D}_k^i \in R^{m_k \times M_k^i}$ $\left(\sum_{i=1}^{n} M_k^i = M_k\right)$ is the training sample set of

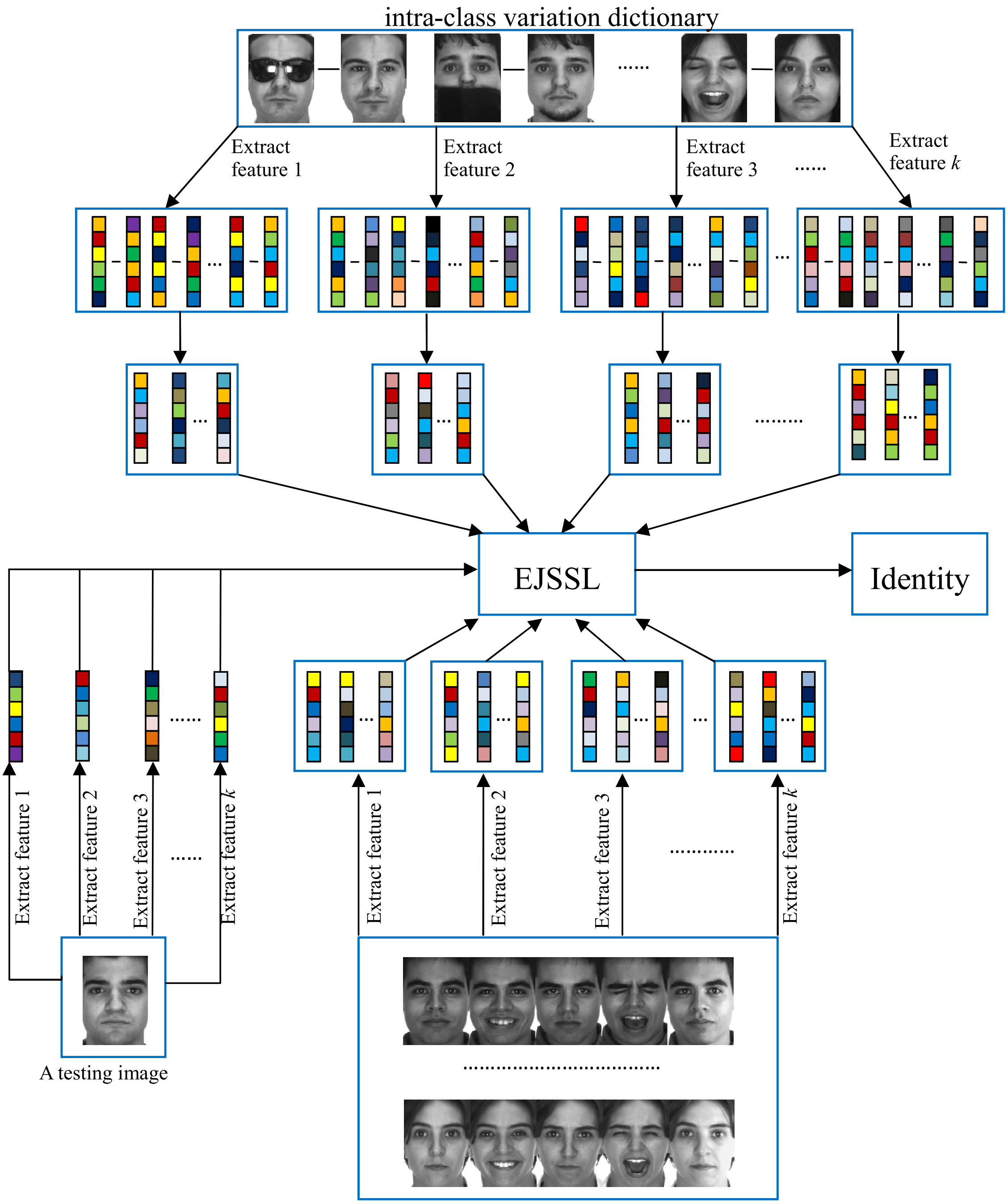


FIGURE 1. **The flowchart of the proposed EJSSL method for multi-feature face recognition.**

the k-th feature of the i-th class. Let $\boldsymbol{y}_k \in R^{m_k}$ be the testing sample. $\boldsymbol{\alpha}_k^c \in R^{M_k}$ is the similar part of the representation coefficient. $\boldsymbol{\alpha}_k^s \in R^{M_k}$ is the specific part of the representation coefficient. JSSL computes $\boldsymbol{\alpha}_k^c \in R^{M_k}$ and $\boldsymbol{\alpha}_k^s \in R^{M_k}$ of $\boldsymbol{y}_k$ over $\boldsymbol{D}_k$ by solving the following equation:

$$\min_{\boldsymbol{\alpha}_k^c, \boldsymbol{\alpha}_k^s} \sum_{k=1}^{K} \left\{ \left\| \boldsymbol{y}_k - \boldsymbol{D}_k \left(\boldsymbol{\alpha}_k^c + \boldsymbol{\alpha}_k^s \right) \right\|_2^2 + \tau \left\| \boldsymbol{\alpha}_k^c - \bar{\boldsymbol{\alpha}}^c \right\|_2^2 + \lambda \left(\left\| \boldsymbol{\alpha}_k^c \right\|_1 + \left\| \boldsymbol{\alpha}_k^s \right\|_1 \right) \right\}. \quad (1)$$

where $\bar{\boldsymbol{\alpha}}^c = \frac{1}{K} \sum_{k=1}^{K} \boldsymbol{\alpha}_k^c$, τ and λ are all the small positive constants.

The reconstruction residual of the i-th class is represented as:

$$e_i = \sum_{k=1}^{K} w_k \left\| \boldsymbol{y}_k - \boldsymbol{D}_k^i \left(\boldsymbol{\alpha}_{k,i}^c + \boldsymbol{\alpha}_{k,i}^s \right) \right\|_2^2. \quad (2)$$

where $\boldsymbol{D}_k^i$ is the dictionary of the k-th feature of the i-th class, $\boldsymbol{\alpha}_{k,i}^c$ is the similar coefficient of the k-th feature of the

i-th class, and $\boldsymbol{\alpha}_{k,i}^{c}$ is the specific coefficient of the k-th feature of the i-th class. w_k is the weight value which is associated with the k-th feature.

The classification scheme of JSSL is determined as:

$$\text{identity}(\boldsymbol{y}) = \arg\min_{i} \{e_i\}. \tag{3}$$

III. THE PROPOSED METHOD

The intra-class variation images of four people are shown in Figure 2. As can be seen from Figure 2, the intra-class variations of four people are similar since the shape of faces are highly correlated, so the intra-class variations of one person can be linearly represented by those of other people. In the section, we extend JSSL and propose the extended joint similar and specific learning (EJSSL) for multi-feature face recognition. There are the testing images with expression changes, illuminations and disguises, so EJSSL constructs the intra-class variant dictionary to represent the probable variation between the training images and the testing images. EJSSL uses the training images and the intra-class variant dictionary to effectively represent the testing images.

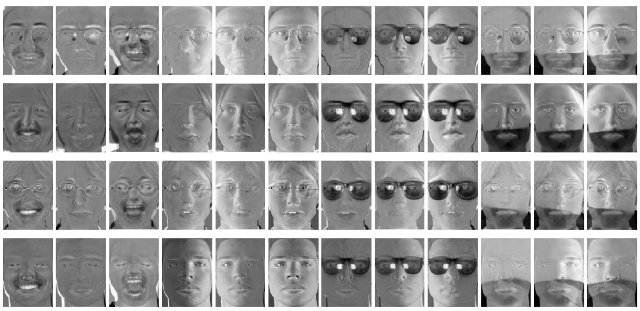


FIGURE 2. The intra-class variation images of four people. Each row represents the intra-class variation images of one person which are achieved by subtracting a natural image from the images with expression changes, illuminations and disguises.

In JSSL, the different features from the same sample can share some similarity, so we think that the representation coefficients of some features (for example, the features are intensity value, low-frequency Fourier [33], local binary patterns [34] and Gabor magnitude [35]) on their associated training samples ought to be similar. This can lead to make the representation stable. In order to realize the goal, the following equation is used to obtain the similarity of the different features.

$$\min_{\boldsymbol{\alpha}_k} \sum_{k=1}^{K} \|\boldsymbol{\alpha}_k - \bar{\boldsymbol{\alpha}}\|_2^2. \tag{4}$$

where $\boldsymbol{\alpha}_k \in R^{M_k}$ $(k = 1, 2, \ldots, K)$ is the representation coefficient of the k-th feature, K is the type number of the different features, and $\bar{\boldsymbol{\alpha}} = \frac{1}{K}\sum_{k=1}^{K}\boldsymbol{\alpha}_k$ is the mean value of all $\boldsymbol{\alpha}_k$. We can easily see that the purpose of Eq. (4) is to decrease the variance of the representation coefficients $\boldsymbol{\alpha}_k$ and make them similar for each other. However, there is also diversity among them. This can lead to make the representation flexible, so it is necessary to utilize the similarity and distinctiveness of the different features. Generally, the balance relationship between stability and flexibility can achieve both stable and accurate representation for multi-feature face recognition.

In order to address the existing problem, the representation coefficient $\boldsymbol{\alpha}_k$ is divided into two parts which are the similar part and the specific part. $\boldsymbol{\alpha}_k$ is represented as $\boldsymbol{\alpha}_k = \boldsymbol{\alpha}_k^c + \boldsymbol{\alpha}_k^s$, where $\boldsymbol{\alpha}_k^c \in R^{M_k}$ is the similar coefficient and denotes similarity, $\boldsymbol{\alpha}_k^s \in R^{M_k}$ is the specific coefficient and denotes distinctiveness. The training sample set of the k-th feature are represented as $\boldsymbol{D}_k = \left[\boldsymbol{D}_k^1, \boldsymbol{D}_k^2, \ldots, \boldsymbol{D}_k^n\right] \in R^{m_k \times M_k}$ $\left(k = 1, 2, \ldots, K\right)$, where n is the number of classes and $\boldsymbol{D}_k^i \in R^{m_k \times M_k^i}$ $\left(\sum_{i=1}^{n} M_k^i = M_k\right)$ is the training sample set of the k-th feature of the i-th class. Denote the intra-class variation dictionary as $\boldsymbol{V}_k$ $\left(k = 1, 2, \ldots, K\right)$, which represents expression changes, illuminations and disguises. Let $\boldsymbol{y}_k \in R^{m_k}$ be a testing sample. The proposed EJSSL is designed as follows:

$$\begin{aligned}\min_{\boldsymbol{\alpha}_k^c, \boldsymbol{\alpha}_k^s, \boldsymbol{\beta}_k} \sum_{k=1}^{K} \Big\{ &\left\|\boldsymbol{y}_k - \boldsymbol{D}_k\left(\boldsymbol{\alpha}_k^c + \boldsymbol{\alpha}_k^s\right) - \boldsymbol{V}_k\boldsymbol{\beta}_k\right\|_2^2 \\ &+\tau\left\|\boldsymbol{\alpha}_k^c - \bar{\boldsymbol{\alpha}}^c\right\|_2^2 + \lambda(\left\|\boldsymbol{\alpha}_k^c\right\|_1 \\ &+ \left\|\boldsymbol{\alpha}_k^s\right\|_1 + \left\|\boldsymbol{\beta}_k\right\|_1)\Big\}.\end{aligned} \tag{5}$$

where $\boldsymbol{\beta}_k$ is the representation coefficient of $\boldsymbol{y}_k$ over $\boldsymbol{V}_k$, $\bar{\boldsymbol{\alpha}}^c = \frac{1}{K}\sum_{k=1}^{K}\boldsymbol{\alpha}_k^c$ is the mean value of all $\boldsymbol{\alpha}_k^c$, τ and λ are all the small positive constants.

The intra-class variation dictionary $\boldsymbol{V}_k$ could be got either from the training sample set themselves (if each person has multiple samples) or from the generic sample set that are outside the training sample set. Given a generic sample set where each person has multiple samples, the generic sample set is denoted by $\boldsymbol{E}_k = \left[\boldsymbol{E}_k^1, \boldsymbol{E}_k^2, \ldots, \boldsymbol{E}_k^l\right] \in R^{m_k \times N_k}$, where $\boldsymbol{E}_k^i \in R^{m_k \times N_k^i}$ $\left(\sum_{i=1}^{l} N_k^i = N_k\right)$ is the generic sample set of the i-th class, $i = 1, 2, \ldots, l$. If there is an image which is labeled as "natural" for each person, the intra-class variation dictionary $\boldsymbol{V}_k$ can be achieved by subtracting the natural image of the k-th class from the other images of the same class:

$$\boldsymbol{V}_k = \left[\boldsymbol{E}_k^{1-} - \boldsymbol{b}_k^{1*}\boldsymbol{p}_k^1, \ldots, \boldsymbol{E}_k^{l-} - \boldsymbol{b}_k^{l*}\boldsymbol{p}_k^l\right] \in R^{m_k \times (N_k - l)}. \tag{6}$$

where $\boldsymbol{p}_k^i = [1, 1, \ldots, 1] \in R^{1 \times \left(N_k^i - 1\right)}$, $\boldsymbol{b}_k^{i*}$ is the natural image of the i-th class, and $\boldsymbol{E}_k^{i-}$ is the reduced image set of the i-th class which removes the natural image. If each person has not the natural image, the intra-class variation dictionary $\boldsymbol{V}_k$ could be obtained as follows:

$$\boldsymbol{V}_k = \left[\boldsymbol{E}_k^1 - \boldsymbol{d}_k^1\boldsymbol{p}_k^1, \ldots, \boldsymbol{E}_k^l - \boldsymbol{d}_k^l\boldsymbol{p}_k^l\right] \in R^{m_k \times N_k}. \tag{7}$$

where $\boldsymbol{p}_k^i = [1, 1, \ldots, 1] \in R^{1 \times N_k^i}$, $\boldsymbol{d}_k^i$ is the centroid of the i-th class.

IV. OPTIMIZATION OF EJSSL

The objective function of Eq. (5) is not jointly convex with respect to $\left(\boldsymbol{\alpha}_k^c, \boldsymbol{\alpha}_k^s, \boldsymbol{\beta}_k\right)$, but it is convex with respect to $\boldsymbol{\alpha}_k^c$ when $\begin{bmatrix} \boldsymbol{\alpha}_k^s \\ \boldsymbol{\beta}_k \end{bmatrix}$ is fixed and it is convex with respect to $\begin{bmatrix} \boldsymbol{\alpha}_k^s \\ \boldsymbol{\beta}_k \end{bmatrix}$ when $\boldsymbol{\alpha}_k^c$ is fixed. Eq. (5) is revised as follows:

$$\begin{aligned}\min_{\boldsymbol{\alpha}_k \boldsymbol{\alpha}_k^n \cdot \boldsymbol{\beta}_k} \sum_{k=1}^{K} \Bigg\{ & \left\| \boldsymbol{y}_k - \boldsymbol{D}_k \boldsymbol{\alpha}_k^c - \begin{bmatrix} \boldsymbol{D}_k & \boldsymbol{V}_k \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_k^s \\ \boldsymbol{\beta}_k \end{bmatrix} \right\|_2^2 \\ & + \tau \left\| \boldsymbol{\alpha}_k^c - \bar{\boldsymbol{\alpha}}^c \right\|_2^2 \\ & + \lambda (\left\| \boldsymbol{\alpha}_k^c \right\|_1 + \left\| \boldsymbol{\alpha}_k^s \right\|_1 + \left\| \boldsymbol{\beta}_k \right\|_1) \Bigg\}. \end{aligned} \tag{8}$$

We alternatively update $\boldsymbol{\alpha}_k^c$ and $\begin{bmatrix} \boldsymbol{\alpha}_k^s \\ \boldsymbol{\beta}_k \end{bmatrix}$. Updating $\boldsymbol{\alpha}_k^c$ by fixing $\begin{bmatrix} \boldsymbol{\alpha}_k^s \\ \boldsymbol{\beta}_k \end{bmatrix}$ and updating $\begin{bmatrix} \boldsymbol{\alpha}_k^s \\ \boldsymbol{\beta}_k \end{bmatrix}$ by fixing $\boldsymbol{\alpha}_k^c$.

1) Updating $\boldsymbol{\alpha}_k^c$: when $\begin{bmatrix} \boldsymbol{\alpha}_k^s \\ \boldsymbol{\beta}_k \end{bmatrix}$ is fixed, the optimization solution of Eq. (8) with respect to $\boldsymbol{\alpha}_k^c$ can be reduced to:

$$\begin{aligned}\boldsymbol{\alpha}_k^c = \arg\min & \left\| \boldsymbol{y}_k - \boldsymbol{D}_k \boldsymbol{\alpha}_k^c - \begin{bmatrix} \boldsymbol{D}_k & \boldsymbol{V}_k \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_k^s \\ \boldsymbol{\beta}_k \end{bmatrix} \right\|_2^2 \\ & + \tau \left\| \boldsymbol{\alpha}_k^c - \bar{\boldsymbol{\alpha}}^c \right\|_2^2 + \lambda \left\| \boldsymbol{\alpha}_k^c \right\|_1 . \end{aligned} \tag{9}$$

Applying the augmented lagrangian method (ALM) algorithm, Eq. (9) can be modified as follows:

$$\begin{aligned}\boldsymbol{\alpha}_k^c = \arg\min \Bigg\{ & \left\| \boldsymbol{y}_k - \boldsymbol{D}_k \boldsymbol{\alpha}_k^c - \begin{bmatrix} \boldsymbol{D}_k & \boldsymbol{V}_k \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_k^s \\ \boldsymbol{\beta}_k \end{bmatrix} \right\|_2^2 \\ & + \tau \left\| \boldsymbol{\alpha}_k^c - \bar{\boldsymbol{\alpha}}^c \right\|_2^2 + \lambda \left\| \boldsymbol{\alpha}_k^{c'} \right\|_1 \\ & + \frac{\mu}{2} \left\| \boldsymbol{\alpha}_k^c - \boldsymbol{\alpha}_k^{c'} + \frac{z_k}{\mu} \right\|_2^2 \Bigg\}. \end{aligned} \tag{10}$$

where $\boldsymbol{\alpha}_k^{c'}$ is the relaxed variable, z_k is the k-th lagrangian multiplier and μ is the value of the step. $\boldsymbol{\alpha}_k^c$ and $\boldsymbol{\alpha}_k^{c'}$ can be optimized alternatively.

(a) First, $\boldsymbol{\alpha}_k^{c'}$ is fixed to obtain $\boldsymbol{\alpha}_k^c$:

$$\begin{aligned}\boldsymbol{\alpha}_k^c = \arg\min \Bigg\{ & \left\| \boldsymbol{y}_k - \boldsymbol{D}_k \boldsymbol{\alpha}_k^c - \begin{bmatrix} \boldsymbol{D}_k & \boldsymbol{V}_k \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_k^s \\ \boldsymbol{\beta}_k \end{bmatrix} \right\|_2^2 \\ & + \tau \left\| \boldsymbol{\alpha}_k^c - \bar{\boldsymbol{\alpha}}^c \right\|_2^2 + \frac{\mu}{2} \left\| \boldsymbol{\alpha}_k^c - \boldsymbol{\alpha}_k^{c'} + \frac{z_k}{\mu} \right\|_2^2 \Bigg\}. \end{aligned} \tag{11}$$

According to [31], the closed-form solution of $\boldsymbol{\alpha}_k^c$ can be accurately calculated as follows:

$$\boldsymbol{\alpha}_k^c = \boldsymbol{\alpha}_{0,k}^c + \frac{\tau}{K} \boldsymbol{P}_k \boldsymbol{Q} \sum_{\eta=1}^{K} \boldsymbol{\alpha}_{0,\eta}^c. \tag{12}$$

where $\boldsymbol{P}_k = \left(\boldsymbol{D}_k^T \boldsymbol{D}_k + \left(\tau + \frac{\mu}{2}\right) \boldsymbol{I}\right)^{-1}$, $\boldsymbol{\alpha}_{0,k}^c = \boldsymbol{P}_k \left(\boldsymbol{D}_k^T \left(\boldsymbol{y}_k - \begin{bmatrix} \boldsymbol{D}_k \boldsymbol{V}_k \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_k^s \\ \boldsymbol{\beta}_k \end{bmatrix} \right) + \frac{\mu}{2} \boldsymbol{\alpha}_k^{c'} - \frac{z_k}{2} \right)$, and $\boldsymbol{Q} = \left(\boldsymbol{I} - \frac{\tau}{K} \sum_{\eta=1}^{K} \boldsymbol{P}_\eta \right)^{-1}$.

(b) Second, when $\boldsymbol{\alpha}_k^c$ is fixed, the optimization solution of Eq. (10) with respect to $\boldsymbol{\alpha}_k^{c'}$ can be reduced to:

$$\boldsymbol{\alpha}_k^{c'} = \arg\min \lambda \left\| \boldsymbol{\alpha}_k^{c'} \right\|_1 + \frac{\mu}{2} \left\| \boldsymbol{\alpha}_k^c - \boldsymbol{\alpha}_k^{c'} + \frac{z_k}{\mu} \right\|_2^2 . \tag{13}$$

$\boldsymbol{\alpha}_k^{c'}$ can be derived by operating threshold $\left(\boldsymbol{\alpha}_k^c + \frac{z_k}{\mu}, \frac{\lambda}{\mu} \right)$. The soft threshold function is shown as follows:

$$\left[\boldsymbol{S}_{\lambda/\mu}(\boldsymbol{\gamma}) \right]_i = \begin{cases} 0, & \left| \boldsymbol{\gamma}_i \right| \le \lambda/\mu \\ \boldsymbol{\gamma}_i - \operatorname{sign}\left(\boldsymbol{\gamma}_i\right) \lambda/\mu, & \text{otherwise} \end{cases} \tag{14}$$

where $\boldsymbol{\gamma}_i$ is the value of the i-th component of $\boldsymbol{\gamma}$. After working out $\boldsymbol{\alpha}_k^c$ and $\boldsymbol{\alpha}_k^{c'}$, z_k and μ can be updated via $z_k = z_k + \mu \left(\boldsymbol{\alpha}_k^c - \boldsymbol{\alpha}_k^{c'} \right)$ and $\mu = 1.2\mu$. The value of μ is too large, so the maximum of μ should be pre-fixed, the initial value of μ is set to 0.01, and $\mu = \min(1.2\mu, 1000)$.

2) Updating $\begin{bmatrix} \boldsymbol{\alpha}_k^s \\ \boldsymbol{\beta}_k \end{bmatrix}$: when $\boldsymbol{\alpha}_k^c$ is fixed, the optimization solution of Eq. (8) with respect to $\begin{bmatrix} \boldsymbol{\alpha}_k^s \\ \boldsymbol{\beta}_k \end{bmatrix}$ can be reduced to:

$$\begin{aligned}\begin{bmatrix} \boldsymbol{\alpha}_k^s \\ \boldsymbol{\beta}_k \end{bmatrix} = \arg\min & \left\| \boldsymbol{y}_k - \boldsymbol{D}_k \boldsymbol{\alpha}_k^c - \begin{bmatrix} \boldsymbol{D}_k & \boldsymbol{V}_k \end{bmatrix} \begin{bmatrix} \boldsymbol{\alpha}_k^s \\ \boldsymbol{\beta}_k \end{bmatrix} \right\|_2^2 \\ & + \lambda \left\| \begin{bmatrix} \boldsymbol{\alpha}_k^s \\ \boldsymbol{\beta}_k \end{bmatrix} \right\|_1 . \end{aligned} \tag{15}$$

The iterative projection method (IPM) [36] can be used to solve Eq. (15). Let $\boldsymbol{X}_k = \begin{bmatrix} \boldsymbol{\alpha}_k^s \\ \boldsymbol{\beta}_k \end{bmatrix}$ $(k = 1, 2, \ldots, K)$. The procedure of the IPM algorithm is shown in Algorithm 1.

Algorithm 1 The update of $\boldsymbol{X}_k$ in EJSSL

Input: $\sigma, \theta = \lambda/2, \quad \boldsymbol{y}_k, \quad \boldsymbol{D}_k, \quad \boldsymbol{V}_k,$ and $\boldsymbol{\alpha}_k^c, \quad k = 1, 2, \ldots, K$

Initialization: $\tilde{\boldsymbol{X}}_k^{(1)} = 0$ and $h = 1$

1: **for** $k = 1, 2, \ldots, K$ **do**

2: **while** the convergence is not reached **do**

3: $h = h + 1$

4: $\tilde{\boldsymbol{X}}_k^{(h)} = \boldsymbol{S}_{\theta/\sigma} \left(\tilde{\boldsymbol{X}}_k^{(h-1)} - \frac{1}{\sigma} \nabla \boldsymbol{F} \left(\tilde{\boldsymbol{X}}_k^{(h-1)} \right) \right)$

where $\nabla \boldsymbol{F} \left(\tilde{\boldsymbol{X}}_k^{(h-1)} \right)$ is the derivative of $\left\| \boldsymbol{y}_k - \boldsymbol{D}_k \boldsymbol{\alpha}_k^c - [\boldsymbol{D}_k \quad \boldsymbol{V}_k] \boldsymbol{X}_k \right\|_2^2$, and $\boldsymbol{S}_{\theta/\sigma}$ is the soft threshold function that defined in Eq. (14)

5: **end while**

6: **end for**

Output: $\boldsymbol{X}_k = \tilde{\boldsymbol{X}}_k^{(h)}, \quad k = 1, 2, \ldots, K$

The procedure of the EJSSL algorithm is shown in Algorithm 2. The values of parameters τ and λ are chosen via the cross validation.

V. THE CLASSIFICATION SCHEME OF EJSSL

After getting the representation coefficient $\boldsymbol{\alpha}_k^c$, $\boldsymbol{\alpha}_k^s$ and $\boldsymbol{\beta}_k$, the reconstruction residual of the i-th class is represented as:

$$e_i = \sum_{k=1}^{K} w_k \left\| \boldsymbol{y}_k - \boldsymbol{D}_k^i \left(\boldsymbol{\alpha}_{k,i}^c + \boldsymbol{\alpha}_{k,i}^s \right) - \boldsymbol{V}_k \boldsymbol{\beta}_k \right\|_2^2 . \tag{16}$$

Algorithm 2 Extended Joint Similar and Specific Learning (EJSSL)

Input: $\tau, \lambda, \boldsymbol{y}_k, \boldsymbol{D}_k, \boldsymbol{V}_k, k = 1, 2, \ldots, K$

Initialization: $\boldsymbol{\alpha}_k^c = 0, \quad \boldsymbol{\alpha}_k^s = 0, \quad \boldsymbol{\beta}_k = 0, \quad z_k = 0$

1: **while** the convergence is not reached **do**

2: update coefficient $\boldsymbol{\alpha}_k^c$ by fixing $\begin{bmatrix} \boldsymbol{\alpha}_k^s \\ \boldsymbol{\beta}_k \end{bmatrix}$

(a) compute $\boldsymbol{\alpha}_k^c$ by Eq. (12)

(b) compute $\boldsymbol{\alpha}_k^{c'}$ by Eq. (13)

(c) $z_k = z_k + \mu \left(\boldsymbol{\alpha}_k^c - \boldsymbol{\alpha}_k^{c'} \right)$

3: update coefficient $\begin{bmatrix} \boldsymbol{\alpha}_k^s \\ \boldsymbol{\beta}_k \end{bmatrix}$ by fixing $\boldsymbol{\alpha}_k^c$, solve $\begin{bmatrix} \boldsymbol{\alpha}_k^s \\ \boldsymbol{\beta}_k \end{bmatrix}$ via **Algorithm 1**

4: **end while**

Output: $\boldsymbol{\alpha}_k^c$ and $\begin{bmatrix} \boldsymbol{\alpha}_k^s \\ \boldsymbol{\beta}_k \end{bmatrix}, k = 1, 2, \ldots, K$

where $\boldsymbol{D}_k^i$ is the dictionary of the k-th feature of the i-th class, $\boldsymbol{\alpha}_{k,i}^c$ is the similar coefficient of the k-th feature of the i-th class, and $\boldsymbol{\alpha}_{k,i}^s$ is the specific coefficient of the k-th feature of the i-th class. w_k is the weight value which is associated with the k-th feature.

The classification scheme of EJSSL is determined as:

$$\text{identity}\,(\boldsymbol{y}) = \arg\min_i \{e_i\}\,. \tag{17}$$

VI. THE ANALYSIS OF COMPUTATIONAL COMPLEXITY

In the EJSSL algorithm, the size of $\boldsymbol{D}_k$ is $m_k \times M_k$, the size of $\boldsymbol{V}_k$ is $m_k \times N_k$ $(k = 1, 2, \ldots, K)$. $\boldsymbol{P}_k$ and $\boldsymbol{Q}$ can be precomputed, the time complexity of updating $\boldsymbol{\alpha}_{0,k}^c$ is $O\left(\sum_{k=1}^K \left(M_k^2 m_k + M_k m_k\right)\right)$ and the time complexity of computing $\boldsymbol{P}_k \boldsymbol{Q} \sum_{\eta=1}^K \boldsymbol{\alpha}_{0,\eta}^c$ is $O\left(\sum_{k=1}^K \left(M_k^3 + M_k^2\right)\right)$, so the time complexity of updating $\boldsymbol{\alpha}_k^c$ is $O\left(\sum_{k=1}^K m_k^2 \left(M_k + N_k\right)^{\varepsilon}\right)$, where $\varepsilon \geq 1.2$ is a constant. In total, the time complexity of EJSSL is $O\left(q \sum_{k=1}^K \left(M_k^2 m_k + M_k m_k + M_k^3 + m_k^2 \left(M_k + N_k\right)^{\varepsilon}\right)\right)$, where q is the number of iteration. The running speed of EJSSL is very fast, its running time on the AR database (see Section VII-B) is 0.06 second.

VII. EXPERIMENTS

To effectively prove the advantage of the proposed EJSSL method, it is compared with NN [37], SRC [6], CRC [16], LRC [38], ProCRC [19], CROC [20], [21], MTJSRC [27], RCR [31], JSSL [32], ESRC [24], and ICS_DLSR [15] methods via some experiments on the public face databases. These databases are the Extended Yale B [39], [40], AR [41], LFW [42] and FERET [43]. There are four features used for the experiments. The features are intensity value, low-frequency Fourier [33], local binary patterns [34] and Gabor magnitude [35]. As a result, the value of K in Eq. (5) is set to 4. The value of w_k $(k = 1, 2, \ldots, K)$ is set to 1. The four different features describe the faces from the different angles. For example, intensity value represents the feature of pixel value, low-frequency Fourier represents the low-frequency feature, local binary patterns represent the local feature, and Gabor magnitude represents the feature of the multi-scales and multi-orientations. The four different features have certain complementarities and conform to the requirement of multi-feature for EJSSL, so the role of them is that they can reveal well the performance of EJSSL. In order to evaluate how the different dictionary sizes affect the performance of EJSSL, firstly, the dictionary $\boldsymbol{U}_k = \left[\boldsymbol{U}_k^1, \boldsymbol{U}_k^2, \ldots, \boldsymbol{U}_k^n\right]$ $(k = 1, 2, \ldots, K)$ is learned from the training samples $\boldsymbol{D}_k = \left[\boldsymbol{D}_k^1, \boldsymbol{D}_k^2, \ldots, \boldsymbol{D}_k^n\right]$ via the Fisher discrimination dictionary learning (FDDL) [44], secondly, the dictionary $\boldsymbol{U}_k$ is used to EJSSL, finally, ProCRC, CROC, MTJSRC and EJSSL are compared in the Extended Yale B, AR and LFW database.

A. EXTENDED YALE B DATABASE

In the section, the experiment is done on the Extended Yale B database [39], [40], which contains 2414 frontal face images of 38 people. One person has about 64 images. The all images were got in the various controlled illumination conditions. In order to make the experiment convenient, the images were cropped and changed to the size of 96 × 84. Figure 3 shows the images of one person which are used for the experiment. In the experiment, the images of the former 32 people of subset 1 (the former 5 images of each person are selected) are selected as the training samples, and the rest images of the subset 3 and subset 4 are used as the testing samples. In order to construct well the intra-class variation dictionary, the images of the remainder 6 people of subset 5 are selected. The intra-class variation dictionary is computed via Eq. (6).

In the experiment, we compare EJSSL with the other competing methods and evaluate the relationship between recognition rates and feature dimension, which can be computed via PCA. When the parameters $\tau = 0.1$ and $\lambda = 0.05$, Table 1 lists the recognition rates of EJSSL and other competing methods and also shows the relationship between recognition rates and feature dimension. As shown in Table 1, EJSSL achieves the best performance in the all competing methods. When the dimension is 50, EJSSL achieves 5.2% improvement of recognition rate over ESRC, which is the second best method. When the dimension is 80 and 110, EJSSL achieves remarkable improvement of recognition rate over LRC, which is the second best method. The recognition rate of EJSSL is increased with the increase of the dimension. Figure 4 shows the recognition rates of ProCRC, CROC, MTJSRC and EJSSL versus the number of dictionary atoms when the dimension is 80. From Figure 4, it can be seen that the recognition rate of EJSSL is at least 6% higher than those of the other competing methods when the number of dictionary atoms is from 32 to 160. The recognition rate of EJSSL basically maintains stable when the number of dictionary atoms is from 64 to 160.

B. AR DATABASE

In the section, the experiment is done on the AR database [41],which contains at least 4,000 frontal images

FIGURE 3. The images of the first line, the images of the second line, and the images of the third line are from subset 1, subset 3, and subset 4, respectively.

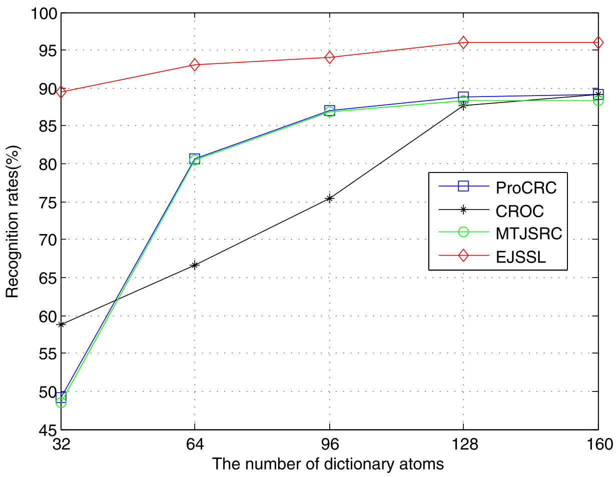


FIGURE 4. The recognition rates versus the number of dictionary atoms on the Extended Yale B database.

TABLE 1. The recognition rates (%) of EJSSL and other competing methods on the Extended Yale B database.

Algorithm	50	80	110
NN [37]	75.0	75.8	75.9
SRC [6]	47.3	49.1	50.1
CRC [16]	76.9	78.3	78.6
LRC [38]	83.4	89.1	91.1
ProCRC [19]	79.7	81.8	83.0
CROC [20]	85.6	88.9	90.9
MTJSRC [27]	76.9	78.3	78.6
RCR [31]	55.7	56.2	57.0
JSSL [32]	58.6	60.5	61.7
ESRC [24]	86.6	87.2	88.2
ICS_DLSR [15]	81.4	81.4	81.6
EJSSL	**91.8**	**92.8**	**93.5**

of 126 people. The 26 face images of each person were got in the two separated sessions. In order to make the experiment convenient, the images were cropped and changed to the size of 165×120. Figure 5 shows the images of one person which are used for the experiment. According to [6], in the experiment, a subset that contains 50 male people and 50 female people is chosen. In the subset, the 90 people are randomly selected for the experiment. For each person, we use the 7 images from session 1 that include expression changes and

TABLE 2. The recognition rates (%) of EJSSL and other competing methods on the AR database.

Algorithm	400	500	600
NN [37]	66.7	66.9	66.8
SRC [6]	63.3	63.6	63.7
CRC [16]	81.2	81.3	81.3
LRC [38]	69.4	69.3	69.2
ProCRC [19]	81.3	81.6	82.0
CROC [20]	75.0	74.9	74.9
MTJSRC [27]	81.4	81.4	81.5
RCR [31]	64.4	64.5	64.6
JSSL [32]	65.6	65.7	65.7
ESRC [24]	78.2	78.6	79.0
ICS_DLSR [15]	80.0	80.5	80.8
EJSSL	**84.5**	**84.5**	**84.4**

illuminations for training, and we use the 13 images from session 2 that include expression changes, illuminations and disguises for testing. In order to construct well the intra-class variation dictionary, the images of the remainder 10 people from session 1 are chosen. The intra-class variation dictionary is computed via Eq. (6).

In the experiment, we compare EJSSL with the other competing methods and evaluate the relationship between recognition rates and feature dimension. When the parameters $\tau = 0.1$ and $\lambda = 0.05$, Table 2 lists the recognition rates of EJSSL and other competing methods and also shows the relationship between recognition rates and feature dimension. From Table 2, we can clearly see that the recognition rate of EJSSL is much higher than those of the other competing methods. When the dimension is 400, EJSSL achieves 3.1% improvement of recognition rate over MTJSRC, which achieves the second highest recognition rate. When the dimension is 500 and 600, EJSSL achieves remarkable improvement of recognition rate over ProCRC, which is the second best method. The recognition rate of EJSSL basically maintains stable with the increase of the dimension. Figure 6 shows the recognition rates of ProCRC, CROC, MTJSRC and EJSSL versus the number of dictionary atoms when the dimension is 500. As shown in Figure 6, we can see that the recognition rate of EJSSL is much higher than those of the other competing methods when the number of dictionary atoms is from 90 to 630. The recognition rate of EJSSL basically maintains stable when the number of dictionary atoms is from 180 to 630.

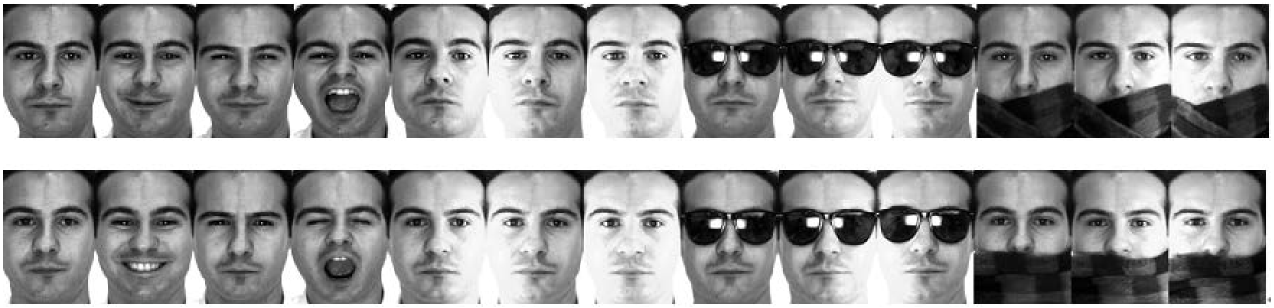


FIGURE 5. The images of the first line are from session 1. The images of the second line are from session 2.

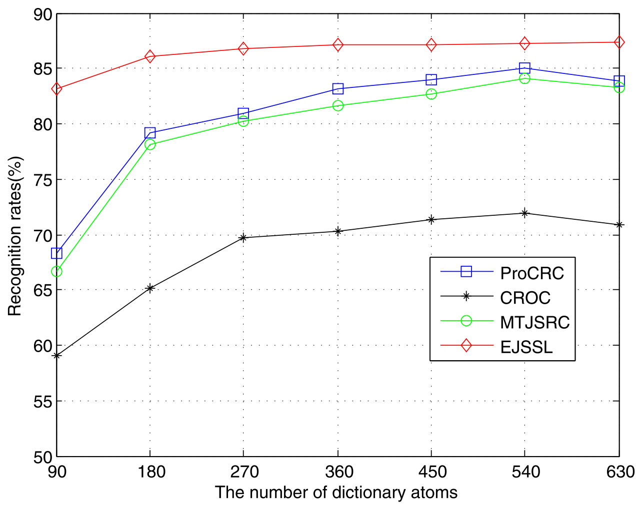


FIGURE 6. The recognition rates versus the number of dictionary atoms on the AR database.

C. LFW DATABASE

In the section, EJSSL is experimented on the aligned labeled face in the wild (LFWa). The scale of LFW [42] database is very large. The all images of the database were got in the uncontrolled conditions and included multi-pose, expression changes, illuminations, age and occlusion. Figure 7 shows the images of one person which are used for the experiment. According to [43], we choose 136 people (each person has no less than 11 samples) and the former 100 people from 136 people for the experiment. For each person, the former 10 images are chosen as the training samples and the remainder images are used as the testing samples. In order to construct well the intra-class variation dictionary, the images of the remainder 36 people from LFWa are chosen. The intra-class variation dictionary is computed via Eq. (6).

In the experiment, we compare EJSSL with the other competing methods and evaluate the relationship between recognition rates and feature dimension. When the parameters $\tau = 0.1$ and $\lambda = 0.001$, Table 3 lists the recognition rates of EJSSL and other competing methods and also shows the relationship between recognition rates and feature dimension. From Table 3, we can clearly see that the recognition rate of EJSSL is much higher than those of the other competing methods when the dimension is 400 and 500. The recognition rate of EJSSL is the same as those of RCR and JSSL and higher than those of the other competing methods except RCR and JSSL when the dimension is 300. The recognition rate of EJSSL is increased with the increase of the dimension. Figure 8 shows the recognition rates of ProCRC, CROC, MTJSRC and EJSSL versus the number of dictionary atoms when the dimension is 400. From Figure 8, it can be seen that the recognition rate of EJSSL is higher than those of the other competing methods when the number of dictionary atoms is from 100 to 1000. The recognition rate of EJSSL basically maintains stable when the number of dictionary atoms is from 200 to 1000.

FIGURE 7. The images of one person from the LFW database.

TABLE 3. The recognition rates (%) of EJSSL and other competing methods on the LFW database.

Algorithm	300	400	500
NN [37]	40.7	41.5	41.8
SRC [6]	72.6	73.7	74.1
CRC [16]	74.9	75.7	76.3
LRC [38]	57.3	58.8	59.1
ProCRC [19]	77.0	78.6	79.4
CROC [20]	57.6	58.9	59.2
MTJSRC [27]	74.9	75.7	76.3
RCR [31]	78.0	79.3	79.5
JSSL [32]	78.0	79.3	79.5
ESRC [24]	74.4	76.1	76.1
ICS_DLSR [15]	75.6	76.9	77.3
EJSSL	**78.0**	**79.5**	**79.9**

D. FERET DATABASE

FERET [45] is a large scale database. It includes some gallery training and probe sets. The experiment is done on

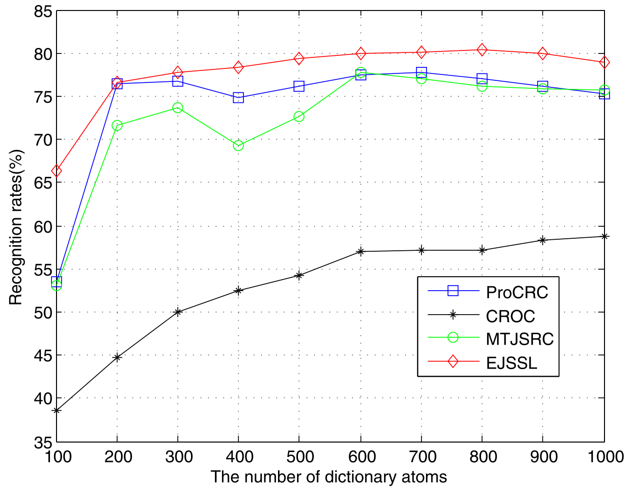


FIGURE 8. The recognition rates versus the number of dictionary atoms on the LFW database.

FIGURE 9. The gallery images of the first line are from the FERET database. The images of the second line are the corresponding probe images.

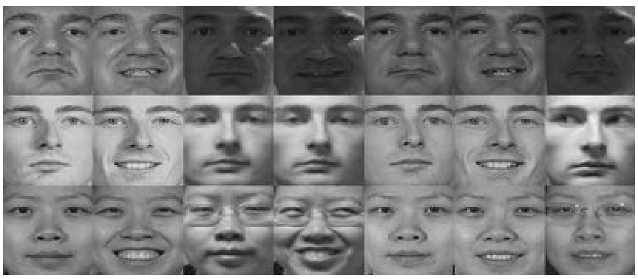


FIGURE 10. Some images from the FRGC V2 database.

the gallery training set, fb probe set and dup2 probe set. The 1196 images of 1196 people are contained in the gallery training set. The 1195 images with an alternative expression are contained in the fb probe set. The 722 images are contained in the dup1 set, and they are captured in the different time. The dup2 probe set is a subset of the dup1 set and consists of 234 images captured beyond one year later. The images were normalized and cropped to the size of 128 × 128 with the pure face region. Figure 9 shows some images from the FERET database. These images have complex intra-class variability. In the experiment, the gallery training set is used as the training samples. The fb probe set and dup2 probe set are used as the testing samples. There are 12766 frontal images of 222 people contained in the FRGC V2.0 database [46]. These images were captured under the uncontrolled conditions. FRGC V2.0 database is independent from the FERET database. Figure 10 shows some images

TABLE 4. The recognition rates (%) of EJSSL and other competing methods on the fb probe set.

Algorithm	200	500	800
NN [37]	73.7	81.4	82.5
SRC [6]	75.3	84.5	85.6
CRC [16]	75.8	86.6	88.7
LRC [38]	75.3	82.5	83.5
ProCRC [19]	78.4	88.1	91.2
CROC [20]	78.9	85.1	87.6
MTJSRC [27]	75.8	86.6	88.7
RCR [31]	73.7	80.9	83.0
JSSL [32]	78.9	85.6	86.1
ESRC [24]	90.2	93.3	94.9
ICS_DLSR [15]	75.3	86.1	89.2
EJSSL	**93.8**	**96.4**	**95.9**

TABLE 5. The recognition rates (%) of EJSSL and other competing methods on the dup2 probe set.

Algorithm	200	500	800
NN [37]	65.4	67.1	68.4
SRC [6]	55.1	60.7	63.3
CRC [16]	59.0	62.0	62.4
LRC [38]	62.4	65.4	65.4
ProCRC [19]	59.0	62.0	62.0
CROC [20]	62.0	66.7	66.2
MTJSRC [27]	59.4	62.4	62.8
RCR [31]	60.7	63.3	63.7
JSSL [32]	58.6	61.1	61.1
ESRC [24]	70.9	73.1	76.1
ICS_DLSR [15]	61.1	66.2	67.5
EJSSL	**78.2**	**81.2**	**83.3**

from the FRGC V2.0 database. In the gallery training set, as each person has only a single image, the standard training set from the FRGC V2.0 database is used for constructing the intra-class variation dictionary. The intra-class variation dictionary is computed via Eq. (7).

In the experiment, we compare EJSSL with the other competing methods and evaluate the relationship between recognition rates and feature dimension. When the parameters $\tau = 0.1$ and $\lambda = 0.05$, Table 4 shows the recognition rates of EJSSL and other competing methods and also shows the relationship between recognition rates and feature dimension on the fb probe set. Table 5 shows the recognition rates of EJSSL and other competing methods and also shows the relationship between recognition rates and feature dimension on the dup2 probe set. From Table 4, we can clearly see that EJSSL achieves the highest recognition rate in the all competing methods. The recognition rate of EJSSL basically maintains stable with the increase of the dimension. From Table 5, we can see that the recognition rate of EJSSL is much higher than those of the other competing methods. The recognition rate of EJSSL is increased with the increase of the dimension.

E. DISCUSSION ON PARAMETERS

In the section, how the number of features affects the performance of EJSSL is discussed. $\boldsymbol{D}_1$ is the dictionary of the first feature. $\boldsymbol{D}_2$ is the dictionary of the second feature. $\boldsymbol{D}_3$ is

TABLE 6. The recognition rate (%) of EJSSL versus the number of features and the single feature.

Dictionary	EB	AR	LFW	fb	dup2
$\boldsymbol{D}_1$	82.2	66.2	61.8	70.6	53.0
$\boldsymbol{D}_2$	71.5	62.1	60.0	48.5	51.3
$\boldsymbol{D}_3$	83.0	81.8	70.3	91.2	83.8
$\boldsymbol{D}_4$	89.7	79.5	70.7	89.7	63.7
$\boldsymbol{D}_{12}$	81.7	64.1	62.8	63.9	55.1
$\boldsymbol{D}_{123}$	88.9	81.7	76.4	92.3	84.2
$\boldsymbol{D}_{1234}$	92.8	84.5	79.5	96.4	81.2

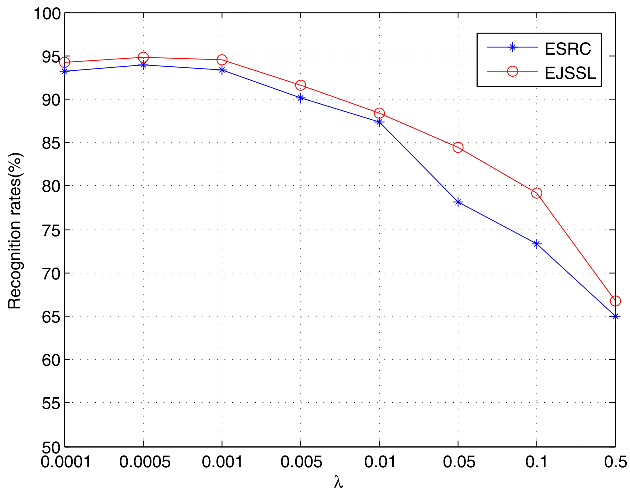


FIGURE 11. The recognition rates versus the parameter λ on the AR database.

the dictionary of the third feature. $\boldsymbol{D}_4$ is the dictionary of the fourth feature. K is the number of features. $\boldsymbol{D}_{12}$ is denoted as $K = 2$ in Eq.(5). $\boldsymbol{D}_{123}$ is denoted as $K = 3$ in Eq.(5). $\boldsymbol{D}_{1234}$ is denoted as $K = 4$ in Eq.(5). Table 6 shows the recognition rate of EJSSL versus the number of features and the single feature. From Table 6, we can see that the recognition rate of EJSSL is increased with the increase of the value of K except the dup2 probe set.

The choice of the parameters is very important. The experimental parameters are discussed in the section. As shown in Eq. (5), the two parameters τ and λ should be discussed. τ controls the variance of similar coefficients, λ controls the sparsity of the representation coefficients. The two parameters are discussed on the AR database. In the discussion, the experiment setting is already given in Section VII-B, the feature dimension is 400. With fixed $\tau = 0.1$, the recognition rates of EJSSL and ESRC versus λ are shown in Figure 11. From Figure 11, it can be clearly seen that EJSSL achieves higher recognition rates than ESRC. With fixed $\lambda = 0.05$, the recognition rates of EJSSL and ESRC versus the big range value of τ are shown in Figure 12, the recognition rates of EJSSL and ESRC versus the small range value of τ are shown in Figure 13. From Figure 12 and Figure 13, it can be clearly seen that EJSSL achieves higher recognition rates than ESRC and the recognition rate of EJSSL is not sensitive to the value of τ.

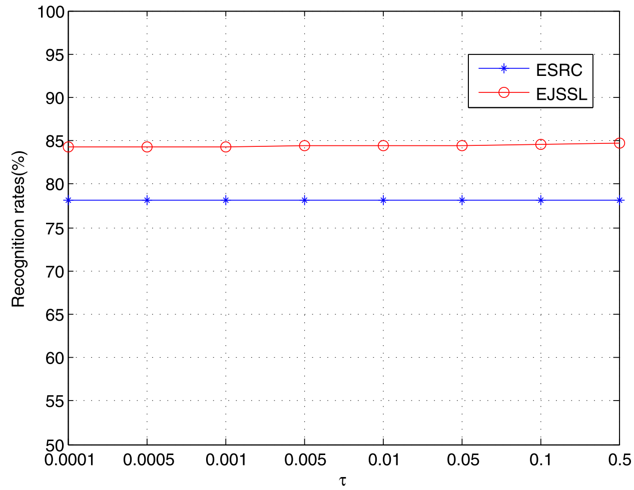


FIGURE 12. The recognition rates versus the big range value of τ on the AR database.

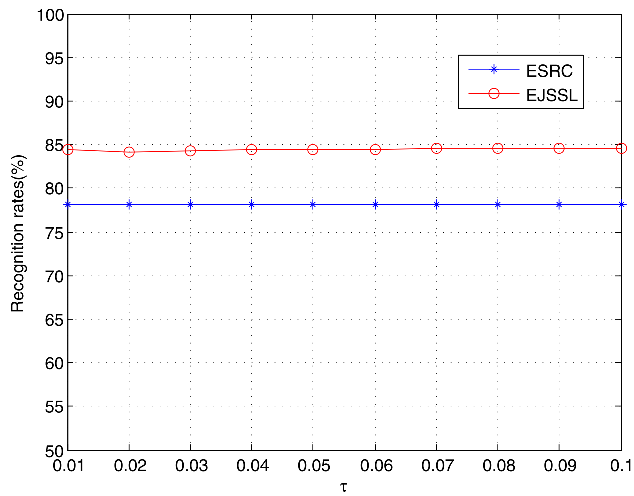


FIGURE 13. The recognition rates versus the small range value of τ on the AR database.

F. DISCUSSION ON EJSSL USING DEEP FEATURES

In the last decade, deep learning has been widely used to learn the deep features of faces and achieved excellent performance in face recognition. The high-level deep features called Deep hidden IDentity features (DeepID) [47] via the deep convolutional neural networks was proposed for face verification. The robust deep feature encoding based discriminative model [48] was proposed for age invariant face recognition. The deep dictionary representation based classification [49] was proposed for occlusion robust face recognition.

The EJSSL using deep features probably achieves better performance than the EJSSL using hand-crafted features, because the deep features have stronger discriminative ability than the hand-crafted features. In the paper, we proposed the EJSSL using hand-crafted features for multi-feature face recognition, because we mainly focus the design of the EJSSL algorithm, which performs better than man current algorithms. We will exploit the EJSSL using deep features in the future.

VIII. CONCLUSION

In the paper, the extended joint similar and specific learning method is proposed for multi-feature face recognition. Four types of different features are extracted. The proposed EJSSL method effectively makes use of the similarity and distinctiveness of the different features. Moreover, the intra-class variant dictionary is constructed and combined with the training images to represent the testing images. The experimental results prove the superiority and effectiveness of our proposed EJSSL method, compared with many current methods. Deep learning has already been successfully applied to extract the deep features from the face images. We will propose EJSSL using the deep features for multi-feature face recognition in the future and hope to achieve better performance.

REFERENCES

[1] J. Yang, K. Yu, Y. Gong, and T. Huang, "Linear spatial pyramid matching using sparse coding for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 1794–1801.

[2] Y. Xu, B. Zhang, and Z. Zhong, "Multiple representations and sparse representation for image classification," *Pattern Recognit. Lett.*, vol. 68, pp. 9–14, Dec. 2015.

[3] Y. Liu, X. Li, C. Liu, and H. Liu, "Structure-constrained low-rank and partial sparse representation with sample selection for image classification," *Pattern Recognit.*, vol. 59, pp. 5–13, Nov. 2016.

[4] 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.

[5] G. Ma, T.-Z. Huang, J. Huang, and C.-C. Zheng, "Local low-rank and sparse representation for hyperspectral image denoising," *IEEE Access*, vol. 7, pp. 79850–79865, 2019.

[6] 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.

[7] Y. Xu, Z. Zhang, G. Lu, and J. Yang, "Approximately symmetrical face images for image preprocessing in face recognition and sparse representation based classification," *Pattern Recognit.*, vol. 54, pp. 68–82, Jun. 2016.

[8] 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.

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

[10] X. Tang, G. Feng, and J. Cai, "Weighted group sparse representation for undersampled face recognition," *Neurocomputing*, vol. 145, pp. 402–415, Dec. 2014.

[11] C.-P. Wei, Y.-W. Chao, Y.-R. Yeh, and Y.-C.-F. Wang, "Locality-sensitive dictionary learning for sparse representation based classification," *Pattern Recognit.*, vol. 46, no. 5, pp. 1277–1287, May 2013.

[12] Y. Zhan, J. Liu, J. Gou, and M. Wang, "A video semantic detection method based on locality-sensitive discriminant sparse representation and weighted KNN," *J. Vis. Commun. Image Represent.*, vol. 41, pp. 65–73, Nov. 2016.

[13] J. Liu, J. Gou, Y. Zhan, and Q. Mao, "Discriminative self-adapted locality-sensitive sparse representation for video semantic analysis," *Multimedia Tools Appl.*, vol. 77, no. 21, pp. 29143–29162, May 2018.

[14] M. Yang and L. Zhang, "Gabor feature based sparse representation for face recognition with Gabor occlusion dictionary," in *Proc. 11th Eur. Conf. Comput. Vis. (ECCV)*, vol. 6316. Heraklion, Greece, Sep. 2010, pp. 448–461.

[15] J. Wen, Y. Xu, Z. Li, Z. Ma, and Y. Xu, "Inter-class sparsity based discriminative least square regression," *Neural Netw.*, vol. 102, pp. 36–47, Jun. 2018.

[16] 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.

[17] M. Yang, L. Zhang, J. Yang, and D. Zhang, "Regularized robust coding for face recognition," *IEEE Trans. Image Process.*, vol. 22, no. 5, pp. 1753–1766, May 2013.

[18] D. M. Vo and S.-W. Lee, "Robust face recognition via hierarchical collaborative representation," *Inf. Sci.*, vol. 432, pp. 332–346, Mar. 2018.

[19] 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.

[20] Y. Chi and F. Porikli, "Connecting the dots in multi-class classification: From nearest subspace to collaborative representation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 3602–3609.

[21] Y. Chi and F. Porikli, "Classification and boosting with multiple collaborative representations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 8, pp. 1519–1531, Aug. 2014.

[22] J. Waqas, Z. Yi, and L. Zhang, "Collaborative neighbor representation based classification using ℓ_2-minimization approach," *Pattern Recognit. Lett.*, vol. 34, no. 2, pp. 201–208, Jan. 2013.

[23] J. Gou, L. Wang, Z. Yi, J. Lv, Q. Mao, and Y.-H. Yuan, "A new discriminative collaborative neighbor representation method for robust face recognition," *IEEE Access*, vol. 6, pp. 74713–74727, 2018.

[24] W. Deng, J. Hu, and J. Guo, "Extended SRC: Undersampled face recognition via intraclass variant dictionary," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 9, pp. 1864–1870, Sep. 2012.

[25] 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. 2018.

[26] G. Lin, M. Xie, and L. Mao, "Extended CRC: Face recognition with a single training image per person via intraclass variant dictionary," *IEICE Trans. Inf. Syst.*, vol. E96.D, no. 10, pp. 2290–2293, 2013.

[27] X.-T. Yuan, X. Liu, and S. Yan, "Visual classification with multitask joint sparse representation," *IEEE Trans. Image Process.*, vol. 21, no. 10, pp. 4349–4360, Oct. 2012.

[28] F. Wu, X.-Y. Jing, and D. Yue, "Multi-view discriminant dictionary learning via learning view-specific and shared structured dictionaries for image classification," *Neural Process. Lett.*, vol. 45, no. 2, pp. 649–666, Apr. 2017.

[29] F. Wu, X. Dong, L. Han, X.-Y. Jing, and Y.-M. Ji, "Multi-view synthesis and analysis dictionaries learning for classification," *IEICE Trans. Inf. Syst.*, vol. E102.D, no. 3, pp. 659–662, Mar. 2019.

[30] X. Deng, P. Song, M. R. D. Rodrigues, and P. L. Dragotti, "RADAR: Robust algorithm for depth image super resolution based on FRI theory and multimodal dictionary learning," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 30, no. 8, pp. 2447–2462, Aug. 2020.

[31] M. Yang, L. Zhang, D. Zhang, and S. Wang, "Relaxed collaborative representation for pattern classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 2224–2231.

[32] J. Li, D. Zhang, Y. Li, J. Wu, and B. Zhang, "Joint similar and specific learning for diabetes mellitus and impaired glucose regulation detection," *Inf. Sci.*, vol. 384, pp. 191–204, Apr. 2017.

[33] Y. Su, S. Shan, X. Chen, and W. Gao, "Hierarchical ensemble of global and local classifiers for face recognition," *IEEE Trans. Image Process.*, vol. 18, no. 8, pp. 1885–1896, Aug. 2009.

[34] T. Ahonen, A. Hadid, and M. Pietikäinen, "Face recognition with local binary patterns," in *Proc. 8th Eur. Conf. Comput. Vis.*, Prague, Czech Republic, May 2004, pp. 469–481.

[35] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced Fisher linear discriminant model for face recognition," *IEEE Trans. Image Process.*, vol. 11, no. 4, pp. 467–476, Apr. 2002.

[36] L. Rosasco, A. Verri, M. Santoro, S. Mosci, and S. Villa, "Iterative projection methods for structured sparsity regularization," MIT, Cambridge, MA, USA, Tech. Rep. MIT-CSAILTR-2009-050, CBCL–282, Oct. 2009.

[37] R. O. Duda, P. E. Hart, and D. G. Stork, *Pattern Classification*. Hoboken, NJ, USA: Wiley, Jan. 2001.

[38] 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.

[39] K.-C. Lee, J. Ho, and D. J. Kriegman, "Acquiring linear subspaces for face recognition under variable lighting," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 5, pp. 684–698, May 2005.

[40] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 23, no. 6, pp. 643–660, Jun. 2001.

[41] A. Martinez and R. Benavente, "The AR face database," CVC Tech. Rep. 24, Jun. 1998. [Online]. Available: http://www.cat.uab.es/Public/Publications/1998/MaB1998/CVCReport24.pdf
[42] L. Wolf, T. Hassner, and Y. Taigman, "Similarity scores based on background samples," in *Proc. 9th Asian Conf. Comput. Vis.*, Xi'an, China, Sep. 2009, pp. 88–97.
[43] G. Lin, M. Yang, J. Yang, L. Shen, and W. Xie, "Robust, discriminative and comprehensive dictionary learning for face recognition," *Pattern Recognit.*, vol. 81, pp. 341–356, Sep. 2018.
[44] M. Yang, L. Zhang, X. Feng, and D. Zhang, "Sparse representation based Fisher discrimination dictionary learning for image classification," *Int. J. Comput. Vis.*, vol. 109, no. 3, pp. 209–232, May 2014.
[45] P. J. Phillips, H. Moon, S. A. Rizvi, and P. J. Rauss, "The FERET evaluation methodology for face-recognition algorithms," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 10, pp. 1090–1104, Oct. 2000.
[46] P. J. Phillips, P. J. Flynn, T. Scruggs, K. W. Bowyer, J. Chang, K. Hoffman, J. Marques, J. Min, and W. Worek, "Overview of the face recognition grand challenge," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2005, pp. 947–954.
[47] Y. Sun, X. Wang, and X. Tang, "Deep learning face representation from predicting 10,000 classes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1891–1898.
[48] M. S. Shakeel and K.-M. Lam, "Deep-feature encoding-based discriminative model for age-invariant face recognition," *Pattern Recognit.*, vol. 93, pp. 442–457, Sep. 2019.
[49] F. Cen and G. Wang, "Dictionary representation of deep features for occlusion-robust face recognition," *IEEE Access*, vol. 7, pp. 26595–26605, 2019.

GUOJUN LIN received the B.S. degree from the Zhejiang University of Technology, in 2001, the M.S. degree from Southwest Jiaotong University, in 2008, and the Ph.D. degree from the University of Electronic Science and Technology of China, in 2014. From May 2008 to October 2008, he was a Software Engineer at Skypine Electronics (Shenzhen) Company Ltd. From 2015 to 2016, he was a Research Assistant with the School of Computer Science and Software Engineering, Shenzhen University. He is currently a Lecturer with the School of Automation and Information Engineering, Sichuan University of Science and Engineering. His research interests include sparse coding, dictionary learning, and face recognition.

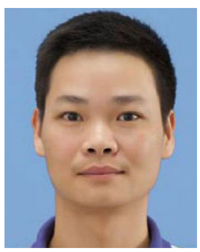
QINRUI ZHANG received the bachelor's degree in electrical engineering from Chengdu Technological University, in 2019. He is currently a Graduate Student with the School of Automation and Information Engineering, Sichuan University of Science and Engineering. His research interests include digital image processing, remote sensing image analysis, and face recognition.

SHUNYONG ZHOU received the B.S. degree in electronic engineering from the Sichuan University of Science and Engineering, Zigong, China, in 2000, and the M.S. degree in signal and information processing from the University of Electronic Science and Technology of China, in 2009. From September 2015 to November 2015, he was a Visiting Scholar with the Department of Engineering, University of Ottawa, Ottawa, ON, Canada. From 2017 to 2018, he was a Visiting Scholar with the Department of Mathematics and Statistics, Texas Tech University, Lubbock, TX, USA. Since 2000, he has been with the Sichuan University of Science and Engineering, where he is currently an Associate Processor. His research interests include image processing and intelligent signal processing.

XINGGUO JIANG received the M.S. degree from Chongqing University, Chongqing, China, in 2003, and the Ph.D. degree from the Institute of Optics and Electronics, Chinese Academy of Sciences, Chengdu, China, in 2007. He is currently working as an Associate Professor with the School of Automation and Electric Information, Sichuan University of Science and Engineering, China. Prior to that, he was an Associate Professor with the School of Information and Communication, Guilin University of Electronic Technology, Guangxi, China. His current research interests include image processing, intelligent information processing, and deep learning.

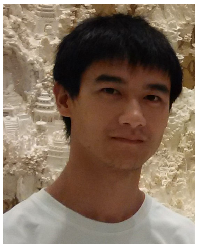
HAO WU received the Ph.D. degree in the automation of electric power systems from Southwest Jiaotong University, Chengdu, China, in 2016. He is currently a Professor with the School of Automation and Information Engineering, Sichuan University of Science and Engineering, Zigong, China. His research interests include power system protection and control, and power systems simulation and modeling.

HAIRONG YOU received the B.E. and B.A. degrees from Shijiazhuang Tiedao University, Shijiazhuang, China, in 2019. She is currently pursuing the M.S. degree with the School of Information Science and Engineering, Northeastern University, Shenyang, China. Her research interests include machine vision and environment perception.



ZUXIN LI was born in Zhejiang, China, in 1972. He received the B.S. degree in industrial automation from the Zhejiang University of Technology, China, in 1995, the M.S. degree in communication and information system from Yunnan University, China, in 2002, and the Ph.D. degree in control theory and control engineering from the Zhejiang University of Technology, in 2008.

From May 2009 to March 2013, he was a Postdoctoral Research Fellow with the Institute of Cyber-Systems and Control, Zhejiang University, China. From August 2013 to November 2013, he was a Visiting Scholar with Dalhousie University, Canada. He is currently a Full Professor with the School of Engineering, Huzhou University, China. His research interests include networked control systems, robust control, estimation, prognostics, and health management.



PING HE was born in Nanchong, China, in November 1990. He received the B.S. degree in automation from the Sichuan University of Science and Engineering, Zigong, China, in 2012, the M.S. degree in control science and engineering from Northeastern University, Shenyang, China, in 2014, and the Ph.D. degree in electromechanical engineering from the Universidade de Macau, Taipa, Macau, in 2017.

From 2015 to 2018, he was an Adjunct Associate Professor with the Sichuan University of Science and Engineering. From 2017 to 2020, he was a Postdoctoral Research Fellow with The University of Hong Kong, Hong Kong, and The Hong Kong Polytechnic University, Hong Kong. Since 2018, he has been a Professor with Jinan University, Guangdong, China. He has authored two books and over 50 articles. His research interests include robot, sensor networks, complex networks, multiagent systems, artificial intelligence, control theory, and control engineering.

Dr. He is a Reviewer Member of *Mathematical Reviews* of American Mathematical Society. He also serves as a Section Editor for *Automatika: Journal for Control, Measurement, Electronics, Computing and Communications*, an Academic Editor for *PLoS ONE*, a Lead Guest Editor for the *International Journal of Distributed Sensor Networks*, and an Associate Editor for *Proceedings of the Institution of Mechanical Engineers—Part E: Journal of Process Mechanical Engineering*, and *The Journal of Engineering* (IET).

HENG LI was born in Hunan, China, in 1963. He received the B.S. and M.S. degrees in civil engineering from Tongji University, Shanghai, China, in 1984 and 1987, respectively, and the Ph.D. degree in architectural science from the University of Sydney, Sydney, NSW, Australia, in 1993.

From 1993 to 1995, he was a Lecturer with James Cook University, Townsville, QLD, Australia. From 1996 to 1997, he was a Senior Lecturer with the Civil Engineering Department, Monash University, Melbourne, VIC, Australia. Since 1997, he has been gradually promoted from an Associate Professor to his current position as a Chair Professor of construction informatics with The Hong Kong Polytechnic University, Hong Kong. He has authored two books and over 500 articles. His research interests include building information modeling, robotics, functional materials, and the Internet of Things.

Dr. Li was a recipient of the National Award from Chinese Ministry of Education, in 2015, and the Gold Prize of Geneva Innovation, in 2019. He is a Reviews Editor of *Automation in Construction*. He is also an Editorial Board Member of *Advanced Engineering Informatics*.