Deep aligned feature extraction for collaborative-representation-based face classification with group dictionary selection

Li Mao, Delei Zhang, Youming Chen, Tao Zhang and Xiaoning Song

Abstract
Face recognition plays an important role in many robotic and human–computer interaction systems. To this end, in recent years, sparse-representation-based classification and its variants have drawn extensive attention in compress sensing and pattern recognition. For image classification, one key to the success of a sparse-representation-based approach is to extract consistent image feature representations for the images of the same subject captured under a wide spectrum of appearance variations, for example, in pose, expression and illumination. These variations can be categorized into two main types: geometric and textural variations. To eliminate the difficulties posed by different appearance variations, the article presents a new collaborative-representation-based face classification approach using deep aligned neural network features. To be more specific, we first apply a facial landmark detection network to an input face image to obtain its fine-grained geometric information in the form of a set of 2D facial landmarks. These facial landmarks are then used to perform 2D geometric alignment across different face images. Second, we apply a deep neural network for facial image feature extraction due to the robustness of deep image features to a variety of appearance variations. We use the term deep aligned features for this two-step feature extraction approach. Last, a new collaborative-representation-based classification method is used to perform face classification. Specifically, we propose a group dictionary selection method for representation-based face classification to further boost the performance and reduce the uncertainty in decision-making. Experimental results obtained on several facial landmark detection and face classification data sets validate the effectiveness of the proposed method.

Keywords
Sparse representation, collaborative-representation-based classification, group dictionary selection, deep aligned feature, face classification

Date received: 20 August 2019; accepted: 24 September 2020

Introduction
In robotic and human–interaction systems, it is crucial for a computer to know more information of a customer, such as the identity, gender, age, behaviour, emotion and so on.1–5 Among these labels/attributes, the identity of a customer might be the most important information for a robot. With
this information, the robot could provide customised services thus to improve users’ experiences significantly. To achieve this goal, a facial recognition system can be deployed on the robot or on a cloud server. For example, as shown in Figure 1, a healthcare robot can capture the facial image of a patient and send the image to the cloud server via a high-speed 5G wireless communication network. Then the cloud server can process the received facial image by searching the database and send the identity and other relevant information of the patient back to the robot. With the information delivered by the cloud server, the robot can provide customised advice or service to the patient. For example, the robot can remind a patient with hypertension to take her/his medicine or show the latest diagnosis report to the patient.

A typical face recognition or classification system has two main steps: feature extraction and face matching. For feature extraction, the use of deep neural networks (DNNs) has become the mainstream in recent years due to their outstanding performance in extracting robust facial features.6,7 Once we obtain the DNN features of a user, we can perform face matching by comparing its facial features with all the existing users’ facial features in a database. To perform face matching, the sparse-representation-based classification (SRC) method has been widely and successfully used in recent years. SRC has also drawn extensive attention in a variety of signal processing and image analysis applications, for example, signal encoding, image compression, feature representation, video analysis and image classification.8–16 For face matching or classification, the key idea of SRC is to obtain the high-fidelity representation of a test sample using a dictionary with sparsity constraints, leading to promising classification results. To be more specific, SRC aims to reconstruct a test sample using a dictionary consisting of all the training samples of all the classes. Meanwhile, the reconstruction coefficients are regularised by the $\ell_1$ norm. The label of the class that has the smallest reconstruction error is assigned as the label of a new test sample. However, it is time-consuming to solve such a $\ell_1$-regularised optimisation problem. To address this issue, Song et al. proposed a quadratic optimisation method with dimensionality reduction to improve the speed of SRC.7 Zhang et al. proposed the collaborative-representation-based classification (CRC) method,17 in which the representation coefficients are regularised by the $\ell_2$ norm. In contrast to SRC, the reconstruction coefficients of CRC can be solved very efficiently with closed-form solutions. Additionally, CRC achieves better performance for face classification in terms of accuracy. Note that, both SRC and CRC use a dictionary that consists of all the samples of all the classes. In this case, all the training samples contribute to the reconstruction of a new test sample. A representation coefficient vector is obtained by solving the $\ell_1$- or $\ell_2$-regularised optimisation problem across all the classes. In contrast to SRC and CRC, Naseem et al. proposed a class-specific method, namely linear-regression-based classification (LRC), in which the training samples of each class are used to form a class-specific dictionary. Those class-specific dictionaries are used to reconstruct a new test sample separately and the label of the one with the smallest reconstruction error is assigned to the test sample. In this article, we proposed a new CRC-based face classification framework due to the efficiency as well as the promising classification accuracy of CRC.

Despite the success of the aforementioned representation-based classification algorithms, it is still a very challenging task to perform robust face classification under unconstrained scenarios in the presence of a wide spectrum of appearance variations, for example, in pose, expression, illumination and occlusion. To address these issues and further strengthen the representative and discriminative capabilities of a representation-based classification method, a number of approaches have been developed in recent years. For example, Deng et al.18 proposed an extended SRC method (ESRC) for face classification, in which an auxiliary intra-class variant dictionary is used to address the small sample size problem as well as the difficulties posed by occlusion and illumination variations. Yang et al.19 and Zhu et al.20 introduced similarity and distinctive of features to present a more general model of CRC. Xu et al. proposed an efficient
SRC by means of an improved norm minimisation.\textsuperscript{21,22} Guo et al. proposed two weighted discriminative collaborative competitive representation methods with the $\ell_1$-norm fidelity for robust image classification.\textsuperscript{23} To improve the performance of a representation-based method for pose-invariant face classification, Song et al. proposed to use a 3D morphable face model for dictionary augmentation, in which a 3D face model is fitted to a 2D face image and the reconstructed 3D face is used to synthesise auxiliary 2D faces with different views.\textsuperscript{11} Mounsef and Karam proposed an augmented SRC framework to improve the performance of the conventional SRC method in the presence of different facial appearance variations.\textsuperscript{24} One common characteristic of those improved algorithms is that they all try to eliminate the difficulties posed by either textural or geometric appearance variations. To further address this issue and improve the performance of a representation-based face classification method, this article presents a new framework using deep align neural network features and a group dictionary selection (GDS) strategy.

As shown in Figure 2, the proposed framework includes three steps: geometric face alignment, robust deep textural feature extraction and collaborative-representation-based face classification with GDS.

The first step, geometric face alignment, is used to address the difficulties posed by geometric appearance variations such as expression, pose and other rigid transformations (translation, scale and rotation). Face alignment, also known as facial landmark detection, plays a very important role in many facial image analysis tasks, for example, face recognition, face tracking, face animation and 3D face modelling (https://www.nist.gov/programs-projects/face-recognition-grand-challenge-frgc).\textsuperscript{25–27} As a preprocessing stage, geometric face alignment influences the performance of a facial image analysis application to a great extent.\textsuperscript{6} However, it is still very challenging to perform robust face alignment under unconstrained scenarios in the presence of large pose variations, abrupt illumination changes, extreme facial expressions and heavy occlusions. To improve the accuracy of face alignment, cascaded shape regression has been proposed and has become very popular in recent years.\textsuperscript{28–32} However, cascaded shape regression usually relies on hand-crafted features and weak regression methods, which cannot address the difficulties posed by appearance variations very well. More recently, DNNs have become the trend in face alignment.\textsuperscript{33–38} For example, Feng et al. proposed a Wing loss function for convolutional neural network (CNN)-based face alignment, which improves the performance of regression-based face alignment with CNNs significantly.\textsuperscript{39} In this article, we use a modified regression visual geometry group (VGG) architecture to obtain fine-grained geometric facial features in the form of a set of 2D facial landmarks. Those facial landmarks are then used to perform 2D geometric normalisation across different face images, using the piecewise affine warp method.

Apart from geometric facial image alignment/normalisation, robust textural facial image feature extraction methods have also been developed to enhance the performance of representation-based classification algorithms. A number of studies have demonstrated that robust image feature extraction methods can promote the performance of pattern classification significantly. Classical representation-based classification methods, such as SRC, CRC and LRC, are usually based on image intensities thus perform poorly in unconstrained scenarios. To address this issue, many robust image feature descriptors, for example, local binary patterns\textsuperscript{40,41} and Gabor\textsuperscript{42,43} have been used in representation-based face classification and demonstrated significant improvements in accuracy. More recently, with the great success of DNNs, CNNs have been proven to be very effective in extracting robust image features for a variety of image classification tasks.\textsuperscript{44–46} However, the use of deep image features in the representation-based

Figure 2. The pipeline of the proposed collaborative-representation-based face classification method with deep aligned feature extraction and group dictionary selection.
classification paradigm has been less investigated by the community. Most existing deep-learning-based image classification methods just simply use the nearest neighbour classifier. This motivates us to further explore the use of deep CNN features for representation-based face classification. To this end, we extract deep CNN features from geometric aligned facial images for CRC-based face classification.

Classical representation-based classification methods, such as SRC and CRC, try to find a representation coefficient for a new test sample using a dictionary consisting of all the training samples of all the classes. In this case, all the classes contribute to the reconstruction of the test sample, which brings uncertainty in decision-making. Although the use of geometric face alignment and DNN features is able to reduce the uncertainty to some extent, the information of a dictionary consisting of all the training samples is redundant. To obtain a compact dictionary and reduce the uncertainty in decision-making, we propose a GDS approach. The proposed GDS approach has two steps. In the first step, we use the classical CRC method to calculate the representation coefficients. Then a measure is used to calculate the response of each class and only the ones with higher responses are selected to form a new dictionary for the final decision-making step. Experimental results demonstrate that the use of our proposed GDS method improves the accuracy of face classification further.

To summary, the main contributions of the proposed method include

- a new collaborative-representation-based face classification framework using robust deep aligned image features;
- an effective GDS approach that reduces information redundancy and uncertainty in decision-making;
- promising experimental results are obtained for both face alignment and face classification on several well-known face benchmarking data sets.

In the next section, we first introduce the classical CRC method, which is the foundation of the proposed algorithm. Then the details of the proposed method are presented in the ‘proposed framework’ section. To validate the performance of the proposed method, we report the experimental results obtained on several well-known face alignment and face classification data sets in the ‘experimental results’ section. Last, in the ‘conclusion and future work’ section, we draw the conclusion of the proposed method and introduce our future work plans.

### Background

In this section, we introduce the classical CRC method, which is the foundation of the proposed method in the next section.

Given $K$ classes and each class has $M$ training samples, we can form a dictionary $X = [x_{1,1}, \ldots, x_{K,M}] \in \mathbb{R}^{P \times KM}$, where $P$ is the dimensionality of each training sample and $KM$ is the total number of samples, also known as atoms, of the dictionary. With this dictionary, a new test sample $y \in \mathbb{R}^P$ can be approximated by a linear combination of all the samples in the dictionary

$$y \approx \sum_{m=1}^{M} \sum_{k=1}^{K} \alpha_{k,m} x_{k,m}$$

where $\alpha_{k,m}$ is the entry of the coefficient vector corresponding to the $m$th training sample of the $k$th class, that is, $x_{k,m} \in \mathbb{R}^P$. The entry $\alpha_{k,m}$ indicates the response of the corresponding training sample to the representation of the test sample $y$. To obtain the reconstruction of the test sample in equation (1), CRC solves the optimisation problem regularised by the $\ell_2$ norm

$$\min \| \alpha \|_2$$

s.t. $y = X\alpha$

where $X \in \mathbb{R}^{P \times KM}$ is the dictionary matrix consisting of all the training samples of all the classes and $\alpha = [\alpha_{1,1}, \ldots, \alpha_{K,M}]^T$ is the coefficient vector estimated by solving the $\ell_2$-norm minimisation problem. The optimisation of equation (2) is a typical least square problem, which can be efficiently achieved by the closed-form solution

$$\alpha = (X^TX + \mu I)^{-1}X^Ty$$

where $\mu$ is a small positive constant controlling the influence of the $\ell_2$ regularisation term, and $I$ is the identity matrix.

Once the coefficient vector is obtained, we can measure the propensity of the $k$th class to the representation of the test sample

$$c_k = \sum_{m=1}^{M} \alpha_{k,m} x_{k,m}$$

where $c_k$ is the signal of the test sample reconstructed using the training samples of the $k$th class. Then we can measure the reconstruction error of the test sample using the $k$th class

$$E(y)_k = \| y - c_k \|_2^2$$

which describes the dissimilarity between the test sample and the $k$th class. Last, the label of the test sample $y$ is assigned by using the label of the class with the smallest reconstruction error

$$Label(y) = \arg \min_k \{E(y)_k\}$$

In spite of the success of CRC-based approaches in face classification, there are some issues of the existing CRC methods. One key issue is that a CRC-based face classification method is sensitive to the geometric variations of a human face. To mitigate this issue and further improve the
face classification accuracy, we propose a new framework for CRC-based face classification in the next section.

**The proposed framework**

To improve the performance of the classical CRC-based face classification method in accuracy, we propose a new geometry aligned facial feature extraction method in this section. In addition, we present a novel group feature selection method for a further performance boost. The proposed CRC-based face classification framework using deep aligned features (DAFs) and GDS has three main steps: geometric face alignment, deep textural image feature extraction and CRC-based face classification with GDS. The pipeline of the proposed method is depicted in Figure 2. To perform robust geometric face alignment, we first use a deep convolutional neural network to predict fine-grained 2D facial landmarks. Then the piecewise affine warp method is applied to both training and test images for geometric face alignment using predicted 2D facial landmarks. Next, we use another deep convolutional neural network to extract robust facial features. Last, the deep aligned facial features are used to perform face classification using a novel CRC-based classification method with GDS.

**Geometric face alignment**

To perform geometric face alignment, we first detect 2D facial landmarks of each training or testing image, using a state-of-the-art deep convolutional neural network, that is, the VGG face model. However, the classical VGG face model is designed for the image classification task, whereas facial landmark detection is a regression task. To meet the requirement of the regression-based 2D facial landmark detection task, we modify the classical VGG network architecture. The classical VGG network has 12 convolution layers, 5 max pooling layers, 3 fully connected layers and a softmax layer. In VGG, each convolution layer is followed by a ReLU non-linear activation layer. We replace the softmax layer and the last fully connected layer by a densely connected regression layer. The architecture of the modified regression VGG network is shown in Figure 3. As depicted in the figure, the input for the network is a 224 × 224 × 3 colour image, , and the output is a 136 × 1 vector, , consisting of the 2D coordinates of 68 facial landmarks.

To train the facial landmark detection network, we use the 300 W face data set. Each face image in the 300W data set has 68 manually annotated 2D facial landmarks. More details of the 300W data set is give in Experimental results Section. As the L2 loss function is sensitive to outliers, we use the L1 loss function for network training

\[
J = \frac{1}{136 \times N} \sum_{n=1}^{N} \sum_{i=1}^{68} |\delta x_{n,i}| + |\delta y_{n,i}| \tag{7}
\]

where \((\delta x_{n,i}, \delta y_{n,i})\) are the differences between the ground-truth coordinates and the predicted coordinates of the network for the \(i\)th facial landmark of the \(n\)th sample in each mini-batch, and \(N\) is the mini-batch size. To train the neural network, we use the stochastic gradient descent optimisation method with momentum. The network is trained for 400 epochs and we set the mini-batch size as \(N = 16\).

After the training of the facial landmark detection network, we apply it to all the training samples and testing samples. Then we apply the piecewise affine warp method to perform geometric face image alignment/normalisation. For more details of the piecewise affine warp method, readers are referred to Matthews and Baker. To be more specific, we map the texture of an input face image, , from its original shape, , to the mean shape, . The mean shape is calculated using the facial landmark of all the training samples in a dictionary

\[
\mathbf{s} = \frac{1}{K \times M} \sum_{m=1}^{M} \sum_{k=1}^{K} s_{k,m} \tag{8}
\]

where \(s_{k,m}\) is the predicted 2D facial landmark of the \(m\)th sample in the \(k\)th class, using our facial landmark detection network.

Some examples of the face image normalisation step using piecewise affine warp are shown in Figure 4. To maintain the background of an input face image, 4 anchor points are added to the obtained 68 landmarks for piecewise affine warp, as shown in the figure. We use the term \(Y'\) for a geometric aligned face image. According to the figure, we can...
see that the original input images of the same subject may vary significantly due to different geometric variations, such as pose variation in the first column and scale variations in the third column. Those geometric variations may cause difficulties in CRC-based face classification. After geometric face alignment, the facial parts with same semantic meanings are very well aligned thus the robustness and accuracy of a face classification method can be improved.

**DAF extraction**

Recently, DNNs especially CNNs have been successfully used for a wide range of image classification tasks. A DNN is able to extract robust facial features for accurate face classification with appearance variations. Instead of the original image intensity feature that has been widely used in many representation-based face classification approaches, we use deep CNN features in our proposed framework. Note that the proposed feature extraction method is applied to a geometric aligned facial image, \( I_0 \), as introduced in the last subsection, rather than the original input image \( I \). We use the term DAFs for the proposed CNN-based feature extraction method.

In many practical applications, such as access control, we may only have very few or even one gallery image of each subject. In this case, it is hard to train or fine-tune a DNN using gallery images and a pre-trained face classification network is used for robust facial feature classification.

**Algorithm 1.** CRC-based face classification with group dictionary selection.

**Require:** A dictionary \( X \), a test sample \( y \) and a pre-fixed dictionary selection proportion \( S \in (0, 1) \);

**Ensure:** The label of the test sample, \( Label(y) \);

1. Recompose the test sample \( y \) using the dictionary \( X \) and obtain the coefficient vector \( \alpha \), using Equation (3);
2. Calculate the contribution of each class in the dictionary to the representation of \( y \) using Equation (10);
3. Rank the contributions of all the classes in the dictionary and select the first \( S \times 100 \) percentage of classes to form a new dictionary \( \tilde{X} \);
4. Recompose the test sample \( y \) using the new dictionary \( \tilde{X} \) and obtain the coefficient vector \( \tilde{\alpha} \), using Equation (3);
5. Obtain the label of the test image using Equation (6);

**return** The label of the test image, \( Label(y) \).

**Table 1.** Facial landmark detection results on the 300W data set.

| Method  | Common set | Challenge set | Full set |
|---------|------------|---------------|----------|
| RCP57   | 6.18%      | 17.26%        | 8.35%    |
| SDM10   | 5.60%      | 15.40%        | 7.52%    |
| CFAN58  | 5.50%      | 16.78%        | 7.69%    |
| ESR31   | 5.28%      | 17.00%        | 7.58%    |
| TCDCN35 | 4.80%      | 8.60%         | 5.54%    |
| CFSS59  | 4.73%      | 9.98%         | 5.76%    |
| LBF56   | 4.95%      | 11.98%        | 6.32%    |
| DDFA60  | 5.53%      | 9.56%         | 6.31%    |
| RAR61   | 4.12%      | 8.35%         | 4.94%    |
| TR-DRN62| 4.10%      | 7.56%         | 4.99%    |
| DSRN63  | 4.12%      | 9.68%         | 5.21%    |
| HORNet64| 4.68%      | 8.77%         | 5.68%    |
| **Our method** | **3.90%** | **7.73%** | **4.67%** |

*The normalised mean error by inter-ocular distance is used as the evaluation metric. The bold font indicates the best results in each subset.*

Figure 4. Geometric face alignment/normalisation using piecewise affine warp. The first and third columns are the original input facial images and the detected 68 facial landmarks. Note that, for each image, 4 anchor points are generated around the facial landmarks to preserve background information. The second and fourth columns are the geometric aligned facial images using the piecewise affine warp approach.
Such a face classification network is usually pre-trained on a large-scale face data set with thousands of identities and it is supposed to generalise well to new identities. To perform face classification for a new subject, the image features of all the gallery images and a probe image are extracted by the pre-trained deep network and the nearest neighbour classifier is usually used to perform face classification. The label of the gallery image with the shortest distance, for example, the cosine distance, to the probe image is assigned to the probe image. However, according to our preliminary experimental results, we found that the combination of cosine distance and the nearest neighbour classifier does not work very well in such a practical application setting.

In this article, we propose to use a representation-based classifier, that is, CRC, for image classification.

To extract deep CNN features, we use the well-known VGG face model. Specifically, for each geometric aligned face image in a dictionary, $I_{k,m}$, we apply the classical VGG face model to extract facial features $x_{k,m} = f(I_{k,m})$ (9) where $f(\cdot)$ stands for the VGG face model that is actually a nonlinear mapping function. We use the 4096-D vector of the second fully connected layer in the classical VGG face model as the facial features. Note that, we also apply the VGG face model to each geometric aligned test image for robust DAF extraction.
GDS for CRC-based face classification

As discussed in the last two subsections, to perform face classification, we use a pre-trained DNN to extract deep aligned CNN features. The main reason to use a pre-trained DNN is due to the fact that we usually have very few training samples for training or fine-tuning a DNN in many practical applications such as access control. However, one issue to use a pre-trained deep CNN is that the network may not generalise well for a new domain. The extracted facial image features are redundant and the standard nearest neighbour classifier cannot address this issue. To mitigate this difficulty, we propose to use a representation-based face classification method, that is, CRC. However, the classical CRC method still has difficulties in addressing the issues posed by information redundancy. Redundant facial image features may lead to uncertainty in decision-making especially when we have a large number of classes. To be more specific, all the samples of all the classes in a dictionary contribute to the reconstruction of a test sample in CRC. Sometimes, a training sample that does not belong to the same label of the test sample may obtain a high response in CRC, which leads to classification errors.

To reduce the information redundancy of the use of a pre-trained deep CNN as well as uncertainty in decision-making, we propose a dictionary optimisation approach, namely GDS.

Given a test image, \( \hat{I} \), we first use our 2D face alignment network to obtain its 68 2D facial landmarks for geometric image alignment that outputs the normalised face image \( \hat{I}_0 \).

Then a pre-trained deep CNN, that is, the classical VGG face model, is used to obtain the DAFs for the test image, which is denoted by \( y \in \mathbb{R}^{R \times P} \). Suppose we have a dictionary \( X \in \mathbb{R}^{P \times KM} \) that consists of the deep aligned CNN features of all the gallery images, we first apply the classical CRC method to obtain the representation coefficient vector, \( \alpha = [\alpha_{1,1}, ..., \alpha_{K,M}]^T \in \mathbb{R}^{KM} \), using equation (3), where \( \alpha_{k,m} \) is the reconstruction coefficient of the \( m \)th sample of the \( k \)th class for the representation of the test sample. Then a \( \ell_2 \)-based measure is applied to the elements of the \( k \)th class which measures the contribution of the training samples from the \( k \)th class to the representation of the test sample.

\[
C_k = \sum_{m=1}^{M} |\alpha_{k,m}| \quad (10)
\]

We use the parameter \( S \in (0, 1) \) for the proportion of the classes selected by the proposed method. In this article, we set the selection proportion as \( S = 0.15 \), which means that only 15% classes are selected to create the optimised dictionary for the final decision making. Note that we use the term GDS for the proposed method because we select all the samples of a class due to the contribution of the whole class to the representation of a new test sample, as described in equation (10). Then the new dictionary is used to perform face classification based on the CRC method introduced in the second section. The proposed CRC-based face classification method with GDS is summarised in Algorithm 1.

**Experimental results**

In this section, we evaluate the performance of the proposed method in terms of accuracy. To be more specific, we first evaluate the performance of the proposed facial landmark detection network on the 300W data set in terms of accuracy. Then we compare the proposed CRC-
based face classification method using deep aligned facial features and GDS with the state-of-the-art approaches on several face classification data sets. In addition, we analyse the effects of each component in the proposed framework, including geometric face alignment, deep CNN feature extraction and GDS, on those face classification data sets. It should be highlighted that the face images from these data sets were captured under a wide spectrum of appearance variations in illumination, pose and expression.

**Evaluation on facial landmark detection**

As the accuracy of 2D facial landmark detection is crucial for our proposed DAF extraction method, we first evaluate the proposed facial landmark detection network on the 300W face data set, compared with a number of state-of-the-art facial landmark detection methods. The 300W data set has a number of facial images selected from different face data sets, including XM2VTS, LFPW, HELEN, Face Recognition Grand Challenge (FRGC) and AFW. 300W has been widely used for benchmarking 2D facial landmark detection algorithms. In this article, we follow the protocol used in Ren et al. This protocol uses AFW, the training sets of LFPW and Helen to create the training set. In total, the training set has 3148 images. The test set of the protocol consists of IBUG, the test sets of LFPW and Helen. In total, the test set has 689 images, which are divided into two subsets: common and challenging subsets. The evaluation metric is the normalised mean error, which is the mean of the Euclidean distance between a predicted landmark and its ground-truth value over all the facial landmarks, normalised by the inter-ocular distance.

We compare the proposed 2D facial landmark detection network with a set of state-of-the-art approaches in Table 1. From the table, we can see that the proposed facial landmark detection network outperforms the state-of-the-art approaches in terms of accuracy on both the common subset and the full set. It achieves competitive results on the challenge set as well, which is only slightly worse than the TR-DRN method. This validates the robustness and accuracy of the proposed method in facial landmark detection.

We show some results of the detected 2D facial landmarks on the 300W test set using our proposed facial landmark detection network in Figure 5. We can see that the proposed facial landmark detection method is very robust to a variety of face appearance variations, such as pose, expression, illumination and so on.

**Evaluation on face classification**

In this part, we first compare the proposed face classification framework with a number of representation-based approaches on several well-known face classification data sets, including FERET, FRGC, LFW and CMU-PIE. Then we conduct an ablation study for the proposed method, in which we analyse each component of the proposed approach, including geometric face alignment, deep CNN feature extraction as well as the proposed GDS method.

**Results on FERET.** The FERET face database is an output of the FERET program, which was sponsored by the US Department of Defence through the DARPA program. This database has become a widely used benchmarking database for the evaluation of face classification techniques. The proposed algorithm was evaluated on a subset of the FERET database, which includes 1400 images of 200 subjects, each has seven different images. Some example images of the FERET data set are shown in Figure 6. To apply the VGG face model to those images for DAF extraction, we converted each FERET image to a colour image by copying the single channel grey-level image to each RGB channel.

| Table 3. A comparison of different face classification methods on the FRGC data set, in terms of face classification accuracy. |
|---------------------------------------------------------------|
| **Method** | **Number of training samples** | |
| | **5** | **10** | **15** |
| CRC | 62.36% | 76.60% | 80.80% |
| LRC | 40.52% | 54.35% | 62.73% |
| ESRC | 66.54% | 79.07% | 82.93% |
| TPTSR | 63.92% | 77.00% | 79.50% |
| SLC-ADL | 36.29% | 48.53% | 59.91% |
| TS-LSRC | 67.95% | 81.18% | 85.88% |
| H-CRC | 79.96% | 93.45% | 95.47% |
| Our method | 97.24% | 98.20% | 98.64% |

Figure 8. Examples images of the PIE data set.
For the FERET database, we used \( q(\theta = 3, 4, 5) \) samples per class as the gallery set, and the remaining samples were used for testing. The classification result of our method is compared to various representation-based methods, including CRC,\(^{17}\) LRC,\(^{68}\) ESRC,\(^{18}\) RRC,\(^{69}\) RCR,\(^{19}\) TPTSR,\(^{70}\) SLC-ADL,\(^{71}\) TS-LSRC,\(^{14}\) Homotopy,\(^{72}\) DALM,\(^{73}\) FISTA\(^{74}\) and DSRL\(^{2,75}\).

As shown in Table 2, we can see that the proposed method consistently achieves much better classification results than the other algorithms, regardless of the number of gallery samples.

**Results on FRGC.** The FRGC version 2 database consists of both constrained and unconstrained facial images. The constrained images have good image quality, whereas the low-quality unconstrained images were captured under complex backgrounds. In this article, we select 100 subjects, each with 30 different images, from FRGC to construct our experimental subset. Some example images of the FRGC database are shown in Figure 7.

For the experiments conducted on the FRGC database, we selected \( q(\theta = 5, 10, 15) \) samples per subject as the gallery set, and the remaining face images were used for testing. The proposed method is compared with CRC,\(^{17}\) LRC,\(^{68}\) ESRC,\(^{18}\) TPTSR,\(^{70}\) SLC-ADL,\(^{71}\) TS-LSRC\(^{14}\) on FRGC in terms of face classification accuracy. Table 3 reports the results of each method using different numbers of training samples. According to the table, the proposed method achieves much better results than the others.

| Method          | Number of training samples |
|-----------------|----------------------------|
| CRC\(^{17}\)    | 31.86% 43.53% 52.94%       |
| LRC\(^{68}\)    | 31.08% 49.41% 52.65%       |
| ESRC\(^{18}\)   | 33.29% 50.92% 54.76%       |
| TPTSR\(^{70}\)  | 59.90% 64.85% 74.41%       |
| SLC-ADL\(^{71}\)| 28.82% 43.02% 45.29%       |
| TS-LSRC\(^{14}\)| 62.55% 67.75% 79.14%       |
| H-CRC\(^{76}\)  | 82.35% 91.29% 93.82%       |
| Our method      | 96.90% 97.85% 97.70%       |

**Results on PIE.** The CMU-PIE face database consists of 41,368 images of 68 individuals with mixed intraclass variations introduced by three types of interference, including 13 different poses, 43 different illumination conditions and 4 different expressions. This database has also become a benchmark database for the evaluation of a face classification algorithm. In this article, the proposed algorithm was evaluated on a subset of the CMU-PIE database, which includes 6800 images of 68 individuals with 20 different images (10 poses and 10 illuminations) of each subject. Some example images of the CMU-PIE database are shown in Figure 8.

For the experiment on the CMU-PIE database, we used \( q(\theta = 5, 10, 15) \) samples per class to construct the gallery set and the remaining images were used for the test. As shown in Table 4, the proposed method achieves better classification accuracy across different sizes of training samples than the other methods including CRC,\(^{17}\) LRC,\(^{68}\) ESRC,\(^{18}\) TPTSR,\(^{70}\) SLC-ADL,\(^{71}\) and TS-LSRC.\(^{14}\)

**Results on LFW.** The LFW face database is one of the most challenging data sets which consists of face images captured under unconstrained scenarios in the presence of pose, illumination, occlusion and expression variations. The LFW database includes 13,233 images of 5749 individuals of different gender, ages and so on. The proposed algorithm is evaluated on a subset of the LFW database, which includes 1580 images of 158 individuals with 10 different images of each subject. Some images from the LFW database are shown in Figure 9.

For the experiment conducted on the LFW database, \( q(\theta = 1, 2, 3, 4) \) samples per subject were randomly selected as gallery images and the remaining images were used for the test. A comparison of the proposed framework with various methods, including CRC,\(^{17}\) LRC,\(^{68}\) ESRC,\(^{18}\) TPTSR,\(^{70}\) SLC-ADL,\(^{71}\) TS-LSRC,\(^{14}\) is presented in Table 5. Again, the proposed method outperforms all the other method to a great extent in terms of face classification accuracy on the LFW data set, which demonstrates the superiority of the proposed method. The key to the success of the proposed method is due to the proposed DAF extraction method as well as the use of the GDS strategy. To analyse the effects of each component of the proposed method, Table 4. A comparison of different face classification methods on the CMU-PIE data set, in terms of face classification accuracy.

**Table 4.** A comparison of different face classification methods on the CMU-PIE data set, in terms of face classification accuracy.

| Method   | Number of training samples |
|----------|----------------------------|
|          | 5  | 10 | 15 |
| CRC\(^{17}\) | 31.86% | 43.53% | 52.94% |
| LRC\(^{68}\) | 31.08% | 49.41% | 52.65% |
| ESRC\(^{18}\) | 33.29% | 50.92% | 54.76% |
| TPTSR\(^{70}\) | 59.90% | 64.85% | 74.41% |
| SLC-ADL\(^{71}\) | 28.82% | 43.02% | 45.29% |
| TS-LSRC\(^{14}\) | 62.55% | 67.75% | 79.14% |
| H-CRC\(^{76}\) | 82.35% | 91.29% | 93.82% |
| Our method | 96.90% | 97.85% | 97.70% |

**Figure 9.** Examples images of the LFW data set.
framework, an ablation study is conducted in the next subsection.

Analysis of the proposed method. The proposed method performs geometric face alignment/normalisation and deep CNN feature extraction for CRC-based face classification.

Table 5. A comparison of different face classification methods on the LFW data set, in terms of face classification accuracy.

| Method       | Number of training samples |
|--------------|----------------------------|
| CRC          | 1  | 2   | 3   | 4   |
| LRC          | 8.34% | 11.58% | 14.03% | 15.63% |
| ESRC         | 4.67% | 7.87% | 10.89% | 13.01% |
| TPTSR        | 9.06% | 14.16% | 17.23% | 19.97% |
| SLC-ADL      | 10.09% | 17.79% | 19.59% | 23.81% |
| TS-LSRC      | 3.86% | 7.69% | 10.84% | 13.82% |
| CRC-ADL      | 11.17% | 18.35% | 21.42% | 25.40% |
| CRC-ADL      | 13.95% | 17.09% | 21.36% | 24.53% |
| CRC          | 20.18% | 34.13% | 44.41% | 52.53% |
| Our method   | 54.52% | 63.23% | 79.22% | 84.07% |

Table 6. Self-analysis of different components of the proposed method, in terms of face classification accuracy.

| Data set | Method       |
|----------|--------------|
| FERET    | CRC          | AF-CRC      | DAF-CRC    | DAF-CRC-GDS |
|          | 50.41% | 80.87% | 94.75% | 96.81% |
| FRGC     | 62.36% | 81.44% | 95.48% | 97.24% |
| CMU-PIE  | 31.86% | 72.06% | 92.64% | 96.90% |
| LFW      | 14.03% | 34.08% | 72.88% | 79.22% |

Figure 10. The YouTube-Faces data set. The first row shows some example faces that are used for the enrolment of our database. The second row shows some example faces that are used for the test of the proposed method.
Simulation. To further evaluate the proposed method in practical applications, we simulate a robotic application scenario on a laptop using the YouTube-Faces data set. To be more specific, we select 200 videos of 100 identities from the YouTube-Faces data set. We enrol these 100 identities to our database as the gallery set using the frames of one video of each identity. For test, the other video of each identity is used, which is different from the video used for identity enrolment of our system. Some example faces from the YouTube-Faces data set are shown in Figure 10.

To simulate the proposed face recognition method in practical scenarios using the YouTube-Faces data set, we use the Single Shot multibox face Detector to detect the face in each frame of a test video. Then the proposed VGG-based facial landmark detection method is used to obtain facial landmarks and perform geometric face alignment. Last, the proposed CRC-based face classification method with GDS is used for face recognition. We show the simulation results of the 10th, 20th, 30th, 40th and 50th frames of a video evaluated on the YouTube-Faces data set in Figure 11. According to the simulation, we can see that the proposed method perform well in practical scenarios.

Conclusion and future work

In this article, we propose a face classification framework using DAFs. The proposed method uses facial landmarks and piecewise affine warp to perform geometric face alignment for robust deep convolutional neural network based facial feature extraction. In addition, a new GDS approach is proposed to further improve the performance of the collaborative-representation-based face classification method. The experiment results obtained on several benchmarking data sets and the simulation results obtained on the YouTube-Faces data set demonstrate the effectiveness of the proposed method. However, based on our simulation results, we find that the proposed method can only perform well when the yaw rotation of a facial image is smaller than around 60°. For the facial images with very large pose variations, the proposed face recognition system becomes unstable, resulting in inaccurate results and failures. The other drawback of the proposed face recognition method is that it cannot deal with motion blur very well. For video-based face recognition applications, we find the fast movement of human faces may lead to severe image blur that degrades the performance of the proposed face recognition system. The above findings motivate us to further improve our face recognition system in the presence of large pose variations and motion blur.

The other challenge or limitation of the proposed method is the deployment of our system in robotics applications. One main reason is that the proposed method is based on large capacity deep CNNs that require high-performance GPU devices to achieve real-time inference speed in practical applications. This can only be done on a cloud server at the current stage due to the energy costs of GPU devices. In our future work, we aim to address this issue by reducing the computational complexity of the proposed method. For example, we intend to design new lightweight DNNs that perform equally well as a large capacity network in facial landmark detection and face classification. Our ultimate goal is to perform accurate and real-time face classification on lightweight edge computing platforms, such as a robot.
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: the National Key Research and Development Program of China (Grant nos. 2017YFC1601800, 2016YFD0401204), the National Natural Science Foundation of China (Grant no. 61876072), China Post-doctoral Science Foundation (Grant no. 2018T110441) and Six Talent Peaks Project in Jiangsu Province (Grant no. XYDXX-012).

ORCID iDs

Li Mao https://orcid.org/0000-0003-3955-1013
Tao Zhang https://orcid.org/0000-0003-3966-4691
Xiaoning Song https://orcid.org/0000-0002-3967-4497

Supplemental material

Supplemental material for this article is available online.

References

1. Ratyal NI, Taj IA, Sajid M, et al. Three-dimensional face recognition using variance-based registration and subject-specific descriptors. Int J Adv Rob Syst 2019; 16(3): 1729881419851716.
2. Zhi Y and Jiang Y. Design of basketball robot based on behavior-based fuzzy control. Int J Adv Rob Syst 2020; 17(2): 1729881420909965.
3. Mohammadi S and Gervei O. Using nonlocal filtering and feature extraction approaches in three-dimensional face recognition by kinect. Int J Adv Rob Syst 2018; 15(4): 172988141877743.
4. Cho H, Roberts R, Jung B, et al. An efficient hybrid face recognition algorithm using PCA and Gabor wavelets. Int J Adv Rob Syst 2014; 11(4): 59.
5. Wu H, Ma X and Li Y. Hierarchical dynamic depth projected difference images–based action recognition in videos with convolutional neural networks. Int J Adv Rob Syst 2019; 16(1): 172988141825093.
6. Taigman Y, Yang M, Ranzato M, et al. Deepface: closing the gap to human-level performance in face verification. In: IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23-28 June 2014. IEEE Computer Society, pp. 1701–1708.
7. Song X, Hu G, Luo LH, et al. Fast SRC using quadratic optimisation in downsized coefficient solution subspace. Signal Process 2019; 161: 101–110.
8. Wright J, Yang AY, Ganesh A, et al. Robust face recognition via sparse representation. IEEE Trans Pattern Anal Mach Intell 2009; 31(2): 210–227.
9. Zhou T, Thung KH, Liu M, et al. Brain-wide genome-wide association study for Alzheimer’s disease via joint projection learning and sparse regression model. IEEE Trans Biomed Eng 2018; 66(1): 165–175.
10. Zhou T, He X, Xie K, et al. Robust visual tracking via efficient manifold ranking with low-dimensional compressive features. Pattern Recognit 2015; 48(8): 2459–2473.
11. Song X, Feng ZH, Hu G, et al. Dictionary integration using 3D morphable face models for pose-invariant collaborative-representation-based classification. IEEE Trans Inf Forensics Secur 2018; 13(11): 2734–2745.
12. Zhong D, Zhu P, Han J, et al. An improved robust sparse coding for face recognition with disguise. Int J Adv Rob Syst 2012; 9(4): 126.
13. Liu S, Zhang G and Liu W. Group sparse representation based dictionary learning for SAR image despeckling. IEEE Access 2019; 7: 30809–30817.
14. Shao C, Song X, Feng ZH, et al. Dynamic dictionary optimization for sparse-representation-based face classification using local difference images. Inf Sci 2017; 393: 1–14.
15. Song X, Feng ZH, Hu G, et al. Half-face dictionary integration for representation-based classification. IEEE Trans Cybern 2017; 47(1): 142–152.
16. Xu T, Feng Z, Wu X, et al. Joint group feature selection and discriminative filter learning for robust visual object tracking. In: IEEE International Conference on Computer Vision, Seoul, Korea (South), October 27-November 2, 2019. IEEE, pp. 7949–7959.
17. Zhang L, Yang M, Feng X, et al. Collaborative representation based classification for face recognition. arXiv preprint arXiv:1204.2358 2012: 1–33.
18. Deng W, Hu J and Guo J. Extended SRC: undersampled face recognition via intraclass variant dictionary. IEEE Trans Pattern Anal Mach Intell 2012; 34(9): 1864–1870.
19. Yang M, Zhang L, Zhang D, et al. Relaxed collaborative representation for pattern classification. In: IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16-21 June 2012. IEEE Computer Society, pp. 2224–2231.
20. Zhu P, Zuo W, Zhang L, et al. Image set-based collaborative representation for face recognition. IEEE Trans Inf Forensics Secur 2014; 9(7): 1120–1132.
21. Xu Y, Zuo W and Fan Z. Supervised sparse representation method with a heuristic strategy and face recognition experiments. Neurocomputing 2012; 79: 125–131.
22. Xu Y, Li X, Yang J, et al. Integrating conventional and inverse representation for face recognition. IEEE Trans Cybern 2014; 44(10): 1738–1746.
23. Guo J, Wang L, Yi Z, et al. Weighted discriminative collaborative competitive representation for robust image classification. Neural Netw 2020; 125: 104–120.
24. Mounsef J and Karam L. Augmented SRC for face recognition under quality distortions. IET Biom 2019; 8(6): 431–442.
25. Kitzler J, Huber P, Feng Z, et al. 3D morphable face models and their applications. In: 9th International Conference on Articulated Motion and Deformable Objects, Palma de Mallorca, Spain, 13-15 July 2016. Springer, pp. 185–206.
26. Chai X, Wang Q, Zhao Y, et al. Robust facial landmark detection based on initializing multiple poses. Int J Adv Rob Syst 2016; 13(5): 1729881416662793.
27. Feng ZH, Huber P, Kittler J, et al. Evaluation of dense 3D reconstruction from 2D face images in the wild. In: IEEE Conference on Automatic Face and Gesture Recognition, Xian, China, 15-19 May 2018. pp. 780–786.
28. Dollár P, Welinder P and Perona P. Cascaded pose regression. In: IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13-18 June, 2010. IEEE Computer Society, pp. 1078–1085.
29. Feng ZH, Huber P, Kittler J, et al. Random cascaded-regression cope for robust facial landmark detection. IEEE Signal Process Lett 2015; 1(22): 76–80.
30. Xiong X and la Torre FD. Supervised descent method and its applications to face alignment. In: IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23-28 June 2013. IEEE Computer Society, pp. 532–539.
31. Cao X, Wei Y, Wen F, et al. Face alignment by explicit shape regression. Int J Comput Vision 2014; 107(2): 177–190.
32. Feng ZH, Hu G, Kittler J, et al. Cascaded collaborative regression for robust facial landmark detection trained using a mixture of synthetic and real images with dynamic weighting. IEEE Trans Image Process 2015; 24(11): 3425–3440.
33. Feng Z, Kittler J, Awais M, et al. Wing loss for robust facial landmark localisation with convolutional neural networks. In: IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18-22 June 2018. IEEE Computer Society, pp. 2235–2245.
34. Feng ZH, Kittler J and Wu XJ. Mining hard augmented samples for robust facial landmark localisation with CNNs. IEEE Signal Process Lett 2019; 26(3): 450–454.
35. Zhang Z, Luo P, Loy CC, et al. Facial landmark detection by deep multi-task learning. In: 13th European Conference on Computer Vision, Zurich, Switzerland, 6-12 September 2014. Springer, pp. 94–108.
36. Zhang R, Mu C, Xu M, et al. Facial component-landmark detection with weakly-supervised LR-CNN. IEEE Access 2019; 7: 10263–10277.
37. Storey G, Bouridane A and Jiang R. Integrated deep model for face detection and landmark localization from “in the wild” images. IEEE Access 2018; 6: 74442–74452.
38. Wu Y and Ji Q. Facial landmark detection: a literature survey. Int J Comput Vision 2019; 127(2): 115–142.
39. Feng ZH, Kittler J, Awais M, et al. Rectified wing loss for efficient and robust facial landmark localisation with convolutional neural networks. Int J Comput Vision 2020; 128: 2126–2145.
40. Min R and Dugelay J. Improved combination of LBP and sparse representation based classification (SRC) for face recognition. In: IEEE International Conference on Multimedia and Expo, Barcelona, Catalonia, Spain, 11-15 July 2011. IEEE Computer Society, pp. 1–6.
41. Song X, Chen Y, Feng ZH, et al. Collaborative representation based face classification exploiting block weighted LBP and analysis dictionary learning. Pattern Recognit 2019; 88: 127–138.
42. Tan X and Triggs B. Fusing Gabor and LBP feature sets for kernel-based face recognition. In: Third International Workshop on Analysis and Modeling of Faces and Gestures, Rio de Janeiro, Brazil, 20 October, 2007. Springer, pp. 235–249.
43. Choi JY and Lee B. Ensemble of deep convolutional neural networks with Gabor face representations for face recognition. IEEE Trans Image Process 2020; 29: 3270–3281.
44. Simonyan K and Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:14091556 2014.
45. He K, Zhang X, Ren S, et al. Deep residual learning for image recognition. In: IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016. IEEE Computer Society, pp. 770–778.
46. He M, Zhang J, Shan S, et al. Deformable face net for pose invariant face recognition. Pattern Recognit 2020; 107: 117113.
47. Parkhi OM, Vedaldi A and Zisserman A. Deep face recognition. In: British Machine Vision Conference, Swansea, UK, 1–8 September 2015. BMVA Press, pp. 41.1–41.12.
48. Sagonas C, Tzimiropoulos G, Zafeiriou S, et al. 300 faces in-the-wild challenge: the first facial landmark localization challenge. In: IEEE International Conference on Computer Vision Workshops, Sydney, Australia, 1–8 December 2013. IEEE Computer Society, pp. 397–403.
49. Girshick RB. Fast R-CNN. In: IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015. IEEE Computer Society, pp. 1440–1448.
50. Matthews I and Baker S. Active appearance models revisited. Int J Comput Vision 2004; 60(2): 135–164.
51. Messer K, Matas J, Kittler J, et al. XM2VTSDB: the extended M2VTS database. In: Second international Conference on Audio and Video-based Biometric Person Authentication, 1999, volume 964. pp. 965–966.
52. Belhumeur PN, Jacobs DW, Kriegman DJ, et al. Localizing parts of faces using a consensus of exemplars. IEEE Trans Pattern Anal Mach Intell 2013; 35(12): 2930–2940.
53. Le V, Brandt J, Lin Z, et al. Interactive facial feature localization. In: 12th European Conference on Computer Vision, Florence, Italy, 7–13 October 2012. Springer, pp. 679–692.
54. Phillips PJ, Flynn PJ, Scruggs WT, et al. Overview of the face recognition grand challenge. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Diego, CA, USA, 20–26 June 2005. IEEE Computer Society, pp. 947–954.
55. Zhu X and Ramanan D. Face detection, pose estimation, and landmark localization in the wild. In: IEEE Conference on Computer Vision and Pattern Recognition, Providence, RI, USA, 16–21 June 2012. IEEE Computer Society, pp. 2879–2886.
56. Ren S, Cao X, Wei Y, et al. Face alignment via regressing local binary features. IEEE Trans Image Process 2016; 25(3): 1233–1245.
57. Burgos-Artizzu XP, Perona P and Dollár P. Robust face landmark estimation under occlusion. In: IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 23–28 June 2013. IEEE, pp. 1513–1520.
58. Zhang J, Shan S, Kan M, et al. Coarse-to-fine auto-encoder networks (CFAN) for real-time face alignment. In: 13th European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014. Springer, pp. 1–16.
59. Zhu S, Li C, Loy CC, et al. Face alignment by coarse-to-fine shape searching. In: IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015. IEEE Computer Society, pp. 4998–5006.
60. Zhu X, Lei Z, Liu X, et al. Face alignment across large poses: A 3D solution. In: IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016. pp. 146–155.
61. Xiao S, Feng J, Xing J, et al. Robust facial landmark detection via recurrent attentive-refinement networks. In European Conference on Computer Vision, Amsterdam, the Netherlands, 11–14 October 2016. Springer, pp. 57–72.
62. Lv J, Shao X, Xing J, et al. A deep regression architecture with two-stage re-initialization for high performance facial landmark detection. In: IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017. IEEE Computer Society, pp. 3691–3700.
63. Miao X, Zhen X, Liu X, et al. Direct shape regression networks for end-to-end face alignment. In: IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018. IEEE Computer Society, pp. 5040–5049.
64. Zhen X, Yu M, Xiao Z, et al. Heterogenous output regression network for direct face alignment. Pattern Recognit 2020; 105: 107311.
65. Phillips PJ, Moon H, Rizvi SA, et al. The FERET evaluation methodology for face-recognition algorithms. IEEE Trans Pattern Anal Mach Intell 2000; 22(10): 1090–1104.
66. Huang GB, Ramesh M, Berg T, et al. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Technical Report 07-49, University of Massachusetts, Amherst, 2007.
67. Sim T, Baker S and Bsat M. The CMU pose, illumination, and expression database. IEEE Trans Pattern Anal Mach Intell 2003; 25(12): 1615–1618.
68. Naseem I, Togneri R and Bennamoun M. Linear regression for face recognition. IEEE Trans Pattern Anal Mach Intell 2010; 32(11): 2106–2112.
69. Yang M, Zhang L, Yang J, et al. Regularized robust coding for face recognition. IEEE Trans Image Process 2013; 22(5): 1753–1766.
70. Xu Y, Zhang D, Yang J, et al. A two-phase test sample sparse representation method for use with face recognition. IEEE Trans Circuits Syst Video Technol 2011; 21(9): 1255–1262.
71. Wang J, Guo Y, Guo J, et al. Synthesis linear classifier based analysis dictionary learning for pattern classification. Neurocomputing 2017; 238: 103–113.
72. Yang AY, Sastry SS, Ganesh A, et al. Fast 11-minimization algorithms and an application in robust face recognition: a review. In: International Conference on Image Processing, Hong Kong, China, 26-29 September 2010. IEEE, pp. 1849–1852.
73. Yang AY, Zhou Z, Balasubramanian AG, et al. Fast 11-minimization algorithms for robust face recognition. IEEE Trans Image Process 2013; 22(8): 3234–3246.
74. Beck A and Teboulle M. A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM J Imag Sci 2009; 2(1): 183–202.
75. Xu Y, Zhong Z, Yang J, et al. A new discriminative sparse representation method for robust face recognition via l1 regularization. IEEE Trans Neural Networks Learn Syst 2017; 28(10): 2233–2242.
76. Shi L, Song X, Zhang T, et al. Histogram-based CRC for 3D-aided pose-invariant face recognition. Sensors 2019; 19(4): 759.
77. Wolf L, Hassner T and Maoz I. Face recognition in unconstrained videos with matched background similarity. In: IEEE Conference on Computer Vision and Pattern Recognition, Colorado Springs, CO, USA, 20-25 June 2011. IEEE Computer Society, pp. 529–534.
78. Liu W, Anguelov D, Erhan D, et al. SSD: single shot multibox detector. In: European Conference on Computer Vision, Amsterdam, the Netherlands, 11–14 October 2016. Springer, pp. 21–37.